Evaluation of Automatic Image Color Theme Extraction Methods

Gianluigi Ciocca, Paolo Napoletano(✉), and Raimondo Schettini

Department of Informatics, Systems and Communication,
University of Milano-Bicocca, Milan, Italy
{ciocca,napoletano,schettini}@disco.unimib.it
http://www.ivl.disco.unimib.it/

Abstract. Color themes are quite important in several fields from visual and graphic art design to image analysis and manipulation. Color themes can be extracted from an image manually by humans or automatically by a software. Plenty of automatic color theme extraction methods, either supervised or unsupervised, have been presented in the state of the art in the last years. Evaluation of a color theme goodness with respect to a reference one is based on visual and subjective comparisons, that may be affected by cultural and social aspects, they are time consuming and not costless. In this paper we experiment several supervised and unsupervised state-of-the-art methods for color theme extraction. To overcome the burden of a subjective evaluation, we experiment the use of a computational metric based on the Earth Mover's distance for goodness evaluation instead of a subjective one. Results show the best color theme is extracted by using a supervised method based on a regression model trained on user-defined color themes and that the computational metric adopted is comparable to a subjective one.

Keywords: Color theme extraction · Color palette evaluation · Earth Mover's Distance

1 Introduction

In the field of visual and graphic art design the choice of an attractive set of colors (also called color theme or color palette) can be a very complex process [21]. The perceived quality of a color decision can be affected by subjective-culture, trending fashions, and individual preference [23,30,32]. Artists and designers often choose colors by taking inspiration from other premade color themes [1, 11,12] or themes extracted from images [23]. A color theme of an image is a finite set of color (usually from 3 to 7) that best represents an image [21].

Color themes are useful in many tasks such as image recoloring, color blending, photo segmentation and editing, image enhancement and manipulation [31]. Color theme can be also adopted as signatures (or feature vectors) for the indexing of images in a content-based image retrieval system [19,25,33,35]. A user

S. Tominaga et al. (Eds.): CCIW 2019, LNCS 11418, pp. 165–179, 2019.
https://doi.org/10.1007/978-3-030-13940-7_13

can query the system by choosing a color, or a set of colors, and then retrieve a set of images that are relevant to that query, or in another way a set of images for which the colors of the query are representative [4,35].

Whatever is the application domain, being able to automatically extract a color theme from an image can facilitate color-related applications such as color picking interfaces, color mood transfer from one image to another, or color harmonization [18,23,26,33]. Human beings are able to recognize millions of colors and more important they are able to describe an image by selecting just a few of them [15,20]. While human beings perform this task quite effortlessly, algorithms does not perform this task easily especially from the computational-cost point of view.

A plenty of automatic color theme extraction methods have been presented in the last years. There are methods based on clustering [8,15,27], that are unsupervised, while other methods are supervised, such as the one by Lin et al. [21]. They presented a regression model trained on user-defined color themes. Very recently a deep-learning based solution has been presented for a discrete-continuous color gamut representation that extends the color gradient analogy to three dimensions and allows interactive control of the color blending behavior [31]. Mellado et al. presented a graph-based palette representation to find a theme as a minimization problem [24].

The evaluation of color theme goodness, that is how much a set of colors is representative of a given image, involves human beings and therefore is highly subjective. Such evaluations are time consuming and not costless. To overcome the limits of subjective evaluations, computational metrics can be adopted. In this paper we experiment several supervised and unsupervised state-of-the-art methods for color theme extraction from images and we exploit a computational metric based on the Earth Mover's distance (EMD) for the automatic evaluation of the goodness of a color theme extracted with respect to a human-based reference theme. Results show that the best method is the one based on a regression model trained on user-defined themes [21] and that the EMD is a quite robust measure of color theme goodness and thus comparable to a subjective one.

2 Methods for Color Theme Extraction

In this section we review the methods for automatic color theme extraction experimented in this paper. Concerning the size of the color theme, previous studies found that the most common value is five [27], therefore the following methods, where possible, will take into account this value as input constraint.

2.1 Color Histogram

This is the simplest algorithm for theme extraction [17]. It is based on the concept of 3D color histogram of an image, that is defined as a distribution over each possible triplet of colors $\mathbf{c_i} = (c_1, c_2, c_3)_i$ of within the color space considered. The distribution is calculated as percentage of the image pixel that assumes a

given color value c_i. Since the number of all possible triplets is enormous, the 3D histogram is usually quantized in small number of volume partitions. We consider a uniform quantization of each color channel of size 3 thus we have $3 \times 3 \times 3 = 27$ partitions v. The color theme is obtained by selecting the K partitions with the highest number of pixels inside and then selecting the K colors $\mathbf{c_v} = (\mu_1, \mu_2, \mu_3)_v$ representing each partition: μ_1 is the mean of the colors belonging to the partition v. Figures 1 and 2 show some color-histogram-based themes obtained with $K = 5$. This method is very simple and at low computational cost. Besides its simplicity, this method may fail because the quantization can aggregate colors that in practice need to be divided, and more important the color representing the volume v may be a color not present in the original image because is obtained as average of all the colors of the partition v. These drawbacks are mostly due to the fact that the quantization is performed uniformly. This method, apart from the choice of the number of volume partitions K, is unsupervised.

