Impressionism, Expressionism, Surrealism: Automated Recognition of Painters and Schools of Art

LIOR SHAMIR*, TOMASZ MACURA, NIKITA ORLOV, D. MARK ECKLEY and ILYA G. GOLDBERG
Image Informatics Group/Laboratory of Genetics/NIA/NIH

We describe a method for automated recognition of painters and schools of art based on their signature styles, and studied the computer-based perception of visual art. Paintings of nine artists, representing three different schools of art - impressionism, surrealism and abstract expressionism - were analyzed using a large set of image features and image transforms. The computed image descriptors were assessed using Fisher scores, and the most informative features were used for the classification and similarity measurements of paintings, painters, and schools of art. Experimental results show that the classification accuracy when classifying paintings into nine painter classes is 77%, and the accuracy of associating a given painting with its school of art is 91%. An interesting feature of the proposed method is its ability to automatically associate different artists that share the same school of art in an unsupervised fashion. The source code used for the image classification and image similarity described in this paper is available for free download.

Categories and Subject Descriptors: I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding

General Terms: Perceptual reasoning

Additional Key Words and Phrases: Art, painting, image similarity

1. INTRODUCTION

The on-going development of computer vision algorithms has been enabling increasingly complex tasks of automated content-based image analysis. Perhaps one of the more challenging tasks for computers is the analysis, evaluation and identification of visual art. This can include associating a specific painting with a painter, classifying a painting by its school of art, authentication of paintings, finding influential links between painters, and more.

The human perception of art has been studied by scientists in the fields of brain sciences, sociology, and both empirical and theoretical psychology. Zeki [1999] showed that different elements of visual art such as colors, shapes, and boundaries

Author's address: Lior Shamir, Laboratory of Genetics/NIA/NIH, 333 Cassell Dr., Baltimore, MD 21224. email: lshamir@mtu.edu
This research was supported entirely by the Intramural Research Program of the NIH, National Institute on Aging.

are processed by different pathways and systems in the brain, designed to interpret each aspect of the art. His work shows that 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. Observations using fMRI show that the experienced painter uses these parts of the brain in a different way than the non-painter [Solso 2000], and analysis of EEG signals also demonstrated functional and topographical differences between artists and non-artists when performing visual perception of paintings [Bhattacharya and Petsche 2002].

The intensive and distributed brain activity supports the contention that the perception of visual art is not only about what the eye can see, but mainly about what the brain can process, so that artists can be classified in respect to how they probe the visual system with pictures [Solso 1994; Latto 1995]. Therefore, it has been stated that the painter does not paint with her eyes, but with her brain [Zeki 1999]. The approach that selective processing of different visual data can answer specific vision queries also led to the "active vision" paradigm in computer vision [Blake and Yuille 1992].

Ramachandran and Herstein [1999] suggested that the key to understanding the perception of art is the identification of the perceptual processes that it activates in the viewer's brain, rather than the identification of its aesthetic properties. They suggest that artists use some form of over-stimulating distortion of reality, and link it to the *peak-shift* effect (a cognitive principle in which visual interest or identification is strengthened by overtly enhanced stimulus).

Wallraven et al. [2007] studied the perceptual processing of the degree of abstraction in visual art by using Robert Pepperell's "indeterminated" paintings, which can be classified by observers as both representational and abstract. This work isolated the psychophysical effect of several different parameters such as the image size, orientation, etc.

While the complex nature of the analysis of visual art makes it one of the more challenging tasks in computer vision, several algorithms for different types of automated analyses of paintings have been proposed and tested. Barnard and Forsyth [2001] proposed a method for associating captions with works of art by modeling the statistics of word and feature occurrence and co-occurrence. Kammerer et al. [2007] presented a method of automated identification of the drawing tools that were used by the artist. Li and Wang [2004] used wavelets and MHMM (Multiresolution Hidden Markov Models) to classify Chinese ink paintings by the creating artists. A relatively high interest had been attracted by the authentication of paintings, especially the work of Jackson Pollock [Taylor et al. 1999; Taylor et al. 2007; Jones-Smith and Mathur 2006], which is motivated by the high value of authentic Pollock paintings. Widjaja et al. [2003] combined several classifiers in order to identify four painters using skin samples, and reported accuracy of 85%. Keren [2002] studied the identification of several painters based on repetitive features, and van den Herik et al. [2000] addressed the problem of selecting informative image features when classifying impressionist paintings obtained from a single source (WebMuseum), and reported classification accuracy above chance level.

Here we describe a method of classifying paintings of nine different artists representing three different schools of art: impressionism, surrealism, and abstract expressionism. The proposed method can classify a given painting to one of the nine painter classes, but is also used for finding similarities between the different styles of the artists, and associating them by their schools of art in an unsupervised fashion. In Section 2 we describe the image data set, in Section 3 we describe the extraction of image features, in Section 4 we discuss the image classification and similarity measures, and in Section 5 the experimental results are described.

2. IMAGE DATASET

The image dataset includes paintings of nine different artists representing three different schools of art, such that each school is represented by three artists. The schools of art are *impressionism*, represented by Vincent Van Gogh, Claude Monet and Pierre-Auguste Renoir; *abstract expressionism*, represented by Mark Rothko, Jackson Pollock, and Wassily Kandinsky; and *surrealism*, represented by Salvador Dali, Max Ernst, and Giorgio de Chirico. For each artist we collected 57 images, such that in each experiment 40 images were used for training and the remaining 17 images were used for testing.

Each artist was represented by different types of images (e.g., portraits, scenery, etc), and no attempt to keep this set homogenous was made. An important exception of this policy is the early work of Kandinsky, which is substaintially different from his signature abstract expressionistic style, and therefore his paintings from that era have been excluded from the dataset.

An important feature of the dataset used in this study is that the images were obtained from various sources using simple internet search, and were different in quality and size. While this policy of constructing the dataset can potentially reduce the classification accuracy, its main purpose was to minimize the source-dependency of the images, and to verify that the images are analyzed based on their actual visual content, rather than the method of acquisition, quality, compression algorithms, or other artifacts that might be a feature of the acquisition and handling of the image by its providing source. Since the authors are not familiar with the exact details of the way the images were acquired and handled, no assumptions can be made regarding their consistency among different artists. For instance, if a certain electronic image gallery (e.g. WebMuseum) obtained Van Gogh images from one source while Monet images were collected from another source, any attempt to classify this dataset might actually classify sources rather than painters, despite the fact that all images were made available for download from a single image gallery.

The use of very many sources leads to different image sizes, which range from 640×640 to 2458×1812. In order to normalize all images into fixed dimensions, we first downsampled each image such that the smallest side of the image is 600 pixels. Then, a 600×600 image block was cropped from the center of each image, resulting in a dataset of images with a standard size of 600×600 pixels. This policy provided a normalized dataset of images without changing the aspect ratio, but with the sacrifice of some of the image content that remained outside the 600×600 block.

3. EXTRACTING IMAGE FEATURES

The perception of visual art is a highly complex cognitive task performed by many different specialized centers in the brain, which perceive the different elements featured in the painting. Therefore, implementation of a system that can process visual art should be based on the analysis of very many different image content descriptors extracted from the paintings. This set of descriptors should be broad enough to sense high contrast features such as shapes and borders, but should also be sensitive to other features that can be perceived by the brain such as color patterns, textures, and intensity variations [Zeki 1999].

