A comparative study of color quantization methods using various image quality assessment indices

María-Luisa Pérez-Delgado[1] · **M. Emre Celebi[2]**

Abstract

This article analyzes various color quantization methods using multiple image quality assessment indices. Experiments were conducted with ten color quantization methods and eight image quality indices on a dataset containing 100 RGB color images. The set of color quantization methods selected for this study includes well-known methods used by many researchers as a baseline against which to compare new methods. On the other hand, the image quality assessment indices selected are the following: mean squared error, mean absolute error, peak signal-to-noise ratio, structural similarity index, multi-scale structural similarity index, visual information fidelity index, universal image quality index, and spectral angle mapper index. The selected indices not only include the most popular indices in the color quantization literature but also more recent ones that have not yet been adopted in the aforementioned literature. The analysis of the results indicates that the conventional assessment indices used in the color quantization literature generate different results from those obtained by newer indices that take into account the visual characteristics of the images. Therefore, when comparing color quantization methods, it is recommended not to use a single index based solely on pixelwise comparisons, as is the case with most studies to date, but rather to use several indices that consider the various characteristics of the human visual system.

Keywords Color quantization · Image quality assessment · Image processing · Color image

1 Introduction

There are numerous image processing operations that require the use of an image quality assessment (IQA) index to evaluate the quality of the images generated. These include image compression, filtering, segmentation, fusion and color quantization. Image quality is subjective in nature, as different people may perceive the same image differently. There are two types of IQA indices: subjective indices and objective indices. Subjective indices are based on the opinion of human observers, while objective indices use mathematical models.

Historically, objective IQA indices were primarily based on simple mathematical measurements, such as the distance between the corresponding pixels in the original and the modified images. One such IQA index is the mean squared error (MSE), which is by far one of the most widely used due to its computational simplicity. However, simple IQA indices like MSE do not always reflect the image distortions perceived by the human visual system (HVS). For example, two images may be considered by a human observer to be nearly identical, but their MSE may be large [1]. For this reason, more elaborate objective IQA indices have been proposed over the past three decades that attempt to model the mechanisms of the HVS [2, 3]. Objective IQA indices can be divided into two groups: full-reference indices [4] and no-reference indices [2, 5]. In the former, the original image is available for comparison with the modified image.

María-Luisa Pérez-Delgado and M. Emre Celebi contributed equally to this work.

✉ María-Luisa Pérez-Delgado
  mlperez@usal.es

  M. Emre Celebi
  ecelebi@uca.edu

[1] Department of Computer Science and Automatics, Universidad de Salamanca, Escuela Politécnica Superior de Zamora, Av. Requejo, 33, Zamora 49022, Spain

[2] Department of Computer Science and Engineering, University of Central Arkansas, 201 Donaghey Ave., Conway, AR 72035, USA

In contrast, in the latter, the quality of the modified image is predicted without having access to the original image.

MSE falls into the category of full-reference indices. Other well-known IQA indices in this group include the mean absolute error (MAE) and the peak signal-to-noise ratio (PSNR). Newer IQA indices in this group attempt to capture visual features of the image that are important to a human observer. Among these IQA indices are the structural similarity index (SSIM), the multi-scale structural similarity index (MS-SSIM), and the visual information fidelity index (VIF).

Color quantization (CQ) is a common image processing operation that not only has practical applications in itself, but is also a preprocessing step for various other image processing techniques [6]. Articles dealing with CQ methods generally use MSE to evaluate the error of the resulting images. MAE and PSNR are also used, although less frequently. There are many articles that use only MSE as the IQA index [7–17]. There are a few articles that use only PSNR [18, 19]. Some articles employ both MSE and PSNR [20, 21], or MSE and MAE [22–25]. It is uncommon to find articles that use HVS-based IQA indices to compare CQ methods, although there are only a few such articles in the recent literature. For example, several articles use SSIM [26–28].

Recent studies on image quality demonstrate the need to use HVS-based IQA indices to better interpret the quality of the images from a human perspective, since such indices can measure the image quality consistently with subjective evaluations [3, 4, 29–31]. For this reason, it is interesting to compare the results of CQ methods using not only the traditional IQA indices, but also newer IQA indices that take into account the HVS characteristics.

The objective of this article is to compare the results obtained with the three most common IQA indices in CQ (MSE, MAE and PSNR) to the results obtained with more recent IQA indices. As a result of the study, it will be possible to determine if these three IQA indices are adequate to assess the quality of a new CQ method or if they should be complemented with newer and more elaborate IQA indices that are less common in the current CQ literature. For this purpose, this article evaluates eight IQA indices to compare the effectiveness of ten CQ methods.

The CQ methods selected for this study are the following: Median-cut (MC) [32], Octree (OC) [33], Variance-based method (WAN) [17], Binary splitting (BS) [34], Greedy orthogonal bi-partitioning method (WU) [35], Neuquant (NQ) [36], Adaptive distributing units (ADU) [7], Variance-cut (VC) [22], WU combined with Ant-tree for color quantization (ATCQ), (WUATCQ) [37] and BS combined with iterative ATCQ (BSITATCQ) [38]. On the other hand, the IQA indices selected are the following: MSE, MAE, PSNR, SSIM, MS-SSIM, VIF, the universal image quality index (UQI) and the spectral angle mapper index (SAM). With the exception of WUATCQ and BSITATCQ, these CQ methods are classical and many authors use them as benchmarks against new CQ methods. WUATCQ and BSITATCQ were included in the analysis because they are methods that try to improve the solution obtained by two classical methods (WU and BS, respectively). The articles that describe both methods compare them with the originals using only MSE, so it seems interesting to compare them using other IQA indices. On the other hand, the selected IQA indices are popular in the image processing literature due in part to their publicly available implementations. Therefore, the selected IQA indices can be easily used by other researchers.

The experiments in this study were performed on images from the publicly available CQ100 dataset [39], which includes 100 color images. The results generated in this study will be published in the upcoming version of CQ100. Since the chosen CQ methods are deterministic, these results may be used by other researchers to assess the effectiveness of new CQ methods.

There are several articles that compare various IQA indices, but do not consider the specific problem of CQ. This includes articles that compare the results of several IQA indices when considering image acquisition and reconstruction strategies [40], image restoration [41], and training deep neural networks for several low-level vision tasks [42]. However, there is a recent work that compares several IQA indices calculated on quantized images [43]. The said article compares nine IQA indices (PSNR, UQI, SSIM, MS-SSIM, VIF, noise quality index, signal-to-noise ratio (SNR), VSNR, weighted SNR) applied to distorted images obtained from two datasets and shows that PSNR outperforms SSIM. The research described in [43] is different from the proposal presented in this article, since it used subjective scores of the distorted images (provided by human observers) to evaluate the results obtained by IQA indices.

Since many IQA indices have been proposed over the years [44], it is necessary to choose some of them to carry out our study. The best would be to use the indices that have generated the best results for the image processing task under consideration, but there is no clear information on this [45]. As an example, the results presented in [46] compare MSE, PSNR, UQI, and SSIM, among other IQA indices, and show that there is no index that generates good results for all the types of distortions that can be present in an image. Therefore, as there is no universal set of indices suitable for CQ, we have used several criteria to make the selection. On one hand, we have selected indices that are commonly used when applying various image processing techniques. On the other hand, a higher priority was given to indices with publicly available implementations, which makes it easier for other researchers to compare their results with ours. In addition, because in the CQ operation both

the original image and the one resulting from processing are available, we consider that it is interesting to apply full-reference quality indices. After considering this first set of general criteria, more specific criteria were considered to choose the specific indices.

MSE, MAE, and PSNR are the most used indices in the literature regarding CQ, so it is necessary to consider them. While these three indices are based on a pixel-by-pixel comparison between the original and the quantized image, the other selected indices use other information. SSIM, UQI, and MS-SSIM quantify the extent to which the image structure has been distorted; VIF is based on the mutual information between the original and quantized images; while SAM compares image spectra.

As noted above, SSIM is the index used in some CQ articles that are not limited to the use of MSE, MAE, or PSNR. SSIM has received a lot of attention since its inception and several variants have been proposed that attempt to improve it. This index has produced good results in various image processing operations and has been shown to improve MSE and PSNR results [47]. On the other hand, the UQI index is a particular case of the SSIM index, which can be calculated with lower computational cost. Therefore, it is interesting to check if UQI generates results similar to those of SSIM, since this would allow the use of an index that can be calculated more quickly, which would allow it to be integrated more efficiently in CQ-related applications. The results published by several authors indicate that UQI outperforms MSE and PSNR for several types of image distortions [4, 48]. As far as MS-SSIM is concerned, it is an improved version of SSIM calculated at a variety of scales and considers a wider range of viewing distances. Furthermore, various authors concluded that it generates better results than SSIM [49].

VIF has been tested across a wide variety of image distortion types, showing good results. In addition, several researchers concluded that VIF generates better results than PSNR, MS-SSIM, SSIM, and UQI [50–52]. A distinctive property of VIF over other IQA indices is that it can handle cases where the distorted image is visually superior to the original image [42].

Spectral information is considered an important aspect of human vision [53]. As noted above, SAM determines the spectral similarity between the original and quantized images. This index is widely used in hyperspectral image analysis, computer vision, and remote sensing applications. The popularity of this index is mainly due to its simplicity and geometrical interpretability. It can be computed easily and quickly, and it is insensitive to illumination.

The main contribution of this article is the comparison of several CQ methods (including very popular methods) using a large test set and several IQA indices that take into account different image characteristics. The article compares the results of ten CQ methods, using eight IQA indices and a test set with 100 images. The tests carried out allow us to conclude that the most widely used IQA index in the CQ literature, the MSE index, does not generate results comparable to those obtained by the other indices analyzed. Therefore, it seems appropriate that when evaluating new CQ methods, several IQA indices are used to compare them with existing CQ methods. On the other hand, to help other researchers in the comparison of CQ methods, the IQA index of the quantized images generated by each CQ method (32000 values in total) is included as supplementary material to the article and will also be published in the upcoming version of CQ100.

The rest of the article is organized as follows. Section 2 briefly describes the CQ methods considered in the article, while Sect. 3 describes the IQA indices that will be used. Then, Sect. 4 discusses the experimental results and finally Sect. 5 presents the conclusions of the investigation.

2 Color quantization methods

Consider a color image whose pixels are distributed over H rows and W columns. If the RGB color space is used to represent the image, each pixel p_{ij}, with $i \in [1, H]$ and $j \in [1, W]$, is defined by three integer values from the interval [0, 255] that represent the amount of red (R), green (G) and blue (B) color of the pixel: $p_{ij} = (R_{ij}, G_{ij}, B_{ij})$.

The goal of a CQ method is to obtain a new image, called a quantized image, that has a limited number of distinct colors, q, but is visually as similar as possible to the original image. In other words, while the dimensions of the two images are identical, the size or number of colors of the palette used to represent the original image is much larger than the size of the palette used to represent the quantized image. Let $p'_{ij} = (R'_{ij}, G'_{ij}, B'_{ij})$ denote the pixel of the quantized image in row i and column j, with $i \in [1, H]$ and $j \in [1, W]$, that is, p'_{ij} occupies in the quantized image the same position as p_{ij} in the original image.

Most of the CQ methods described below are divisive methods. In general, they are fast methods that perform successive divisions of the color space. They apply an iterative process to divide the color space into q regions and add a color to the quantized palette representing each of these regions. The methods differ mainly in the region selected for the next split, the splitting axis and the splitting point used.

2.1 The median-cut method

MC generates a quantized palette where each color represents approximately the same number of pixels in the original image [32]. Each iteration of the splitting process splits

the region with the most pixels to obtain two new regions. The selected region is divided along the longest axis at the median point. When the splitting process is complete, the centroids of the resulting regions define the colors of the quantized palette.

2.2 The octree method

OC uses an octree structure to define the quantized palette [33]. The nodes of this tree can have 8 children and the tree needs to include a maximum of 8 levels to store the colors of an RGB image. This method includes two stages. The first stage uses the original image pixels to build the tree. The second stage merges the leaves of the tree that represent a few pixels, and this operation continues until the number of leaves equals the size of the quantized palette.

2.3 The variance-based method

WAN is based on the same idea as MC, but uses a different criterion to select the region to split and also considers different splitting axis and splitting point [17]. The selected region is the one with the largest weighted variance. This region is divided along the axis with the least weighted sum of projected variances at the point that minimizes the marginal squared error.