2.2 Median Cut

This algorithm is based on the concept of 3D color histogram [17]. In this case, the color space is divide not uniformly by taking into account the distribution of pixels within the three color channels. The algorithm takes as input a maximum number of groups K and works as follows. At the beginning, all the image pixels belong to the same group or partition. The color channel with the highest range of values is chosen as reference c_R. The median value m_{c_R} of the selected color channel is computed and two subgroups are formed by selecting pixels that are higher or lower than the median value m_{c_R}. If the number of subgroups is equal to K then the algorithm stops, otherwise it starts again from the selection of the subgroup with the highest color range. The algorithm stops when the number of subgroups equals K.

The final color theme is obtained, as for the color histogram, by selecting the K volumes with the highest number of pixels inside and then selecting then colors representing each volume. This method is considered unsupervised. Figures 1 and 2 show some examples obtained with $K = 5$.

The color space is partitioned in a smarter way than the color histogram, but in those cases where a huge partition is composed of similar colors, the algorithm splits this partition in two or more partitions (see the greenish colors in Fig. 1 column 3) penalizing those colors that are not so present in the image but are semantically important (look at the bluish color that is reported in the r2 of third column of Fig. 1.

2.3 K-Means and C-Means

Unsupervised clustering methods are largely adopted in machine-learning-based applications. One of the most famous and adopted clustering method is the k-means [22,34]. Given m points $\{x_1, \ldots, x_m\}$ belonging to a $n-$dimensional space R^n, and a given number of clusters K, the k-means algorithm search for

the centroids of the clusters $\{c_1, \ldots, c_K\}$ belonging to R^n such that the sum of the distances between each point x_i belonging to a cluster l and its centroid c_l is minimized:

$$\min_{\{c_1,\ldots,c_K\}} = \sum_{i=1}^{m} \min_{l=1\ldots K} \|x_i - c_l\|$$

K-means is influenced by the initialization step. In this paper we consider two variations of k-means. The first variant (k-means1) is the one adopted by Lin et al. [21] where the initial seeds are stratified randomly sampled within the CIELab color representation of the image. The second (k-means2) is the original k-means with a special initialization where the initial seeds are chosen uniformly over a set of colors ordered from the brightest to darkest.

One of the drawbacks of the k-means is that it does not take into account spatial arrangement of the colors in the image, and can thus wash out important image regions [21].

Fuzzy c-means clustering is quite similar to k-means, except on how the pixels are assigned to the clusters. Here the assignment is soft instead of hard [10]. This makes the algorithm less subject to the outliers problem and so more robust to catch colors related to small (but important) details of the image. For both k-means and c-means the number K is set to 5. Figures 1 and 2 show some examples of color themes obtained with k-means1, k-means2 and c-means.

2.4 ISODATA

ISODATA clustering [16] is an unsupervised classification method alternative to k-means. Unlike the latter, it is not necessary to set in advance the number of final clusters because, during execution, the algorithm checks whether to merge or divide the clusters in order to better fit the distribution of input data. The execution in fact continues until the preset thresholds of variance and distance between the clusters are not below the parameters established by the user. The best set of parameters are not easy to be found and it requires a "trial and error" stage. Even though we need not to set the number K, the disadvantage of the algorithm is that the output color theme may include a higher number of colors than k-means. Figures 1 and 2 show some examples of color themes obtained with ISODATA. This algorithm is considered unsupervised.

2.5 Mean Shift

Mean shift is a clustering algorithm based on the concept of kernel density estimation. The algorithm considers the color space as a empirical probability density function $f(x)$ and the image pixels are considered as sampled from the underlying probability density function. Dense regions (or clusters) in the color space correspond to the mode (or local maxima) of the probability density function. The aim of the algorithm is to find those local maxima [9]. Given the pixels

x_i of the input image, the algorithm applies a kernel $K(x_i - x)$ to each pixel around an initial pixel x. For each pixel, the algorithm defines a window around it and computes the weighted mean $m(x)$ of the pixel as:

$$m(x) = \frac{\sum_{x_i \in N(x)} K(x_i - x) x_i}{\sum_{x_i \in N(x)} K(x_i - x)}$$

where $N(x)$ is the set of neighbours of x for which $K(x_i) \neq 0$. Now the algorithm shifts the center of the window to the mean ($x \leftarrow m(x)$) and repeats the estimation until $m(x)$ converges. The difference $m(x) - x$ is called mean shift. Given a kernel K, bandwidth parameter h, we chose a Gaussian kernel: $K(x_i - x) = e^{-\frac{\|x_i - x\|^2}{h}}$. The parameter h influences the width of the kernel and then the granularity of the cluster and implicitly the number of clusters.

