Expert Systems With Applications 97 (2018) 83–94

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (QArt-Learn)

Zoe Falomir[a,*], Lledó Museros[b], Ismael Sanz[b], Luis Gonzalez-Abril[c]

[a] Bremen Spatial Cognition Centre (BSCC), University of Bremen, Enrique-Schmidt-Str. 5, Bremen, 28359, Germany
[b] Engineering and Computer Science Department, Universitat Jaume I, Av. Vicent Sos Baynat s/n, E-12071, Castellón, Spain
[c] Applied Economics I Department, Universidad de Sevilla, Avda. Ramon y Cajal, 1, Sevilla, E-41018, Spain

ARTICLE INFO

Article history:
Received 31 July 2017
Revised 17 October 2017
Accepted 30 November 2017
Available online 2 December 2017

Keywords:
Qualitative modelling
Color naming
Color similarity
Support vector machines
Art
Machine learning

ABSTRACT

The QArt-Learn approach for style painting categorization based on Qualitative Color Descriptors (QCD), color similarity (*SimQCD*), and quantitative global features (i.e. average of brightness, hue, saturation and lightness and brightness contrast) is presented in this paper. k-Nearest Neighbor (k-NN) and support vector machine (SVM) techniques have been used for learning the features of paintings from the Baroque, Impressionism and Post-Impressionism styles. Specifically two classifiers are built, and two different parameterizations have been applied for the QCD. For testing QArt-Learn approach, the Painting-91 dataset has been used, from which the paintings corresponding to Velázquez, Vermeer, Monet, Renoir, van Gogh and Gauguin were extracted, resulting in a set of 252 paintings. The results obtained have shown categorization accuracies higher than 65%, which are comparable to accuracies obtained in the literature. However, QArt-Learn uses qualitative color names which can describe style color palettes linguistically, so that they can be better understood by non-experts in art since QCDs are aligned with human perception.

1. Introduction

Everyday perception usually involves a task in a real scenario in which the degree of performance matters (i.e. find a cup to drink water from) and typically uses prior knowledge to solve the ambiguous sensory information (i.e. cups usually are located in the kitchen area, typically inside cupboards nearby to the sink/dishwasher). In contrast, in visual art perception there is not a defined task to carry out, and ambiguous sensory information is resolved by the use of conventions, which can be different across periods and cultures (Mamassian, 2008). For example, not knowing the nature of the illuminant leads to ambiguities on object color. However, it is usually assumed as a convention that there is only one stationary light source in the scene (Mamassian, Knill, & Kersten, 1998). However, through history, new styles have appeared that attacked these conventions; for example, *Impressionism* broke the tradition of reproducing shapes accurately, and introduced blurred boundaries of objects producing movement effects on observers. Also *Fauvism* broke the tradition of reproducing the colors as faithfully as possible, and *Cubism* broke the convention of using a unitary point of view (Mamassian, 2008). All these ambiguities show that automatic painting categorization is a challenge to overcome.

Art works are appreciated by humans primarily for their beauty or emotional power[1], but computers cannot feel beauty or emotions yet. In the literature, there are approaches for identifying aesthetics in paintings (Li & Chen, 2009) or for calculating harmonious color combinations (Museros, Sanz, Falomir, & Gonzalez-Abril, 2016). Still, the field of art is especially challenging for computers.

Art critics and experts draw upon a broad experience and knowledge to succeed in art categorization. Therefore, there must be some cognitive properties in paintings that are easier for people to identify than they are for computers. This research focuses on the challenge of identifying qualities in paintings that aligned to human cognition, with color as the starting point: Are color palettes relevant to identify art styles? Which level of detail is necessary for this color representation? Is color naming enough?

* Corresponding author.
E-mail addresses: zfalomir@uni-bremen.de (Z. Falomir), museros@uji.es (L. Museros), isanz@uji.es (I. Sanz), luisgon@us.es (L. Gonzalez-Abril).

[1] Oxford dictionary: https://en.oxforddictionaries.com/definition/art.

https://doi.org/10.1016/j.eswa.2017.11.056

Are those features easy to identify also for artificial agents? These questions are challenges and our paper is a little step towards answering them.

Previous research have dealt with the problem of art paintings categorization (Jiang, Huang, Ye, & Gao, 2006; Karayev, Hertzmann, Winnemoeller, Agarwala, & Darrell, 2013; Mensink & van Gemert, 2014; Shamir, Macura, Orlov, Eckley, & Goldberg, 2010; Shamir & Tarakhovsky, 2012; Shen, 2009; Siddiquie, Vitaladevuni, & Davis, 2009). Jiang et al. (2006) classified traditional Chinese paintings using colors and SVMs using a dataset of 1799 Gongbi and 1889 Xieyi paintings; they obtained 94.61% as overall classification accuracy in these 2 categories. Shamir et al. (2010); Shamir and Tarakhovsky (2012) automated the recognition of nine painters and three schools of art based on their signature styles. The features they used were those by Orlov, Johnston, Macura, Shamir, and Goldberg (2007) for analyzing microscopy images (i.e., Radon Transform features, Chebyschev statistics, Gabor filters, multi-scale histograms, first 4 moments, Tamura texture features, Edge statistics, object statistics, Zernike features, Haralick's features, Chebyshev-Fourier features) and they applied the weighted nearest neighbor (WNN), and SVMs with linear, polynomial and RBF kernels for learning. Their comparison is based on a 2-way classification of Renoir/Monet, Dali/Ernst, Dali/Kandinsky, Renoir/Rothko, and Pollock/Ernst. They used a dataset of 57 paintings per author. The accuracy obtained when classifying paintings into the nine painter classes was 77%, and the accuracy of associating a given painting with its school of art (impressionism, abstract expressionism and surrealism) was 91%. Siddiquie et al. (2009) used Boost-based SVMs, an alternate method to select training data instances using AdaBoost for each of the SVM base kernels, to classify painting styles. A dataset of 498 paintings (81 Abstract Expressionist, 84 Baroque, 84 Cubist, 82 Graffiti, 89 Impressionist and 78 Renaissance was used. And the accuracies obtained were 72.6% in Abstract expressionist, 92.5% in Baroque, 75.6% in Cubist, 89.6% in Graffiti, 73.9% in Impressionist and 81.9% in Renaissance. Mensink and van Gemert (2014) classified paintings in Rijksmuseum according to their author using Fisher vectors of local SIFT descriptors and SVMs, obtaining 76.3% as mean class accuracy (MCA) in a dataset of 112039 photographic reproductions of artworks.

Convolutional neural networks produced further advances in AI and art. For example, an artificial neural system achieved a separation of image content from style using deep neural networks, which allowed to recast the content of one image in the style of another image (Gatys, Ecker, & Bethge, 2015). In the literature, convolutional neural networks have been also used also for categorising paintings. Shen (2009) categorised paintings by artist using a radial basis function (RBF) neural network classifier obtaining an accuracy of 69.7% in a dataset of 25 different painters with more than 20 pieces authored by each of them. Karayev et al. (2013) applied deep neural networks trained on object recognition for style recognition in order to classify artworks according to their period, including Baroque (81.45%), Impressionism (82.15%) and Post-Impressionism (74.51%) styles.

Moreover, in the literature wavelets of brush strokes have also been analyzed in paintings to find out artist identity which had very successful results in the case of identifying Vicent van Gogh's paintings (Johnson et al., 2008; Li, Yao, Hendriks, & Wang, 2012).

However, none of the features used in the previous classification works were qualitative concepts. Qualitative descriptors have been shown to be successful in managing incomplete, imprecise and ambiguous information (Cohn & Renz, 2007; Freksa, 2013). Moreover, they use linguistic concepts which align with human perception and can be easily used to generate explanations and give feedback to the users. Some research works, such as that by Yelizaveta, Tat-Seng, and Irina (2005) used semantic categories (i.e. *warm, cold*) and color names showing to be effective for painting retrieval in databases. In previous works by the authors, qualitative color descriptors have been defined, a measure to calculate the color similarity of paintings has been studied (Falomir, Museros, & Gonzalez-Abril, 2015a) and inspired the current categorizing study. A previous pilot test (Falomir, Museros, Sanz, & Gonzalez-Abril, 2015b) proved the feasibility of a classification of painting styles based on their qualitative color palette.

The implications of using machine learning with methods that cannot be interpretable can have counter-intuitive properties (Szegedy et al., 2013). This the main reason why this paper focus on qualitative features, and a human expert is interviewed regarding painting categorization. Our research goals are to provide interpretable definitions of art styles, using qualitative terms that provide reasons to explain the commonalities and patterns followed in those styles. In this way, human experts could correct those systems, and even teach them using natural language, which requires extracting and using qualitative features. In this paper, qualitative features of colors are obtained and compared to quantitative ones in order to investigate if there is a loss in accuracy when only qualitative features are used.

Specifically, *Qualitative Color Descriptors (QCDs)* are obtained for each painting using two different parameterizations: 37-QCD-palette (Falomir et al., 2015a) and 28-QCD-palette (Sanz, Museros, Falomir, & Gonzalez-Abril, 2015). The QArt-Learn approach is defined here for categorizing paintings into art styles implemented using machine learning techniques such as *k*-Nearest Neighbor (*k*-NN) and Support Vector Machines (SVMs). It combines qualitative color palettes with average color features and analyses their performance. For testing the QArt-Learn approach, the Painting-91 dataset by Khan, Beigpour, van de Weijer, and Felsberg (2014) — actively applied in the literature — was used.