This approach is implemented by first extracting a large set of image features, from which the most informative features are selected. Image features are computed not only from the raw pixels, but also from several transforms of the image, and transforms of transforms. These compound transforms have been found highly effective in classification and similarity measurement of biological and biometric image datasets [Shamir et al. 2008; Orlov et al. 2008].

For image feature extraction we use the following algorithms, described more throughly in [Orlov et al. 2007]:

1. **Radon transform features** [Lim 1990], computed for angles 0, 45, 90, 135 degrees, and each of the resulting series is then convolved into a 3-bin histogram, providing a toal of 12 image features.
2. **Chebyshev Statistics** [Gradshtein and Ryzhik 1994] - A 32-bin histogram of a 1×400 vector produced by Chebyshev transform of the image with order of N=20.
3. **Gabor Filters** [Gabor 1946], where the kernel is in the form of a convolution with a Gaussian harmonic function [Gregorescu et al. 2002], and 7 different frequencies are used (1,2...,7), providing 7 image descriptor values.
4. **Multi-scale Histograms** computed using various number of bins (3, 5, 7, and 9), as proposed by [Hadjidementriou et al. 2001], and providing 3+5+7+9=24 image descriptors.
5. **First 4 Moments**, of mean, standard deviation, skewness, and kurtosis computed on image "stripes" in four different directions (0, 45, 90, 135 degrees). Each set of stripes is then sampled into a 3-bin histogram, providing 4×4×3=48 image descriptors.
6. **Tamura Texture features** [Tamura et al. 1978] of *contrast*, *directionality* and *coarseness* , such that the coarseness descriptors are its sum and its 3-bin histogram, providing 1+1+1+3=6 image descriptors.
7. **Edge Statistics features** computed on the Prewitt gradient [Prewitt 1970], and include the mean, median, variance, and 8-bin histogram of both the magnitude and the direction components. Other edge features are the total number of edge pixels (normalized to the size of the image), the direction homogeneity [Murphy et al. 2001], and the difference amongst direction histogram bins at a certain angle α and $\alpha + \pi$, sampled into a four-bin histogram.
8. **Object Statistics** computed on all 8-connected objects found in the Otsu binary mask of the image [Otsu 1979]. Computed statistics include the Euler Number [Gray 1971], and the minimum, maximum, mean, median, variance, and a 10-bin histogram of both the objects areas and distances from the to the image centroid.
9. **Zernike features** [Teague 1979] are the absolute values of the coefficients of

the Zernike polynomial approximation of the image, as described in [Murphy et al. 2001], providing 72 image descriptors.
10. **Haralick features** [Haralick et al. 1973] computed on the image's co-occurrence matrix as described in [Murphy et al. 2001], and contribute 28 image descriptor values.
11. **Chebyshev-Fourier features** [Orlov et al. 2007] - 32-bin histogram of the polynomial coefficients of a Chebyshev–Fourier transform with highest polynomial order of N=23.

In order to extract more image descriptors, the algorithms are applied not only on the raw pixels, but also on several transforms of the image and transforms of transforms [Orlov et al. 2008; Rodenacker and Bengtsson 2003; Gurevich and Koryabkina 2006]. The image transforms are FFT, Wavelet (Symlet 5, level 1) two-dimensional decomposition of the image, and Chebyshev transform. Color transform was applied by classifying each pixel into one of 16 noticeably different color classes using fuzzy logic modeling of the human perception of colors [Shamir 2006], and then assigning each pixel with an intensity value based on the relative wavelength of the classified color, normalized to [0,255] interval. I.e., pixels classified as red are assigned with 0, pixels classified as purple are assigned with 255, and pixels classified as other colors are assigned with values between 0 and 255, based on their relative wavelength. Another transform that was used is Edge Transform, which is simply the magnitude component of the image's Prewitt gradient, binarized by Otsu global threshold [Otsu 1979]. The various combinations of the compound image transforms are described in Figure 1.

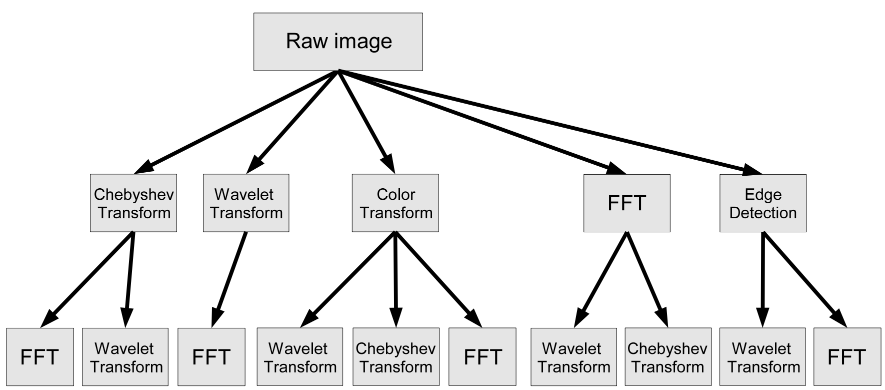

Fig. 1. Image transforms and paths of the compound image transforms.

In the proposed method, different image features are extracted from different image transforms or compound transforms. The image features that are extracted from all transforms are the statistics and texture features, which include the first 4 moments, Haralick textures, multiscale histograms, Tamura textures, and Radon features. Polynomial decomposition features, which include Zernike features, Chebyshev statistics, and Chebyshev-Fourier polynomial coefficients, are also extracted from all transforms, except from the Fourier and Wavelet transforms of the Chebyshev transform, and the Wavelet and Chebyshev transforms of the Fourier transform. In addition, high contrast features (edge statistics, object statistics, and Gabor filters) are extracted from the raw pixels and from the color transform. The entire set of image features extracted from all image transforms described in Figure 1 consists of a total of 3658 numeric image descriptors.

In this paper, each image was divided into 16 equal-sized (150×150 pixels) tiles such that features are computed on each of the 16 tiles, providing 16 feature vectors for each painting in the dataset.

4. IMAGE CLASSIFICATION AND IMAGE SIMILARITY

Since analysis of visual art can be considered a fairly complex task, one can reasonably assume that very many different image features can be used for assessing paintings, painters, and schools of art. However, not all image features are assumed to be equally informative, and some of these features are expected to represent noise. In order to select the most informative features while rejecting noisy features, each feature is assigned with a Fisher score [Bishop 2006], described by Equation 1,

$$W_f = \frac{\sum_{c=1}^{N}(\overline{T_f} - \overline{T_{f,c}})^2}{\sum_{c=1}^{N}\sigma_{f,c}^2} \tag{1}$$

where W_f is the Fisher Score, N is the total number of classes, $\overline{T_f}$ is the mean of the values of feature f in the entire training set, and $\overline{T_{f,c}}$ and $\sigma_{f,c}^2$ are the mean and variance of the values of feature f among all training images of class c. The Fisher Score can be conceptualized as the ratio of variance of class means from the pooled mean to the mean of within-class variances. All variances used in the equation are computed after the values of feature f are normalized to the interval $[0, 1]$. After Fisher scores are assigned to the features, the weakest 85% of the features (with the lowest Fisher scores) are rejected, resluting in a feature space of 548 image features.