As in the case of ISODATA, the number of clusters can be determined automatically. As drawbacks, the choice of the right value for h is a "trial and error" process, and the number of clusters is usually higher than 5. Figures 1 and 2 show some examples of color themes obtained with Mean shift. This algorithm is considered unsupervised.

2.6 Diffused Expectation Maximization

Diffused expectation maximisation (DEM) is an algorithm for image segmentation. The method models an image as a finite mixture, where each mixture component corresponds to a region class and uses a maximum likelihood approach to estimate the parameters of each class, via the expectation maximisation algorithm, coupled with anisotropic diffusion on classes, in order to account for the spatial dependencies among pixels [5,6]. Each image is conceived as drawn from a mixture of Gaussian density function, so that for any pixel we have:

$$p(x_i|\theta) = \sum_{k=1}^{K} p(x_i|k, \theta) P(k)$$

and the likelihood of the image data is

$$\mathcal{L} = p(x|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$$

Image segmentation can be achieved by finding the set of labels that maximise the likelihood $\mathcal{L}$. The final set of labels are used to select the Gaussians. The mean values of the Gaussians are the colors of the final color theme. The algorithm takes the number of Gaussians K as input and uses k-means for the initialization of the Guassian' parameters. One of the problem of this method is that due to the anisotropic diffusion spatial dependencies between pixels are taken into account. It causes the lost of those pixels that are associated to very small (but important) regions. Figures 1 and 2 show some examples of color themes obtained with DEM. This algorithm, apart from the choice of $K = 5$ is considered unsupervised.

2.7 Regression

This method has been presented by Lin et al. [21]. It is based on a regression model trained on 1,600 color themes extracted from a set of 40 images by 160 different human subjects recruited through the Amazon Mechanical Turk platform. They were asked to pick 5 different colors that would "best represent the image and order them in a way that would capture the image well." Part of these color themes (1,000) have been used to fit a linear model of the training set through the LASSO regression and the remaining for testing. Before training, target scores for each theme on how close it is to human-extracted themes has been computed. 79 features have been extracted from each color theme by considering six types of features to describe each theme: saliency, coverage error both for pixels and for segments, color diversity, color impurity, color nameability, and cluster statistics. To score the goodness of a color theme with respect to the human-based ones during the training process, the authors defined the following distance:

$$score(p) = 1 - \frac{1}{|H|} \sum_{h \in H} \frac{dist(p,h)}{maxDist}$$

where p is the given theme, H is the set of human-extracted themes, $dist$ is the Euclidean error between the two themes in the CIELab color space, and $maxDist$ is some maximum possible distance between two themes. The theme scores are then scaled between 0 and 1 for each image, so that each image gets equal weight in training. Themes with scores closer to 1 are more perceptually similar to human themes on average than themes with scores closer to 0. Given the distance metric and the human-extracted themes for each image, an optimal oracle color theme, that is closest on average to all the human-extracted themes is calculated. Due to the fact that this method requires a training process, it is considered a supervised approach. Figures 1 and 2 show some examples of color themes obtained with the regression method.

2.8 Clarifai

Clarifai (https://clarifai.com/) is an online service for visual recognition. Given an image is possible, through a set of API, to get tags related to the contents of the image, and the color theme. The number of colors of the theme can not be set, thus the output is variable in lengths. Figures 1 and 2 show some examples of color themes obtained with the Clarifai API.

2.9 Random Color Theme

To evaluate the goodness of the color theme methods discussed above, as a baseline, we consider also a random color theme approach. It extracts $K = 5$ colors from a 3D histogram representation of the input image. To reduce the number of possible themes we consider 27 volumes of the 3D histogram by performing

a $3 \times 3 \times 3$ channels quantization. The random algorithm is executed 10 times and the themes showed in Figs. 1 and 2 is the best in terms of similarity with respect to the ground truth.

3 Evaluation Metrics

The qualitative analysis of the color themes extracted using the methods described in the previous section is of limited utility, because it is based on visual and subjective comparisons (see Figs. 1 and 2). To score the goodness of the methods analyzed and quantitatively calculate the similarity between the palettes it is necessary to define a suitable metric.

Lin et al. [21] adopted a subjective metric by asking human subjects to evaluate a set of color themes and rate "how well they represent the color theme of the image". Subjective evaluations strictly depends on the number of subjects and may be affected by cultural and social aspects related to the profiles of the human subjects involved. More important, they are not costless and are time consuming.

To overcome these problems, and so to ease the evaluation step, we introduce an objective metric based on the Earth Mover's Distance (EMD) [28,29]. EMD considers each color theme as a probability distribution, so given two distributions, it performs a quantitative measure of their dissimilarity. EMD measures the minimal cost that must be paid to transform one distribution into the other. As we have seen in the previous section, each theme extraction method outputs a variable number of colors, usually from 3 to 7, so most of the time distributions to be compared are of different length. The EMD by definition can operate on variable-length representations of the distributions and thus is more suitable than traditional histogram matching metrics, like euclidean distance. Moreover, to make the EMD metric capable of approximating perceptual dissimilarity as well as possible, we perform evaluations in the CIE-Lab color space that is more perceptually linear than other color spaces [36]. Within the CIE-Lab color space, a change of the same amount in a color value should produce a change of about the same visual importance.

