Automatically Classifying Paintings with Perceptual Inspired Descriptors

Răzvan George Condorovici RCONDOROVICI@IMAG.PUB.RO
Image Processing and Analysis Laboratory
University "Politehnica" of Bucharest, Romania, Address Splaiul Independenţei 313

Corneliu Florea[1] CORNELIU.FLOREA@UPB.RO
Image Processing and Analysis Laboratory
University "Politehnica" of Bucharest, Romania, Address Splaiul Independenţei 313

Constantin Vertan CONSTANTIN.VERTAN@UPB.RO
Image Processing and Analysis Laboratory
University "Politehnica" of Bucharest, Romania, Address Splaiul Independenţei 313

Abstract

We propose a framework for the automatic recognition of artistic genre in digital representations of paintings. As we aim to contribute to a better understanding of art by humans, we extensively mimic low-level and medium-level human perception by relying on perceptually inspired features. While Gabor Filter Energy has been applied in art description, Dominant Color Volume (DCV) and frameworks extracted using anchoring theory are novel in this field. To perform the actual genre recognition, we rely on a late fusion scheme based on combining Multi–Layer Perceptron (MLP) classified data with Support Vector Machines (SVM). The performance is evaluated on an extended database containing more than 4000 paintings from 8 different genres, outperforming the reported state of the art.

1. Introduction

George Bernard Shaw said that "without art, the crudeness of reality would make the world unbearable" acknowledging that art has accompanied the human evolution through his entire history. With the late growth of computers usage in daily life, the art world began to be dissected by artificial, intelligent systems. Tremendous efforts were put lately into creating automatic image processing solutions that facilitate a better understanding of art (Cornelis et al., 2011), either by obtaining high-quality and high-fidelity digital versions of paintings (Martinez et al., 2002), either by targeting subjects like image analysis and diagnostics, virtual restoration, color rejuvenation, pigment analysis, brush stroke analysis, lightning incidence, perspective anomalies detection, three dimensional space recovery, craquelure analysis or painting authentication, etc. as discussed in the review of Stork et al. (Stork, 2009). While gathering more than 20 years of intensive research, digital investigation of visual art has not yet answered all questions.

A crucial aspects for artwork understanding is to successfully place it into a context. Typically, two cases are envisaged: a narrower one which is to nominate the painter and a broader one, namely recognizing the artistic genre. The state of the art in automatic identification of the context of a painting, although witnessed noticeable results, still offers space for improvements. The current proposal lies into the second category, as we describe a system for automatic identification of artistic genres.

1.1. Related work

In reviewing solutions to the artistic context for recognition problem (both painter and genre), we identify the very typical pattern recognition approach: first, using features, digitized paintings are described, than a learning scheme is employed to extract common and respectively discriminative traits among envisaged classes. A condensate overview of the state of the art methods may be followed in Table 1.

Automatic identification of the painter proved to be more popular in the early stages. Thus, adopting the cosine transform for extract repetitive texture features linked to a

Table 1. Artistic genre/painter recognition methods: main differences.

Method	Recognizes	No. of classes	Desc. level	Features	Learning scheme
Keren (Keren, 2002)	Painter	5	High	Spectral (Cosine)	Naive Bayes
Li (Li and Wang, 2004)	Painter	5	High	Wavelet	2D-MHMM
Widjaja (Widjaja et al., 2003)	Painter	4	High	Color, Skin texture	SVM
Khan (Khan et al., 2010)	Painter	10	Low	Color, Shape	BoW
Gunsel (Gunsel et al., 2005)	Genre	3	Low	Luminance, Color	PCA-SVM
Zujovic (Zujovic et al., 2009)	Genre	5	Low	Texture, Edge, Color	AdaBoost
Shamir (Shamir et al., 2010)	Genre/Painter	3/9	High	Edge, Texture	WNN
Arora (Arora et al., 2012)	Genre	7	High	Classemes	BoW/SVM
Condorovici (Condorovici et al., 2013)	Genre	6	High	Dominant Color, Anchors (Shape), Gabor	Bagged ensemble of trees
Proposed	Genre	8	High	Dominant Color, Anchors (Shape), Gabor	Late fusion

Naive Bayes Classifier (NBC), Keren (Keren, 2002) identified several painters. In the same line, Li and Wang (Li and Wang, 2004) used 2–dimensional Multiresolution Hidden Markov Models (MHMM) over wavelet extracted features to classify five Chinese ink painters. Widjaja et al. (Widjaja et al., 2003) identified four painters based on selected skin samples (described from both color and texture point of view) with a reported accuracy of 85%. More recently, Khan et al. (Khan et al., 2010) combined color and shape information in a Bag of Word (BoW) approach to recognize among 10 painters out of 400 images. Yet, taking into account that a painter is consistently more conservative in the approached themes and in the techniques than his peers, even from the same current, the painter recognition is rather less intricate when compared with the genre recognition, which requires an extended level of abstraction. Furthermore, due to the finite work capacity of any human, the amount of paintings authored by a single artist is also limited, thus in painter recognition cases, the database are confined to less than 500 examples (with a maximum of 50 examples per class/painter).

The other direction of the context recognition, namely the artistic genre recognition, is more difficult, sometimes even for the specialists, due to the natural variation within the artistic genres. In this direction, we identify two types of systems: relying on low level features (such as pixel luminances and color means or as total edginess) and relying on high level features. Systems with low-level features were proposed by Gunsel et al. (Gunsel et al., 2005), which dissociate three genres based on six basic features extracted only from the luminance image and by Zujovic et al. (Zujovic et al., 2009) who relies on a set of gray-level features for a five–genre classification. The downside of these methods is the reduced number of paintings used to test the systems (107 paintings for (Gunsel et al., 2005) and 353 paintings for (Zujovic et al., 2009)).

Acknowledging the task difficulty, the solutions from the second class introduce larger sets and higher complexity of the features. More recently, Shamir et al. (Shamir et al., 2010) adopted an extensive set of 548 features, out of which, by means of the Fisher criterion filtering, selected the most discriminative 83 ones, coupled with a weighted nearest neighbor (WNN) classifier; as a result they discriminated among 9 schools of art within 3 artistic currents for a reported accuracy of 77% within a database of 517 images. In the same line, Arora and Elgammal (Arora et al., 2012) described paintings with Classemes (Torresani et al., 2010) framework and distributed them in 7 currents by means of a Bag of Words (BoW) schema with a Support Vector Machine (SVM) as classifier. Yet, the use of complex features opens the way for high accuracy only in narrow cases (e.g. specific artistic identification) and within confined variation.