The rest of the paper is organized as follows. Section 2 summarizes Qualitative Color Descriptors. Section 3 presents a measure of similarity defined for qualitative colors (*SimQCD*) and a similarity measure between complete paintings based on this *SimQCD*. Section 4 presents average quantitative global features regarding the hue, saturation, lightness and brightness parameters of a digital painting which have been shown to be useful to categorize paintings in the literature. Section 5 presents a description of Baroque, Impressionism and Post-Impressionism art styles based on their visual color palette. Section 6 explains the SVM machine learning method. Section 7 presents the experimentation and the results. Section 8 discusses the results and, finally, conclusions and future work are provided.

2. A qualitative color descriptor (*QCD*)

The QCD model (Falomir et al., 2015a; Falomir, Museros, Gonzalez-Abril, & Sanz, 2013) defines a reference system (RS) in the Hue, Saturation, and Lightness (HSL) color space, which assigns interval values to qualitative colors (QCs). The resulting Qualitative Color Reference System (QCRS) is built according to Fig. 1 and defined as:

$QCRS = \{uH, uS, uL, QC_{NAME1..5}, QC_{INT1..5}\}$ where uH is the unit of Hue; uS is the unit of Saturation; uL is the unit of Lightness; $QC_{NAME1..5}$ refers to the color names; and $QC_{INT1..5}$ refers to the intervals of HSL coordinates associated with each color. The chosen QC_{NAME} are:

QC_{NAME_1} = {*black, dark_grey, grey, light_grey, white*}

QC_{NAME_2} = {*red, orange, yellow, green, turquoise, blue, purple, pink*}

QC_{NAME_3} = {*pale_* + QC_{NAME_2}}

QC_{NAME_4} = {*light_* + QC_{NAME_2}}

QC_{NAME_5} = {*dark_* + QC_{NAME_2}}

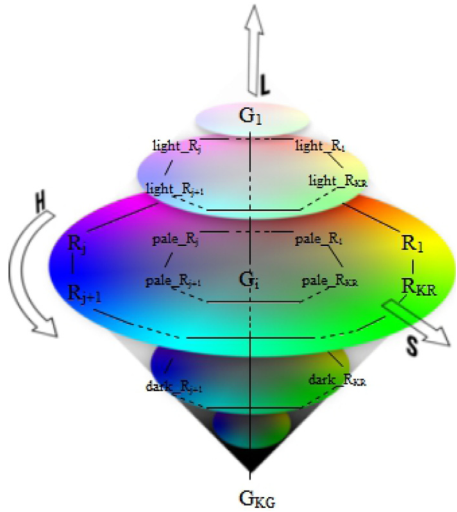

Fig. 1. Diagram for describing QCD: discretization of the HSL color space.

In order to determine the interval of values associated to the Qualitative Color Reference System (QCRS), a test was carried out on 534 participants (students and teachers) at Universitat Jaume I and Universidad de Sevilla in Spain. A computer application was implemented (Falomir et al., 2015a; Sanz et al., 2015) which showed 10 different colors selected randomly and uniformly using their HSL coordinates. For each color selected, participants were asked if they considered the color to be in the *grey* or *rainbow* scale. For those colors classified in the *grey* scale, participants were asked if the color was *white, light_grey, grey, dark_grey* or *black*, that is, $card(QC_{NAME_1}) = 5$. For those colors classified in the *rainbow* scale, participants were asked if the color was *red, orange, yellow, green, turquoise, blue, purple* or *pink*, that is, $card(QC_{NAME_{2..5}}) = 8$, and if it was *light, pale* or *dark*. Thus, a total of 37 color names were considered.

Let us justify the parameters selected: (i) $card(QC_{NAME_1}) = 5$ because the less saturated and extreme colors in luminance are *white* and *black* and, according to the *M* sets defined, there are two more gradations in lightness *light-* and *dark-* and one more in saturation *pale-*, which correspond to *light-grey, dark-grey*, and *grey*, respectively; and (ii) $card(QC_{NAME_{2..5}}) = 8$ since the rainbow/spectral colors are 7 and the majority of the participants of the test suggested to add also *pink*[2].

The QCRS was calibrated according to the results obtained from those studies (Falomir et al., 2015a; Sanz et al., 2015). The selected QC_{INT} are given in Table 1 showing the HSL values assigned to each color name.

It is possible to obtain tailored parameterizations of the QCD theory, with different thresholds and/or sets of adjectives. An example of another experimentally validated parameterization for a specific user community (including less color experts) is provided by Sanz et al. (2015). The main differences with respect to the previous QCD parameterization are (i) slightly different hue intervals (ii) a smaller threshold between the gray scale and the chromatic scale and (iii) the *pale-* and *light-* modifiers are merged into a single *palelight* adjective. This alternative parameterization is specified in Table 2.

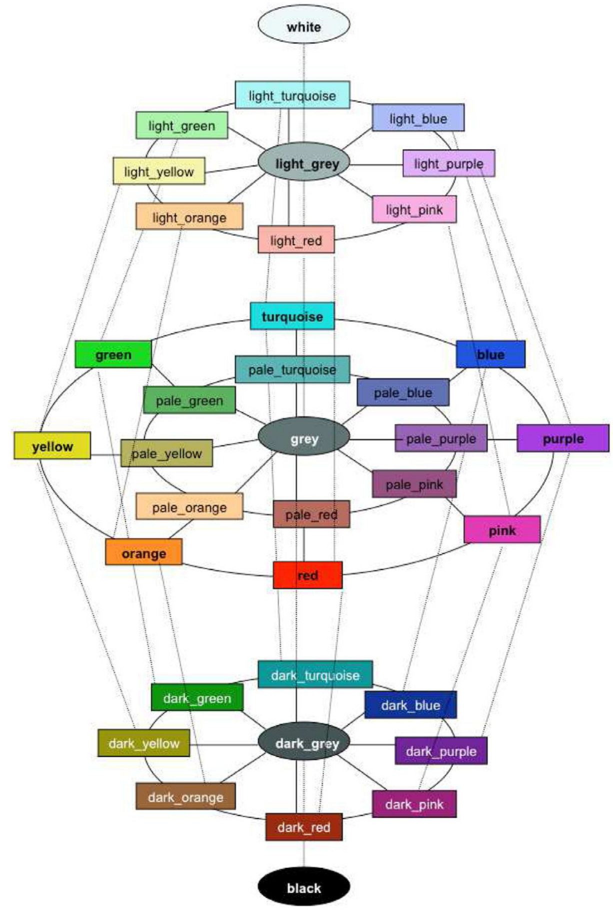

Fig. 2. Diagram for describing QCD: CND corresponding to QCD, w_i correspond to the weights used to calculate the similarity between colors.

3. A similarity measure between paintings using QCD

The QCD model has a relational structure since it is build on a 3-dimensional color space where the luminosity, saturation and hue of the colors change continuously. The colors defined by the QCD model can be organized in a Conceptual Neighborhood Diagram (CND) (Freksa, 2013) according to how one can be transformed into another by changing its luminosity, saturation or hue. For example, the colors *red* and *orange* are conceptual neighbors since a continuous change in hue causes a direct transition from *red* to *orange*. However, the colors *blue* and *red* are not conceptual neighbors since a continuous transformation of hue from *blue* to *red* finds other colors in between. A CND for the computational QCD model has been built for the original parameterization (37 colors) and it is shown in Fig. 2. A similar CND can be built for the adapted parameterization. The nodes of this CND correspond to the color names, where the path connecting neighboring colors are drawn as lines. In order to calculate the similarity between the color names in the QCD, an interval distance as defined below has been used, since it can be calculated for any of the used parameterizations (Falomir et al., 2013).

Given two intervals[3], $I_1 = (a_1,b_1) = B_{r1}(c_1)$ and $I_2 = (a_2,b_2) = B_{r2}(c_2)$, a family of distances between them was defined in Gonzalez-Abril, Velasco, Ortega, and Cuberos (2009), which de-

[2] Note that the selected values for the cardinalities of the sets depend on the current use case and that different values of those parameters could have produced different outcomes in the survey.

[3] An interval is written indistinctly in the classic notation, $I = (a, b)$ or in the Borelian notation, $B_r(c)$ (open ball), that is, $I = (a, b) = (c - r, c + r)$, where $c = (a + b)/2$ and $r = (b - a)/2$. If $c \in \mathbb{R}^2$, then $B_r(c)$ represent an open ball in $\mathbb{R}^2$.