After computing the image features as described in Section 3, each of the 16 feature vectors is classified using a simple weighted nearest neighbor rule, such that the feature weights are the Fisher scores. The classification of each feature vector provides a vector of the size N (N is the total number of classes), such that each entry c in the vector represents the computed similarity of the feature vector to the class c, deduced using Equation 2,

$$M_{f,c} = \frac{1}{\min(D_{f,c}) \cdot \sum_{i=1}^{N} \frac{1}{\min(D_{f,i})}} \tag{2}$$

where $M_{f,c}$ is the computed similarity of the feature vector f to the class c, and $\min(D_{f,c})$ is the shortest weighted Euclidean distance between the feature vector f and any tile in the training set that belongs to the class c. This assigns each of the 16 feature vectors of any image in the test set with N values within the interval $[0, 1]$, representing its similarity to each class. The entire image is then classified by averaging the 16 similarity vectors of the tiles, providing a single similarity vector. Naturally, the predicted class for each test image is the class that has the highest value in the similarity vector.

Averaging the similarity vectors of all images of a certain class provides the similarities between that class and any of the other classes in the dataset. Repeating this action for all classes results in a similarity matrix that represents the full set of similarities between all classes. The similarity matrix contains two similarity values for each pair of classes. I.e., the cell n, m is the similarity value between class n to class m, which may be different from the cell m, n. Although these two values are expected to be close, they are not expected to be fully identical due to the different images used when comparing n to m and m to n. Averaging the two values provides a single distance between each pair of classes, which can be used for visualizing the class similarity using phylogenies (evolutionary trees) inferred automatically by using the Phylib package [Felsenstein 2004].

As will be discussed in Section 5.7, the Weighted Nearest Neighbor classification with Fisher Scores used as weights provides an effective method for computing distances between samples in the feature space, where the informativeness of the different features varies significantly. It also provided a better classification accuracy figures than linear classification methods such as SVM (Support Vector Machine) [Vapnik 1995], which can become less effective when the variation in the discriminative power of the features is large.

5. EXPERIMENTAL RESULTS

5.1 Classification accuracy

The efficacy of the method described in Section 3 was tested using the dataset described in Section 2 by building several different classifiers: A 9-way classifier that simply classifies a given painting into one of the nine painter classes, three 3-way classifiers for the painters within each school of art, one 3-way classifier for the schools of art, and a two-stage classifier that first classifies paintings to schools of art, and then classifies to the painter within that school. The purpose of using the two-stage classifier was to improve the selection and scoring of the relevant image features for each of the classification problems. The similarities between the paintings and the painters were also tested to show that the method can unsupervisely associate painters of the same artistic style.

In the first experiment, 360 images (40 for each of the nine artists) were used for training and 153 (17 per artist) images for testing. In order to classify paintings by their creating artists, we randomly split each class in the dataset into training and test images, and repeated the experiment 50 times. The average accuracy of all 50 runs was 71% ($P < 0.001$). The P value is computed automatically by the software used for the experiments [Shamir et al. 2008], and simply reflects the probability of a distribution that is equal or better than the classification accuracy, comparing to the null hypothesis which is that the paintings cannot be classified with accuracy higher than random. Further description of the P values can be found in [Shamir et al. 2008].

5.2 Image similarity

Except from the classification accuracy, the classifier provided also a similarity matrix in the form described in Section 4, normalized such that the similarity of each artist to itself is set to 1. The similarity values are listed in Table I.

Transforming the similarity matrix into a phylogeny using the Phylip package

Table I. Similarity matrix of the painters

	Dali	De Chirico	Ernst	Kandinsky	Monet	Pollock	Renoir	Rothko	Van Gogh
Dali	1.00	0.96	0.97	0.92	0.87	0.78	0.94	0.90	0.93
De Chirico	0.99	1.00	0.98	0.91	0.84	0.77	0.95	0.92	0.94
Ernst	0.95	0.94	1.00	0.90	0.84	0.76	0.93	0.87	0.92
Kandisnky	0.89	0.88	0.90	1.00	0.87	0.83	0.94	0.84	0.91
Monet	0.90	0.88	0.90	0.91	1.00	0.93	0.98	0.74	0.98
Pollock	0.77	0.74	0.84	0.85	0.88	1.00	0.84	0.59	0.91
Renoir	0.94	0.92	0.94	0.92	0.97	0.86	1.00	0.81	0.98
Rothko	0.95	0.95	0.95	0.93	0.81	0.72	0.90	1.00	0.88
Van Gogh	0.89	0.86	0.93	0.92	0.95	0.94	0.96	0.72	1.00

[Felsenstein 2004] provided the evolutionary tree of Figure 2. As can be easily seen in the figure, the proposed algorithm clustered together the three impressionist artists, and featured all surrialist artists in another cluster, indicating that it can automatically associate artists within the same school of art. The abstract expressionists, however, were strongly separated from each other. This demonstrates the high diversity of the signature styles of the different painters, and the absence of a "typical" artistic style in that school of art.

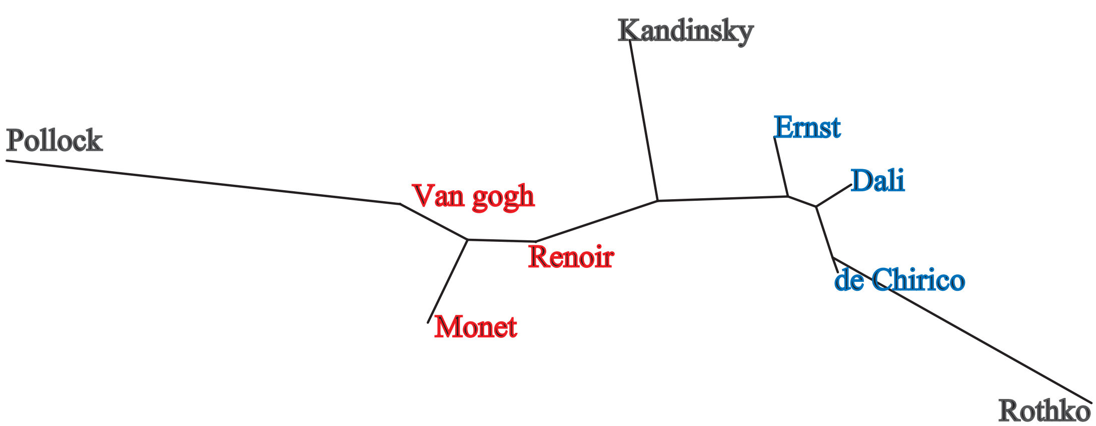

Fig. 2. Phylogeny of the similarities between painters.

While the phylogeny of Figure 2 features just nine artists, the similarity values (Table I) were determined by averaging the similarity values of all test images of 50 different runs, as described in Section 4. Figure 3 is an example of a phylogeny of the individual paintings of a single run. The phylogeny is based on the weighted Euclidean distances of 153 feature vectors (17 per artist), normalized to the $[0, 1]$ interval. That is, each image is assigned with a similarity value to each of the other test images, deduced by the weighted Euclidian distance. The training images in this case were used only for determining the weights of the different features, as described in Section 4.