In all the mentioned methods, the results are somehow restricted in generalization due to the limited size of the database (i.e. less than 1000 examples).

1.2. Paper structure

To motivate our construction we recall that Michelangelo wrote down in Middle Ages that “a man paints with his brains and not with his hands”. Furthermore, although computer based discrimination among artistic genre is difficult, Wallraven et al. (Wallraven et al., 2009) noted that

non expert humans still achieve considerably larger scores than computers. Thus, we claim that the key to better accuracy is to rely on features compatible with human perception.

We addressed the problem from a perceptual point of view and we constructed the descriptors to be highly correlated with human perception, thus encoding the major classes of perceptual features: luminance and shapes, color and, respectively, texture and edge.

To ensure proper coverage of this problem, we propose a new color descriptor named Dominant Color Volume and for the dominant luminance levels, we introduce the anchoring theory into the art digital analysis. For recognition, we employed a late fusion scheme, as the human process first each category of data and then aggregate the results. The efficiency of the proposed system is tested on a unrestrictive database of some 4200 paintings from 8 artistic genres yielding high within-current and cross-current variation.

In the continuation of this paper, the motivational overview of the proposed system and the descriptive features are presented in section 2; the data set and the classifier details are given in section 3. Finally, the results obtained with the proposed system are discussed in section 4, while the last section is dedicated to conclusions and perspectives.

2. Feature Extraction

There were many attempts to unravel the human understanding of art from a neuro-scientific point of view. The first significant results were disclosed by Zeki (Zeki, 1999), who showed that different elements of visual art, such as shapes, colors, and boundaries, are processed by different pathways and systems in the brain, designed to interpret each aspect of the art and there is no single central mechanism that receives and interpret visual art, but instead, pieces of information received from a painting are selectively redistributed to more specialized centers for processing.

Ramachandran and Herstein (Ramachandran and Herstein, 1999) identified as the key for understanding the art perception to be the identification of the perceptual processes, rather than the analysis of the aesthetic properties, augmenting Zeki's tweak on Michelangelo statement ("the painter does not paint with his eyes, but with his brain"). Thus we divided our set of features into three categories, each closely connected with one of the important perceptual elements: lightness perception and shape extraction, color distribution and, respectively, texture and edge analysis.

For the image shapes and lightness description we relied on the anchoring complex image decomposition, derived from the gestalt (shape) theory; for the color, we computed the Minimum Volume Enclosing Ellipsoid over the 3D Lab Color Histogram to get the Dominant Color Volume (DCV), while for textures and edges we employed the Gabor energy. These features are presented in Table 2 and are extracted for each painting.

2.1. Anchoring Theory and Frameworks

Although many studies attempted to explain and to mimic the human perception of lightness and scene decomposition, no definite model exists. However, the reformulation by Gilchrist et al. (Gilchrist et al., 1999) of the anchoring theory for complex scene proved to pass many perceptual tests and explained many phenomena.

The anchoring theory states that when depicting a scene, the relation between the representation luminance and the scene lightness can be correctly perceived only through a mapping between the luminance value and the value on the scale of perceived gray shades, process called *anchoring*. Once the anchor is established, the lightness value for each luminance is determined as the ratio between the value and the anchor.

For determining the anchoring there are two known theories: the average luminance rule and the highest luminance rule. The first approach states that the average luminance in the visual field is perceived as middle gray, such that luminance values should be anchored by their average value to middle gray. The highest luminance rule anchors the highest luminance level to a lightness value perceived as white. However, some experiments (Li and Gilchrist, 1999) proved that if a small white area is surrounded by dark areas it appears to be self-luminous, being perceived as lighter than white. As a consequence, the highest luminance rule was redefined such that the highest dominant luminance within the field of view becomes a stable anchor. However, when the darker areas become dominant, the anchor is determined as a weighted average of the luminance proportionally to the occupied area.

Li and Gilchrist (Li and Gilchrist, 1999) provided rich experimental evidence that favored the highest luminance rule in comparison to the average luminance rule.

When it comes to increasingly complex scenes, the anchoring theory asserts that scenes are perceived by the humans in terms of consistent areas, called *frameworks*. A framework is defined as a region of common illumination. Regarding the image perception, the human brain estimates the lightness within each framework through the anchoring to the luminance perceived as white, followed by the computation of the global lightness. We claim that scene decomposition in frameworks is crucial for unravelling the

Table 2. Features extracted from paintings and used for classification.

Feat. No.	Type Feature	Description	Interpretation	Number of values
1.1-1.2	*Anchor:* Region Cardinality	No. of pixels belonging to the current framework	Indication about the spatial relationship in the painting	2 Fmwks $\times$1=2
1.3-1.4	*Anchor:* Articulation	Measure of the dynamic range in the framework	The luminance variation inside the current region	$2 \times 1 = 2$
1.5-1.6	*Anchor:* Mean Value	The average luminance of the current framework	Indicates the luminance value of the current region	$2 \times 1 = 2$
1.7-1.10	*Anchor:* Center pos.	Position of the region center inside the image	Indicates the position of the framework in the painting	$2 \times 2 = 4$
2.1-2.9	*DCV:* Directions	Principal directions of the ellipsoid, computed as the eigenvectors of the matrix representing the ellipsoid	Indication about the main colors used in the painting	$3 \times 3 = 9$
2.10-2.12	*DCV:* Geometric Center	Coordinates of the ellipsoid center	Indication about the painting dominant color	$1 \times 3 = 3$
2.13-2.15	*DCV:* Axis length	Length of the ellipsoid axis, computed as the eigenvalues of the matrix representation	Indication about the color gamut size	$3 \times 1 = 3$
2.16-2.18	*DCV:* Center of mass	Center of mass coordinates, computed as the weighted average of the current mode bins	The painting dominant color; together with geometric center, offers information about the mode's shape	$1 \times 3 = 3$
2.19	*DCV:* Geometric Volume	Ellipsoid's volume	Information about the gamut's size	$1 \times 1 = 1$
2.20	*DCV:* Mass	Number of pixels belonging to the current mode	Information about the colors variation in the painting	$1 \times 1 = 1$
2.21	*DCV:* Punctual Density	Average bin load for the current mode, computed as the ratio between the number of bins and the number of pixels belonging to the current mode	Information about the colors variation (e.g. a high percentage of pixels located in a small volume mode indicates a small number of colors used in the painting)	$1 \times 1 = 1$
2.22	*DCV:* Volumetric Density	The mass to the geometric volume ratio	Indication about the shape of the mode	$1 \times 1 = 1$
2.23	*DCV:* Number of modes	Number of significant modes in the histogram	Information about the color variation in the painting	$1 \times 1 = 1$
3.1-3.24	*Gabor* Energy	The normalized bins of the Gabor energies	Magnitude of the specific orientation and band	3 scales $\times$ 8 orient = 24

painter scene composition procedure, as artistic genres do differentiate themselves by the scene composition technique.