Table 1
HSL intervals for color names (QC_{INT}) corresponding to the 37-QC-palette by Falomir et al. (2015a).

	color	UH	US	UL
QC_{LAB_1}	*black*	[0, 360]	[0, min{20, 2 *UL*, 200 − 2 *UL*}]	(0, 20]
	dark_grey			(20,40]
	grey			(40, 60]
	light_grey			(60, 80]
	white			(80, 100]
QC_{LAB_2}	*red*	(335, 360]∧ [0, 20]	(50, min{100, 2 *UL*, 200 − 2 *UL*}]	(40, 60]
	orange	(20, 50]		
	yellow	(50, 80]		
	green	(80, 160]		
	turquoise/cyan	(160, 200]		
	blue	(200, 239]		
	purple	(239, 297]		
	pink/magenta	(297, 335]		
QC_{LAB_3}	*pale_QC_{LAB_2}*	Idem	(20, 50]	(40, 60]
QC_{LAB_4}	*light_QC_{LAB_2}*	Idem	(20, 200 − 2*UL*]	(60, 90]
QC_{LAB_5}	*dark_QC_{LAB_2}*	Idem	(20, 2*UL*]	(10, 40]

Table 2
HSL intervals for color names (QC_{INT}) in a 28-QC-palette (Sanz et al., 2015).

	color	UH	US	UL
QC_{LAB_1}	*black*	[0 − 360]	[0, min{18.5, 200 − 2 *UL*}]	(0, 27.5]
	dark_grey			(27.5, 45.5]
	grey			(45.5, 78.5]
	white			(78.5, 100]
QC_{LAB_2}	*red*	(345.5, 360] ∧ [0, 10]	(53.5, min{100, 2 *UL*, 200 − 2 *UL*}]	(45.5, 78.5]
	orange	(10, 42.5]		
	yellow	(42.5, 66.5]		
	green	(66.5, 151.5]		
	turquoise/cyan	(151.5, 193.5]		
	blue	(193.5, 256.5]		
	purple	(256.5, 305.5]		
	pink/magenta	(305.5, 345.5]		
QC_{LAB_3}	*palelight_QC_{LAB_2}*	Idem	(18.5, 53.5]	(45.5, 90]
QC_{LAB_4}	*dark_QC_{LAB_2}*	Idem	(18.5, 2*UL*]	(10, 45.5]

pends on three positive parameters, w_1, w_2 and w_3, as follows:

$$d^2(I_1, I_2) = (\Delta c \quad \Delta r) A \begin{pmatrix} \Delta c \\ \Delta r \end{pmatrix}$$

where $\Delta c = c_2 - c_1$, $\Delta r = r_2 - r_1$ and $A = \begin{pmatrix} w_1 & w_2 \\ w_2 & w_3 \end{pmatrix}$ is a symmetrical 2×2 matrix of weights, which must be a positive definite matrix, that is, $w_1 w_3 - w_2^2 > 0$ and $w_1 + w_3 > 0$. From the A matrix, the weights given to the position of the intervals and to the radius can be controlled.

A choice for the A matrix is to consider a diagonal matrix, that is, $A = \begin{pmatrix} w_1 & 0 \\ 0 & w_3 \end{pmatrix}$ which provides the next distance:

$$d_{(w_1, w_3)}(I_1, I_2) = \sqrt{w_1(\Delta c)^2 + w_3(\Delta r)^2}. \quad (1)$$

Hence, the qualitative color names (QC) are defined in HSL from three intervals $I^H \times I^S \times I^L$, as follows: $QC = [h_0, h_1] \times [s_0, s_1] \times [l_0, l_1]$. It follows that:

- In QC_{LAB_1}, the interval values obtained in HSL coordinates correspond to cylinders in the Cartesian coordinate system, described as:

  $$QC = [0, 360] \times [0, s_1] \times [l_0, l_1] = B_{s_1}(0, 0) \times B_{l_r}(l_c)$$

- In all other cases, the interval values in HSL, $QC = [h_0, h_1] \times [s_0, s_1] \times [l_0, l_1]$, correspond to wedges in the Cartesian axis (see Fig. 3). Note that, as it is not possible to represent them as precise open balls, the largest open ball within the wedge corresponding to the UH and US coordinates will be selected (see

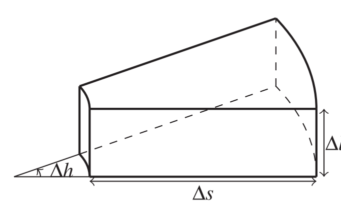

Fig. 3. Wedge corresponds with the interval values in HSL, $QC = [h_0, h_1] \times [s_0, s_1] \times [l_0, l_1]$, where $\Delta h = h_1 - h_0$, $\Delta s = s_1 - s_0$ and $\Delta l = l_1 - l_0$.

Fig. 4). That is, the largest cylinder within the wedge. For this some reference values are calculated:

$$h_c = \frac{h_0 + h_1}{2}, \quad h_r = \frac{h_1 - h_0}{2}, \quad s_c = \frac{s_1}{1 + sin(h_r)}$$

$$s_r = s_1 - s_r, \quad l_c = \frac{l_0 + l_1}{2}, \quad l_r = \frac{l_1 - l_0}{2}$$

and, hence,

$$QC \equiv B_{s_r}(s_c\, cos(h_c), s_c\, sin(h_c)) \times B_{l_r}(l_c)$$

Hence, given two color names $QC_A \equiv B_{r_1}(x_1, y_1) \times B_{lr_1}(lc_1)$ and $QC_B \equiv B_{r_2}(x_2, y_2) \times B_{lr_2}(lc_2)$, the interval distance between them is calculated from (1) as:

$$\begin{aligned} dQCInt^2(QC_A,\ QC_B) = {} & d^2_{(1,0.5)}(B_{r_1}(x_1), B_{r_2}(x_2)) \\ & + d^2_{(1,0.5)}(B_{r_1}(y_1), B_{r_2}(y_2)) \\ & + d^2_{(2,2)}(B_{lr_1}(lc_1), B_{lr_2}(lc_2)) \end{aligned}$$

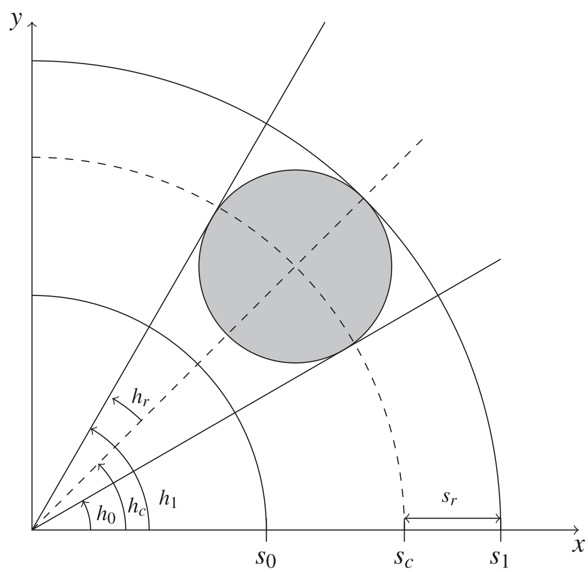

Fig. 4. The largest open ball within the wedge corresponding to the UH and US coordinates.

Note that the weights given in the distances are different. This is due that to HSL color space being formed by two inverted cones joined at their bases (spindle) in which the distance between their vertices is 100 (UL axis). However, the diameter length in the equator of the cone (US axis) is 200, that is, the ratio is 1:2.

It is clear that *dQCInt* is a distance since it is defined from an addition of distances. Hence, given two color names QC_A and QC_B, a normalized similarity *SimQCDInt* between them is defined as:

$$SimQCDInt(QC_A, QC_B) = 1 - \frac{dQCInt(QC_A, QC_B)}{max(dQCInt)} \quad (2)$$

where $max(dQCInt)$ denotes the maximum distance for all colors defined by $\mathrm{QC}_{LAB1..5}$.

The similarity measure defined between the qualitative colors in the QCD model is used to compute the similarity of two compositions (digital images) based on the colors appearing in them and their percentage of appearance. Furthermore, a distance measure and a dot product are also defined between compositions.

Let us denote the set of the representative color names of the QCD model as: $\mathcal{C} = \{QC_1, \cdots, QC_n\}$, where $n = 28$ or 37 in this paper. Thus, the similarity SimQCD: $\mathcal{C} \times \mathcal{C} \longrightarrow [0, 1]$ provides a matrix $S = \{s_{i,j}\} = \{SimQCD(QC_i, QC_j)\}_{i,j=1}^{n}$ which is symmetric and whose main diagonal is filled with 1 values.

Let us consider $\mathcal{Y}$ as the set of the color compositions/images to compare. If *Image* represents a color composition, the system obtains a color histogram: $Image = (f_1, f_2, \cdots, f_n)$ where f_i corresponds to the percentage of the color QC_i within the *Image* ($f_i \geq 0$). Therefore, each image is assigned a unique vector, $\mathcal{Y} \longrightarrow \mathbb{R}^n$ that is, $Image \equiv \mathbf{I}$ where $\mathbf{I} \in \mathbb{R}^n$. Note that two images or color compositions are equal in the presented system if they have the same representation as an $\mathbb{R}^n$ vector.

In order to define a similarity measure, let us consider the following matrix S^* associated to S and defined as follows:

$$S^* = \{s^*_{ij}\}_{i,j=1}^{n} \quad \text{where}^{4}\ s^*_{ij} = 0.5 \cdot s_{ij} \text{ if } i \neq j \text{ and } s^*_{ij} = s_{ij} \text{ otherwise.}$$

Furthermore, the S^* matrix is positive definite.