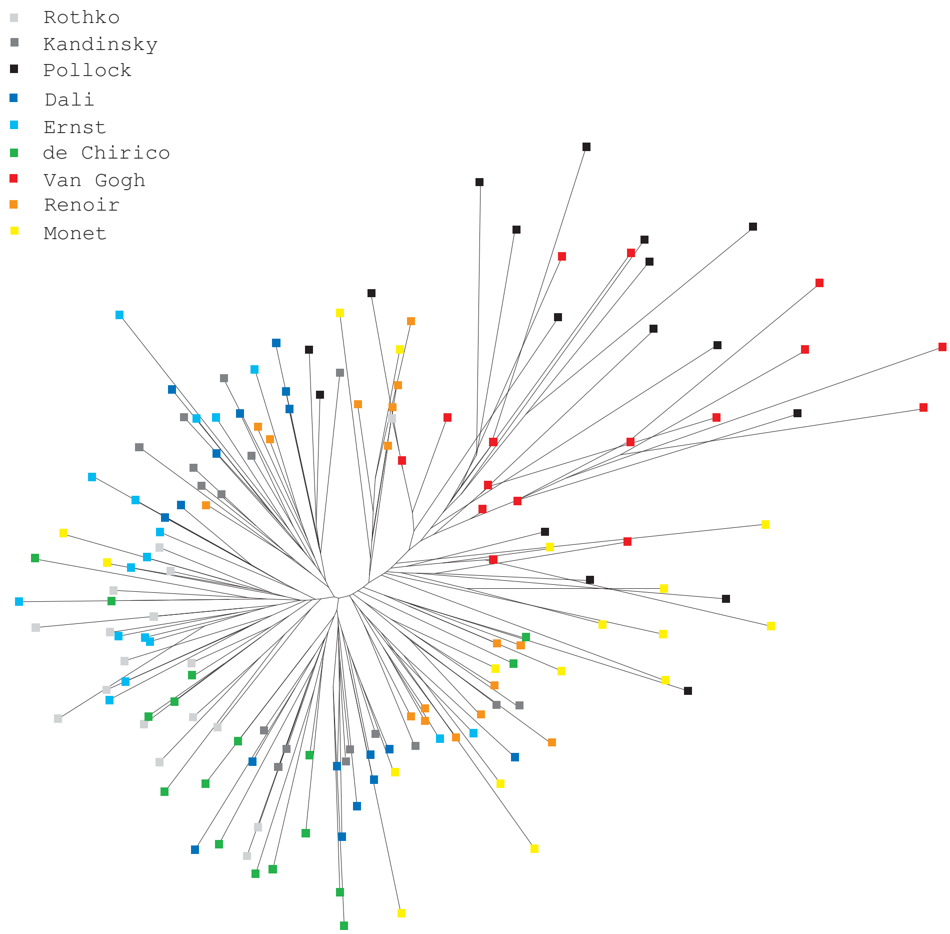

Fig. 3. Phylogeny of the similarities between individual paintings.

As the figure shows, most impressionist paintings are located around the top right part of the phylogeny, while surrealist paintings are usually clustered around the bottom left. Abstract expressionist paintings are distributed around the phylogeny, and can be clustered only by painters, but not as a school of art.

5.3 Two-way classification between artists

Another experiment was based on using the proposed method for building a 2-way classifier for each pair of artists in the dataset. I.e., the same method described in Sections 3 and 4 was used, but instead of using nine painters, each classifier was built and tested using two painters only. The classification accuracy (average accuracy of 50 random splits, where 40 paintings of each artist are used for training and 17 for testing) for each pair of artists is listed in Table II.

Table II. Classification accuracies (%) of all 2-way classifiers

	Monet	Renoir	Dali	Ernst	De Chirico	Kandinsky	Rothko	Pollock
Van Gogh	81	93	100	100	100	100	100	92
Monet	-	87	100	100	100	100	100	100
Renoir	-	-	100	100	100	94	100	100
Dali	-	-	-	86	81	92	95	100
Ernst	-	-	-	-	88	95	94	100
de Chirico	-	-	-	-	-	100	96	100
Kandisnky	-	-	-	-	-	-	100	100
Rothko	-	-	-	-	-	-	-	100

As the table suggests, the proposed algorithm can generally achieve better classification accuracy when classifying paintings of artists from different schools of art, rather than painters within the same school. That is, surrealist and impressionist painters can be accurately differentiated from painters representing other schools of art, but classification accuracy within the same school is significantly weaker. An exception is introduced by the abstract expressionist painters, who can be differentiated from each other in a very high level of accuracy, presumably due to the unstructured nature of this artistic movement, leading to very different signature styles of its members.

5.4 Using a two-stage classifier

Since the proposed classifier is sensitive to the school of art, the accuracy of classifying paintings by their creating artists can be improved by using a two-stage classification: In the first stage the given painting is classified by its school of art, and in the second stage the painting is classified by the painters within that school to find the specific artist. I.e., instead of using one 9-way classifier, we use one 3-way classifier to find the school of art of a given test painting, and three 3-way classifiers (one for each school) to find the specific artist within that school.

A classifier for the schools of art was built by merging 40 paintings of each of the three painters of each school of art into one class, so that each of the three classes - impressionism, expressionism and surrealism - contained 120 training images. Testing the classifier using the remaining 51 images (17 for each painter) from each class provided classification accuracy of 91% ($P< 0.001$). Each of the three classifiers of artists within the same school of art was built using the same 120 images, such that each painter class contained 40 training samples. The performance of these classifiers was generally weaker than the one associating a specific painting with a school of art, and was 78% ($P< 0.001$) when classifying the three impressionist artists (Van Gogh, Monet and Renoir), 71% ($P< 0.001$) when classifying the surrialist artists (Dali, Ernst and De Chirico), and 100% ($P< 0.001$) for the abstract expressionist artists (Kandinsky, Rothko and Pollock).

The performance of the two-stage classifier was tested using 153 images (17 per artist), and was repeated 50 times such that in each run the training and test images were selected randomly. The performance of the two-stage classifier was significantly better than the single-stage 9-way classifier, and produced classification accuracy of ∼76%. The statistical significance of this experiment is <0.001, when the null hypothesis is that the two experiments should perform equally well.

5.5 Using different numbers of images and tiles

To evaluate the effect of the size of the training set on the performance, we measured the performance of the classifiers using different number of training images. In all experiments the test set contained 17 images for each class, and the allocation for test and training sets was performed by randomly selecting images from the pool of 57 paintings of each artist, and repeating each experiment 50 times. In all cases 0.15 of the image features were used, and the reported values are the means of the 50 accuracy figures. The standard error in all cases was smaller than 0.5.