2.4 The binary splitting method

BS uses a binary tree to define the colors of the quantized palette [34]. Each iteration selects the leaf node with the highest distortion and creates two children for that node, distributing the parent's pixels between the two. The process ends when the number of leaves in the tree reaches q. The average color of the pixels associated with each leaf defines a color of the quantized palette.

2.5 The greedy orthogonal bi-partitioning method

This method applies the same idea as WAN but the splitting axis considered is different. WU uses the axis that minimizes the sum of the variances of both sides to divide the selected region [35].

2.6 The neuquant method

NQ is based on the use of a one-dimensional self-organizing feature map [36]. This neural network includes as many neurons as the size of the quantized palette. The image pixels are used to train this neural network. The final weights of the network define the quantized palette.

2.7 The variance-cut method

VC applies the binary splitting strategy like MC, WAN, BS, and WU [22]. In each iteration, this method selects the region with the largest sum of squared errors. That region is then divided along the coordinate axis with the largest variance. In this case, the splitting point is the mean point. When the splitting process is complete, the centroids of the final set of regions define the colors of the quantized palette.

2.8 The adaptive distributing units method

ADU is a clustering-based method that applies competitive learning [7]. The centroid of each cluster defines a color of the quantized palette. All pixels in the image initially define a single cluster, so the quantized palette only includes one color. The method applies an iterative process that increases the number of clusters and, therefore, the size of the quantized palette, until a palette of desired size is obtained. In each iteration, a pixel of the original image is associated with the closest color in the current quantized palette. When the number of pixels associated with a color reaches a threshold, a new color is added to the palette.

2.9 WU combined with ATCQ

WUATCQ [37] is a two-stage method that applies a swarm-based algorithm (ATCQ) to improve the quantized palette obtained by WU. ATCQ is a CQ method that represents each pixel of the original image by an ant and builds a tree where the ants progressively connect, taking into account the similarity between their color and the color of the tree node they are trying to connect to. The second level nodes of the resulting tree define the colors of the quantized palette [12]. WUATCQ first applies the WU method, resulting in a quantized palette. Next, the ATCQ operations are applied, but using a tree whose second level nodes initially represent the colors of the quantized palette generated by WU.

2.10 BS combined with ITATCQ

BSITATCQ [38] is another two-stage method where the iterative variant of ATCQ (ITATCQ) is applied to the quantized palette generated by BS. In this case, the methods used in the two stages use a tree structure to carry out the quantization process. ITATCQ applies a first phase in which the ATCQ operations define an initial tree. An iterative process

is then applied to improve the quality of the quantized palette, where all ants are disconnected from the tree and then reconnected [21].

3 Image quality assessment indices

This section describes the IQA indices that are used in this study to compare the results of the selected CQ methods. All of them are full-reference IQA indices.

3.1 Mean squared error

The MSE compares each pixel of the original image to the corresponding pixel in the quantized image (Eq. 1). The interval for this error is $[0, 3 \times 255^2]$. The smaller the MSE value, the better the quantized image. The error is zero when the original and quantized images are exactly the same [54, 55].

$$MSE = \frac{1}{WH} \sum_{i=1}^{H} \sum_{j=1}^{W} ||p_{ij} - p'_{ij}||_2^2 = \frac{1}{WH} \sum_{i=1}^{H} \sum_{j=1}^{W} (R_{ij} - R'_{ij})^2 + (G_{ij} - G'_{ij})^2 + (B_{ij} - B'_{ij})^2 \tag{1}$$

3.2 Peak signal-to-noise ratio

The PSNR of a quantized image can be computed from its MSE by Eq. 2. Larger values of this error correspond to quantized images more similar to the original image. In the numerator of the logarithm function's argument, the value 255 represents the dynamic range of the pixel values for an 8-bit-per-channel RGB image [55].

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE/3} \right) \tag{2}$$

3.3 Mean absolute error

MAE is computed by Eq. 3. It is based on the calculation of the absolute value of the difference between the corresponding pixels in the original and the quantized images [55]. The quantized image is more similar to the original as this error approaches zero. The interval for this error is $[0, 3 \times 255]$.

$$MAE = \frac{1}{WH} \sum_{i=1}^{H} \sum_{j=1}^{W} ||p_{ij} - p'_{ij}||_1 = \frac{1}{WH} \sum_{i=1}^{H} \sum_{j=1}^{W} |R_{ij} - R'_{ij}| + |G_{ij} - G'_{ij}| + |B_{ij} - B'_{ij}| \tag{3}$$

3.4 Universal image quality index

The UQI models any image distortion as a combination of three factors: loss of correlation, luminance distortion and contrast distortion [48]. It is computed by Eq. 4, where $\bar{x}$ and σ_x are the average color and the standard deviation, respectively, of the pixels in the original image; $\bar{y}$ and σ_y represent the same quantities, but for the pixels of the quantized image; σ_{xy} is the correlation coefficient between the two images. This index takes values in the range $[-1, 1]$. The value 1 is obtained when both images are the same.

$$UQI = \left(\frac{\sigma_{xy}}{\sigma_x \sigma_y} \right) \left(\frac{2\bar{x}\bar{y}}{\bar{x}^2 + \bar{y}^2} \right) \left(\frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \right) \tag{4}$$

The developers of the index proposed to apply it to local regions of the image. With this purpose, a sliding window of size $B \times B$ is used to compute the index for the pixels inside that window. This sliding window operates in raster fashion. Therefore, the overall index is computed by Eq. 5, where UQI_k is the local index computed within the sliding window at step k and W denotes the total number of steps.

$$UQI = \frac{1}{W} \sum_{k=1}^{W} UQI_k \tag{5}$$

3.5 Structural similarity index measure

SSIM is computed by Eq. 6, based on the differences in the luminance (l), the contrast (c) and the structure (s) between the original and quantized images [47]. The index values range from 0 (completely different images) to 1 (identical images).

$$SSIM = l^\alpha \cdot c^\beta \cdot s^\gamma \tag{6}$$

where α, β and γ are positive constants that define the relative importance of each component. The values of l, c and s are computed by Eqs. 7, 8 and 9, respectively, where symbols $\bar{x}$, σ_x, $\bar{y}$, σ_y and σ_{xy} have the same meanings as in UQI. In addition, T_1, T_2 and T_3 are positive constants introduced to avoid instabilities. The authors of the index proposed the following values for these constants [47]: $T_1 = (0.01L)^2$, $T_2 = (0.03L)^2$, $T_3 = T_2/2$, $\alpha = \beta = \gamma = 1$, where L represents the dynamic range of the pixel values and it is set to 255 for 8-bit-per-channel RGB images. SSIM was proposed by the same authors that developed the UQI index. As can be observed, Eq. 4 is a special case of Eq. 6 with $T_1 = T_2 = T_3 = 0$ and $\alpha = \beta = \gamma = 1$.

$$l = \left(\frac{2\bar{x}\bar{y} + T_1}{\bar{x}^2 + \bar{y}_y^2 + T_1} \right) \tag{7}$$

$$c = \left(\frac{2\sigma_x\sigma_y + T_2}{\sigma_x^2 + \sigma_y^2 + T_2} \right) \tag{8}$$

$$s = \left(\frac{\sigma_{xy} + T_3}{\sigma_x\sigma_y + T_3} \right) \tag{9}$$

The authors of the index proposed a mean SSIM index (MSSIM) to assess the overall quality of the entire image [47]. This is computed in the same raster fashion as UQI. Then, the index for the entire image is computed by Eq. 10, where W is the number of windows processed in the image and $SSIM_k$ is the SSIM index computed within the sliding window k.

$$MSSIM = \frac{1}{W} \sum_{k=1}^{W} SSIM_k \tag{10}$$

3.6 Multi-scale structural similarity index

The MS-SSIM [49] is based on SSIM and calculates luminance, contrast, and structure on multiple scales. SSIM is a single-scale index, while MS-SSIM is a multi-scale index based on the idea that the correct scale depends on viewing conditions. For this reason, MS-SSIM simulates different spatial resolutions by iterative downsampling and weighting the values of the three SSIM components at different scales.

MS-SSIM is computed on several scales of the images (Eq. 11). Each scale is obtained by downscaling the original and quantized images by a power of two followed by suitable filtering to avoid aliasing. The original image corresponds to the case with $j = 1$, and the highest scale corresponds to the case with $j = M$. In other words, $j = 1$ represents the original resolution of the image, while M corresponds to the lowest resolution considered (the number of times the image is downsampled). The global error is then obtained by combining the measurements at different scales.

$$MS\text{-}SSIM = (l_M)^{\alpha_M} \cdot \prod_{j=1}^{M} (c_j)^{\beta_j} \cdot (s_j)^{\gamma_j} \tag{11}$$

where l_M denotes the luminance component, which is calculated for the highest scale only, while c_j and s_j represent contrast component and structure component at the j-th scale, respectively. The parameters α_M, β_j, γ_j control the relative importance of each component. The authors recommend $\alpha_M = \beta_j = \gamma_j$ for all j, to simplify parameter selection. They also suggest normalizing the cross-scale settings such that $\sum_{j=1}^{M} \gamma_j = 1$. The best results presented by the authors were obtained for $M = 2$. A value of the index closer to 1 indicates better image quality, whereas a value closer to 0 indicates poorer quality.

3.7 Visual information fidelity

The VIF index is based on the idea that the IQA problem can be viewed as an information fidelity problem [50]. This IQA index quantifies the information present in the original image and how much of this information can be extracted from the quantized image.

VIF uses three models to measure the visual information: the Gaussian scale mixture model, the distortion model, and the HVS model. To calculate the index, the images are first decomposed into several sub-bands and each sub-band is analyzed in blocks. The visual information is measured by computing mutual information in the different models in each block and each sub-band. The final value of the IQA index is computed by integrating visual information for all the blocks and all the sub-bands.

VIF is computed on multiple sub-bands that are completely independent of each other with respect to the parameters of the Gaussian and distortion models. It is computed by Eq. 12, where k is the sub-band index and b is the block index.

$$VIF = \frac{\sum_{k \in sub-band} \sum_b I(C_{kb}, F_{kb})}{\sum_{k \in sub-band} \sum_b I(C_{kb}, E_{kb})} \tag{12}$$

where $I(C_{kb}, E_{kb})$ (Eq. 13) and $I(C_{kb}, F_{kb})$ (Eq. 14) represent the information that could ideally be extracted by the brain from a particular sub-band and block in the original and the quantized images, respectively. C_{kb} represents the original image in a sub-band and block, while E_{kb} and F_{kb} denote the cognitive output of the original and quantized images extracted from the brain, respectively. They are created by applying the HVS model in one sub-band and block to the original and the quantized images, respectively.

$$I(C_{kb}, E_{kb}) = \frac{1}{2} \log_2 \left(1 + \frac{s_{kb}^2 C_U}{\sigma_N^2} \right) \tag{13}$$

$$I(C_{kb}, F_{kb}) = \frac{1}{2} \log_2 \left(1 + \frac{g_{kb}^2 s_{kb}^2 C_U}{\sigma_{v_{kb}}^2 + \sigma_N^2} \right) \tag{14}$$

where σ_N^2, $\sigma_{v_{kb}}^2$, g_{kb}, s_{kb}^2, and C_U must be estimated in advance. The HVS parameter σ_N^2 (the variance of the visual noise) can be adjusted based on trial-and-error to get the optimum VIF estimation, but it can also be set to the constant value 2, as proposed in [50]. On the other hand, $s_{kb}^2 C_U$ can be estimated from the local variance of the original image pixels based on maximum likelihood criteria by Eq. 15, where σ_{kb}^r is the standard deviation of the original image in the block b at the sub-band k.

$$s_{kb}^2 C_U = (\sigma_{kb}^r)^2 \tag{15}$$

The distortion channel parameter g_{kb} and $\sigma_{v_{kb}}^2$ can be estimated using simple linear regression based on the coefficients of the original and quantized image bands by Eqs. 16 and 17, respectively, where σ_{kb}^{rd} is the covariance of the original image and the quantized image in the block b at the subband k.

$$g_{kb} = \frac{\sigma_{kb}^{rd}}{(\sigma_{kb}^r)^2} \tag{16}$$

$$\sigma_{v_{kb}}^2 = (\sigma_{kb}^d)^2 - g_{kb}^2(\sigma_{kb}^r)^2 \tag{17}$$

The index takes values in the interval [0, 1]. The quality of the quantized image is better as the error approaches 1. The value 0 indicates that all information from the original image has been lost, while the value 1 indicates that the images are the same.