Hence, a dot product and a quadratic form[5] are considered as follows:

$$\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad \langle \mathbf{x}, \mathbf{y} \rangle_{QC} = \mathbf{x} S^* \mathbf{y}^t$$
$$QF : \mathbb{R}^n \longrightarrow \mathbb{R}, \qquad QF(\mathbf{x}) = \mathbf{x} S^* \mathbf{x}^t,$$

and given two images $\mathbf{I_x} = (f_1, f_2, \cdots, f_n)$ and $\mathbf{I_y} = (g_1, g_2, \cdots, g_n)$:

$$\langle \mathbf{I_x}, \mathbf{I_y} \rangle_{QC} = \mathbf{I_x} \cdot \mathbf{I_y} = \sum_{i=1}^{n} \sum_{j=1}^{n} f_i\, g_j\, s^*_{ij}, \qquad \text{(dot product)}$$

$$QF(\mathbf{I_x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_i\, f_j\, s^*_{ij}, \quad \text{(quadratic form)}$$

Hence, *QF* defines a norm in $\mathbb{R}^n$ as follows: $\|\mathbf{x}\| = \sqrt{QF(\mathbf{x})}$ for any $\mathbf{x} \in \mathbb{R}^n$. Hence, a 'quasi'-distance[6] in $\mathcal{Y}$, called QC-distance, $d_{QCD} : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$, is defined as:

$$d_{QC}(Image1, Image2) = \|\mathbf{I}_1 - \mathbf{I}_2\| \quad (3)$$

where $Image1 = \mathbf{I}_1 = (f_1, \cdots, f_n)$ and $Image2 = \mathbf{I}_2 = (f'_1, \cdots, f'_n)$.

4. Average global features for paintings

Images composed of color pixels try to reproduce objects in space which represent a scene in the real world. In our digital society, paintings can also be accessed as computerized images composed of pixels. Therefore, a characteristic that defines an image is how similar is to another with respect to their colors. The previous section explained how to calculate this feature using the QCD previously defined. This section presents some quantitative average global features for describing a painting regarding hue, saturation, lightness and brightness, which have been shown to be useful in the literature (Li & Chen, 2009).

The following set of quantitative features are selected to complement a global description of a digital painting regarding its color composition:

1. The arithmetic average of the hue coordinate.
2. The circular average of the hue coordinate (Jammalamadaka & Sengupta, 2001), which is convenient for hues around 0° (i.e. hues at 359° and 1° are perceptually almost identical shades of red; its circular mean is 0°, while its arithmetic mean is 180°, a completely different color).
3. Arithmetic average of the lightness coordinate.
4. Arithmetic average of the saturation coordinate in the HSL cylinder representation.
5. Arithmetic average of the saturation coordinate in the HSL bi-cone representation.
6. Average brightness, according to the following formula:

   $$\frac{1}{MN} \sum_{n} \sum_{m} L(m, n)$$

   where M and N are the dimensions of the image and $L(m, n)$ refers to the value of the Lightness coordinate at point (m, n) (formula (14) by Li and Chen (2009)).
7. Logarithmic average brightness,

   $$\frac{100}{MN} \exp\left(\sum_{n} \sum_{m} \log(\epsilon + \frac{L(m, n)}{100} \right)$$

   where M, N and $L(m, n)$ are defined as above and ϵ is added to prevent computing $\log(0)$. This is a simple scaling of formula (15) in Li and Chen (2009).
8. The brightness contrast is defined as $b - a$ where $[a, b]$ is the interval in the brightness histogram that centralizes 98% of the total energy; the precise calculation is detailed in Li and Chen (2009).

[4] A 0.5 factor is needed in order to avoid the duplicity of $f_i \cdot f_j$ when $i \neq j$.
[5] $\mathbf{x}^t$ means the transpose vector of $\mathbf{x}$.
[6] The distance condition $d(\mathbf{x}, \mathbf{y}) = 0 \Rightarrow \mathbf{x} = \mathbf{y}$ is not true.

Fig. 5. Paintings corresponding to the Baroque style (authored by Velázquez (a and b) and Vermeer (c)), Impressionism style (authored by Monet (d and e) and Renoir (f)) and Post-Impressionism style (authored by van Gogh (g and h) and Gauguin(i)) are shown.

5. Describing color composition of paintings in art styles

The styles selected as a baseline for this paper were the Baroque, Impressionism and Post-Impressionism. In Fig. 5, some art pieces corresponding to these styles are shown together with their authors.

The Baroque style started around 1600 in Rome, Italy, and it spread to the rest of Europe. The Baroque painting style exaggerated lighting, created by contrasting *dark* colors to *light-pale* colors. Mainly the paintings correspond to indoor scenes. They are also characterized by including ferroxide-based *yellows, oranges* and *reds* (Grygar, Hradilová, Hradil, Bezdička, & Bakardjieva, 2003).

The Impressionist style is conventionally taken to start with an exhibition organized in Paris in 1874 by Claude Monet, Edgar Degas and Camille Pissarro among others (Leroy, 1874). Impressionist painters captured the effects of sunlight by painting *en plein air* (outdoors); grays and dark tones were produced by mixing complementary colors since pure impressionism avoided the use of black paint. Rather than neutral *white, grays*, and *blacks*, impressionist painters often rendered shadows and highlights in color. The development of synthetic pigments provided the artists with vibrant shades of *blue, green*, and *yellow*. In paintings made *en plein air, bright* and *light* colors appear and shadows are boldly painted with the blue of the sky as it is reflected onto surfaces, giving a sense of freshness previously not represented in paintings (blue shadows on snow inspired this technique). Landscapes were also brought up to date with innovative compositions, light effects, and use of color (Samu, 2000).

The Post-Impressionist style (a name coined in 1910) broke with the conventions of the impressionist painters' concern to reproduce naturalistic light and color. Post-Impressionist painters continued using vivid colors, but they were more inclined to emphasize geometric forms, and to use unnatural or arbitrary color. Post-Impressionist paintings are mostly influenced by color contrast, specially *red* vs. *green* and *blue* vs. *yellow* (Mamassian, 2008). They also used complementary colors to create vibrant contrast and mutual enhancement when juxtaposed[7] and also to create shadows in adjacent objects since using the right hue, the grey color is obtained[8].

Note that some color logics can be extracted from these descriptions, which are intuitive for humans to understand.

It is also important to notice that the Impressionism and the Post-Impressionism styles overlapped in time, and that the latter appear as a consequence of the first one. In contrast, there is a difference of about 200 years between the Baroque style and both the Impressionism and the Post-Impressionism. Therefore, our hypothesis is that the paintings in Impressionism and Post-Impressionism styles will be more challenging to distinguish between them and that it will be easier to distinguish them from those in the Baroque style.

6. Machine learning using support vector machines

SVMs are learning machines which implement the structural risk minimization inductive principle to obtain good generalization on a limited number of learning patterns (Vapnik, 1998). This theory was developed on the basis of a separable binary classification problem where the optimization criterion is the width of the margin with ℓ_2-norm[9] between the positive and negative examples. A SVM with a large margin separating two classes has a small Vapnik-Chervonenkis dimension, which provides good generalization performance, as demonstrated in several applications (Cristianini & Shawe-Taylor, 2000). Let us present the standard SVM.

Let $\mathcal{Z} = \{z_i\}_{i=1}^{n} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ be a training set, with $x_i \in \mathcal{X}$ as the input space and $y_i \in \mathcal{Y} = \{\theta_1, \theta_2\} = \{+1, -1\}$ the output space. Let $\phi : \mathcal{X} \to \mathcal{F}$ be a feature mapping with a dot product denoted by $\langle \cdot, \cdot \rangle$. A linear classifier $f_{\mathrm{w}}(x) = \langle \mathrm{x}, \mathrm{w} \rangle + b$ is sought in $\mathcal{F}$, with $b \in \mathbb{R}$. Outputs are obtained in the form $h_{\mathrm{w}}(x) = \mathrm{sign}(f_{\mathrm{w}}(x))$.

For the standard SVM formulation (González, Angulo, Velasco, & Català, 2006; Vapnik, 1998), the optimization problem becomes

$$\begin{aligned} &\min_{\mathrm{w} \in \mathcal{F}, b \in \mathbb{R}} \frac{1}{2}\|\mathrm{w}\|^2 + C\sum_i \xi_i \\ &\text{s.t. } y_i(\langle \mathrm{x}_i, \mathrm{w} \rangle + b) + \xi_i \geq 1, \xi_i \geq 0, z_i \in \mathcal{Z} \end{aligned} \tag{4}$$

where C is the regularization term and ξ_i are slack variables. The solution can be written as $\mathrm{w} = \sum_i \alpha_i y_i \mathrm{x}_i$ where α_i are Lagrange multipliers for the dual problem of (4). Furthermore, $\sum_i \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, \alpha_i(y_i(\langle \mathrm{x}_i, \mathrm{w} \rangle + b) - 1 + \xi_i) = 0, i = 1, \cdots, n$.

Term b is calculated a posteriori (Gonzalez-Abril, Angulo, Velasco, & Ortega, 2008), and the classifier can be written as:

$$f(x) = \sum_i \alpha_i y_i \langle \mathrm{x}_i, \mathrm{x} \rangle + b$$

Although some joint SVM methods exist as the extension of binary classification to multi-classification (Angulo, Ruiz, González, & Ortega, 2006; Vapnik, 1998; Wang, Xue, & Chan, 2008), the binary ad-hoc methods of K one-versus-rest (1-v-r) or $K(K-1)/2$ one-versus-one SVMs for the solution of the multi-class problem still prevail due, in general, to their good performance and manageable optimization. In this paper, the 1-v-r approach is used which is briefly presented as follows.