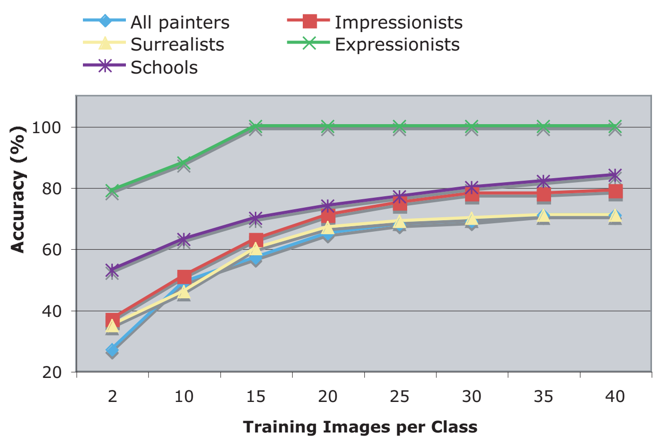

Fig. 4. Classification accuracy as a function of the size of the training set.

As the figure shows, the classification accuracy improves as the size of the training set gets larger, but marginally increases when the size of the training set is larger than 25 images per class.

As described in Section 4, image feature were extracted from 16 equal-sized tiles of each image. This policy improved the classification accuracy comparing to using the entire images, as described in Figure 5. The contribution of the tiling to the performance is in agreement with the observation that not all pixels in an image are equally important to the human perception [Change et al. 2000]. For instance, the center of an image tend to fire more neurons in the parts of the brain devoted to perception than areas closer to the border [Change et al. 2000]. However, the graph also shows that using more than 16 tiles leads to performance degradation. This can be explained by the observation that the size of the image has a substantial effect on the ability of a human observer to classify paintings into schools of art [Wallraven et al. [2007].

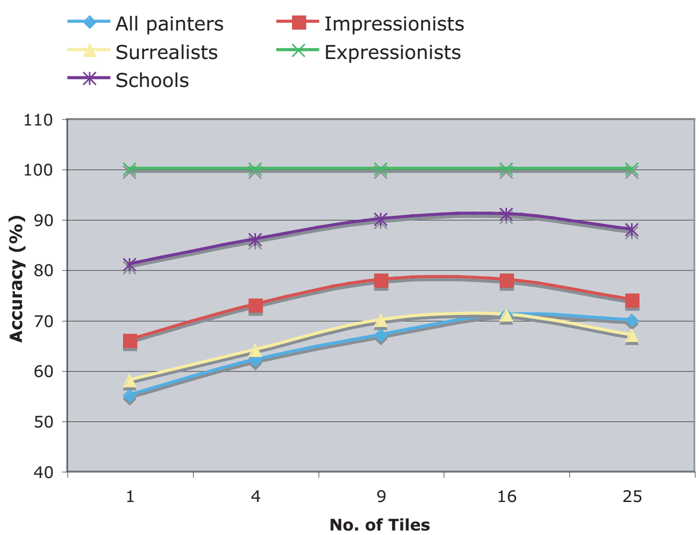

Fig. 5. Classification accuracy as a function of the number of tiles.

5.6 Assessment of the Image Features and Image Transforms

In order to assess the contribution of the image features extracted from the transforms and compound transforms, we measured the classification accuracies of the experiments with and without using the transforms and compound transforms. The performance figures are listed in Table III.

Table III. Classification accuracies (%) when using images features extracted from raw pixels only, raw pixels and image transforms, and raw pixels, transforms and compound transforms

	Raw Pixels	Raw Pixels + Transforms	Raw Pixels + Transforms + Compound Transforms
All painters (9-way)	61	68	71
Impressionist painters (3-way)	70	76	78
Surrealist painters (3-way)	67	68	71
Expressionist painters (3-way)	100	100	100
Schools (3-way)	86	90	91

As the table shows, extracting image features from the transforms and compound transforms substantially improves the performance of the classifier. The only exception is the classification of the three abstract expressionist painters, which provided a perfect accuracy when using the image features extracted from the raw pixels, so that no improvement could be made by using transforms.

The brain perceives art in a fashion that makes use of very many different elements of the art. These can include elements such as shapes, borders, colors and textures. In order to cover such a broad range of elements, very many different image features are extracted from each painting. Since the number of features is large, it can be assumed that some features are more informative than others. The discriminativeness of the different image features is reflected by their Fisher Scores, defined by Equation 1. Figure 6 shows the Fisher Scores assigned to the different feature groups, extracted from the raw pixels, image transforms, and image compound transforms for each of the four classifiers listed in Table III. The values in the graph are the sum of the scores assigned to all bins of each group of image features.

As the figure shows, the absolute Fisher Score values are much higher for the classification of the expressionists painters, which is an easier classification problem, than the absolute scores of the more difficult classification of the impressionist and surrealist painters.

5.6.1 *Estimating feature informativeness by elimination.* The large set of image features is not orthogonal, and many of the image features are sensitive to the same type of image content, which leads to a significant redundancy of the image features. To test the informativeness of single groups of image features while eliminating the effect of redundancy, we tested the four classifiers listed in Table III, such that in each test one group of features was eliminated. The classification accuracy was then computed by averaging 50 runs, and compared to the classification accuracy when the full set of image features was used. This process was repeated for all groups of features described in Section 3. Figure 7 shows the decrease in classification accuracy when each group of features is eliminated.

As the figure shows, eliminating some groups of features can lead to a higher decrease in classification accuracy than others. The redundancy of the image features is evident by the absence of full correlation between the Fisher Scores of the feature groups shown in Figure 6 and the decrease in classification accuracy when each group of features is ignored. This can be easily noticed for the expressionist painters, where eliminating any single group of features does not reduce the classification accuracy to less than 100%.

To determine which combinations of image features are informative, we sequentially added groups of image content descriptors, ordered by their Fisher scores. For instance, for the abstract expressionists we first tested the classifier using the Haralick texture features (which have the highest Fisher score), then added first four moments features, then the Zernike features, and so on. For the sake of simplicity, we used only the image features computed on the raw pixels. Table IV shows the classification accuracies of the four classifiers when adding groups of image features based on their Fisher scores.

As the table shows, using Haralick texture alone is fairly informative for the classification of the abstract expressionist paintings, and provides 88% of accuracy. Adding the first four moments and the Zernike features does not significantly improve the accuracy until adding the object statistics, which elevates the classification accuracy to perfection. This shows that the Zernike and first four moments are in this case redundant to the Haralick features, but the higher-level object statistics features add new information that is not reflected by the Haralick texture. For instance, the object statistics can be used to discriminate between Rothko's paintings, which typically feature few large shapes, and the work of Kandinsky, which composites numerous smaller objects.

The impressionist paintings can also be differentiated by using the Haralick tex-

Table IV. Classification accuracies (%) when adding groups of image features by their Fisher scores

Expressionists		Impressionists		Surrealists		Schools	
Haralick	88	Haralick	60	Zernike	44	Obj. Statistics	52
First 4 Moments	90	First 4 Moments	66	Edge Statistics	55	Haralick	63
Zernike	90	Edge Statistics	70	Obj. Statistics	59	First 4 Moments	67
Obj. Statistics	100	Obj. Statistics	70	Haralick	66	Edge Statistics	74
Edge Statistics	100	Zernike	70	Multiscale Hist.	67	Zernike	83
Tamura	100	Multiscale Hist.	70	Chebyshev Stat.	67	Multiscale Hist.	86
Gabor Filters	100	Tamura	70	First 4 Moments	67	Tamura	86
Multiscale Hist.	100	Chebyshev Stat.	70	Gabor Filters	67	Gabor Filters	86
Chebyshev Stat.	100	Gabor Filters	70	Tamura	67	Chebyshev Stat.	86
Radon	100	Radon	70	Radon	67	Radon	86

ture, and the classification improves when using the first four moments and the edge statistics. As explained by Keren [2002], the pixel statistics can be used for the recognition of painter signatures. The edge statistics can reflect the style of the paint strokes, which noticeably varies from different members of the impressionist movement.