In terms of images, the first computational model for the anchoring theory was provided by Krawczyk et al. in (Krawczyk et al., 2005) for rendering high dynamic images. We followed the same procedure, that relyed on segmentation with mean-shift (Comaniciu and Meer, 2002) for initial decomposition, followed by bilateral filtering (Durand and Dorsey, 2002) for removing very small anchors. An example of a painting and the extracted frameworks is shown in Fig. 1.

While the frameworks are the shapes that holistically describe a painting, for matters of classification, we extracted specific parameters for each of the detected frameworks. The first parameter describing the framework was its cardinality, computed as the number of pixels belonging to the current framework. The next feature considered for describing a framework was its articulation, a measure proposed by Krawczyk in (Krawczyk et al., 2005). The articulation represents the dynamic range of the current region and is modelled using a Gaussian function:

$$A_i = 1 - e^{-\frac{9}{2}(\max Y_i - \min Y_i)^2}, \quad (1)$$

where A_i is the articulation for the framework i and $\min Y_i$ and $\max Y_i$ represents the minimum and maximum luminance in the current framework.

The other two parameters extracted for each framework were the average luminance and its center of mass within the image. As the number of frameworks may vary from painting to painting, we restricted the description to the the most significant two frameworks (showing the relationship between the two most important areas of the painting).

2.2. Color Description

Rappaport (Rappaport and Rapaport, 1984) noted that different artistic currents approximately match different historic periods and techniques known at that time and each of the currents is formally described as inspiring specific sentiments that are subsequently associated with colors. The color palette of a painting may be accurately depicted by the 3D Color Histogram, which was shown to be precisely connected to a scene (thus painting) structure (Novak and Shafer, 1992). The natural choice for computation is the Lab color space, as it exhibits the most perceptual–like inter–color distances. Although the whole 3D histogram might offer plenty of information about the color palette, it would have been a difficult task to describe the most representative modes. Thus, we opted for finding the Minimum Volume Enclosing Ellipsoid (MVEE) that contains color modes, constructing the so-called *Dominant Color Volume* (DCV).

For the representation of the color palette, we opted for a reduced set of parameters, extracted as described below (and also exemplified in Fig. 2):

1. Compute the 3D Histogram with N^3 bins. The histogram is a hyper-volume (4D volume). Depending on the actual data, this volume may be sparse (made from smaller, disjoint, volumes) or more compact.
2. Ignore bins smaller than certain threshold.
3. Label the histogram modes (which are in fact compact volumes) and disregard the modes with fewer than a certain threshold bins. A histogram mode consists in adjacent histogram bins (compact volumes).
4. Keep only the biggest histogram mode and compute the Minimum Volume Enclosing Ellipsoid as defined in (Kumar and Yildirim, 2005) for the cloud of 3D points.
5. For the obtained ellipsoid compute the descriptive parameters presented in Table 2.

While the computation of the Lab 3D color histogram is straightforward, for determination of the MVEE, we recalled that Moshtagh (Moshtagh, 2005) provided a solution for representation of color data from a selected object.

In (Moshtagh, 2005), it was noted that having a set of points $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\} \in \mathbb{R}^3$ in the 3D space, the enclosing ellipsoid ε in center form is defined as:

$$\varepsilon = \{\mathbf{x}_i \in \mathbb{R}^3 | (\mathbf{x}_i - \mathbf{c})^T E (\mathbf{x}_i - \mathbf{c}) \leq 1\}, \forall i = 1, \ldots, m \quad (2)$$

where $\mathbf{c} \in \mathbb{R}^3$ is the center of the ellipsoid and E contains the ellipsoid parameters. The problem was to determine the center, $\mathbf{c}$ and the ellipsoid parameters E, and it has been shown (Moshtagh, 2005) that a solution is found by the standard gradient ascent algorithm.

Due to the fact that the main discrimination between arts movement is not necessarily offered by the colors used, but by their distribution, high classification performance is achievable by keeping only the largest mode. For example, a scarce mode containing few pixels reflects the usage of a wide range of colors, while a large and dense mode corresponds to the usage of a single main color. The ellipsoid containing the greatest number of image pixels was described through a set of parameters presented in Table 2.

Various values for N (number of histogram bins) were tested and it turned out that for values higher than $N = 6$,

Figure 1. Original Painting (left), the extracted frameworks (center) and the two largest frameworks marked with green/yellow (shades of gray) used in classification (right).

the computation times drastically increased, despite the fact the performance remained virtually constant (less than 1% improvement of the detection rate).

2.3. Texture and Edge Features

Edginess is the complementary aspect in terms of both human perception and art theory (Melcher and Cavanagh, 2011) found to be relevant for detecting a genre. Here we envisaged both magnitude (as for example, in Impressionism objects tend to have softer edges, while in Cubism edges are sharper) and orientation.

To assess edginess, we used Gabor filters, as they are known to mimic human perception in terms of contour analysis (Daugman, 1985). We used two banks of Gabor filters, as described in (Grigorescu et al., 2002): one with symmetric and one with antisymmetric Gabor kernels. As they have 8 orientations and 3 scales, the filter bank lead to 24 Gabor feature images $r(x, y)$, computed as follows:

$$r(x,y) = \int_D I(\xi,\eta) g_{\lambda,\theta,\varphi}(x-\xi, y-\eta) d\xi d\eta, \quad (3)$$

where $I(x, y)$ is the input image and $g_{\lambda,\theta,\varphi}(x, y)$ is the 2-D Gabor function described by:

$$g_{\lambda,\theta,\varphi}(x,y) = e^{-((x'^2+\gamma^2 y'^2)/2\sigma^2)} \cos(2\pi \frac{x'}{\lambda} + \varphi), \quad (4)$$

with $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$, $\sigma = 0.56\lambda$ and $\gamma = 0.5$.