Let $\{\theta_1, \cdots, \theta_K\}$ be a set of labels with $K \geq 2$ and $\mathcal{Z}_k = \{(x_i, y_i) : y_i = \theta_k\}$. In the 1-v-r SVM scheme, in a first decomposition phase, K binary classifiers are trained to generate functions $f_j(x) =$

[7] Tate Gallery:http://www.tate.org.uk/art/art-terms/c/complementary-colours

[8] Post-Impressionism Survey by Ami Werbel: http://academics.smcvt.edu/awerbel/Survey~of~Art~History~II/PostImpressionism.htm

[9] A generalization is given in Gonzalez-Abril, Velasco, Ortega, and Franco (2011).

$\langle w_j, x\rangle + b_j,\ 1 \le j \le K$, by separating training vectors $\mathcal{Z}_j$ with label θ_j from the rest of the training vectors $\mathcal{Z} \setminus \mathcal{Z}_j$ (j-v-r SVM).

In the reconstruction phase, a map $\Theta : \mathcal{X} \to \{\theta_1, \cdots, \theta_K\}$ is defined from the set of classifiers $F = \left\{f_j(x)\right\}_{j=1}^{K}$ such that, given an input vector x, it assigns a label to it as follows:

$$\Theta(x) = \arg \max_{j=1,\cdots,K} f_j(x)$$

7. Experimentation and results

In order to test the QArt-Learn approach, three painting styles were selected: Baroque, Impressionism and Post-Impressionism (described in Section 5). Two representative authors were selected for each style, Velázquez and Vermeer as Baroque painters, Monet and Renoir as Impressionists, and van Gogh and Gauguin as Post-Impressionists.

The dataset selected for the experimentation was the Painting-91[10] dataset by Khan et al. (2014). From this dataset, the paintings by the selected authors were extracted, which resulted in 252 images: 74 for Baroque style (39 by Velázquez and 35 by Vermeer), 85 as Impressionist paintings (36 by Monet and 49 by Renoir) and 93 as Post-Impressionist paintings (42 by van Gogh and 51 by Gauguin)[11].

Those images were normalized to 10.000 pixels. For each painting image, their QCD color features were extracted, for both parameterizations: (i) the original one with 37 color names and (ii) the modified parameterization with 28 color names (Section 2). Other quantitative global parameters were also obtained to further analyze the hue, lightness, saturation and brightness of these paintings (Section 4).

Furthermore, to avoid problems with the units of the variables, both qualitative and quantitative features were normalized as follows:

- Each one of the eight average global parameters are normalized to a maximum of 1. That is, given the values of the feature, $x_1, \cdots, x_m$, the normalized feature, denoted by x^N, is defined as: $x_i^N = \frac{x_i}{\max_i x_i}$, and hence, $0 \le x_i^N \le 1$ for any $i = 1, \cdots, m$ since the eight features considered are positive. Therefore, the weight of all quantitative feature is less to 8 for each painting.
- For the $n = 37$ or the $n = 28$ qualitative color features, the relative frequencies for each painting were obtained and then they were normalized to 8. That is, if $f_1, \cdots, f_n$ denote the relative frequencies then $\sum_{i=1}^{n} f_i = 1$, and the normalized frequencies, denoted by f_i^N, is defined as: $f_i^N = 8\, f_i$

Therefore, the quantitative and qualitative features have the same weight when a joint vector is built with them. Let us denote this vector as follows:

$$\mathbf{x} = (\mathbf{f}, \mathbf{q}) = (\mathbf{f_1}, \cdots, \mathbf{f_n}, \mathbf{q_1}, \cdots, \mathbf{q_8})$$

where $\mathbf{f} = (f_1, \cdots, f_n)$ are the relative frequencies ($n = 28$ or 37) and $\mathbf{q} = (q_1, \cdots, q_8)$ are the quantitative global parameters.

Let us indicate that a reference value for the accuracy of the classifiers can be obtained from two simple classifiers: a random classifier (that is, one that takes into account the number of paintings belonging to each style) can obtain an accuracy of $33.62\% = (\frac{74}{252})^2 + (\frac{85}{252})^2 + (\frac{93}{252})^2$ and a naive classifier (i.e. one that always says the style with the most of paints, in this cases, Post-Impressionism) will obtain an accuracy of $36.9\% = \frac{93}{252}$.

In the QArt-Learn approach, two machine learning methods were used to obtain an automatic classification of the paintings: k-Nearest Neighbor (KNN) and Support Vector Machines (SVMs). Let us indicate that these classifiers were chosen to contrast the performance of the K-NN, which uses a very simple algorithm, to the performance of the SVM, which applies a more sophisticated classification algorithm.

Regarding the K-NN (Nearest Neighbor) classifier, let us highlight that:

- if only the eight quantitative global parameters are considered, $\mathbf{x} = \mathbf{q} = (q_1, \cdots, q_8)$, then the QArt-Learn approach uses the Euclidean distance, $d_E(\cdot,\cdot)$.
- if only the 37 or the 28 qualitative color features are considered, $\mathbf{x} = \mathbf{f} = (f_1, \cdots, f_n)$, then the QArt-Learn approach applies the QC distance, $d_{QC}(\cdot,\cdot)$, from (3).
- if all features are considered, $\mathbf{x} = (\mathbf{f}, \mathbf{q}) = (f_1, \cdots, f_n, q_1, \cdots, q_8)$, then the QArt-Learn approach applies the distance, $d_{QC+E}(\cdot,\cdot)$ which is given from the following a dot product: $\mathbb{R}^{n+8} \times \mathbb{R}^{n+8} \longrightarrow \mathbb{R}$, $\langle \mathbf{x}, \mathbf{y}\rangle_{QC+E} = \mathbf{x} A, \mathbf{y}^t$, where $A = \left(\begin{array}{c|c} S^*_{n\times n} & 0_{n\times 8} \\ 0_{8\times n} & I_{8\times 8} \end{array}\right)$ with $I_{8\times 8}$ the identity matrix, as the same form as has been defined the QC distance, $d_{QC}(\cdot,\cdot)$, from (3).

In this experimentation, the number of considered nearest neighbors was $k = 2, \cdots, 20$. The number of training vectors, denoted by *NTraining*, was randomly chosen, where $NTraining = \{90 + u \cdot 15$, for $u = 0, 1, \cdots, 6\}$, and the rest of instances were the set of test vectors. This procedure was repeated 100 times in order to ensure good statistical behavior.

Regarding the SVM (Support Vector Machine) classifier, let us indicate that:

- if only the eight quantitative global parameters are considered, $\mathbf{x} = \mathbf{q} = (q_1, \cdots, q_8)$, then the QArt-Learn approach applies the Euclidean dot product, $\langle\cdot,\cdot\rangle_E$.
- if only the 37 or the 28 qualitative color features are considered, $\mathbf{x} = \mathbf{f} = (f_1, \cdots, f_n)$, then the QArt-Learn approach uses the QC dot product, $\langle\cdot,\cdot\rangle_{QC}$.
- if all features are considered, $\mathbf{x} = (\mathbf{f}, \mathbf{q}) = (f_1, \cdots, f_n, q_1, \cdots, q_8)$, then the QArt-Learn approach applies a mixed product, $\langle\cdot,\cdot\rangle_{QC} + \langle\cdot,\cdot\rangle_E$.

In this experimentation, the SVM was designed following a similar experimental framework to Angulo et al. (2006). The accuracy performance for the 1-versus-rest SVM was evaluated on models using a Radial Basis Function (RBF) kernel with σ (RBF width) and C (regularization term) initially explored on a two-dimensional grid, $\sigma = \left[2^{-4}, 2^{-3}, \ldots, 2^5, 2^6\right]$ and $C = \left[2^{-2}, 2^{-1}, \ldots, 2^6, 2^7\right]$, after a finer grid is chosen. The criteria selected to estimate the generalized accuracy was 10-fold cross-validation on the whole set of training data which is the same as that used in the k-NN. Similarly to k-NN, this procedure is repeated 100 times in order to ensure good statistical behavior.

The next sections show the results obtained using k-NN or SVMs methods and both parameterizations:

I the 37-QC-palette (Falomir et al., 2015a) and the 8 average global parameters (Section 4) and
II the 28-QC-palette (Sanz et al., 2015) and the 8 average global parameters (Section 4).

Then, the confusion matrix for $NTraining = 180$ using 6-NN and the 37-QC color palette is analyzed in relation to the hypothesis stated in Section 5.

[10] http://www.cat.uab.cat/~joost/painting91.html

[11] Note that the resulting dataset is slightly unbalanced with respect to the number of paintings belonging to each style. This is why, we wanted to preserve the original composition in the Painting-91 dataset and to avoid biased discards.

Table 3
Accuracy values obtained by k-NN using the 37-QC-palette and/or the 8 average color features.