For the surrealist paintings, the most informative set of features is the Zernike features, which gives just 44% accuracy. The edge and object features can in this case correlate with the contrast and the amount of details in the painting. For instance, Giorgio de Chirico's signature geometric style typically features straight lines of cityscape structures and shades, as well as large "flat" areas that evoke the mysterious and haunted mood. These characteristics can be reflected by the edge and object features.

5.6.2 *The effect of the number of features.* Due to the complex nature of differentiating between artists and schools of art, the image analysis is performed using a relatively large number of image features. Since image features are weighted by their informativeness, the effect of each feature is expected to be smaller as its discriminative power gets weaker. However, if very many non-informative image features are used, their large number can be weighed against their low Fisher Scores, leading to an undesirable degradation of the performance. Figure 8 shows how the classification accuracy changes as more features are used.

As the graph shows, the classification accuracy improves as more image features are used. However, when more than $\sim$0.25 of the features are included in the analysis, the performance generally degrades due to the undesirable effect of non-informative image features. Another observation is that each experiment reaches its peak accuracy using a different percentage of features. This can be explained by the different elements that can discriminate between the paintings of each school of art, and therefore different image features are used for each of the classifiers. Although the differences are usually marginal, the performance of the two-stage classifier described in this section can be improved if 0.25 of the features are used when classifying the schools of art, and 0.1 and 0.15, for the classification of the impressionist and surrealist painters, respectively. Using these values improves the accuracy of the two-stage classifier to 77%.

5.7 Comparing Weighted Nearest Neighbor to Other Classification Methods

Using Weighted Nearest Neighbor for classifying the feature values provides a fast and effective method of classifying and measuring distances between feature vectors. The motivation for selecting this simple method was based on the fact that while many image features are used, not all features are equally informative, and the discriminative power of the different features can vary significantly. This can be handled effectively when using WNN with the weights assigned to the different features. Table V shows the classification accuracy of WNN and SVM using linear, polynomial, and RBF kernel functions implemented in the SVMPerf software package [Johachims 2006]. Since SVM was originally designed for binary classification problems, the comparison is based on the 2-way classification of Renoir/Monet, Dali/Ernst, Dali/Kandinsky, Renoir/Rothko, and Pollock/Ernst.

Table V. Accuracies (%) of 2-way classification problem when using Weighted Nearest Neighbor, and VM with linear, polynomial and RBF kernels.

	WNN	Linear kernel	Polynomial kernel (d=4)	RBF kernel (γ=10)
Renoir/Monet	87	67	71	69
Dali/Ernst	86	63	65	63
Dali/Kandinsky	92	70	73	71
Renoir/Rothko	100	100	100	100
Pollock/Ernst	100	100	100	100

The better performance of the WNN classifier can be explained by the fact that Weighted Nearest Neighbor classification can effectively handle the large variance in the informativeness of the different image features reflected by their Fisher scores, while SVM works by using a threshold that determines which features are used for the classification and which features are ignored. The effectiveness of the weighted feature space in comparison to the non-weighted feature space is also discussed in [Orlov et al. 2008].

6. CONCLUSION

In this paper we studied the automated analysis of paintings, and described a method for classifying paintings by their creating artists and schools of art. The proposed classifier also provides similarity measures between a given test image to each of the painter classes, which was used for assessing the similarities between different painters and automatically associating artists and paintings that are part of the same artistic movements. The ability of the computer to analyze paintings can raise important theoretical questions about the applied perception of art. Can a computer system criticize literature? Can it tell between a good movie and a bad movie? Another question is whether computers can not just evaluate art, but also create it.

The method used here makes use of a very large set of image features, covering as many aspects as possible and computed from several different parts of the painting, making it difficult to correlate the machine vision features with the visual cues used by the human eye. The extraction of very many different types of data is an attempt to follow the perception of the artist, who does not focus on just one technique or specific part of the painting at a time, but perceives the entire painting as a whole at the time of creation. This approach of painting is reflected by the advice of impressionist Camille Pissaro to a young artist: "Don't paint bit by bit, but paint everything at once by placing tones everywhere, with brushstrokes of the right colour and value, while noticing what is alongside. Use small brushstrokes and try to put down your perceptions immediately. The eye should not be fixed on one spot, but should take in everything, while observing the reflections which the colours produce on their surroundings" [Rewald 1963].

The best performance in terms of automatically identifying the artist of a given test painting was achieved by using a two-stage image classifier, such that in the first stage the school of art was determined, and in the second stage the specific painter was recognized among the painters of that school.

The full source code used for this study is available for free download via CVS at *www.openmicroscopy.org* (recommended), or as a *tar* archive at *http://www.phy.mtu.edu/~lshamir/downloads/ImageClassifier* in a convenient form of a command line utility [Shamir et al. 2008]. The code was initially developed for analyzing and understanding physiological processes and mechanisms that can be studied using microscopy or other imaging devices. It has also been found useful for several other image classification and image similarity problems, and we encourage scientists and engineers to download and use this code for their needs.

REFERENCES

BARNARD, K., AND FORSYTH, D.A. 2001. Clustering Art. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2, 434–439.

BHATTACHARYA, J., AND PETSCHE, H. 2002. Shadows of artistry: cortical synchrony during perception and imagery of visual art. *Cognitive Brain Research*, 13, 179–186.

BISHOP, C.M. 2006. Pattern Recognition and Machine Learning. Springer Press.

BLAKE, A., AND YUILLE, A. 1992. Active Vision. MIT Press. Cambridge, MA.

CHANG, E.Y., Li, B., AND Li, C. 2000. Toward perception-based image retrieval. In *Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries*, 101–105.

FELSENSTEIN, J. 2004. PHYLIP Phylogeny Inference Package, Version 36.

GABOR, D. 1946. Theory of communication. *Journal of IEEE*, 93, 429–457.

GRADSHTEIN, I., AND RYZHIK, I. 1994. Table of integrals, series and products. 5 ed. Academic Press, 1054.

GRAY, S.B. 1971. Local properties of binary images in two dimensions. *IEEE Trans. on Computers*, 20, 551–561.

GREGORESCU, C., PETKOV, N., AND KRUIZINGA, P. 2002. Comparison of Texture Features Based on Gabor Filters, *IEEE Trans. on Image Proc.*, 11, 1160–1167.

GUREVICH, I.B., AND KORYABKINA, I.V. 2006. Comparative analysis and classification of features for image models. *Pattern Recognition and Image Analysis*, 16, 265–297.

HADJIDEMENTRIOU, E., GROSSBERG, M., AND NAYAR, S. 2001. Spatial information in multiresolution histograms. *IEEE Conf. on Computer Vision and Pattern Recognition*, 1, 702.