The Gabor Energy Image was formed by the combination of the feature images filtered by the two mentioned banks. The vector of Total Gabor Energy was, then, computed for each of the 24 vector–filters.

3. Database and Implementation

3.1. Database

Cornelis et al. (Cornelis et al., 2011) noted that one of the most important challenges when it comes to painting genre analysis is the lack of a standard public database. Most state of the art methods were evaluated on databases acquired either from the public domain, either selecting specific images from Artchive[2] or in partnership with art museums or galleries. Yet, to the moment, the state of the art databases are not public and, as mentioned in the introduction, are limited in size.

We evaluated the performance of our system on a database containing 4119 paintings belonging to 8 different art movements (Baroque, Cubism, Renaissance, Byzantine Icons, Impressionism, Greek Pottery Paintings, Rococo and Romanticism), from more than 600 authors. Considered pairwise, the chosen genres includes pairs that are both separable and easy to discern classes (like Icons vs. Rococo) but also mixed and hard to separate (e.g. Baroque vs. Renaissance). The images were acquired from various sources (e.g. scanning art albums, Internet) hence deliberately lacking cohesion in the acquisition conditions.

3.2. Classifier Design

To recognize the artistic genre when given an image, we recall that Zeki (Zeki, 1999) proved that different elements of visual art such as shapes, colors, and boundaries are processed by different pathways and systems in the brain. In classifying terms this corresponds to a *late fusion* classification scheme (Snoek et al., 2005). The late fusion paradigm states that unimodel features are first reduced to separately learned concept scores, then these scores are integrated to learn concepts, as opposed to early fusion schemes that integrate uni–modal features before learning concepts (i.e. the feature set forms a single vector that is sent to a single classifier).

[2] http://www.artchive.com/

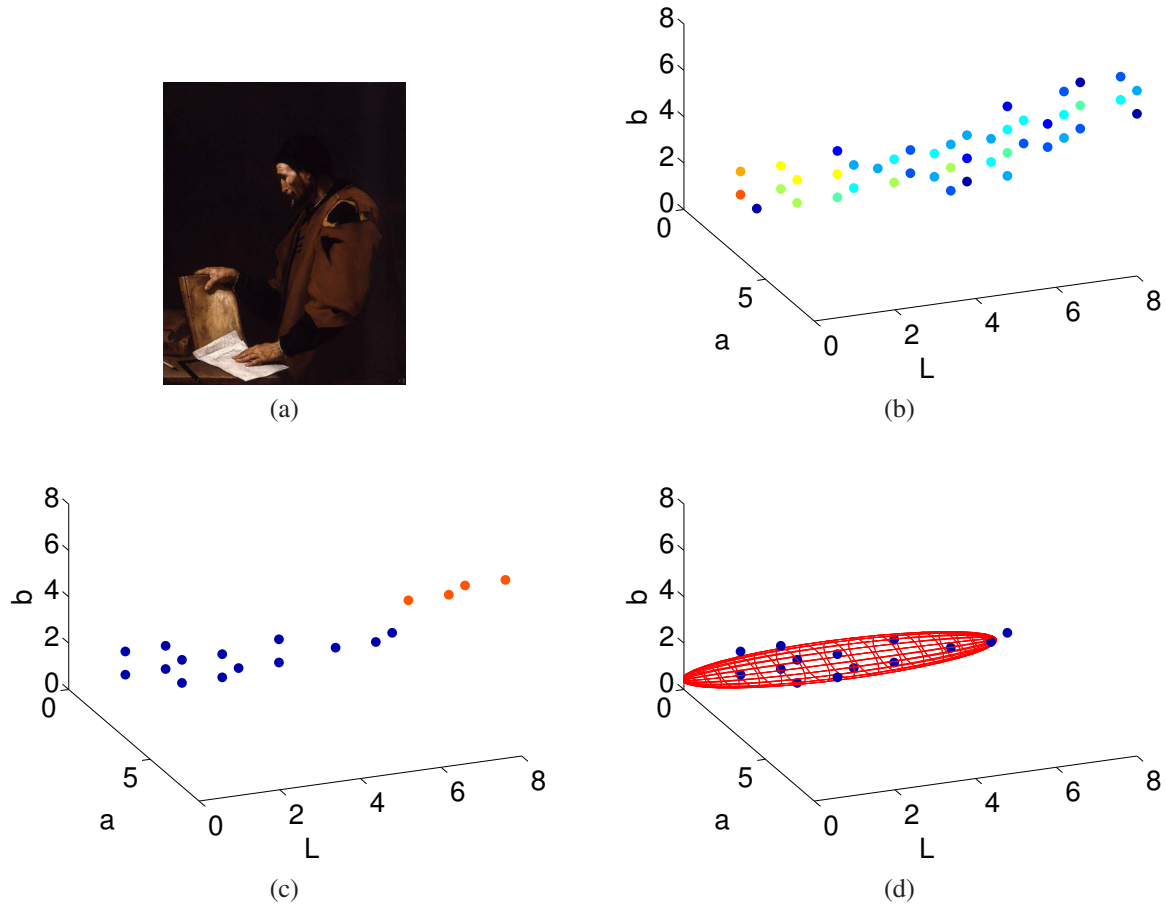


Figure 2. Example of 3D Histogram Modelling. (*a*) Original input image. (*b*) Initial $8 \times 8 \times 8$ 3D *Lab* Histogram (redness of dots means larger bins). (*c*) The two histogram modes obtained after segmentation. (*d*) Dominant Color Volume for the most representative histogram mode.

Thus, instead of merging the three feature categories (composition-anchoring, color palette-Dominant Color Volume and texture/edginess-Gabor Energy) into a final feature vector that is fed to a classifier (according to the early fusion paradigm), we trained three independent Multi–Layer Perceptron (MLP) classifiers, one for each of the feature categories. Each MLP had one hidden layer with 20 neurons. The outputs of the three classifiers are then fed to a final SVM classifier (trained with the sequential minimal optimization method) that will offer the final result. The MLPs were implemented in Matlab, while for the SVM we relied on the implementation from WEKA (Hall et al., 2009).