The EMD definition is based on the notion of distance between the basic features that are aggregated into the distributions, in our case the single colors of the theme. This distance is called ground distance and measures dissimilarity between individual colors in the CIE-Lab space: $\mathbf{c} = \{c_L, c_a, c_b\}$. Given two colors $\mathbf{c_i}$ and $\mathbf{c_j}$, the ground distance d_{ij} is the euclidean distance between the two colors,

$$d_{ij} = \sqrt{(c_L^i - c_L^j)^2 + (c_a^i - c_a^j)^2 + (c_b^i - c_b^j)^2}.$$

The concept of EMD is based on the concept of signature that coincides with the concept of color theme itself. Each signature is made of m clusters and then each color of the theme is the centroid of each of the m clusters. Let us consider two signatures $P = \{(\mathbf{p}_1, w_{\mathbf{p}_1}), \ldots, (\mathbf{p}_m, w_{\mathbf{p}_m})\}$ and $Q = \{(\mathbf{q}_1, w_{\mathbf{q}_1}), \ldots, (w_{\mathbf{q}_n})\}$ of size m and n respectively, where $\mathbf{p}_i$ and $\mathbf{q}_j$ are two different colors or centroids of the clusters, $w_{\mathbf{p}_i}$ and $w_{\mathbf{q}_j}$ are the weights associated to each cluster.

The computation of the EMD between two signatures is obtained as linear programming optmization. The aim of the linear programming solver is, given two signatures made of different sets of clusters and related cluster weights, to fill in the cluster weights of the second signature with the cluster weights of the first signature in a way that the work needed to move these weights from a signature to another is minimum. Intuitively, the solution of this linear programming optimization is the amount of work needed to transform a signature into another.

More formally, the aim is to find a flow $\mathbf{F} = [f_{ij}]$, with f_{ij} the flow between $\mathbf{p}_i$ and $\mathbf{p}_j$, that minimizes the overall cost

$$WORK(P, Q, \mathbf{F}) = \sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij},$$

subject to the following constraints:

$$f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n \tag{1}$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{\mathbf{p}_i}, 1 \leq i \leq m \tag{2}$$

$$\sum_{j=1}^{n} f_{ij} \leq w_{\mathbf{q}_j}, 1 \leq j \leq n \tag{3}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = min\left(\sum_{i=1}^{m} w_{\mathbf{p}_i} \sum_{j=1} w_{\mathbf{q}_j}\right) \tag{4}$$

The first constraint allows to move weights only from P to Q but not the opposite. The second two constraints limit the weights that can be moved from a cluster of P and Q respectively. The last constraint forces to move the maximum amount of weights from P to Q.

Once the optimal flow $\mathbf{F}$ is found, the EMD is defined as follows:

$$EMD(P, Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}$$

The denominator of the EMD formula is a normalization factor that is the total weight of the smaller signature. The ground distance d_{ij} can be any distance, but if it is a metric and the total weights of two signatures are equal, the EMD is a true metric [29].

4 Experiments

In this Section we present the results of our experiments. In Subsect. 4.1 we discuss the data employed for the evaluation of the methods and all the parameters adopted for the themes extraction process. In Subsect. 4.2 we present the color themes extracted using each of the considered methods and we present the results of the evaluation metric adopted.