	All features		37-QC-palette		Average features	
NTraining	K	%	K	%	K	%
180	6	**67.27**	8	66.69	6	65.86
165	6	**66.70**	5	65.79	4	65.57
150	6	65.31	11	**66.14**	6	64.71
135	4	64.70	9	**65.16**	4	64.45
120	4	64.43	6	**64.93**	4	63.79
105	4	63.94	8	**64.74**	4	62.75
90	9	63.93	7	**63.98**	4	62.13

7.1. Learning art styles using k-NN, SVMs and parameterization I

For each painting, 45 different features were extracted: 37 regarding their qualitative color (QC) palette, and 8 regarding quantitative global parameters. The results obtained by the k-NN are given in Table 3. The higher accuracy for each set of features is highlighted by a grey background and the best result for each *NTraining* time is indicated in bold.

Let us indicate that the accuracy results are similar independently of the number of neighbors used (k). Generally, the higher the training (*NTraining*), the higher the accuracy, for any training set. Note also that, by using 180 training sets, a higher accuracy is obtained, but it only differs in 3 points from that obtained by 90 training sets. Thus, as training further is not efficient, the QArt-Learn approach stops.

It is worth noting that, for any *NTraining*, the obtained accuracy is always higher when considering only the 37-color-palette than when using the eight average global parameters. Note also that sometimes the results are better when considering only the 37-color-palette than when considering all the features together.

The results obtained by the SVMs using the best cross-validation mean rate are shown in Table 4. The higher accuracy for each set of features is highlighted by a grey background and the best result for each *NTraining* time is indicated in bold.

Analyzing the results for any *NTraining*, note that the accuracy is always higher when using only the 37-QC-palette than when all features or the eight average global parameters are considered.

Both classifiers, k-NN and SVM, performed with an accuracy over 65% using the 37-QC-palette which shows that there is a qualitative color palette that can describe each style. The eight quantitative global average features are obtaining the lower accuracy results for both classifiers k-NN and SVM. Note that when adding these quantitative global average features the accuracy results obtained by SVM are not improved, but the results by the k-NN are improved slightly. This shows that both the 37-QC-palette and the global quantitative average features are quite equivalent for this classification problem. However, the 37-QC-palette, as it is based on color names, can be used to give a definition of the style in color names or to provide an explanation of the classification outliers, while the quantitative features does not have a correspondent alignment with linguistic concepts.

7.2. Learning art styles using k-NN, SVMs and parameterization II

In this experimentation, an adapted parameterization of the QCD with 9 less features than that applied in the previous section is used, that is, a 28-QC-palette. Thus, in this case, for each painting, 36 different features were extracted: 28-QC-palette, and 8 average global features.

The results obtained by the k-NN are shown in Table 5. The higher accuracy for each set of features is highlighted by a grey background and the best result for each *NTraining* time is indicated in bold.

Fig. 6. Dataset of Baroque-style paintings by Velázquez and Vermeer.

Note that a similar analysis to that provided in the previous section can be obtained from Table 5 and Table 3. It is worth noting than the accuracy obtained using 37-QC-palette is slightly higher than that obtained using the 28-QC-palette.

The results provided by the SVM using the best cross-validation mean rate are shown in Table 6.

Note also that the results show provide similar conclusions as those obtained before for Table 6 and Table 4. It is worth noting than the accuracy is the same when using the 37-QC-palette or the 28-QC-palette. Thus, this is an improvement since similar results are obtained with less features.

In this case, similar conclusions are obtained. Both classifiers, k-NN and SVM, performed with an accuracy over 66% using the 28-QC-palette. The eight quantitative global average features obtained the lower accuracy results for both classifiers k-NN and SVM. These additional global average features did not improved the accuracy results obtained by SVM, but improved the k-NN accuracy slightly in some cases. This shows that the data provided by these global quantitative average features is not relevant, while the qualitative color palettes can be used to obtain the same accuracy and to provide a definition of a style using color names.

7.3. Confusion matrix

As Section 5 hypothesizes, paintings belonging to Impressionism and Post-Impressionism styles are more difficult to distinguish for the human eye by only taking into account color features. The reader can experience this effect by observing the set of paintings shown in Figs. 7 and 8. In contrast, paintings belonging to Baroque style are clearly darker to the human eye. The reader can experience this effect by comparing the set of paintings in Fig. 6 with the set of paintings in Fig. 7 and 8.

In order to prove this intuitive hypothesis, the confusion matrix showing the best classification result (for $NTraining = 180$ using 6-NN and 37-QC-palette) has been obtained and analyzed (see Table 8).

Our hypothesis is supported by the data shown in Table 8. Note that, for any of the 3 configurations (all features, 37-QC-palette, Average Features), the percentage of success predicting the paintings belonging to Baroque style is higher in than 77% in all the cases, even getting higher than 83% when only the qualitative color features are used.

In contrast, all the configurations obtain worst results when distinguishing between paintings belonging to the Impressionism and

Table 4
Accuracy values obtained by SVMs using the 37-color-palette and/or the 8 average color features.

	All features		37-QC-palette		Average features	
NTraining	(σ, C)	%	(σ, C)	%	(σ, C)	%
180	(1.68, 1.19)	64.87	(1.68, 1.00)	**65.34**	(0.21, 1.00)	63.02
165	(2.00, 1.00)	64.28	(1.68, 2.00)	**64.82**	(0.25, 1.00)	62.84
150	(2.00, 1.00)	64.55	(1.69, 1.41)	**65.13**	(0.30, 0.84)	62.37
135	(1.68, 1.41)	63.57	(2.00, 0.71)	**65.56**	(0.30, 1.00)	62.29
120	(2.00, 0.84)	63.52	(2.00, 1.19)	**64.91**	(0.35, 0.70)	62.61
105	(2.00, 0.84)	63.07	(2.00, 0.84)	**64.59**	(0.29, 1.00)	62.38
90	(1.68, 1.68)	62.80	(1.68, 1.00)	**64.18**	(0.35, 0.84)	61.51

Table 5
Accuracy values obtained by k-NN using the 28-QC-palette and/or 8 average color features.

	All features		28-QC-palette		Average features	
NTraining	K	%	K	%	K	%
180	7	**66.41**	5	66.33	6	65.86
165	6	**66.66**	5	66.55	4	65.57
150	6	66.43	4	**66.55**	6	64.71
135	6	65.34	5	**65.63**	4	64.45
120	6	**65.49**	6	64.29	4	63.79
105	6	**63.93**	10	63.91	4	62.75
90	4	**63.44**	6	63.20	4	62.13

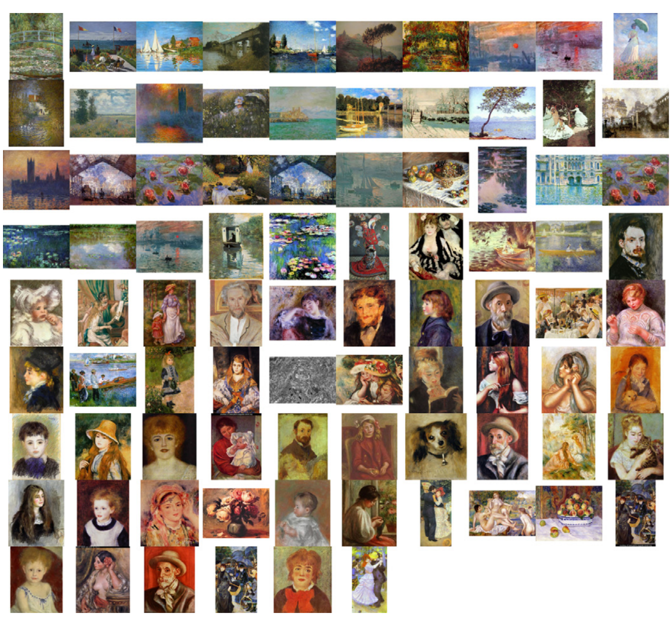

Fig. 7. Dataset of Impressionism-style paintings by Monet and Renoir.

Fig. 8. Dataset of Post-Impressionism-style paintings by van Gogh and Gauguin.

Post-Impressionism styles since the number of misclassification errors is high.

It can be seen from the experiments that there are similarities between the Impressionist and Post-Impressionist styles. Thus, checked whether the accuracy is improved by merging both styles together. We ran the experimentation again using a k-NN classifier to distinguish between Baroque and Impressionist/Post-Impressionist styles. The results are given in Table 8. and the confusion matrix showing the best classification result (for $NTraining = 180$ using 6-NN and 37-QC-palette) is shown in Table 9. From these two tables it can be seen that the accuracy is improved when distinguishing between Baroque and the combination of Impressionism/Post Impressionism (from 63 to 65% with three classes to 88–89% with two). However, the error percentage for Baroque has risen, which is motivated by the class imbalance in this experiment: there are 178 Impressionist/Post Impressionist paintings, but only 74 in the Baroque style.

Table 6
Accuracy obtained by SVMs using the 28-QC-palette. The higher accuracy for each set of features is highlighted by a grey background and the best result for each *NTraining* time is indicated in bold.