Training/testing scheme For testing, a 10-fold cross validation technique was assumed; the 10-fold rule was applied for each artistic current.

Table 3. 3D Histogram Color Spaces Performance.

Color space	RGB	Lab	HSV
DR	70.27 ± 2.7	73.32 ± 2.36	68.82 ± 2.29

Performance measure The detection rate (DR) for a genre is defined as the number of correctly identified images from a given genre normalized to the total number of paintings of the genre. When the overall results were presented, the detection rate was averaged over all envisaged images and genres.

3.3. Feature Choice and Parameter Tuning

In order to asses the optimal features parameters we performed a series of experiments on a subset of 1000 images using a 10-fold classification with a late fusion scheme. As it will be discussed further, the scheme was made out of a first level of 3 Multi-Layer Perceptrons, aggregated by an

Table 4. Features categories importance

Features	Detection Rate
Anchoring	47.05%
Dominant Color Volume	61.16%
Gabor Energy	42.81%
Anchoring + DCV	67.66%
Anchoring + Gabor	59.93%
DCV + Gabor	70.21%
Anchoring + DCV + Gabor	73.32%

ensemble of 10 bagged classification trees with unlimited depth.

Color The performance of the color features (Dominant Color Volume) was evaluated for three color spaces using these features alone. We considered three of the most common color spaces: RGB, HSV and Lab. While the RGB color space is the most common representation for digital images, the HSV color space mimics the way human describe colors. The last choice, the Lab color space, has the property of being perceptually uniform for human vision, meaning that a change of a certain amount in one of the color values produces a change of a similar importance for the human visual system. As Table 3 shows, the best average performance was obtained for the Lab representation, hence proving the perceptual claim.

Edges/Texture As a replacement for Gabor energy we considered the Histogram of Oriented Gradient (HOG) as it was found to offer a basis for reliable aesthetic measures (Redies et al., 2012). Yet the Gabor Energy features offered a marginally better ($+2\%$) detection rate.

Feature importance In order to assess the contribution of each feature, two tests were performed: first we identified the contribution of each feature, by removing the feature and re-classifying the database; secondly we considered the Fisher score of that feature.

The results for individual contribution are shown in figure 3 (a). Noting that the worst feature was one of the Gabor energies, we concluded that redundancy had low values in the proposed solution.

The results on feature importance calculated with the Fisher criterion can be seen in figure 3 (b). One may see that the Dominant Color Volume features play a more important role. In order to validate this we performed another series of experiments in which we removed one or two whole categories of features. As can be seen in Table 4, the Dominant Color Volume features had, indeed, the highest contribution.

Classification choice It should be now accepted as common knowledge the fact that there is no a priori ubiquitous optimal classifier, and depending on the specific problem, different classifiers perform best (Michie et al., 1994) (ch.9,10), (Forman and Cohen., 2004). To determine if the proposed choice of classification is the best compared with other variants, based on both late fusion and early fusion paradigms, we have considered seven classifiers for early fusion and four of them for late fusion. For early fusion, we relied on the implementations offered in the open-source machine learning library Weka (Hall et al., 2009). For the initial classifier choice, WEKA offers a pre-optimization that in most of the cases leads to near-optimal result.

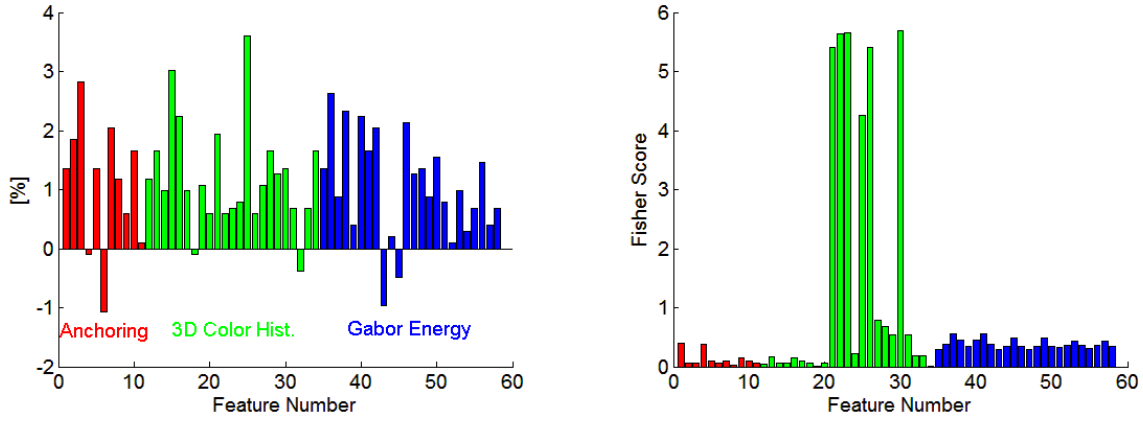
For the late fusion, we first took a decision based on each of the three categories of features considered using the presented MLP scheme. The late decision is taken with various classification schemes. As presented in Table 5, the late fusion classification scheme offers clear advantages when compared with in the the early fusion classification. This was expected, taking into account that human process separately each category of information and the classification target, namely the division in genres, is introduced by humans. From a machine learning point of view, the rather large dimensionality (i.e. 54) of the feature space compared with the database population size (i.e. maximum 4119), produces a sparse dataset which imposes an upper boundary for the machine learning performance. This fact is also proved by the significantly larger accuracy achieved by the Bagging (Bagged Ensemble of Tree) when compared to the Random Forest (RF) in the early fusion case, pointing to the insufficiently large population for the randomization part of the RF.

The best performance was achieved with a SVM having gaussian (RBF) kernel, with penalty parameter $C = 4$ and kernel width $\Gamma = 1$. The performance of the linear kernel (which was shown to be a degenerated case of RBF kernel (Keerthi and Lin, 2003)) presented a decrease with -2.5% when compared with the basic version.

Motivated by the superior performance of the SVM in the second layer of classification, we have tested a late fusion scheme where Support Vector Regressor (SVR) are used on the first layer, to aggregate each of the features types. Results are in table 6. Yet, the MLP performance was still better, pointing to the existence of a linear dependence between feature's dimensions.