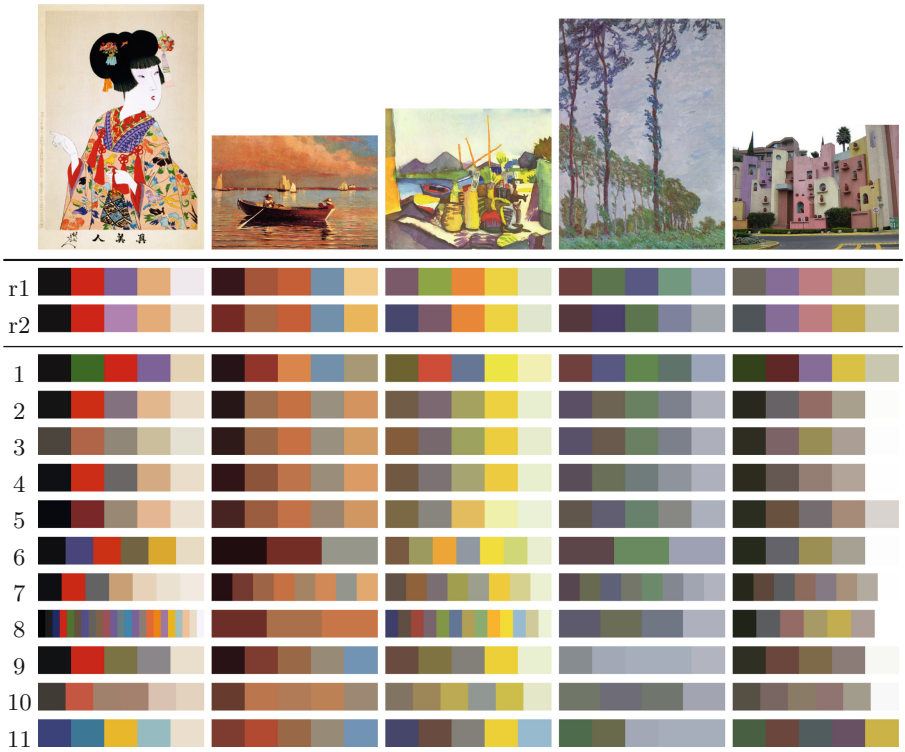

Fig. 1. 1-Regression, 2-c-means, 3-k-means1, 4-k-means2, 5-Median Cut, 6-Clarifai, 7-ISODATA, 8-Mean Shift, 9-Color Histogram, 10-DEM, 11-Random. Color themes extracted by each method from the first 5 images of the test set. From the top, color theme are listed from the best to the worst. The first two rows represent the color themes achieved throughthe Mechanical Turk (r1) or provided by the artists (r2). (Color figure online)

4.1 Data and Experimental Setup

For the evaluation we employ the test part of the dataset presented by Lin et al. [21]. The dataset has been collected by asking 160 different human subjects recruited through the Amazon Mechanical Turk platform, to extract themes from a set of 40 images. These images consisted of 20 paintings and 20 photographs. The paintings were chosen from five artists with different artistic styles (Impressionist, Expressionist, Pointillist, Realist, and Ukiyoe prints). The photographs were Flickr Creative Commons images chosen from the categories Landscape, Architecture, Interior, Closeup, and Portrait.

The authors required participants to choose exactly 5 colors from candidate color swatches generated by running k-means clustering on the image. As we have discussed above, color theme of size 5 are considered the most common size on online theme sharing sites. Each participant extracted themes from either

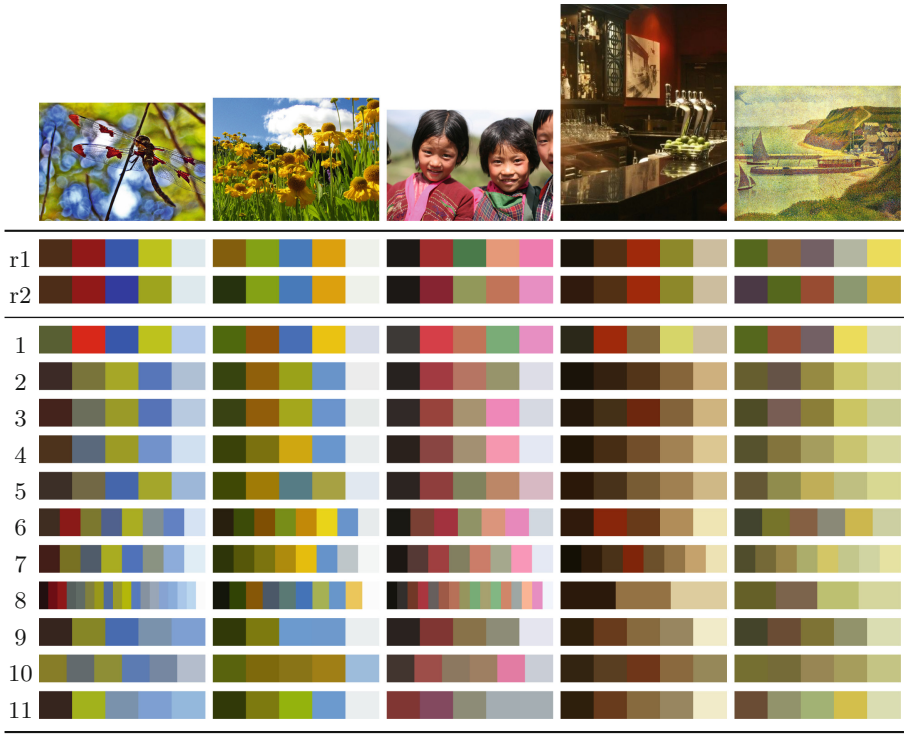

Fig. 2. 1-Regression, 2-c-means, 3-k-means1, 4-k-means2, 5-Median Cut, 6-Clarifai, 7-ISODATA, 8-Mean Shift, 9-Color Histogram, 10-DEM, 11-Random. Color themes extracted by each method from the last 5 images of the test set. From the top, color theme are listed from the best to the worst. The first two rows represent the color themes achieved through the Mechanical Turk (r1) or provided by the artists (r2). (Color figure online)

10 paintings or 10 photographs. The authors asked the participants to pick 5 different colors that would "best represent the image and order them in a way that would capture the image well." The total number of themes collected was 1,600 (40 themes for each image).