HARALICK, R.M., SHANMUGAM, K., AND DINSTEIN, I. 1973. Textural features for image classification. *IEEE Trans. on Syst., Man and Cyber.*, 6, 269–285.

JOACHIMS, T. 2006. Training linear SVMs in linear time, In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*, 217–226.

JONES-SMITH, K., AND MATHUR, H. 2006. Fractal Analysis: Revisiting Pollock's drip paintings. Nature, 444. 9–10.

KEREN, D., 2002. Painter identification using local features and naive Bayes. In: *Proceeding of the 16th International Conference on Pattern Recognition*, vol. 2, 474–477.

KAMMER, P., LETTNER, M., ZOLDA, E., AND SABLATNIG, R. 2007. Identification of drawing tools by classification of textural and boundary features of strokes, *Pattern Recognition Letters* 28, 710–718.

KIRSCH, J.L., AND KIRSCH, R.A. 1988. The anatomy of painting style: Description with computer rules. *Leonardo - Journal of Art Science and Technology*, 21, 437–444.

LATTO, R. 1995, The brain of the beholder, in: *The Artful Eye*, Gregory, R.L., Harris, S., ed. Oxford University Press, Oxford, UK.

Li, J., AND WANG, J. 2004. Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Trans. on Image Proc.* 13 (3), 340–353.

LIM, J.S. 1990. Two-Dimensional signal and image processing. Prentice Hall, 42–45.

MURPHY, R.F., VELLISTE, M., YAO, J., AND PORRECA, G. 2001. Searching online journals for fluorescence microscopy images depicting protein subcellular location patterns. In: *Proceedings of the 2nd IEEE International Symposium on Bioinformatics and Biomedical Eng.*, 119–128.

ORLOV, N., Shamir, L., MACURA, T., SHAMIR, L., JOHNSTON, J., ECKLEY, M., AND GOLDBERG, I. G. 2008. 2008. WND-CHARM: Multi-purpose image classification using compound image transforms, *Pattern Recognition Letters*, 29, 1684–1693.

ORLOV, N., JOHNSTON, J., MACURA, T., SHAMIR, L., AND GOLDBERG, I.G. 2007. Computer vision for microscopy applications, In: *Vision Systems - Segmentation and Pattern Recognition*. 221–242. Obinata, G., and Dutta, A. (Eds.). Vienna, Austria: ARS Pub.

OTSU, N. 1979. A threshold selection method from gray level histograms. *IEEE Trans. on Syst., Man and Cyber.*, 9, 62–66.

PREWITT, J.M. 1970. Object enhancement and extraction. Picture processing and psychopictoris, B. S. Lipkin and A. Rosenfeld, Eds. New York: Academic, 75–149.

RAMACHANDRAN, V.S., AND HERSTEIN, W. 1999. The Science of Art: A Neurological Theory of Aesthetic Experience. *Journal of Consciousness Studies*, 6, 15–51.

REWALD, J. 1963. Pissarro. Thames & Hudson, 118.

RODENACKER, K., AND BENGTSSON, E. 2003. A feature set for cytometry on digitized microscopic images. *Anal. Cell. Pathol.*, 25, 1–36.

SHAMIR, L. 2006. Human Perception-based color segmentation using fuzzy logic. *Intl. Conf. on Image Processing, Computer Vision and Pattern Recognition*, vol. II, 496–505.

SHAMIR, L., ORLOV, N., MACURA, T., ECKLEY, JOHNSTON, J.M., AND GOLDBERG, I. G. 2008. Wndchrm - An Open Source Utility for Biological Image Analysis. *BMC - Source Code for Biology and Medicine*, 3, 13.

SOLSO, R.L.. 1994. Cognition and the Visual Arts. MIT Press. Cambridge, MA.

SOLSO, R.L. 2000. The Cognitive Neuroscience of Art: A Preliminary fMRI Observation. *Journal of Consciousness Studies*, 7, 75–86.

TAMURA, H., MORI, S., AND YAMAVAKI, T. 1978. Textural features corresponding to visual perception. *IEEE Trans. on Syst., Man and Cyber.*, 8, 460–472.

TAYLOR, T.P., MICOLICH, A.P., AND JONAS, D. 1999. Fractal analysis of Pollock's drip paintings, *Nature*, 399, 422

TAYLOR et al. 2007. Authenticating Pollock paintings using fractal geometry, *Pattern Recognition Letters* 28, 695–702.

TEAGUE, M.R. 1979. Image analysis via the general theory of moments. *Journal of the Optical Society of America*, 70, 920–920.

VAN DEN HERIK, J.H., AND POSTMA, E.O. 2000. Discovering the visual signature of painters. In: *Kasabov, N. (Ed.), Future directions for intelligent systems and information sciences*. The future of speech and image technologies, brain computers, WWW, and bioinformatics, Physica Verlag (Springer-Verlag), Heidelberg, Germany. 129–147.

VAPNIK, V. 1995. The nature of statistical learning theory. Springer-Verlag, Heidelberg, Germany.

WALLRAVEN, C., KAULARD, K., KUMER, C., PEPPERELL, R., AND BULTHROFF, H.H. 2007. Psychophysics for perception of (in)determinate art, In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*, p. 115–122.

WIDJAJA, I., LEOW , W.K., AND WU, F.C. 2003. Identifying painters from color profiles of skin patches in painting images, In: *Proceedings of the International Conference on Image Processing*, vol. 1, p. 845–848.

ZEKI, S. 1999. Inner Vision, Oxford University Press, Oxford, UK.