	All features		28-QC-palette		Average features	
NTraining	(σ, C)	%	(σ, C)	%	(σ, C)	%
180	(2.00, 0.71)	64.54	(2.00, 0.71)	**65.36**	(0.21, 1.00)	63.02
165	(2.00, 1.00)	64.13	(2.00, 0.71)	**66.05**	(0.25, 1.00)	62.84
150	(2.00, 1.00)	63.78	(2.00, 0.71)	**64.94**	(0.30, 0.84)	62.37
135	(2.00, 1.41)	63.02	(2.00, 0.84)	**64.08**	(0.30, 1.00)	62.29
120	(2.00, 1.41)	61.92	(2.00, 1.00)	**63.96**	(0.35, 0.70)	62.61
105	(2.00, 1.41)	62.05	(2.00, 1.41)	**63.24**	(0.29, 1.00)	62.38
90	(2.00, 1.41)	**62.56**	(2.83, 0.35)	61.98	(0.35, 0.84)	61.51

Table 7
Confusion matrix for *NTraining* = 180 using 6-NN and the 37-QC-palette.

	Real	Prediction			Total	Percentage of success		
		Post-Imp.	Imp.	Baroque	7200	Post-Imp.	Imp.	Baroque
All features	Post-Impressionism	1837	546	265	2648	69.37	20.62	10.01
	Impressionism	1018	1215	191	2424	42.00	50.12	7.88
	Baroque	176	202	1750	2128	8.27	9.49	82.24
37-QC	Post-Impressionism	1763	668	270	2701	65.27	24.73	10.00
	Impressionism	945	1269	177	2391	39.52	53.07	7.40
	Baroque	126	228	1754	2108	5.98	10.82	83.21
Only Average features	Post-Impressionism	1859	599	190	2648	70.20	22.62	7.18
	Impressionism	1089	1148	185	2422	44.96	47.40	7.64
	Baroque	241	231	1658	2130	11.31	10.85	77.84

Table 8
Accuracy values obtained by *k*-NN using the 37-QC-palette and/or the 8 average color features by considering Baroque versus Impressionist/Post-Impressionist paintings.

	All features		37-QC-palette		Average features	
NTraining	K	%	K	%	K	%
180	8	88.88	6	89.18	4	89.43
165	8	88.53	8	88.76	8	88.98
150	8	88.77	8	88.82	6	89.03
135	8	88.68	8	88.75	4	88.58
120	6	88.43	6	88.64	4	88.15
105	9	88.39	6	88.37	4	87.15
90	8	88.23	6	88.42	9	87.31

Results obtained by 28-QC-palette and SVM learning are not provided since they are similar to those in Table 8, and the same conclusions can be extracted.

8. Discussion

In this section, we comment the results obtained in the experimentation and other methods that appeared in the literature.

8.1. Regarding the results obtained by QArt-Learn

The quantitative average global features did not increased the accuracy of the classification. In general, low-level features that are usually exploited for photographic image analysis are not as effective for artistic image identification, since artists may use color and textures patterns in an atypical way; for example, orange or red grounds, or a purple sky, appear in some Post-Impressionist paintings (see Fig. 5). Moreover, painters usually try to differentiate their art pieces so that each of them becomes unique. This is why it could be easier to find similarities in the color used by authors in the same style (time period) that in the paintings produced by the same painter in all their life, since an artist can paint in different styles in different time periods.

Regarding the different parameterizations, the best classification results (66.05%) are obtained using the 28-QC-palette. This indicates that the level of detail that the classifiers needs is not high, since more data does not result in a higher accuracy. Note also that, even using relatively simple classifiers, such as *k*-NN classifiers, the representation used is strong enough to obtain satisfactory accuracies.

There is still research to be done in order to enrich the classification with variables showing the color style description presented in Section 5. These color logics enriched by art experts' opinion can provide the starting point to provide feedback to the users, so that the classification criteria followed can be explained and the classifiers can give hints about the reason of misclassifications.

8.2. Regarding other methods in the literature

Convolutional neural network (CNN) methods have been successful in different applications, including painting classification; e.g. Karayev et al. (2013) when classifying Baroque (81.45%), Impressionism (82.15%), and Post-Impressionism (74.51%). However, whereas their expressiveness is their main advantage, it is also the reason that make them learn non-interpretable solutions that could have counter-intuitive properties (Szegedy et al., 2013).

That motivates the current push towards a next-generation of learning algorithms that are more cognitive or intuitive to people (Vellido, Martín-Guerrero, & Lisboa, 2012). That is, they must generate explanations to communicate their knowledge while learning. They need to self-explain, in the same way children can tell us their reasoning arguments after they arrived to a weird conclusion for us. When children use language (concepts and semantics), we can understand their deductive reasoning line, then we can show them where they are wrong and how they can improve their reasoning deduction by adding other things (or parameters) that they can verify by their own if they would like to. With that explanation, we can identify which is the child's wrong assumption or which is our wrong assumption, and we can correct misunderstandings.

Table 9
Confusion matrix for *NTraining* = 180 using 6-NN and the 37-QC-palette.

	Real	Prediction		Total	Percentage of success	
		Impressionism/Post-Imp.	Baroque	7200	Impressionism/Post-Imp.	Baroque
All features	Impressionism/Post-Imp.	4727	313	5040	93.79	06.21
	Barroque	513	1647	2160	23.75	76.25
37-QC	Impressionism/Post-Imp.	4780	305	5085	94.00	6.00
	Baroque	537	1568	2105	25.51	74.49
Average features	Impressionism/Post-Imp.	4943	219	5162	95.76	4.24
	Baroque	580	1458	2038	28.46	71.57

Machine learning algorithms are very convenient to learn repetitive tasks, even they can help optimizing the actions for the users. However, most of them are 'black boxes', that is, they cannot provide explanations of their result. We are interested in having machine learning approaches that are able to provide explanations; otherwise they would not be applicable to human-machine interaction scenarios. Qualitative descriptors are key to provide those explanations.

The strength of qualitative representations is that the use concepts aligned with human understanding. According to Iwasaki (1997), qualitative reasoning “produces predictions directly in terms of salient qualitative characteristics of the behavior which requires much less effort to interpret.”. This is used to capture features in the way that human experts in the field would express (e.g. levels of contrast, color intensity, the qualities of the used palette, ...) instead of raw, precise numerical values.

8.3. Discussing the results with experts in art

We presented our results to an expert on art at the Contemporanean Art Museum *Vicent Aguilera Cerni*, MACVAC[12] in Vilafamés, Spain. Here we summarize our discussion.

The expert explained to us that the basic information for categorising a piece of art in a style is its year of painting. However, the year by itself is not enough to provide a precise classification (styles overlap in time, Impressionism and Post-Impressionism being particular examples; and, additionally, it is not uncommon for artists to use anachronistic painting styles), and in many cases this datum is unknown. Thus, experts usually analyse the composition to try to infer its style. For that, the used colors and the contrast between lightning areas and shadows are essential. In the Baroque style, there is no gradation between light and dark colors, and the contrasting of lighting and darkness is very characteristic of this style. However, in the Impressionism and Post-impressionism style, there are very few shadows and lighting colors are very usual in the paintings. The expert also explained that the variability of author's style in the Post-impressionism style is very high.

Other aspects to take into account in the categorization of pieces of art were also highlighted by the expert, e.g. the composition lines. In the Baroque style the use of a diagonal composition is very common, while in other styles it is not. Moreover, in the Impressionism style, painters wanted the observer to perceive the painting as a subjective composition, this is why their brush strokes where spaced, so that the whole scene would be “assembled” by the observers' brain.

In conclusion, colors take an important role in painting categorization, as QArt-Learn approach shows. It is more challenging to differentiate between styles which are close in time, which it is also showed by the results presented in this paper. Moreover, there are other qualitative aspects that must be taken into account when categorising paintings such as the form of the brushstroke or the geometry of the whole composition. This aspects will be taken into account in our studies in the near future.

9. Conclusions

In this paper, the QArt-Learn approach is presented, which automatize the color categorization of paintings in three art styles: Baroque, Impressionism and Post-Impressionism. The QArt-Learn approach is based on a Qualitative Color Descriptor (QCD), its associated similarity (*SimQCD*) and some average global color features. The QCD can be customized depending on the context of the problem.

Two machine learning classifiers, K-Nearest Neighbor (K-NN) and Support Vector Machines (SVMs) were built and tested together with two different parameterizations of QCD: 37-QC-palette and 28-QC-palette. The results of this experimentation show that both classifiers, *k*-NN and SVM, performed with an accuracy over 65% using the 37-QC-palette or the 28-QC-palette, showing that a color palette may describe a painting style. The quantitative global average features did not improve the accuracy results obtained significantly.

Let us highlight that, using the color names used by the QArt-Learn approach, an artificial agent could describe a painting color style to a non-expert human.

The results provided by a confusion matrix confirmed our hypothesis that the Impressionism and the Post-Impressionism styles are more challenging to differentiate using QArt-Learn approach because the color palettes are more similar.

As future work, we will intend to extend the approach to more painters and styles, and to provide an explanation of the classification outliers using also the colors appearing or not appearing in the paintings. We would also like to improve the categorization of styles in accuracy but also in understandability by studying other classification methods (i.e. decision tress, learning rule methods (Thabtah, 2007)). We also intend to compare the results of classification to that provided by human users (experts and non-experts), by running a set of questionnaires.

Finally, we are currently exploring several applications that could be based on an extension of the work presented in this paper. One is the estimation of the market price of a painting, which is currently done solely by experts without any support by automated tools.