For comparison purposes, the authors choose a subset of 10 images (5 paintings and 5 photographs) and asked 11 art students to extract themes from those images. This set of 10 images is the test set adopted by Lin et al. [21] and also adopted in this paper. Using the distance $score(p)$ defined in Sect. 2.7 and the human-extracted themes for each image, an optimal oracle color theme can be found, that is the closest on average to all the human-extracted themes. Thus, for the test set we have two oracles: the Amazon-Mechanical-Turk (r1) one and the Artists one (r2). See the first two rows of the Figs. 1 and 2.

We extract from the 10 test images all the color themes using the 11 methods discussed in Sect. 2: 1-Regression, 2-c-means, 3-k-means1, 4-k-means2, 5-

Median Cut, 6-Clarifai, 7-ISODATA, 8-Mean Shift, 9-Color Histogram, 10-DEM, 11-Random. For all the methods that require an input number of cluster we consider $K = 5$. The regression, the c-means and k-means1 methods are obtained from the author resources that are available at https://github.com/sharondl/color-themes.

Table 1. EMD measured between each color theme obtained using the evaluated methods and the color theme achieved through Amazon Mechanical Turk or provided by the Artists. The last column shows the sum of the average over the two references and the standard deviation. The methods are sorted by the minimum distance.

Method	Mech. Turk		Artist		
	avg (max)	±std	avg (max)	±std	avg + std
Regression	**14.29** (24.14)	5.25	**17.63** (24.93)	5.11	**21.14**
c-means	18.81 (27.79)	4.89	18.85 (29.07)	6.43	24.49
k-means1	19.39 (28.75)	5.39	20.14 (28.81)	5.59	25.26
k-means2	20.32 (29.43)	5.17	20.91 (30.42)	6.42	26.41
Median Cut	21.93 (28.12)	4.34	21.91 (28.93)	5.26	26.72
Clarifai	20.48 (29.14)	4.91	22.37 (37.27)	7.13	27.44
ISODATA	21.98 (32.79)	4.83	22.52 (34.63)	6.04	27.68
Mean Shift	24.87 (30.83)	**4.02**	23.57 (32.47)	**4.78**	28.62
Color Histogram	26.05 (39.17)	7.44	25.24 (39.81)	7.08	32.90
DEM	25.97 (40.44)	6.70	26.44 (41.21)	7.46	33.29
Random	32.76 (40.23)	6.43	31.50 (43.64)	6.43	38.56

To check how many colors the extracted color themes have in common with the ground truth we adopt the color names mapping defined by the ISCC–NBS system proposed in the 1955 by the Inter-Society Color Council and the National Bureau of Standards (NBS, now NIST) [2]. The system was designed to describe colours in non-technical, everyday terms that anybody could understand. The backbone of the system is based on the following 13 names: Pink, Red, Orange, Brown, Yellow, Olive, Yellow green, Green, Blue, Purple, White, Gray, Black. From these names, other subcategories have been derived and the final number of colors is 267 [7]. This color mapping allows to represent each color theme (also the ground truth) with a set of a limited number of color. In this way it is more likely that two color themes can share the same colors.

4.2 Results

Qualitative results are showed in Figs. 1 and 2. The first row of each figure are the test images, the following two rows are the oracle themes r1 (Mechanical Turk) and r2 (Artists) respectively. The remaining 11 rows are the color themes

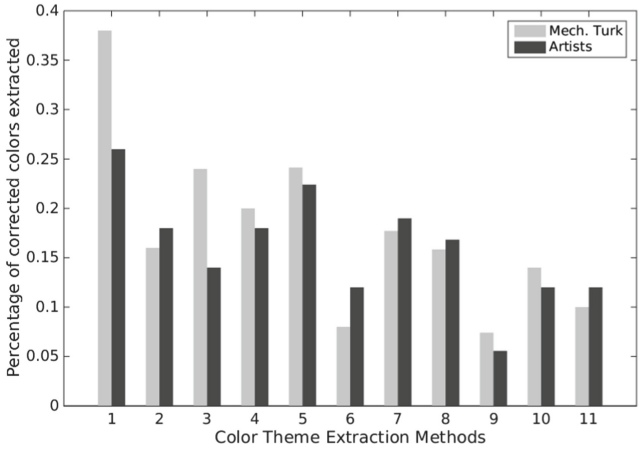

Fig. 3. 1-Regression, 2-c-means, 3-k-means1, 4-k-means2, 5-Median Cut, 6-Clarifai, 7-ISODATA, 8-Mean Shift, 9-Color Histogram, 10-DEM, 11-Random. Average percentage of colors in common between the given method and Mechanical Turk or Artists ground truth (light and dark grey respectively). The percentage is the number of colors in common divided by the number of colors extracted by the given method.

extracted using the corresponding methods. The method are sorted in terms of goodness measured through the EMD. It is quite evident also from a visual analysis that the regression method in most of the cases outputs a color theme that is closer to the reference ones than other methods. Clarifai, ISODATA and mean shift found in most of the cases higher colors than 5.