3.8 Spectral angle mapper

SAM is an indicator of the spectral quality of the quantized image. It determines the spectral similarity between two spectra (the reference spectrum and the image spectrum) by computing the angle formed between the two spectra, which are treated as vectors in a space of dimensionality equal to the number of bands [56, 57]. SAM is computed by Eq. 18, where X is the test spectrum (i.e., quantized image) and Y the reference spectrum (i.e., original image). The angle α defines the similarity between the two spectra. The operation computes the arccosine of the dot product of the spectra and it can also be expressed as Eq. 19, where nb is the number of spectral bands.

$$\alpha = \cos^{-1}\left(\frac{X \cdot Y}{||X|| \cdot ||Y||}\right) \tag{18}$$

$$\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{nb} X_i Y_i}{\sqrt{\sum_{i=1}^{nb}(X_i)^2 \sum_{i=1}^{nb}(Y_i)^2}}\right) \tag{19}$$

The similarity between the images increases as the angle computed decreases. The angle α presents a variation anywhere between 0 and 90 degrees. If the equation is expressed as $\cos\alpha$, the value of the index ranges in [0, 1] and the best estimate takes values close to 1.

4 Experimental results and discussion

This section presents various results of our experimental study. It has been divided into several subsections to give it a better structure. The first subsection describes the experiments performed and the remaining subsections describe the analysis of the results. First, various box plots are given, which allow analyzing the distribution of the results. Then, the mean values obtained for each CQ method and IQA index are analyzed. Next, a ranking of the methods is established, in order to determine if different IQA indices result in comparable rankings. Finally, statistical tests are performed to support the conclusions presented in the previous subsections.

4.1 Experiments

The experiments were carried out on the images included in the CQ100 dataset [58]. This is a large, diverse, and high-quality color image dataset collected in order to compare CQ methods. The dataset includes 100 RGB images, each with dimensions 768×512 or 512×768. Researchers can download the original images as well as the quantized images obtained by various CQ methods. The MSE values for all quantized images are also publicly available [39]. A table is included, as supplementary material to this article, with the names of the images and the number of distinct colors in each one (Online Resource 1). For thumbnails of the images in the dataset, see [58].

Table 1 lists the IQA indices used in the experiments, together with the ranges of those indices and the best value. It also shows the values of the parameters used in the experiments (whenever parameters are required). For SSIM, the calculated value is the average value obtained by Eq. 10, although SSIM is used to represent that value (instead of MSSIM) for simplicity. In general, the value calculated in the literature is the one obtained with the said equation and is usually referred to as SSIM. It should be noted that MSE and MAE are not necessarily comparable across images, as their values are unnormalized.

The ten CQ methods considered in the experiments are WU, BS, OC, MC, VB, VC, WAN, NQ, WUATCQ, BSITATCQ. These are all deterministic methods, so the same result is obtained when applied to the same original image under identical settings.

The selected CQ methods were executed to generate quantized images with 32, 64, 128 and 256 colors, which are common palette sizes in the CQ literature. Therefore, each CQ method generated 400 quantized images, resulting in a total of 4000 quantized images. Then, the eight IQA indices were calculated for each quantized image. The

Table 1 IQA indices used in the study and their ranges, best possible values, and parameters

IQA index	Range	Best	Parameters
MSE	$[0, 3 \times 255^2]$	0	
MAE	$[0, 3 \times 255]$	0	
PSNR	$[0, 48.13]$	48.13	
UQI	$[-1, 1]$	1	Sliding window size $= 8$
SSIM	$[0, 1]$	1	Sliding window size $= 11$, $T_1 = (0.01L)^2$, $T_2 = (0.03L)^2$, $L = 255$
MS-SSIM	$[0, 1]$	1	Sliding window size $= 11$, #Scales$=5$
			Weights for the scale $= [0.0448, 0.2856, 0.3001, 0.2363, 0.1333]$
			T_1, T_2 and L with the same values defined for SSIM
VIF	$[0, 1]$	1	Variance of the visual noise $= 2$
SAM	$[0, 1]$	0	

IQA index value for each image, CQ method, palette size, and IQA index are included as supplementary material to this article (Online Resource 1). The following subsections discuss the results obtained and summarize the data.

4.2 Box plots

This subsection analyzes the box plots of the results (Figs. 1, 2, 3, 4). Each figure shows the results for a palette size and includes a subfigure for each IQA index. The figures exhibit some clear trends.

MC is the worst method and WAN is the second-worst method. The box corresponding to MC has the worst median, and 25th and 75th percentile in all subfigures. In the case of WAN, these three statistics are worse than the corresponding ones for all methods except MC. On the other hand, the boxes corresponding to the BSITATCQ and WUATCQ methods are very similar in all the subfigures, so the three statistics obtained by these hybrid methods for each IQA index are very similar.

ADU is the best method. In general, the box corresponding to the ADU method is the one with the smallest amplitude (i.e., difference between the 75th and 25th percentiles) for all IQA indices except VIF so its results have the least dispersion. Certainly, ADU always has the box with the smallest amplitude for the MSE, MS-SSIM, and SAM indices, while for the other four indices the box has either the smallest or the second-smallest amplitude. ADU has the best median in most cases, being the second-best in the six cases in which it is not the best. In those six cases, the best median is obtained by BSITATCQ. Finally, it should be noted that in the case of the VIF index, MC has the least dispersion.

Among the four IQA indices for which the best result is 1, UQI and MS-SSIM are the ones that generate larger values and vary in a smaller interval (this especially in the case of MS-SSIM). This makes many of the boxes very small and therefore difficult to differentiate. In any case, it is clearly observed that the results obtained for UQI present an asymmetrical distribution, but this does not occur with the results

of MS-SSIM. Conversely, VIF is the IQA index that generates smaller values and varies over a larger interval.

4.3 Average results

Tables 2 and 3 summarize the results of the tables included in Online Resource 1. They show the average and standard deviation of each IQA index and CQ method for each palette size. The best average value obtained for each IQA index and palette size is underlined. Because a limited number of decimal places can be included in the tables, in some cases the same result is highlighted for multiple methods. In these cases, italics are used to identify the best of all these values (taking into account more decimals than those shown in the table).

It is clearly observed that MC generates the worst average results for all the IQA indices and palette sizes. In addition, it is observed that WAN is the second-worst method. On the contrary, ADU, BSITATCQ, WUATCQ and NQ are the subset of methods attaining the best results.

ADU is the method that generates the overall best results. This method is the best according to SSIM, MS-SSIM and VIF for all palette sizes; it is also the best for three palette sizes when MAE and UQI are considered (32, 64, and 256 colors for MAE; 64, 128, and 256 colors for UQI). On the other hand, MSE indicates that ADU is the best only for the case with 32 colors and the second-best for the other palette sizes. The SAM index values are very similar (even almost the same) for the ADU, WUATCQ and BSITATCQ methods, which are the best methods according to this IQA index. Finally, ADU is the third-best method based on PSNR (for all four palette sizes).