4. Results and Discussions

The overall detection rate over the database containing 4119 images from 8 genres is 72.24%, marginally lower than the 73.32% obtained in the 1000 images training experiment. In order to asses how the proposed solution be-



(a) (b)

Figure 3. Contribution of individual features (as named in table 2): (a) the amount of decrease in DR if the specified feature is removed form the system (higher values show more significant contribution). (b) Fisher Score according to the Fisher features selection criterion

Table 5. Average Detection Rate (ADR) for tested classifiers (Bayesian Network – BN, Multilayer Perceptron – MLP, Support Vector Machine – SVM, Bagging – Ba, LogitBoost – LB, Multiclass Classifier – MCC, Decision Table – DT, Random Forest – RF). Details regarding the implementation of the classifiers are to be found in (Hall et al., 2009) and references therein.

Classifiers	BN	MLP	SVM	Ba	LB	MCC	DT	RF
Early fus. [%]	58.19	65.42	68.51	68.9	64.16	71.25	50.37	63.34
Late fus. [%]	–	–	73.32 ± 2.24	72.11 ± 2.43	–	71.62 ± 2.28	–	69.66 ± 2.02

Table 6. Average Detection Rate (ADR) achieved for various regressors on the first level (as they return continuous values and not categorical) of the early fusion schema: Multilayer Perceptron – MLP (1 layer, with 20 neurons), Support Vector Regressor – SVR with linear kernel (lin) and Gaussian (RBF) kernel ($C = 4$, $\Gamma = 1$). The second level classifier was a SVM -RBF.

Classifiers	MLP	SVR-lin	SVR-RBF
DR [%]	73.32 ± 2.43	72.10 ± 2.85	72.20 ± 2.55

haves for each tested art movement, we computed the confusion matrix which is detailed in Table 7 and visually presented in figure 4 (a).

Next, we tested the behavior of our system with respect to the variation of the number of classes. As expected, the overall detection rate decreased with the increase of possible artistic genres, but the results remained acceptable for all tested genres. All possible combinations of the following eight genres were tested: Baroque, Cubism, Renaissance, Byzantine Icons, Impressionism, Greek Pottery Paintings, Rococo and Romanticism. In Fig. 4 (b) the average, the lowest and the highest detection rates are presented for each possible number of classes.

It can be seen that for some number of classes (genres) the difference between the best and the worst detection rate can be significant. This can be explained through the genres that were chosen; if the paintings belonged to similar genres, (e.g. Baroque, Rococo and Romantism - 63.87 %) the detection rate can be smaller, while if the paintings belonged to more distinct genres (Cubism, Renaissance and Romantism - 88.97 %), a higher detection rate is obtained.

Figure 5 shows examples of incorrectly classified paintings. In general, the miss–classification occurred between more similar genres, harder to discriminate even for a human user. The similarities between different artistic genres can also be observed from the confusion matrix presented in Table 7.

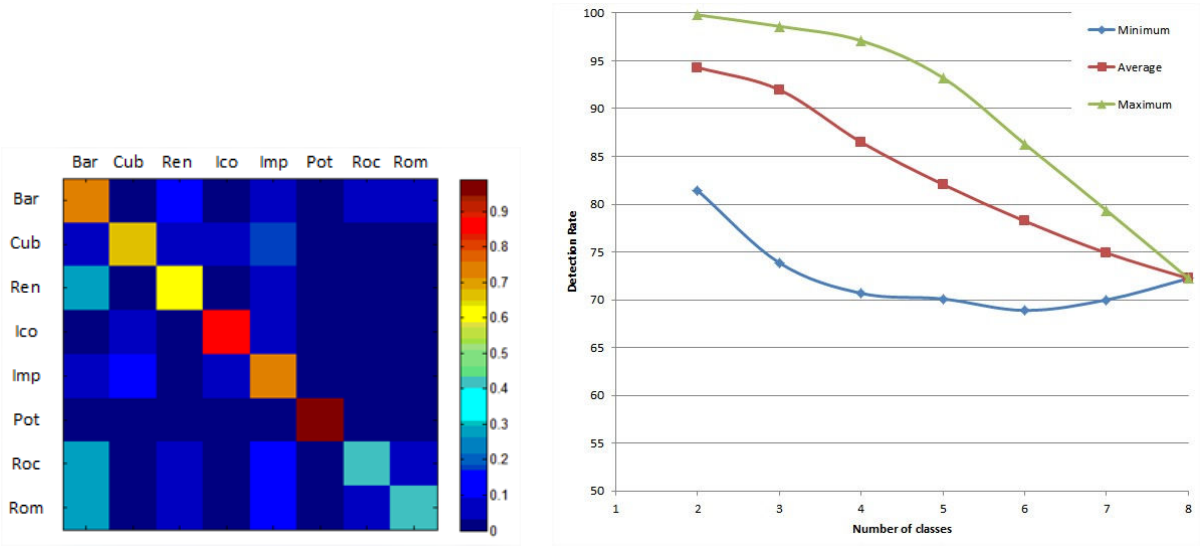
While the computation time is not of interest at the current stage of the work, we note that a painting classification query takes in average 5400 msec on Intel 2.7 GHz in Matlab code, out of which some 5000 msec are required by the bilateral filtering used in determining the image anchors.

4.1. Comparison with state of the art

Two main difficulties in comparing the proposed approach with the state of the art solutions are encountered: the use of different and non-public databases, doubled by the lack of author approved training and testing code. Thus, differentiating factors remain the size of the database and the average detection rate when the same number of currents is assumed.

Table 7. Database content and Confusion Matrix computed for a 10-fold run for the entire database.

	Bar	Cub	Ren	Ico	Imp	Pot	Roc	Rom	Total	Error	DR
Baroque	**492**	8	115	2	44	0	35	35	731	239	67.31
Cubism	26	**360**	34	44	99	1	5	6	575	215	62.61
Renaissance	121	13	**316**	4	21	0	6	4	485	169	65.15
Icons	7	22	3	**564**	29	0	0	0	625	61	90.24
Impress.	27	77	16	31	**600**	0	17	21	789	189	76.05
Pottery	1	2	0	1	0	**328**	0	0	332	4	98.80
Rococo	87	3	8	1	33	0	**132**	28	292	160	45.21
Romanticism	74	0	13	0	40	0	29	**134**	290	156	46.21



(a) (b)

Figure 4. (a) Confusion matrix (b) Average, best and worst detection rate (DR) with respect to the number of classes.