Table 1 shows the quantitative results expressed in terms of EMD between each method and the two reference color palette. For each method it is showed the average over the 10 test images, the maximum distance and the standard deviation. The last column is the sum between the average and the standard deviation of all the distance computed with the two references. The list of the methods are sorted by the minimum distance, that is by similarity to the reference themes. The results show that the regression method is the best performing in terms of similarity with respect to the ground truth. In particular, the regression method is able to produce palette more similar to the Mechanical Turk ground truth than the Artist ground truth. C-means, k-means1 and k-means2 are quite similar both in terms of visual output and similarity with respect to the ground truth. Color histogram, DEM and random themes are the worst with a EMD value much higher than the value achieved by the regression method. Mean shift achieves the standard deviation lower than other methods. It depends mostly on the fact that the size of the generated palette is, on average, larger than the others. They more likely contains more colors of the reference themes than the other methods, that means that the distance between the mean shift themes and the reference themes are quite similar between each other.

The sorting obtained by the EMD metric is similar to the sorting obtained by Lin et al. [21] using a subjective metric. In this paper, the authors asked humans to evaluate how well a set of themes represented the color themes of the test set. The themes compared included the themes extracted by the regression method, ground truth and k-means1. This further demonstrates the goodness of the EMD as metric for color theme goodness evaluation.

Figure 3 shows the percentage of colors that are in common between the 11 methods and the two reference themes. The percentage is computed by dividing the number of colors in common between the two color themes and the length of the color theme computed with the given state-of-the-art method. Bars represent the average behavior across all the test images. This figure confirms that the regression method is the best performing also in terms of number of colors in common with both ground truth. Following this evaluation the worst ones are Clarifai and color histogram. Bad evaluation of Clarifai color themes is influenced by the fact that the size of the themes is often slightly higher than 5. This also happens to the mean shift and ISODATA methods that output more colors than Clarifai and so they have higher probability to match the ground-truth colors.

5 Conclusion

In this paper we compared several state-of-the-art methods for color theme extraction from images: supervised, unsupervised and a commercial service. The methods have been evaluated by adopting an objective metric that measures the similarity between user-defined themes and color themes achieved by the methods. The metric is based on the Earth Mover's distance (EMD) that allows to handle the variable size of the color palettes obtained by the methods. The EMD is based on the definition of the ground distance that is the distance between two colors belonging to two different themes. We have adopted the Euclidean distance as ground distance measured in the CIE-Lab color space that models the perceptual dissimilarity between colors better than other color spaces. Results on a test set of 10 images demonstrated that a supervised method based on a regression model trained on a set of user-defined color theme performs better than the other methods although unsupervised clustering methods are also quite good in terms of performance. A comparison between the EMD metric and subjective one adopted in [21] shows that the two metrics are quite similar for color theme goodness evaluation. As a future work, we plan to perform a more extensive evaluation with a larger set of images. More specifically, a direct comparison between the objective metric and a human-based/subjective metric will be carried out. It would be interesting in the future to explore the use of deep learning for color theme extraction [3] and other alternative machine learning approaches [14]. Another aspect that would deserve to be taken into account is how image complexity influences color theme extraction [13].

Acknowlegment. The authors wish to thank Marco Verna because part of this work has been developed during his thesis. Published in the context of the projects: FooDesArt - Food Design Arte - L'Arte del Benessere, CUP (Codice Unico Progetto): E48I16000350009 - Call "Smart Fashion and Design", cofunded by POR FESR 2014–2020 (Programma Operativo Regionale, Fondo Europeo di Sviluppo Regionale); E4S: ENERGY FOR SAFETY Sistema integrato per la sicurezza della persona ed il risparmio energetico nelle applicazioni di Home & Building Automation, CUP: E48B17000310009 - Call "Smart Living".