BSITATCQ is the best according to MSE, PSNR and SAM for palettes with more than 32 colors; it is the second-best based on PSNR and SAM for 32 colors, and the third-best based on MSE for 32 colors. Based on VIF this is the second-best method for 32 colors, and the third-best method for the other palette sizes. MAE indicates that this is the third-best method for all palette sizes. For the other three
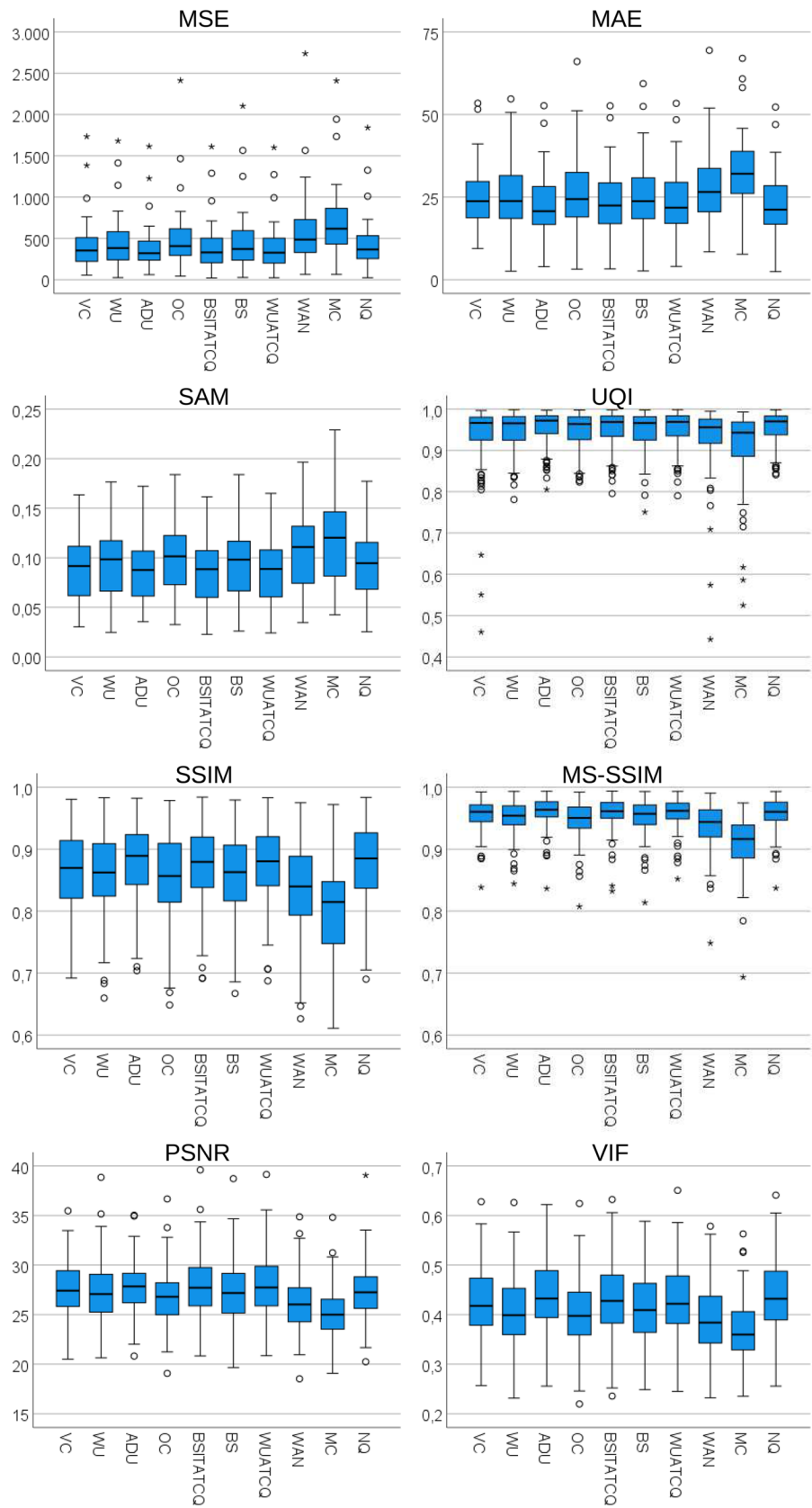
Fig. 1 Box plots for 32-color images
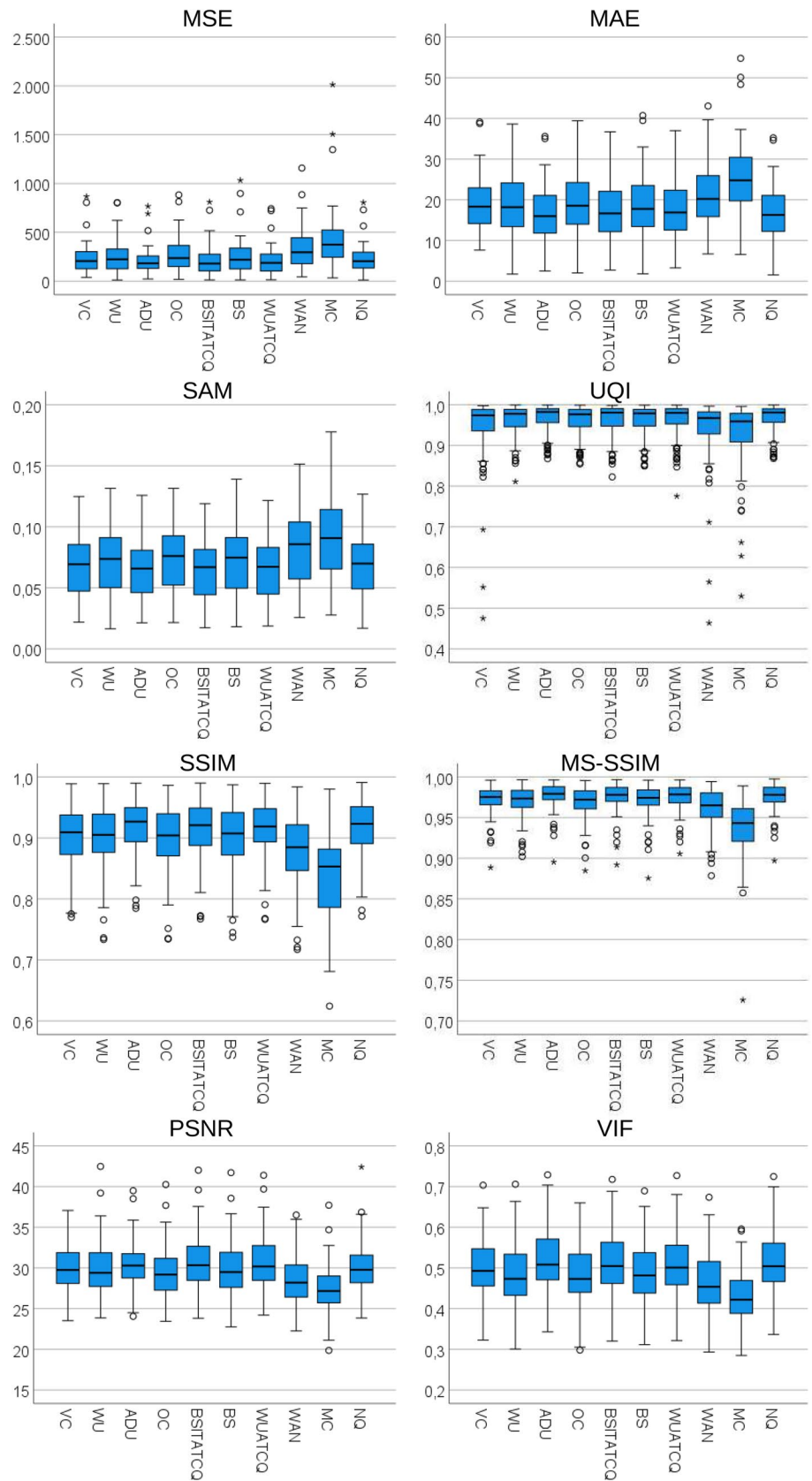
Fig. 2 Box plots for 64-color images
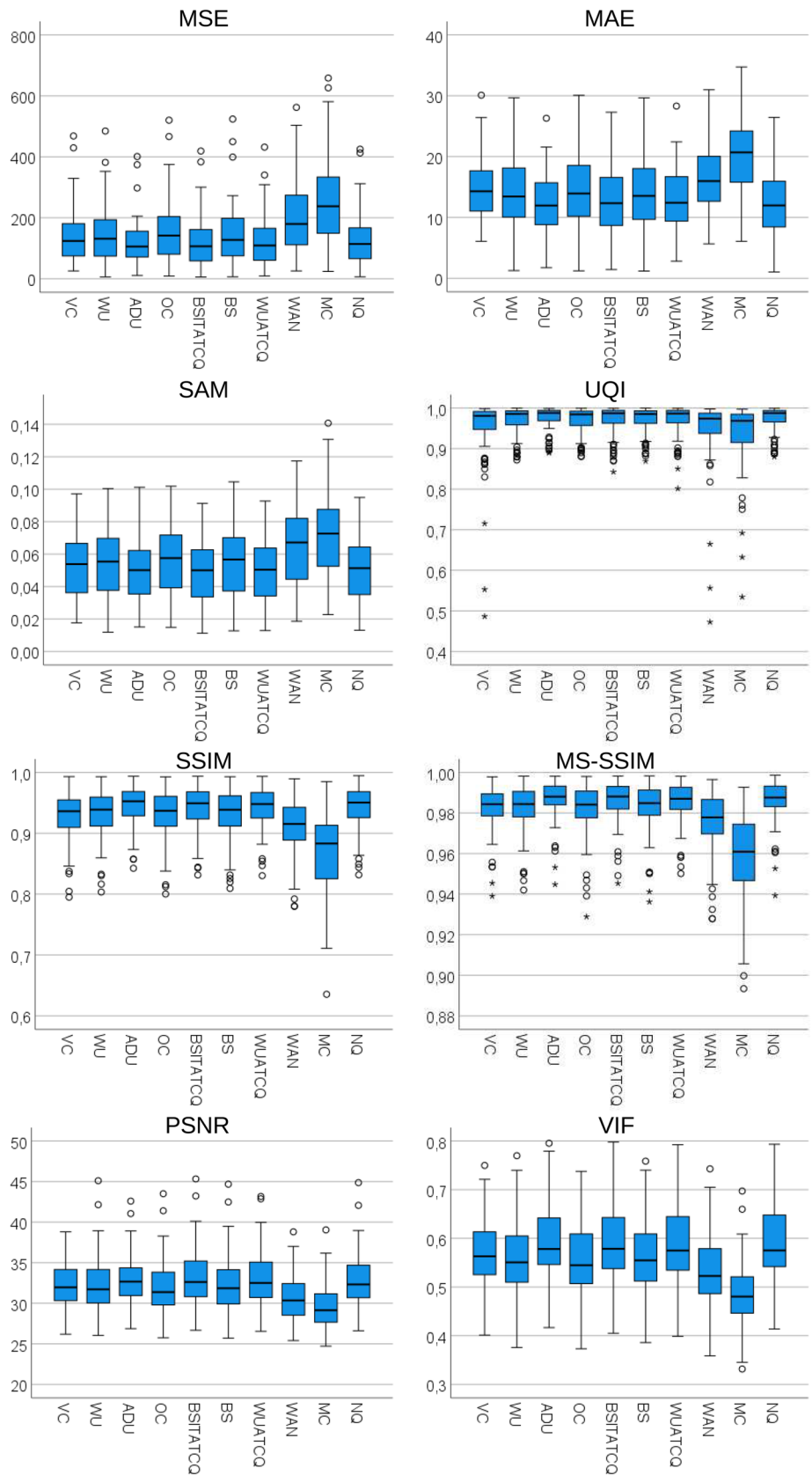
Fig. 3 Box plots for 128-color images

Fig. 4 Box plots for 256-color images
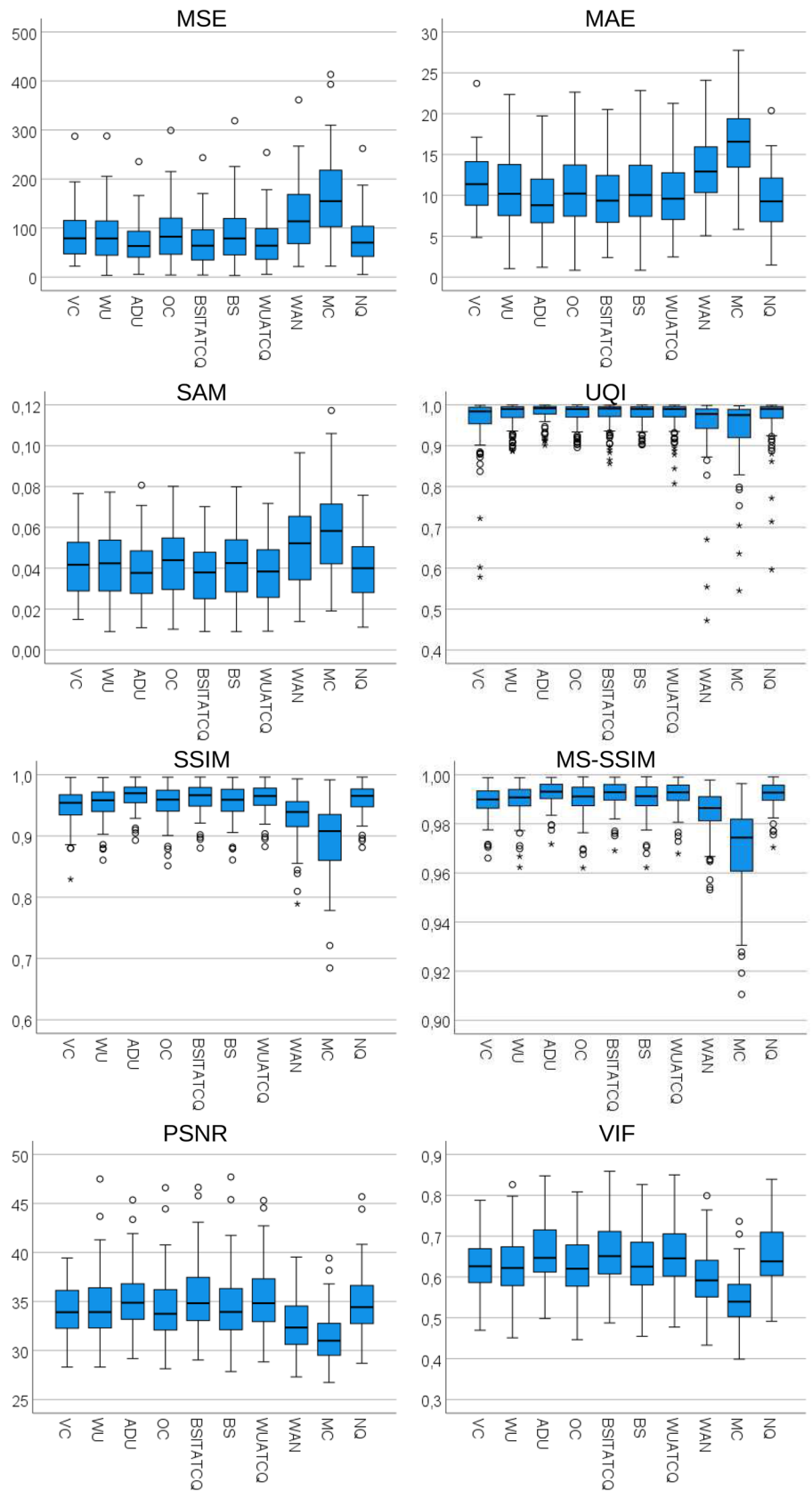
Table 2 Average and standard deviation of the IQA indices for each CQ method—32 and 64 colors