Received Month Year; revised Month Year; accepted Month Year

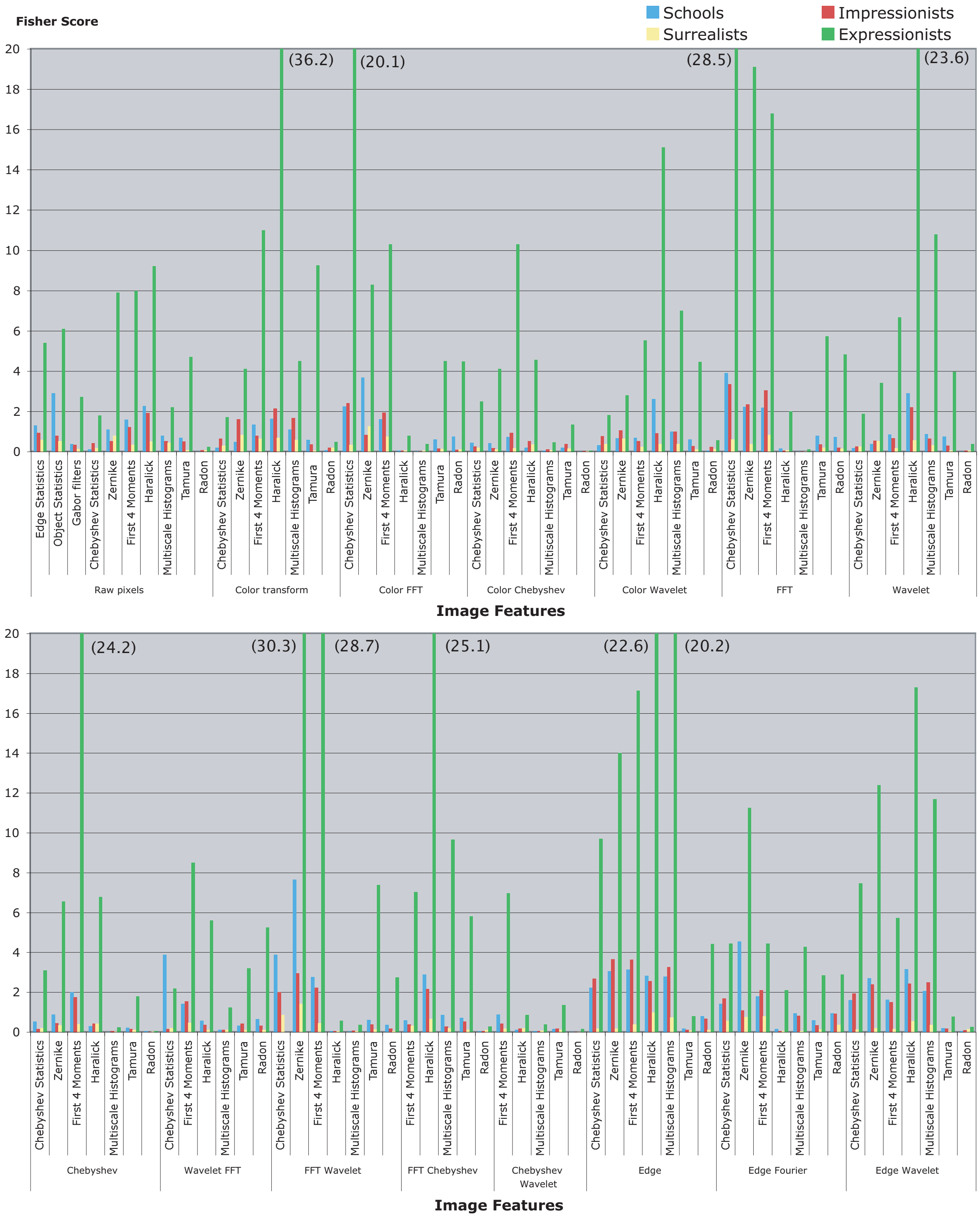

Fig. 6. Fisher Scores of the groups of features.

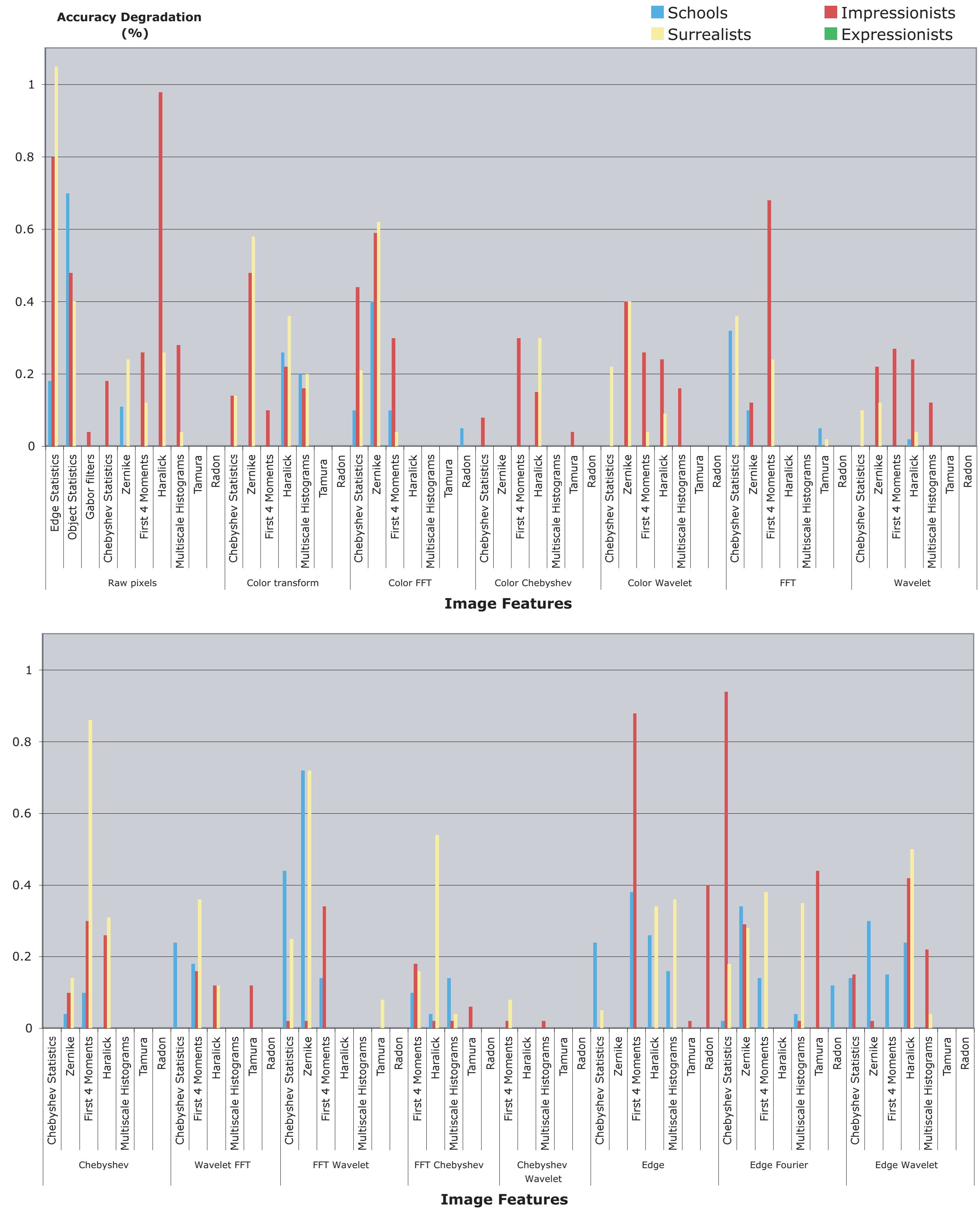

Fig. 7. Decrease in classification accuracy (%) when eliminating groups of features.

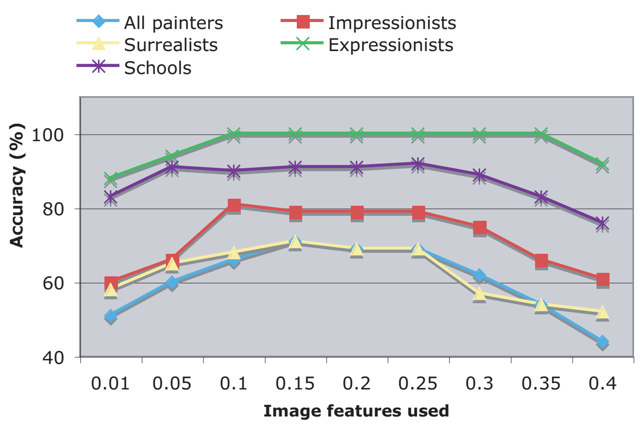

Fig. 8. Classification accuracy as a function of the percentage of image features used.