References

1. Adobe Color CC: Adobe 2017 (2017). https://color.adobe.com
2. Agoston, G.A.: Color Theory and Its Application in Art and Design, vol. 19. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-540-34734-7
3. Bianco, S., Cadene, R., Celona, L., Napoletano, P.: Benchmark analysis of representative deep neural network architectures. IEEE Access **6**, 64270–64277 (2018)
4. Bianco, S., Ciocca, G.: User preferences modeling and learning for pleasing photo collage generation. ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) **12**(1), 6 (2015)
5. Boccignone, G., Ferraro, M., Napoletano, P.: Diffused expectation maximisation for image segmentation. Electron. Lett. **40**(18), 1 (2004)
6. Boccignone, G., Napoletano, P., Caggiano, V., Ferraro, M.: A multiresolution diffused expectation-maximization algorithm for medical image segmentation. Comput. Biol. Med. **37**(1), 83–96 (2007)
7. Centore, P.: sRGB centroids for the ISCC-NBS colour system. Munsell Colour Sci. Painters (2016)
8. Chang, H., Fried, O., Liu, Y., DiVerdi, S., Finkelstein, A.: Palette-based photo recoloring. ACM Trans. Graph. (TOG) **34**(4), 139 (2015)
9. Cheng, Y.: Mean shift, mode seeking, and clustering. IEEE Trans. Pattern Anal. Mach. Intell. **17**(8), 790–799 (1995)
10. Chuang, J., Stone, M., Hanrahan, P.: A probabilistic model of the categorical association between colors. In: Color and Imaging Conference. vol. 2008, pp. 6–11. Society for Imaging Science and Technology (2008)
11. Colormind: Colormind.io (2018). http://colormind.io/
12. COLOURlovers: Colourlovers (2017). http://www.colourlovers.com
13. Corchs, S., Ciocca, G., Bricolo, E., Gasparini, F.: Predicting complexity perception of real world images. PLoS ONE **11**(6), e0157986 (2016)
14. Cusano, C., Napoletano, P., Schettini, R.: Remote sensing image classification exploiting multiple kernel learning. IEEE Geosci. Remote Sens. Lett. **12**(11), 2331–2335 (2015)
15. Delon, J., Desolneux, A., Lisani, J.L., Petro, A.B.: Automatic color palette. In: IEEE International Conference on 2005 Image Processing, ICIP 2005, vol. 2, pp. II-706. IEEE (2005)
16. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. **3**, 32–57 (1973)
17. Gonzalez, R.C., Woods, R.E., et al.: Digital Image Processing. Prentice Hall, Upper Saddle River (2017)
18. Greenfield, G.R., House, D.H.: Image recoloring induced by palette color associations. J. WSCG **11**, 189–196 (2003)

19. Gudivada, V.N., Raghavan, V.V.: Content based image retrieval systems. Computer **28**(9), 18–22 (1995)
20. Hubel, D.H.: Eye, Brain, and Vision. Scientific American Library/Scientific American Books, New York (1995)
21. Lin, S., Hanrahan, P.: Modeling how people extract color themes from images. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3101–3110. ACM (2013)
22. MacQueen, J., et al.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, vol. 1, pp. 281–297 (1967)
23. Meier, B.J., Spalter, A.M., Karelitz, D.B.: Interactive color palette tools. IEEE Comput. Graph. Appl. **3**, 64–72 (2004)
24. Mellado, N., Vanderhaeghe, D., Hoarau, C., Christophe, S., Brédif, M., Barthe, L.: Constrained palette-space exploration. ACM Trans. Graph. (TOG) **36**(4), 60 (2017)
25. Napoletano, P.: Hand-crafted vs learned descriptors for color texture classification. In: Bianco, S., Schettini, R., Trémeau, A., Tominaga, S. (eds.) CCIW 2017. LNCS, vol. 10213, pp. 259–271. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56010-6_22
26. Obrador, P.: Automatic color scheme picker for document templates based on image analysis and dual problem. In: Digital Publishing, vol. 6076, p. 607609. International Society for Optics and Photonics (2006)
27. O'Donovan, P., Agarwala, A., Hertzmann, A.: Color compatibility from large datasets. ACM Trans. Graph. (TOG) **30**, 63 (2011)
28. Rubner, Y., Tomasi, C.: The earth mover-distance. Perceptual Metrics for Image Database Navigation. The Springer International Series in Engineering and Computer Science (Robotics: Vision, Manipulation and Sensors), vol. 594, pp. 13–28. Springer, Boston (2001). https://doi.org/10.1007/978-1-4757-3343-3_2
29. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vis. **40**(2), 99–121 (2000)
30. Battiato, S., Ciocca, G., Gasparini, F., Puglisi, G., Schettini, R.: Smart photo sticking. In: Boujemaa, N., Detyniecki, M., Nürnberger, A. (eds.) AMR 2007. LNCS, vol. 4918, pp. 211–223. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-79860-6_17
31. Shugrina, M., Kar, A., Singh, K., Fidler, S.: Color sails: discrete-continuous palettes for deep color exploration. arXiv preprint arXiv:1806.02918 (2018)
32. Walch, M., Hope, A.: Living Colors: The Definitive Guide to Color Palettes Through the Ages. Chronicle Books, San Francisco (1995)
33. Wang, B., Yu, Y., Wong, T.T., Chen, C., Xu, Y.O.: Data-driven image color theme enhancement. ACM Trans. Graph. (TOG) **29**, 146 (2010)
34. Weeks, A.R., Hague, G.E.: Color segmentation in the HSI color space using the k-means algorithm. In: Nonlinear Image Processing VIII, vol. 3026, pp. 143–155. International Society for Optics and Photonics (1997)
35. Wong, K.M., Chey, C.H., Liu, T.S., Po, L.M.: Dominant color image retrieval using merged histogram. In: Proceedings of the 2003 International Symposium on Circuits and Systems, 2003 ISCAS 2003, vol. 2, p. II. IEEE (2003)
36. Wyszecki, G., Stiles, W.S.: Color Science, vol. 8. Wiley, New York (1982)