IQA		BS	WU	OC	ADU	VC	WAN	MC	NQ	WUAT.	BSITAT.
		32 colors									
MSE	av	435.65	425.64	469.55	<u>357.36</u>	385.96	542.58	666.64	408.94	360.19	362.10
	de	296.59	261.12	297.89	217.93	245.21	340.90	361.53	246.03	235.09	233.72
MAE	av	24.614	24.766	25.795	<u>22.142</u>	24.283	27.433	32.366	22.293	23.093	23.029
	de	9.228	6.379	9.013	8.050	7.940	9.432	9.550	8.105	8.463	8.358
PSNR	av	27.378	27.420	26.895	28.033	27.773	26.272	25.283	27.482	<u>28.190</u>	28.136
	de	2.890	2.856	2.594	2.453	2.590	2.591	2.459	2.620	2.898	2.845
UQI	av	0.946	0.946	0.946	0.955	0.936	0.929	0.913	<u>0.956</u>	0.951	0.951
	de	0.051	0.049	0.045	0.041	0.084	0.082	0.087	0.039	0.046	0.046
SSIM	av	0.858	0.856	0.854	<u>0.881</u>	0.866	0.834	0.802	0.879	0.873	0.873
	de	0.065	0.065	0.066	0.057	0.060	0.071	0.076	0.058	0.060	0.060
MS-SSIM	av	0.952	0.950	0.947	<u>0.960</u>	0.956	0.938	0.910	0.957	0.958	0.957
	de	0.029	0.028	0.030	0.025	0.026	0.037	0.044	0.025	0.025	0.027
VIF	av	0.414	0.408	0.406	<u>0.441</u>	0.427	0.393	0.368	0.439	0.430	0.433
	de	0.074	0.075	0.072	0.074	0.074	0.072	0.064	0.076	0.077	0.077
SAM	av	0.095	0.095	0.100	<u>0.087</u>	0.090	0.107	0.118	0.094	<u>*0.087*</u>	<u>0.087</u>
	de	0.035	0.034	0.034	0.030	0.032	0.037	0.041	0.033	0.032	0.031
		64 colors									
MSE	av	244.53	238.67	260.64	199.70	222.16	317.40	417.12	221.99	201.03	<u>199.65</u>
	de	163.04	144.20	147.91	117.64	134.84	178.81	277.47	128.36	127.63	128.90
MAE	av	18.205	18.414	18.961	<u>16.273</u>	18.477	20.853	25.378	16.306	17.138	16.940
	de	6.982	6.810	6.952	6.002	5.908	6.812	7.768	6.062	6.342	6.306
PSNR	av	29.951	29.998	29.485	30.610	30.153	28.595	27.464	30.186	30.777	<u>30.801</u>
	de	3.066	3.046	2.757	2.630	2.574	2.640	2.672	2.779	3.039	3.023
UQI	av	0.962	0.960	0.961	<u>0.968</u>	0.946	0.941	0.926	0.967	0.962	0.962
	de	0.039	0.040	0.037	0.032	0.081	0.079	0.083	0.034	0.042	0.040
SSIM	av	0.901	0.899	0.898	<u>0.920</u>	0.902	0.878	0.840	0.917	0.913	0.913
	de	0.052	0.052	0.052	0.043	0.048	0.056	0.070	0.044	0.046	0.046
MS-SSIM	av	0.971	0.970	0.969	<u>0.977</u>	0.972	0.961	0.937	0.976	0.975	0.975
	de	0.019	0.019	0.019	0.016	0.017	0.023	0.037	0.016	0.016	0.017
VIF	av	0.490	0.485	0.481	<u>0.518</u>	0.500	0.465	0.429	0.516	0.509	0.514
	de	0.078	0.079	0.074	0.076	0.074	0.076	0.064	0.078	0.079	0.079
SAM	av	0.071	0.071	0.074	0.065	0.068	0.082	0.092	0.069	0.065	<u>0.064</u>
	de	0.027	0.026	0.026	0.023	0.025	0.030	0.033	0.025	0.024	0.024

IQA indices, BSITATCQ obtains results very similar to WUATCQ; the two methods share the second or third rank in many cases. The only case in which BSITATCQ is below the third rank corresponds to the UQI index and palette size of 256, where it ranks fourth.

NQ ranks between second and fourth in many cases. It is the best method only based on MAE for 128 colors and UQI for 32 colors. It is the second-best based on MAE and for the other palette sizes, and also based on UQI for 64 and 128 colors, but the seventh-best based on this IQA index for 256 colors. In addition, it is the second-best method for all palette sizes when considering SSIM and for three palette sizes when considering MS-SSIM and VIF (64, 128, and 256 colors for MS-SSIM; 32, 64, and 128 colors for VIF). Indeed, NQ ranks the same as WUATCQ or BSITATCQ in some cases when considering the SSIM and MS-SSIM indices. The MSE and PSNR indices assign the worst value to this method (it ranks fifth for 32 colors and fourth for the rest of the palette sizes). Based on the SAM index, NQ is the second-best method for 128 colors (tied with ADU) and ranks third or fourth for the rest of the palette sizes.

WUATCQ oscillates between the second and fourth ranks in most cases, since it is the best method only based on PSNR and SAM for 32 colors. It is the second-best method based on PSNR and SAM for palettes with more than 32 colors. It should be noted that based on the SAM index, this method generates results comparable to ADU and BSITATCQ. On the other hand, WUATCQ is the second-best method based on the MSE index for 32 colors and the third-best for the other palette sizes. It is the third-best method

Table 3 Average and standard deviation of the IQA indices for each CQ method—128 and 256 colors

IQA		BS	WU	OC	ADU	VC	WAN	MC	NQ	WUAT.	BSITAT.
		128 colors									
MSE	av	140.70	137.86	149.38	115.37	133.64	193.93	248.93	120.42	115.86	<u>113.50</u>
	de	89.89	82.57	87.02	66.18	75.70	104.05	124.42	72.60	72.17	72.03
MAE	av	13.680	13.875	14.132	12.182	14.391	16.310	20.106	<u>12.145</u>	12.883	12.674
	de	5.239	5.080	5.301	4.500	4.339	4.942	5.540	4.611	4.741	4.756
PSNR	av	32.401	32.416	31.965	33.022	32.308	30.721	29.557	32.944	33.202	<u>33.323</u>
	de	3.233	3.138	2.909	2.730	2.486	2.625	2.495	2.984	3.125	3.213
UQI	av	0.971	0.970	0.970	<u>0.976</u>	0.954	0.948	0.936	0.974	0.970	0.970
	de	0.032	0.034	0.031	0.027	0.078	0.079	0.079	0.030	0.038	0.035
SSIM	av	0.932	0.931	0.930	<u>0.946</u>	0.928	0.910	0.871	0.944	0.941	0.941
	de	0.039	0.038	0.039	0.031	0.038	0.045	0.063	0.033	0.033	0.034
MS-	av	0.983	0.982	0.982	<u>0.987</u>	0.983	0.976	0.958	0.986	0.985	0.986
SSIM	de	0.012	0.011	0.013	0.009	0.011	0.014	0.022	0.010	0.010	0.010
VIF	av	0.562	0.558	0.555	<u>0.592</u>	0.567	0.533	0.486	0.590	0.584	0.589
	de	0.080	0.078	0.076	0.075	0.071	0.076	0.065	0.077	0.079	0.079
SAM	av	0.054	0.054	0.056	0.050	0.053	0.064	0.073	0.050	<u>0.049</u>	<u>*0.049*</u>
	de	0.021	0.020	0.020	0.018	0.019	0.024	0.026	0.019	0.019	0.019
		256 colors									
MSE	av	82.32	80.72	86.03	67.36	83.11	123.57	165.19	73.50	67.95	<u>65.96</u>
	de	51.14	47.24	48.77	36.66	42.75	64.49	79.81	41.71	41.19	39.82
MAE	av	10.375	10.625	10.618	<u>9.188</u>	11.489	13.131	16.366	9.368	9.856	9.658
	de	4.021	3.834	4.036	3.457	3.186	3.645	4.265	3.504	3.603	3.561
PSNR	av	34.754	34.718	34.406	35.352	34.282	32.635	31.307	35.049	35.507	<u>35.666</u>
	de	3.340	3.102	3.058	2.767	2.314	2.512	2.421	2.924	3.118	3.221
UQI	av	0.979	0.976	0.977	<u>0.982</u>	0.960	0.952	0.944	0.969	0.975	0.976
	de	0.025	0.030	0.027	0.023	0.069	0.079	0.077	0.058	0.035	0.032
SSIM	av	0.955	0.953	0.954	<u>0.965</u>	0.947	0.932	0.895	0.961	0.961	0.961
	de	0.027	0.026	0.028	0.020	0.028	0.037	0.055	0.023	0.022	0.023
MS-	av	0.990	0.990	0.990	<u>0.993</u>	0.989	0.985	0.969	0.992	0.992	0.992
SSIM	de	0.007	0.007	0.007	0.005	0.006	0.009	0.018	0.005	0.006	0.005
VIF	av	0.631	0.626	0.627	<u>0.660</u>	0.628	0.597	0.545	0.654	0.652	0.659
	de	0.078	0.073	0.075	0.071	0.066	0.072	0.063	0.073	0.075	0.076
SAM	av	0.042	0.041	0.043	0.038	0.042	0.051	0.059	0.040	0.038	<u>0.037</u>
	de	0.017	0.016	0.016	0.014	0.015	0.019	0.022	0.015	0.015	0.015

according to SSIM and the fourth-best method according to MAE and VIF. MS-SSIM indicates that this method is the fourth-best for 128 colors and the second-best for all other palette sizes. Finally, UQI indicates that this method is the third-best for palettes with less than 256 colors, but sharing this rank with other methods (BSITATCQ in all cases and BS, WU and OC in one case). This IQA index gives the worst result for 256 colors, since it indicates that WUATCQ is the sixth-best method.

Based on the above analysis, we can conclude that the quality assessment obtained when analyzing the MSE index does not coincide with that of the rest of the IQA indices. Next we will analyze a concrete example to show the differences. The detailed results included in Online

Resource 1 indicate that the best MSE value (over the 100 images) is obtained for *shopping_bags* image reduced to 256 colors by the BS method (*MSE* = 3.308). Figures 5 and 6 show the quantized image with 256 colors obtained by the ten CQ methods. Each subfigure shows the values obtained for each of the eight IQA indices; the best value of each IQA index appears in bold. It can be seen that BS generates the best image according to MSE, PSNR, SAM, MS-SSIM and UQI. On the other hand, according to SSIM, VIF and MAE, the best image is generated by three different methods (NQ, BSITATCQ and OC, respectively). Therefore, the best value of all IQA indices is not obtained by BS.

(a) BS (MSE: **3.31**, MAE: 0.84, PSNR: **47.71**, SAM: **0.0090**, SSIM: 0.9959, MS-SSIM: **0.9992**, VIF: 0.8069, UQI: **0.9998**)

(b) WU (MSE: 3.47, MAE: 1.04, PSNR: 47.50, SAM: 0.0090, SSIM: 0.9950, MS-SSIM: 0.9988, VIF: 0.7983, UQI: 0.9998)

(c) OC (MSE: 4.26, MAE: **0.83**, PSNR: 46.61, SAM: 0.0102, SSIM: 0.9961, MS-SSIM: 0.9992, VIF: 0.7864, UQI: 0.9998)

(d) ADU (MSE: 8.97, MAE: 1.20, PSNR: 43.37, SAM: 0.0150, SSIM: 0.9932, MS-SSIM: 0.9984, VIF: 0.7107, UQI: 0.9996)

(e) VC (MSE: 32.55, MAE: 8.45, PSNR: 37.78, SAM: 0.0213, SSIM: 0.9826, MS-SSIM: 0.9957, VIF: 0.7753, UQI: 0.9908)

(f) WAN (MSE: 37.65, MAE:9.42, PSNR: 37.14, SAM: 0.0230, SSIM: 0.9839, MS-SSIM: 0.9965, VIF: 0.7643, UQI: 0.9900)

Fig. 5 shopping_bags image with 256 colors

4.4 Ranking of the methods

To better understand the global results obtained, a ranking can be established to order the methods, taking into account the objective quality of the images generated by each of them. For this, the eight IQA indices calculated for the quantized images generated by each CQ method were analyzed independently. Furthermore, the results obtained for

(a) MC (MSE: 22.25, MAE: 5.84, PSNR: 39.43, SAM: 0.0218, SSIM: 0.9836, MS-SSIM: 0.9899, VIF: 0.7052, UQI: 0.9964)

(b) NQ (MSE: 5.26, MAE: 1.48, PSNR: 45.69, SAM: 0.0118, SSIM: **0.9962**, MS-SSIM: 0.9992, VIF: 0.7946, UQI: 0.9991)

(c) WUATCQ (MSE: 5.74, MAE: 2.47, PSNR: 45.31, SAM: 0.0110, SSIM: 0.9949, MS-SSIM: 0.9987, VIF: 0.7906, UQI: 0.9997)

(d) BSITATCQ (MSE: 4.22, MAE: 2.39, PSNR: 46.65, SAM: 0.0094, SSIM: 0.9954, MS-SSIM: 0.9989, VIF: **0.8244**, UQI: 0.9991)

Fig. 6 shopping_bags image with 256 colors

each palette size were considered independently. An index between 1 and 10 was assigned to the results generated by the ten CQ methods for an image, where 1 represents the method that produces the best result and 10 the method that produces the worst result. In the event that several methods generate exactly the same result for an image, they are all assigned the same index (for example, if two methods generate the best result in the set, they are both assigned an index of 1, and the second-best method is then assigned an index of 3). This process was applied for each of the 100 images.