Acknowledgments

Dr.-Ing. Zoe Falomir gratefully acknowledges the project *Cognitive Qualitative Descriptions and Applications* (CogQDA) project funded bythe Universität Bremen through the *04-Independent Projects for Postdocs action*.

Dr. Gonzalez-Abril acknowledges funding by the Spanish Ministry of Economy and Competitiveness (HERMES,TIN2013-46801-C4-1-r) and by the Andalusian Regional Ministry of Economy, Innovation and Science (Simon, TIC-8052).

Dr. Museros and Dr. Sanz acknowledge funding by the Spanish Ministry of Economy and Competitiveness (TIN2014-55335-R), by Generalitat Valenciana (GV/2015/102) and by Universitat Jaume I (P1 · 1B2013-29).

Appendix

The QArt-Learn-Dataset of paintings used in this paper corresponding to Baroque style, Impressionism style and Post-Impressionism style are shown in Fig. 6, Fig. 7 and Fig. 8, respectively.

References

Angulo, C., Ruiz, F., González, L., & Ortega, J. A. (2006). Multi-classification by using tri-class SVM. *Neural Proceeding Letters, 23*(1), 89–101.
Cohn, A. G., & Renz, J. (2007). *Qualitative spatial reasoning, handbook of knowledge representation*. Wiley-ISTE, London: Elsevier.
Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University press 2000.
Falomir, Z., Museros, L., & Gonzalez-Abril, L. (2015a). A model for colour naming and comparing based on conceptual neighbourhood. an application for comparing art compositions. *Knowledge-Based Systems, 81*, 1–21. doi:10.1016/j.knosys.2014.12.013.
Falomir, Z., Museros, L., Gonzalez-Abril, L., & Sanz, I. (2013). A model for qualitative colour comparison using interval distances. *Displays, 34*, 250–257. doi:10.1016/j.displa.2013.07.004.

[12] Museum MACVAC http://macvac.vilafames.es/

Falomir, Z., Museros, L., Sanz, I., & Gonzalez-Abril, L. (2015b). Guessing art styles using qualitative colour descriptors, SVMS and logics. . In E. Armengol, D. Boixader, & F. Grimaldo (Eds.), *Artificial Intelligence Research and Development*. In *Frontiers in Artificial Intelligence and Applications: 277* (pp. 227–236). IOS Press. doi:10.3233/978-1-61499-578-4-227.

Freksa, C. (2013). Spatial computing–how spatial structures replace computational effort. In M. Raubal, D. Mark, & A. F. (Eds.) (Eds.), *Cognitive and linguistic aspects of geographic space*. Heidelberg: Springer.

Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. arXiv:1508.06576.

González, L., Angulo, C., Velasco, F., & Català, A. (2006). Dual unification of bi-class support vector machine formulations. *Pattern Recognition, 39*(7), 1325–1332.

Gonzalez-Abril, L., Angulo, C., Velasco, F., & Ortega, J. (2008). A note on the bias in SVMs for multi-classification. *IEEE Transactions on Neural Networks, 19*(4), 723–725.

Gonzalez-Abril, L., Velasco, F., Ortega, J., & Franco, L. (2011). Support vector machines for classification of input vectors with different metrics. *Computers & Mathematics with Applications, 61*(9), 2874–2878.

Gonzalez-Abril, L., Velasco, F., Ortega, J. A., & Cuberos, F. J. (2009). A new approach to qualitative learning in time series. *Expert Systems with Applications, 36*, 9924–9927.

Grygar, T., Hradilová, J., Hradil, D., Bezdička, P., & Bakardjieva, S. (2003). Analysis of earthy pigments in grounds of baroque paintings. *Analytical and Bioanalytical Chemistry, 375*(8), 1154–1160. doi:10.1007/s00216-002-1708-x.

Iwasaki, Y. (1997). Real-world applications of qualitative reasoning. *IEEE Expert: Intelligent Systems and Their Applications, 12*(3), 16–21. doi:10.1109/64.590068.

Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in circular statistics* (Har/dskt). World Scientific Pub Co Inc.

Jiang, S., Huang, Q., Ye, Q., & Gao, W. (2006). An effective method to detect and categorize digitized traditional chinese paintings. *Pattern Recognition Letters, 27*(7), 734–746. doi:10.1016/j.patrec.2005.10.017.

Johnson, C. R., Hendriks, E., Berezhnoy, I., Brevdo, E., Hughes, S., Daubechies, I., et al. (2008). Image processing for artist identification - computerized analysis of Vincent van Gogh's painting brushstrokes. *IEEE Signal Processing Magazine, Special Issue on Visual Cultural Heritage, 25*(4), 37–48. doi:10.1109/MSP.2008.923513.

Karayev, S., Hertzmann, A., Winnemoeller, H., Agarwala, A., & Darrell, T. (2013). Recognizing image style. *CoRR, abs/1311.3715*.

Khan, F. S., Beigpour, S., van de Weijer, J., & Felsberg, M. (2014). Painting-91: A large scale database for computational painting categorization. *Machine Vision and Applications, 25*(6), 1385–1397. doi:10.1007/s00138-014-0621-6.

Leroy, L. (1874). The exhibition of the impressionists. *Le Charivari*.

Li, C., & Chen, T. (2009). Aesthetic visual quality assessment of paintings. *Selected Topics in Signal Processing, IEEE Journal of, 3*(2), 236–252.

Li, J., Yao, L., Hendriks, E., & Wang, J. Z. (2012). Rhythmic brushstrokes distinguish van Gogh from his contemporaries: Findings via automated brushstroke extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(6), 1159–1176. doi:10.1109/TPAMI.2011.203.

Mamassian, P. (2008). Ambiguities and conventions in the perception of visual art. *Vision Research, 48*(20), 2143–2153. Vision Research Reviews doi: 10.1016/j.visres.2008.06.010.

Mamassian, P., Knill, D. C., & Kersten, D. (1998). The perception of cast shadows. *Trends in Cognitive Sciences, 2*(8), 288–295. doi:10.1016/S1364-6613(98)01204-2.

Mensink, T., & van Gemert, J. (2014). The Rijksmuseum challenge: Museum-centered visual recognition. *ACM international conference on multimedia retrieval (icmr)*.

Museros, L., Sanz, I., Falomir, Z., & Gonzalez-Abril, L. (2016). A qualitative color harmony theory. In À. Nebot, X. Binefa, & R. L. de Mántaras (Eds.), *Artificial intelligence research and development - proceedings of the 19th International Conference of the Catalan Association for Artificial Intelligence (CCIA), Barcelona, Catalonia, Spain, october 19–21, 2016*. In *Frontiers in Artificial Intelligence and Applications: 288* (pp. 98–107). IOS Press. doi:10.3233/978-1-61499-696-5-98.

Orlov, N., Johnston, J., Macura, T., Shamir, L., & Goldberg, I. (2007). Computer vision for microscopy applications. In G. Obinata, & A. D. Eds (Eds.), *Vision systems–segmentation and pattern recognition* (pp. 221–242). Viena, Austria: ARS Pub..

Samu, M. (2000). *Impressionism: Art and modernity*. New York: The Metropolitan Museum of Art.

Sanz, I., Museros, L., Falomir, Z., & Gonzalez-Abril, L. (2015). Customising a qualitative colour description for adaptability and usability. *Pattern Recognition Letters, 67*, 2–10. doi:10.1016/j.patrec.2015.06.014.

Shamir, L., Macura, T., Orlov, N., Eckley, D. M., & Goldberg, I. G. (2010). Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Transactions on Applied Perception, 7*(2), 8:1–8:17. doi:10.1145/1670671.1670672.

Shamir, L., & Tarakhovsky, J. A. (2012). Computer analysis of art. *Journal on Computing and Cultural Heritage, 5*(2), 7:1–7:11. doi:10.1145/2307723.2307726.

Shen, J. (2009). Stochastic modeling western paintings for effective classification. *Pattern Recognition, 42*(2), 293–301. doi:10.1016/j.patcog.2008.04.016.

Siddiquie, B., Vitaladevuni, S., & Davis, L. (2009). Combining multiple kernels for efficient image classification. In *Applications of computer vision (WACV), 2009 workshop on* (pp. 1–8). doi:10.1109/WACV.2009.5403040.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. J., & Fergus, R. (2013). Intriguing properties of neural networks. *CoRR, abs/1312.6199*.

Thabtah, F. (2007). A review of associative classification mining. *Knowledge Engineering Review, 22*(1), 37–65. doi:10.1017/S0269888907001026.

Vapnik, V. (1998). *Statistical learning theory*. John Wiley & Sons, Inc.

Vellido, A., Martín-guerrero, J., & Lisboa, P. J. G. (2012). Making machine learning models interpretable. *In proc. European symposium on artificial neural networks, computational intelligence and machine learning*.

Wang, L., Xue, P., & Chan, K. (2008). Two criteria for model selection in multiclass support vector machines. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on, 38*(6), 1432–1448.

Yelizaveta, M., Tat-Seng, C., & Irina, A. (2005). Analysis and retrieval of paintings using artistic color concepts. In *Multimedia and expo, 2005. ICME 2005. IEEE international conference on* (pp. 1246–1249). doi:10.1109/ICME.2005.1521654.