We stress that our database is about 10 times larger than any reported solution. Each author has used, apart the different number of image examples per class, different classes. The exact number of genres, as well as the number of paintings in the used database and the average detection rate are presented in Table 8. To directly compare the performance, we also report the detection rate achieved, in average, for the same number of genres. In interpreting these results, we shall recall the results reported by Wallraven et al. (Wallraven et al., 2009) for non-expert human user manual classification, which set the "reasonable" achievement of an automatic image classification system immediately above 70% average detection rate for around 8 artistic genres.

Considering the presented results, one may easily notice that even using much smaller databases, only Gunsel et al. (Gunsel et al., 2005) outperformed the here-proposed method in terms of accuracy. For a second test, we have implemented and tested the Gunsel solution on our extensive database. In such a case, Gunsel's solution offered an overall DR of 36.63%. The explanation lies in the larger database used here with more paintings, acquired with more variability. The results of the comparison are shown in Table 9, proving that, indeed, the here-proposed solution outperformed the solution in (Gunsel et al., 2005).

5. Conclusion

In this paper we proposed a perceptually inspired system for the automated analysis of paintings, applied for the discrimination among eight artistic currents. Both the extracted features and the classifier were specifically selected as being relevant for human perception. We tested our system on more images than any reported state of the art methods and we still provided higher values in terms of classification accuracy. Although the results are satisfactory, a 100% DR is practically impossible to achieve yet, as long as the separation between some genres is not always very clear even for art historians.

As continuation paths, we envisage two directions. First,

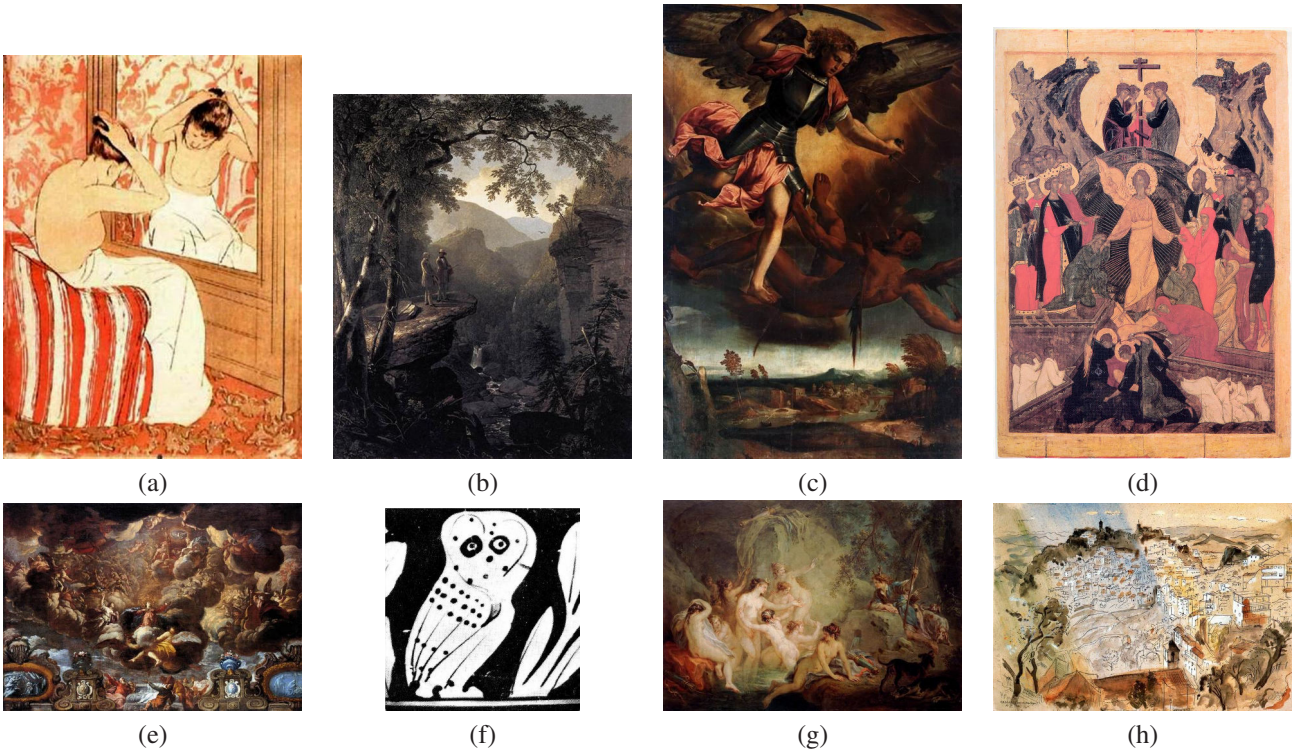


Figure 5. Examples of incorrectly classified paintings. (a) Impressionism as Cubism: Mary Cassatt - "The Coiffure Study" (b) Romanticism as Pottery: Asher Brown Durand - "Kindred Spirits" (c) Renaissance as Baroque: Raphael - "St. Michael Vanquishing Satan" (d) Ortodox Icon as Rococo: Russian Icon - The Descent into Hell (e) Baroque as Renaissance: Palomino - "Assumption of the Virgin" (f) Pottery as Cubism: Greek Pottery Painting (g) Rococo as Romanticism: Martin Johann Schmidt - "Diana in Akteon" (h) Cubism as Impressionism: Picasso - Watercolor

the system has to be further refined for even more improved accuracy and tested for more artistic genres and paintings. The second direction is to adapt the perceptual background of the feature and classifiers used in this work towards the analysis of other artistic-related image types (drawings, artistic photography, comics, etc.).

6. Acknowledgements

This work was supported by the Romanian Sectoral Operational Programme Human Resources Development 2007-2013 through the European Social Fund Financial Agreement POSDRU/107/1.5/S/76903 and POSDRU/159/1.5/S/134398.

References

Arora, R. S., and Elgammal, A. (2012). Towards automated classification of fine–art painting style: a comparative study, *International Conference on Pattern Recognition (ICPR)*, pp. 3541–3544.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5): 603–619.

Condorovici, R., Florea, C. and Vertan, C. (2013). Perceptually-inspired artistic genre identification system in digitized painting collections, *Proc. Scandinavian Conference on Image Analysis (SCIA)*, pp. 687–696.