Table 4 shows the average rank obtained for the set of 100 images considering each palette size independently and also the average value for the four palette sizes. The table does not show integers between 1 and 10, but real values, with the aim of being able to determine if multiple methods obtain a close value in the ranking. An attempt has been made to place the columns of the table in ascending order, so that the interpretation of the results is easier. Figure 7 shows the rank of each method for each IQA index and palette size. Ten

integer values are used in the subfigures to plot the results, so that the order assigned to the ten CQ methods can be seen more clearly. In addition, Fig. 8 shows the overall results for each IQA index and each CQ method, without separating the results by the palette size.

Table 4 shows that BSITATCQ is the best method according to MSE, PSNR and SAM, but ADU is the best method according to the other five IQA indices. The values for ADU and BSITATCQ are very close with respect to MSE and PSNR for 32 and 64 colors, but the difference increases for SAM for all the palette sizes.

Based on the MSE and PSNR values, the second-best method is ADU for the two largest palettes; however, WUATCQ is the second-best method for the two smallest palettes. WUATCQ is also the second-best according to SAM and UQI. On the other hand, NQ is the second-best method according to MAE, SSIM, and VIF. In the case of MS-SSIM, WUATCQ is the second-best for 32 colors, but BSITATCQ is the second-best for all other palette sizes.

Table 4 Ranking of each CQ method for each IQA index (c: number of distinct colors. The label 'all' represents the average of the four ranks corresponding to 32, 64, 128, and 256 colors)

IQA index	l	ADU	BSITAT.	WUAT.	NQ	VC	WU	BS	OC	WAN	MC
MSE	32	2.36	2.23	2.26	5.49	4.14	5.93	6.09	7.64	8.94	9.92
	64	2.32	2.11	2.25	4.91	4.82	5.92	6.05	7.70	8.97	9.95
	128	2.51	1.81	2.60	3.74	5.62	6.01	6.11	7.65	9.02	9.93
	256	2.45	1.54	2.68	4.25	6.23	5.72	5.84	7.37	8.99	9.93
	all	2.41	1.92	2.45	4.60	5.20	5.90	6.02	7.59	8.98	9.93
MAE	32	1.62	3.44	3.54	2.01	5.66	6.32	5.99	7.58	8.88	9.96
	64	1.61	3.25	3.74	1.89	6.13	6.34	5.78	7.38	8.91	9.97
	128	1.74	3.18	3.87	1.63	6.55	6.37	5.57	7.20	8.94	9.95
	256	1.41	3.13	4.04	2.16	7.20	6.47	5.19	6.53	8.94	9.93
	all	1.60	3.25	3.80	1.92	6.39	6.38	5.63	7.17	8.92	9.95
PSNR	32	2.36	2.23	2.26	5.49	4.14	5.93	6.09	7.64	8.94	9.92
	64	2.32	2.11	2.25	4.91	4.82	5.92	6.05	7.70	8.97	9.95
	128	2.51	1.81	2.60	3.74	5.62	6.01	6.11	7.65	9.02	9.93
	256	2.45	1.54	2.68	4.25	6.23	5.72	5.84	7.37	8.99	9.93
	all	2.41	1.92	2.45	4.60	5.20	5.90	6.02	7.59	8.98	9.93
UQI	32	2.49	3.77	3.12	3.53	5.37	5.96	5.39	6.97	8.64	9.76
	64	2.25	3.68	3.06	3.42	6.50	5.69	5.29	6.57	8.72	9.82
	128	2.20	3.58	3.28	3.35	6.94	5.71	5.14	6.16	8.87	9.76
	256	2.10	3.58	3.25	4.24	7.34	5.65	4.56	5.70	8.97	9.61
	all	2.26	3.65	3.18	3.64	6.54	5.75	5.10	6.35	8.80	9.74
SIM	32	1.88	3.49	3.59	2.44	4.45	6.89	6.33	7.11	8.95	9.87
	64	1.60	3.27	3.42	2.46	5.41	6.66	6.29	7.00	8.97	9.92
	128	1.56	3.19	3.53	2.38	6.17	6.49	6.04	6.73	8.96	9.95
	256	1.46	2.96	3.32	3.11	6.71	6.62	5.81	6.10	8.96	9.95
	all	1.63	3.23	3.47	2.60	5.69	6.67	6.12	6.74	8.96	9.92
MS-SSIM	32	2.01	3.25	3.16	3.51	4.40	6.62	6.01	7.31	8.74	9.99
	64	1.80	2.94	3.16	3.00	5.35	6.61	6.08	7.19	8.90	9.97
	128	1.81	2.90	3.42	2.41	6.03	6.78	5.92	6.84	8.94	9.95
	256	1.65	2.67	3.45	2.82	6.74	6.90	5.75	6.09	8.96	9.97
	all	1.82	2.94	3.30	2.94	5.63	6.73	5.94	6.86	8.89	9.97
VIF	32	1.95	3.25	3.85	2.40	4.39	7.03	6.29	7.32	8.74	9.78
	64	1.89	2.65	3.61	2.70	4.96	6.89	6.27	7.26	8.88	9.89
	128	1.84	2.55	3.55	2.48	5.41	6.80	6.27	7.26	8.89	9.95
	256	1.81	2.11	3.55	2.97	5.98	6.84	6.05	6.78	8.94	9.97
	all	1.87	2.64	3.64	2.64	5.19	6.89	6.22	7.16	8.86	9.90
SAM	32	2.51	2.15	2.25	5.51	3.98	6.02	6.12	7.69	8.92	9.85
	64	2.46	2.06	2.19	5.01	4.54	5.96	6.13	7.74	8.99	9.91
	128	2.62	1.69	2.57	3.83	5.44	6.09	6.10	7.68	9.05	9.93
	256	2.61	1.52	2.56	4.28	6.03	5.80	5.83	7.41	9.03	9.93
	all	2.55	1.86	2.39	4.66	5.00	5.97	6.05	7.63	9.00	9.91

For cases (palette sizes and IQA indices) not listed above, the third rank is attained by one of the following methods: ADU, BSITATCQ, WUATCQ, and NQ.

For all the IQA indices, MC is the worst method and WAN is the second-worst. In addition, OC is the third-worst method for all IQA indices but UQI, in which case it is the seventh method. On the other hand, for 256 colors, OC achieves results similar to those of WU for MAE, UQI, SSIM, and VIF.

Table 4 shows that in some cases the ranks obtained by various methods are very similar. For example, the second-best and third-best methods identified by SAM (WUATCQ and ADU, respectively) have ranks that exceed 2 but do not reach 3. In order to show a summary of the ranking of CQ methods for each IQA index showing a clear order for the ten

Fig. 7 Rank of each CQ method for each IQA index (lower ranks are better)—results for each palette size
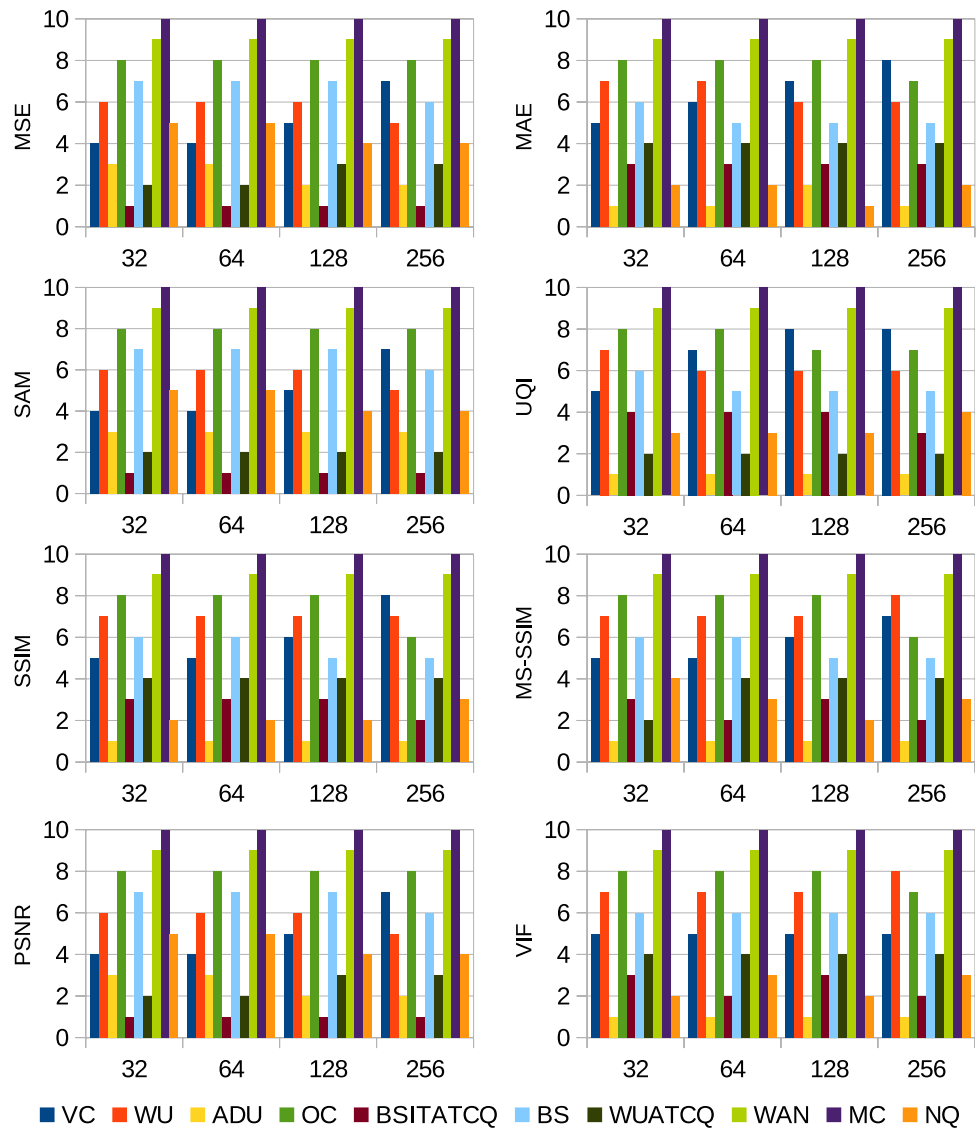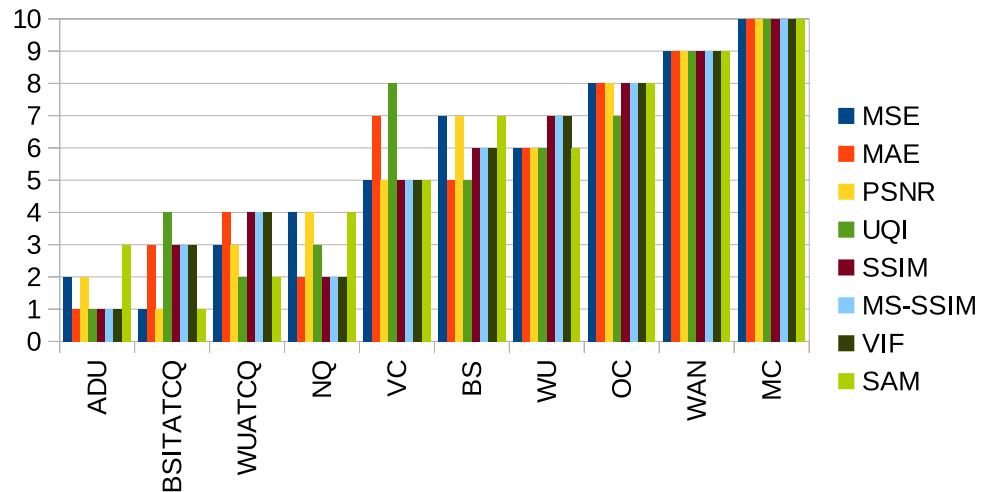

Fig. 8 Rank of each CQ method for each IQA index (lower ranks are better)

CQ methods, Table 5 shows the ranking obtained considering the results of the 400 images generated by each method as a single set (that is, without grouping the images by palette size). The methods that rank second and third according to VIF attain the same rank (2.64) in Table 4. NQ was assigned the best of both ranks because its rank is slightly lower (2.6375) than that of the other method (2.6400). In any case, for all practical purposes, we can consider both methods to have the same rank. The same is true for the same methods with respect to the MS-SSIM index. Table 5 includes superscripts to mark the methods that attain similar ranks with respect to a certain IQA index, so they could have been assigned the same rank in the said ranking. Table 5 and Fig. 8 summarize the classification of the CQ methods obtained for each IQA index analyzed.

The analysis of Table 5, Fig. 8, and the results included in the rows labeled 'all' in Table 4 allows us to interpret the final ranking and establish a clear order among the ten CQ methods analyzed:

- It is possible to identify two groups of IQA indices that generate the same ranking for the CQ methods: a group made up of SSIM, MS-SSIM, and VIF (we will call it group A), and a second group made up of MSE, PSNR, and SAM (we will call it group B). It can be observed that the methods that rank second and third according to the group-B indices could be interchanged, since the corresponding ranks are very similar.

    The rankings obtained for the other two IQA indices (MAE and UQI) are not directly comparable with any other, except for the two worst methods, which are the same for all IQA indices. The ranking defined for MAE is the same as for group-A indices except for ranks 5, 6, and 7. Although the ranking defined according to UQI shows more differences with the other IQA indices, the subset that defines the four best methods includes the same cases as for the other IQA indices.

- ADU, BSITATCQ, WUATCQ and NQ attain the best four ranks for all the IQA indices.

    ADU is the best method according to group A, but the second or third according to group B. BSITATCQ is the best method according to group B, but the third according to group A. The second rank according to group A corresponds to NQ, but this method ranks fourth according to group B. On the other hand, WUATCQ ranks better according to the group-B indices (second and third) than the group-A indices (fourth). It is observed that ADU and BSITATCQ obtain very similar results for the IQA indices in group B, so we can consider both methods to have the same rank.

    Although both MAE and UQI indicate that ADU is the best method, the results of these two IQA indices for the next three ranks are not the same. Certainly, MAE generates the same results as group-A indices for the top four ranks.

- VC ranks the same (fifth) according to the IQA indices in groups A and B but it ranks worse according to MAE and UQI (between sixth and seventh).

- BS and WU share ranks sixth and seventh, respectively, according to the IQA indices in groups A and B. In the case of group B, it is observed that both methods generate very similar results and attain the same rank. In the case of group A, BS is better than WU and also for two of the IQA indices in this group (SSIM and MS-SSIM), the rank of WU is comparable to that of OC.

- OC ranks eighth according to all IQA indices except UQI, according to which it ranks seventh although with a result very similar to that of the method that ranks eighth (VC).

- MC and WAN attain the worst ranks (tenth and ninth, respectively) according to all the IQA indices.

Table 5 Ranking of methods for each IQA index (lower ranks are better)

rank	MSE	PSNR	SAM	MAE	SSIM	MS-SSIM	VIF	UQI
1	BSITAT.	BSITAT.	BSITAT.	ADU	ADU	ADU	ADU	ADU
2	ADU[1]	ADU[1]	WUAT.[4]	NQ	NQ	NQ[0]	NQ[0]	WUAT.
3	WUAT.[1]	WUAT.[1]	ADU[4]	BSITAT.	BSITAT.	BSITAT.[0]	BSITAT.[0]	NQ[1]
4	NQ	NQ	NQ	WUAT.	WUAT.	WUAT.	WUAT.	BSITAT.[1]
5	VC	VC	VC	BS	VC	VC	VC	BS
6	WU[3]	WU[3]	WU[2]	WU[1]	BS	BS	BS	WU
7	BS[3]	BS[3]	BS[2]	VC[1]	WU[2]	WU[3]	WU	OC[4]
8	OC	OC	OC	OC	OC[2]	OC[3]	OC	VC[4]
9	WAN	WAN	WAN	WAN	WAN	WAN	WAN	WAN
10	MC	MC	MC	MC	MC	MC	MC	MC

A superscript is used to mark cases with similar ranking (0: same value, 1: difference less than 0.05, 2: difference less than 0.10, 3: difference less than 0.15, 4: difference less than 0.20)

4.5 Statistical analysis

A statistical analysis was performed to complete the discussion of the results. The Friedman test was executed to determine if the applied CQ method has an effect on the objective quality of the quantized image. This test makes it possible to compare several related samples, indicating whether there are significant differences between any of the pairs of the samples. The test was applied independently to the results obtained for each of the eight IQA indices. The *p*-value used for the test was 0.05. Table 6 shows the test statistic obtained in each case. The significance obtained in all cases is 0, so it is not included in the table. This value indicates that the Friedman test is significant in all eight cases.

The Friedman test indicates that the quality of the quantized image obtained is not the same for the ten CQ methods compared. This statement is true for any of the eight IQA indices considered. This means that at least two of the methods compared are significantly different, but the test does not indicate which ones. Therefore, it is necessary to carry out an additional test to determine which methods present differences. For this purpose, the Wilcoxon post-hoc test was conducted, with the Bonferroni correction to control the probability of occurrence of a type I error. Since we have 10 CQ methods, it was necessary to compare 45 pairs of CQ methods for each IQA index. The test was significant for most of the 360 total cases that were evaluated. Therefore, to reduce the table size, only the corrected p-values associated with non-significant cases were included (Table 7). It can be seen that, for each IQA index, the differences are not

significant for a number of method pairs ranging between 3 and 6 cases. The smallest value corresponds to VIF and MAE. Rows 2–4 show cases for which only one IQA index shows no significant differences between two methods. Rows 5–6 show pairs of methods with non-significant differences only regarding two IQA indices.

The differences between WAN and all other CQ methods are significant with respect to all IQA indices. The same is true for MC. When considering ADU, the differences are significant compared to all methods except BSITATCQ and WUATCQ regarding MSE, PSNR, and SAM, and compared to NQ regarding MAE. It can be observed that the differences are not significant for BSITATCQ and WUATCQ regarding all IQA indices except VIF. The same is true for BS and WU except for MS-SSIM and MAE indices.

UQI, SSIM, MS-SSIM, and VIF do not show significant differences for the pairs of methods WU-OC and NQ-BSITATCQ, although the other IQA indices show the contrary. In addition, there are no significant differences between the methods for the pairs ADU-WUATCQ, ADU-BSITATCQ, and VC-NQ regarding MSE, PSNR, and SAM, but there are differences regarding the other five IQA indices. Regarding MSE and PSNR, the cases that show no significant differences correspond to the same pairs of CQ methods. On the other hand, the results of SAM are the same as those of MSE and PSNR except for the VC-WU pair.

We can see that the results of the statistical test for the IQA indices in group B defined in the previous section are very similar. The only difference is in the pair of methods VC-WU according to the SAM index. Regarding the indices

Table 6 Friedman test results (400 valid cases, 9 degrees of freedom, and p-value 0 for all tests)

MSE	MAE	PSNR	SAM	SSIM	MS-SSIM	VIF	UQI
3016.231	3136.419	3016.231	3024.308	2948.671	2903.731	3007.637	2342.081

Table 7 Corrected significance of the post-hoc test (only the pairs of methods that generate some non-significant result are given)

	MSE	PSNR	SAM	UQI	SSIM	MS-SSIM	VIF	MAE
NQ-ADU								1
OC-BS					0.177			
OC-VC				1				
BS-VC					1	1		
NQ-WUATCQ				1		1		
ADU-WUATCQ	1	1	1					
ADU-BSITATCQ	1	1	0.053					
VC-NQ	0.212	0.212	1					
VC-WU	0.055	0.055						1
WU-OC				0.237	1	1	1	
NQ-BSITATCQ				1	0.146	1	1	
BS-WU	1	1	1	0.096	0.475		0.079	
WUATCQ-BSITATCQ	0.639	0.639	0.542	1	1	1		0.475

in group A, the results are not so homogeneous, as can be seen in Table 7.

5 Conclusion

MSE is the most widely used IQA index in the CQ literature. MSE and its close relatives MAE and PSNR are conventional IQA indices that are calculated based on a pixel-by-pixel comparison between the original and quantized images. However, there are more recent IQA indices that take into account the HVS characteristics. Therefore, it is interesting to determine if the conventional IQA indices generate comparable results with those generated by these new IQA indices or if, on the contrary, it is necessary to combine several IQA indices for a more comprehensive analysis of CQ methods. With this purpose, this study compares images generated by ten CQ methods (WU, BS, OC, MC, VB, VC, WAN, NQ, WUATCQ, and BSITATCQ) using eight IQA indices (MSE, MAE, PSNR, UQI, SSIM, MS-SSIM, VIF, and SAM).

The results obtained indicate that the IQA index used determines the ranking of the CQ methods. It was observed that the results of the MSE index are comparable to those obtained with PSNR and SAM. On the other hand, the results obtained by the SSIM, MS-SSIM, and VIF indices are also comparable. However, the results of these two sets of IQA indices are different. On the other hand, the results obtained for MAE are comparable to a certain extent to those obtained by SSIM, MS-SSIM, and VIF, mainly with regard to the best and worst CQ methods determined by each IQA index. Finally, the results generated by UQI are not directly comparable to those of other IQA indices.

As a result of the analysis carried out it is recommended that the CQ studies not only include results from MSE, but also from other IQA indices that take into account the visual characteristics of the images, such as SSIM, MS-SSIM, and VIF. It is important to use several IQA indices to determine if a given CQ method produces better results than the others.

Unlike what happens in other articles that analyze CQ methods, the results included in this article are calculated for a large set of quantized images (400 quantized images generated by each CQ method) and various IQA indices are used that take into account different characteristics of the images. The results generated in this study will be published, together with the original image set, in the upcoming version of CQ100. This information will help other researchers to compare a new CQ method with the ten CQ methods considered in this article. Since the selected IQA indices are popular in the image processing literature, other authors can use the publicly available implementations to compute results for the new method and compare them to the results included in CQ100. This will speed up and simplify the work of other researchers in the comparison process.

MSE, MAE, and PSNR compare the original and quantized images pixelwise and do not take the HVS into account, so they do not always agree with human perception. It is reasonable to expect the results of PSNR to be comparable to those of MSE, since PSNR is calculated based on MSE. Although both MSE and MAE compare images pixelwise, MAE is less sensitive to outliers than MSE. The MSE calculation emphasizes large errors, while small errors have little effect. This is probably why the results of the two indices do not agree.

Unlike the aforementioned pixelwise indices, UQI, SSIM, and MS-SSIM consider changes in structural information taking into account three factors (loss of correlation, luminance distortion, and contrast distortion). Various studies show that SSIM performs remarkably well across a wide variety of image and distortion types, and the scores of this index are much more consistent with human perception than the MSE scores [50]. The SSIM index was proposed by the same authors as the UQI index as an improvement on the latter, and our study shows that these two indices do not generate comparable results when applied to the CQ problem. On the other hand, several studies indicate that the multi-scale MS-SSIM performs better than the single-scale SSIM [49, 59]. This seems logical since images contain structures that occur over a range of spatial scales. In addition, the HVS decomposes visual data into multi-scale spatial channels in the early stages of image perception [59]. On the other hand, VIF is computed based on a statistical model for natural scenes, a model for image distortions, and an HVS model, which probably influences its greater correlation with subjective assessment scores. Several studies indicate that VIF generates better results than some of the other indices considered in our study, including UQI, SSIM, MS-SSIM, and PSNR [50–52].

Our experiments do not allow us to determine which of the IQA indices is the best for CQ. As discussed in the introduction, determining the best IQA index for a specific image processing operation is non-trivial, and thus studies carried out by different researchers present mixed results. If we limit our attention to the CQ operation considered in this study, we can compare our results with those presented in [43]. As mentioned in the introduction, the said article compares several IQA indices calculated on quantized images, and concludes that the best results are provided by the VIF index, followed by MS-SSIM, SSIM, and PSNR, while the worst results are provided by UQI. The remaining IQA indices considered in both articles are different, so they cannot

be compared. Therefore, it is observed that our results are compatible to a certain extent with those presented in [43], since the three best indices identified in the said study show similar results according to our study. On the other hand, in our study the UQI index does not generate results comparable to the rest of the IQA indices analyzed.

Therefore, taking into account our results and those presented in [43], we can suggest using MSE and VIF or MS-SSIM to evaluate CQ methods. On the one hand, VIF and MS-SSIM seem promising indices not only for CQ but also for other image processing operations. On the other hand, MSE allows comparisons with much of the published CQ literature.