Cornelis, B., Dooms, A., Cornelis, J., Leen, F. and Schelkens, P. (2011). Digital painting analysis, at the cross section of engineering, mathematics and culture, *Proc. of European Signal Processing Conference*, pp. 1254–1259.

Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *Journal of the Optical Society of America A,* **2**(7): 1160–1169.

Durand, F. and Dorsey, J. (2002). Fast bilateral biltering for the display of high-dynamic-range images, *ACM Transactions on Graphics* **21**(3): 257–266.

Forman, G. and Cohen., I. (2004). Learning from little: Comparison of classifiers given little training, *Proc. of*

Table 8. Average detection rate comparison with state of the art: number of genres, number of paintings. Higher DR values stand for better classification.

Author	Genres	Database	DR reported	Proposed method for the same no. genres
Proposed	*8*	*4119*	*72.24%*	*72.24*
Gunsel et al. (Gunsel et al., 2005)	3	107	91.66%	91.93
Zujovic et al. (Zujovic et al., 2009)	5	353	68.3%	82.1
Shamir et al. (Shamir et al., 2010)	3	517	77%	91.93
Arora et al. (Arora et al., 2012)	7	490	65.4%	74.9
Non-experts humans (Wallraven et al., 2009)	11	275	68%	not available

Table 9. Detection rate comparison between the here-proposed solution and other similar solutions from the recent literature; all tests are performed on the current 4119 image database.

Genre	Bar	Cub	Ren	Ico	Imp	Pot	Roc	Rom	Overall
Proposed	*67.31*	*62.61*	*65.15*	*90.24*	*76.05*	*98.8*	*45.21*	*46.21*	*72.24*
Gunsel et al. (Gunsel et al., 2005)	31.2	43.3	43.9	51.2	46.6	46.8	13.8	16.3	36.63

European Conference on Machine Learning (ECML), pp. 161–172.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V. and Economou, E. (1999). An anchoring theory of lightness perception, *Psychological Review* **106**(4): 795–834.

Grigorescu, S., Petkov, N. and Kruizinga, P. (2002). Comparison of texture features based on Gabor filters, *IEEE Transactions on Image Processing* **11**(10): 1160–1167.

Gunsel, B., Sariel, S. and Icoglu, O. (2005). Content-based access to art paintings, *Proc. of International Conference on Image Processing (ICIP)*, pp. 558–561.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2009). The weka data mining software: an update, *ACM SIGKDD Explorations Newsletter* **11**(1): 10–18.

Keerthi, S. S. and Lin, C.-J. (2003). Asymptotic behaviors of support vector machines with gaussian kernel, *Journal Neural Computation* **15(7)**: 1667–1689.

Keren, D. (2002). Painter identification using local features and naive Bayes, *Proc. of Internation Conference on Pattern Recognition (ICPR)*, Vol. 2, pp. 474–477.

Khan, F. S., van de Weijer, J. and Vanrell., M. (2010). Who painted this painting?, *Proceedings of The CREATE 2010 Conference*, pp. 329–333.

Krawczyk, G., Myszkowski, K. and Seidel, H.-P. (2005). Lightness perception in tone reproduction for high dynamic range images, *Proc. of EUROGRAPHICS, Computer Graphics Forum*, Vol. 24.

Kumar, P. and Yildirim, E. A. (2005). Minimum-volume enclosing ellipsoids and core sets, *Journal of Optimization Theory and Applications* **126(1)**: 1–21.

Li, J. and Wang, J. (2004). Studying digital imagery of ancient paintings by mixtures of stochastic models, *IEEE Transactions on Image Proccessing* **13**(3): 340–353.

Li, X. and Gilchrist, A. (1999). Relative area and relative luminance combine to anchor surface lightness values, *Perception and Psychophysics* **61**: 771–785.

Martinez, K., Cupitt, J., Saunders, D. and Pillay, R. (2002). Ten years of art imaging research, *Proceedings of the IEEE* **90**(1): 28–41.

Melcher, D. and Cavanagh, P. (2011). *Art and the senses*, Oxford University Press, chapter Pictorial cues in art and in visual perception, pp. 359–394.

Michie, D., Spiegelhalter, D. J., Taylor, C. C. and Campbell, J. (1994). *Machine learning, neural and statistical classification*, Ellis Horwood Upper Saddle River.

Moshtagh, N. (2005). Minimum volume enclosing ellipsoid, *Convex Optimization* .

Novak, C. and Shafer, S. (1992). Anatomy of a color histogram, *Proc. of Computer Vision and Pattern Recognition*, pp. 599–605.

Ramachandran, V. and Herstein, W. (1999). The science of art: A neurological theory of aesthetic experience, *Journal of Consciousness Studies* **6**: 15–51.

Rappaport, A. and Rapaport, A. (1984). Color preferences, color harmony, and the quantitative use of colors, *Empirical Studies of the Arts* **2**(2): 95–111.

Redies, C., Amirshahi, S. A., Koch, M. and Denzler, J. (2012). PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects, *Proc. of European Conference on Computer Vision (ECCV)*, pp. 522–531.

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

Snoek, C., Worring, M. and Smeulders, A. (2005). Early versus late fusion in semantic video analysis, *ACM International Conference on Multimedia*, pp. 339–402.

Stork, D. (2009). Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature, *Proc. of Computer Ananlysis of Images and Patterns (CAIP)*, pp. 9–24.

Torresani, L., Szummer, M. and Fitzgibbon, A. (2010). Efficient object category recognition using classemes, *Proc. of European Conference on Computer Vision (ECCV)*, pp. 776–789.

Wallraven, C., Fleming, R. W., Cunningham, D. W., Rigau, J., Feixas, M. and Sbert, M. (2009). Categorizing art: Comparing humans and computers, *Computers & Graphics* **33**(4): 484–495.

Widjaja, I., W.K., L. and Wu, F. (2003). Identifying painters from color profiles of skin patches in painting images, *Proc. of International Conference on Image Processing (ICIP)*, pp. 845–848.

Zeki, S. (1999). *Inner Vision: an exploration of art and the brain*, Oxford University Press.

Zujovic, J., Gandy, L., Friedman, S., Pardo, B. and Pappas, T. (2009). Classifying paintings by artistic genre: An analysis of features & classifiers, *Proc. of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–5.