Based on the results presented in the article, it can be concluded that it is necessary to use several IQA indices when comparing CQ methods. Using only one IQA index can lead to erroneous conclusions about the relative quality of the compared methods, so it is better to use several IQA indices. We also suggest that IQA indices with very high correlations should not be used together. However, if two IQA indices have a low correlation, we cannot automatically assume they can be used together in a CQ application because one of the indices (or even both) may be irrelevant for this application. In any case, to conclude whether a specific IQA index is relevant for CQ, it is necessary to perform subjective experiments to quantify how well the index agrees with the results of such experiments.

A future line of research to expand the work presented in this article will consist of conducting subjective experiments to compare the results of those experiments with the results presented in this article. This information will allow us to suggest the most promising IQA indices for evaluating CQ methods.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE Signal Process Mag. **26**(1), 98–117 (2009). https://doi.org/10.1109/MSP.2008.930649
2. Kamble, V., Bhurchandi, K.: No-reference image quality assessment algorithms: a survey. Optik. **126**(11–12), 1090–1097 (2015). https://doi.org/10.1016/j.ijleo.2015.02.093
3. Zhai, G., Min, X.: Perceptual image quality assessment: a survey. Sci China Inf Sci. **63**, 1–52 (2020). https://doi.org/10.1007/s11432-019-2757-1
4. Zhang, L., Zhang, L., Mou, X., Zhang, D.: A comprehensive evaluation of full reference image quality assessment algorithms. In: 2012 19th IEEE International Conference on Image Processing. IEEE; 2012. p. 1477–1480 (2012)
5. Xu, S., Jiang, S., Min, W.: No-reference/blind image quality assessment: a survey. IETE Tech Rev. **34**(3), 223–245 (2017). https://doi.org/10.1080/02564602.2016.1151385
6. Celebi, M.E.: Forty years of color quantization: a modern, algorithmic survey. Artif Intell Rev. **56**, 13953–14034 (2023). https://doi.org/10.1007/s10462-023-10406-6
7. Celebi, M.E., Hwang, S., Wen, Q.: Colour quantisation using the adaptive distributing units algorithm. Imaging Sci J. **62**(2), 80–91 (2014). https://doi.org/10.1179/1743131X13Y.0000000059
8. Hsieh, S., Fan, K.C.: An adaptive clustering algorithm for color quantization. Pattern Recognit Lett. **21**(4), 337–346 (2000). https://doi.org/10.1016/S0167-8655(99)00165-8
9. Lei, M., Zhou, Y., Luo, Q.: Color image quantization using flower pollination algorithm. Multimed Tools Appl. **79**, 32151–32168 (2020). https://doi.org/10.1007/s11042-020-09680-1
10. Omran, M.G., Engelbrecht, A.P., Salman, A.: A color image quantization algorithm based on particle swarm optimization. Informatica. **29**(3), 261–269 (2005)
11. Özdemir, D., Akarun, L.: A fuzzy algorithm for color quantization of images. Pattern Recognit. **35**(8), 1785–1791 (2002). https://doi.org/10.1016/S0031-3203(01)00170-4
12. Pérez-Delgado, M.L.: Colour quantization with Ant-tree. Appl Soft Comput. **36**, 656–669 (2015). https://doi.org/10.1016/j.asoc.2015.07.048
13. Pérez-Delgado, M.L.: Artificial ants and fireflies can perform colour quantisation. Appl Soft Comput. **73**, 153–177 (2018). https://doi.org/10.1016/j.asoc.2018.08.018
14. Pérez-Delgado, M.L.: Color image quantization using the shuffled-frog leaping algorithm. Eng. Appl. Artif. Intell. **79**, 142–158 (2019). https://doi.org/10.1016/j.engappai.2019.01.002
15. Pérez-Delgado, M.L.: The color quantization problem solved by swarm-based operations. Appl. Intell. **49**(7), 2482–2514 (2019). https://doi.org/10.1007/s10489-018-1389-6

16. Pérez-Delgado, M.L., Román-Gallego, J.Á.: A two-stage method to improve the quality of quantized images. J. Real-Time Image Process. **17**(3), 581–605 (2020). https://doi.org/10.1007/s11554-018-0814-8

17. Wan, S., Prusinkiewicz, P., Wong, S.: Variance-based color image quantization for frame buffer display. Color Res. Appl. **15**(1), 52–58 (1990). https://doi.org/10.1002/col.5080150109

18. Chang, C.H., Xu, P., Xiao, R., Srikanthan, T.: New adaptive color quantization method based on self-organizing maps. IEEE Trans. Neural Netw. **16**(1), 237–249 (2005). https://doi.org/10.1109/TNN.2004.836543

19. Kasuga, H., Yamamoto, H., Okamoto, M.: Color quantization using the fast K-means algorithm. Syst. Comput. Jpn. **31**(8), 33–40 (2000). https://doi.org/10.1002/1520-684X(200007)31:8<33::AID-SCJ4>3.0.CO;2-C

20. Celebi, M.E.: Improving the performance of k-means for color quantization. Image Vis. Comput. **29**(4), 260–271 (2011). https://doi.org/10.1016/j.imavis.2010.10.002

21. Pérez-Delgado, M.L.: An iterative method to improve the results of ant-tree algorithm applied to colour quantisation. Int. J. Bio-Inspir. Comput. **12**(2), 87–114 (2018). https://doi.org/10.1504/IJBIC.2018.094199

22. Celebi, M.E., Wen, Q., Hwang, S.: An effective real-time color quantization method based on divisive hierarchical clustering. J. Real-Time Image Process. **10**(2), 329–344 (2015). https://doi.org/10.1007/s11554-012-0291-4

23. Park, H.J., Kimy, K.B., Cha, E.Y.: An effective color quantization method using color importance-based self-organizing maps. Neural Netw World. **25**(2), 1 (2015). https://doi.org/10.14311/NNW.2015.25.006

24. Pérez-Delgado, M.L., Günen, M.A.: A comparative study of evolutionary computation and swarm-based methods applied to color quantization. Expert Syst. Appl. **231**, 120666 (2023). https://doi.org/10.1016/j.eswa.2023.120666

25. Wen, Q., Celebi, M.E.: Hard versus fuzzy c-means clustering for color quantization. EURASIP J. Adv. Signal Process. **2011**(1), 1–12 (2011). https://doi.org/10.1186/1687-6180-2011-118

26. Ramella, G., Sanniti Di Baja, G.: A new technique for color quantization based on histogram analysis and clustering. Int. J. Pattern Recognit. Artif. Intell. **27**(03), 1360006 (2013). https://doi.org/10.1142/S0218001413600069

27. Roberto e Souza, M., Carlos Sousa e Santos, A., Pedrini, H.: In: Sourav De SB Sandip Dey, editor. A Hybrid Approach Using the k-means and Genetic Algorithms for Image Color Quantization. Wiley Online Library; p. 151–171 (2020)

28. Ueda, Y., Koga, T., Suetake, N., Uchino, E.: Color quantization method based on principal component analysis and linear discriminant analysis for palette-based image generation. Opt Rev. **24**, 741–756 (2017). https://doi.org/10.1007/s10043-017-0376-1

29. Duanmu, Z., Liu, W., Wang, Z., Wang, Z.: Quantifying visual image quality: a Bayesian view. Ann. Rev Vis Sci. **7**, 437–464 (2021). https://doi.org/10.1146/annurev-vision-100419-120301

30. Varga, D.: An optimization-based family of predictive, fusion-based models for full-reference image quality assessment. J Imaging. **9**(6), 116 (2023). https://doi.org/10.3390/jimaging9060116

31. Yang, P., Sturtz, J., Qingge, L.: Progress in blind image quality assessment: a brief review. Mathematics. **11**(12), 2766 (2023). https://doi.org/10.3390/math11122766

32. Heckbert, P.: Color Image Quantization for Frame Buffer Display. In: Proceedings of the 9th Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '82. New York, NY, USA: ACM; p. 297–307 (1982)

33. Gervautz, M., Purgathofer, W.: A Simple Method for Color Quantization: Octree Quantization. In: Glassner, A.S. (ed.) Graphics Gems, pp. 287–293. Academic Press Professional, Inc., San Diego, CA, USA (1990)

34. Orchard, M.T., Bouman, C.A.: Color quantization of images. IEEE Trans. Signal Process. **39**(12), 2677–2690 (1991). https://doi.org/10.1109/78.107417

35. Wu, X.: Efficient statistical computations for optimal color quantization. In: Arvo, J. (ed.) Graphics Gems II, pp. 126–133. Academic Press (1991)

36. Dekker, A.H.: Kohonen neural networks for optimal colour quantization. Network: Comput Neural Syst. **5**(3), 351–367 (1994). https://doi.org/10.1088/0954-898X/5/3/003

37. Pérez-Delgado, M.L., Román-Gallego, J.Á.: A hybrid color quantization algorithm that combines the greedy orthogonal bi-partitioning method with artificial ants. IEEE Access. **7**, 128714–128734 (2019). https://doi.org/10.1109/ACCESS.2019.2937934

38. Pérez-Delgado, M.L.: A mixed method with effective color reduction. Appl Sci. **10**(21), 7819 (2020). https://doi.org/10.3390/app10217819

39. Celebi, M.E., Pérez-Delgado, M.L.: CQ100: A High-Quality Image Dataset for Color Quantization Research. Mendeley Data, V3. Accessed 5 June 2023. Available from: https://data.mendeley.com/datasets/vw5ys9hfxw/3

40. Mason, A., Rioux, J., Clarke, S.E., Costa, A., Schmidt, M., Keough, V., et al.: Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. IEEE Trans. Med. Imaging. **39**(4), 1064–1072 (2019). https://doi.org/10.1109/TMI.2019.2930338

41. Jinjin, G., Haoming, C., Haoyu, C., Xiaoxing, Y., Ren, J.S., Chao, D.: PIPAL: a large-scale image quality assessment dataset for perceptual image restoration. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16. Springer; p. 633–651 (2020)

42. Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Comparison of full-reference image quality models for optimization of image processing systems. Int. J. Comput. Vis. **129**, 1258–1281 (2021). https://doi.org/10.1007/s11263-020-01419-7

43. Ramella, G.: Evaluation of quality measures for color quantization. Multimed. Tools Appl. **80**(21–23), 32975–33009 (2021). https://doi.org/10.1007/s11042-021-11385-y

44. Pedersen, M., Hardeberg, J.Y.: Full-reference image quality metrics: classification and evaluation. Found Trends Comput. Graph Vis. **7**(1), 1–80 (2012)

45. Wang, Z., Bovik, A.C., Lu, L.: Why is image quality assessment so difficult? In: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing. vol. 4. IEEE; p. IV–3313 (2002)

46. Samajdar, T., Quraishi, M.I.: Analysis and evaluation of image quality metrics. In: Information Systems Design and Intelligent Applications: Proceedings of Second International Conference INDIA 2015, Volume 2. Springer; p. 369–378 (2015)

47. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861

48. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process Lett. **9**(3), 81–84 (2002). https://doi.org/10.1109/97.995823

49. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. vol. 2. IEEE; p. 1398–1402 (2003)

50. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. **15**(2), 430–444 (2006). https://doi.org/10.1109/TIP.2005.859378

51. Chandler, D.M., Hemami, S.S.: VSNR: A wavelet-based visual signal-to-noise ratio for natural images. IEEE Trans. Image Process. **16**(9), 2284–2298 (2007). https://doi.org/10.1109/TIP.2007.901820

52. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A feature similarity index for image quality assessment. IEEE Trans. Image Process. **20**(8), 2378–2386 (2011). https://doi.org/10.1109/TIP.2011.2109730

53. Akins, K.A., Hahn, M.: More than mere colouring: The role of spectral information in human vision. Br. J. Philosophy Sci. (2014)

54. González, R.C., Woods, R.E.: Digital Image Processing. Pearson Education Ltd. (2008)

55. Burger, W., Burge, M.J.: Principles of Digital Image Processing, vol. 111. Springer (2009)

56. Kruse, F.A., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., et al.: The spectral image processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data. Remote Sens. Environ. **44**(2–3), 145–163 (1993). https://doi.org/10.1016/0034-4257(93)90013-N

57. Yuhas, R.H., Goetz, A.F., Boardman, J.W.: Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In: JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop; p. 147–149 (1992)

58. Celebi, M.E., Pérez-Delgado, M.L.: CQ100: A High-Quality Image Dataset for Color Quantization Research. J. Electron. Imaging. **32**(3), 033019 (2023). https://doi.org/10.1117/1.JEI.32.3.033019

59. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Process. **15**(11), 3440–3451 (2006)