Classifying paintings by artistic genre: An analysis of features & classifiers

Conference Paper · November 2009

DOI: 10.1109/MMSP.2009.5293271 · Source: IEEE Xplore

5 authors, including:

Lisa M. Gandy
Central Michigan University

Scott E. Friedman
Smart Information Flow Technologies

Bryan Pardo
Northwestern University

Thrasyvoulos N. Pappas
Northwestern University

Classifying Paintings by Artistic Genre: An Analysis of Features & Classifiers

Jana Zujovic, Lisa Gandy, Scott Friedman, Bryan Pardo, Thrasyvoulos N. Pappas

EECS Department, Northwestern University
2145 Sheridan Road, Evanston, IL 60208, USA
{jana.zujovic, lisa.gandy, scott.friedman, pardo, pappas}@eecs.northwestern.edu

***Abstract*—This paper describes an approach to automatically classify digital pictures of paintings by artistic genre. While the task of artistic classification is often entrusted to human experts, recent advances in machine learning and multimedia feature extraction has made this task easier to automate. Automatic classification is useful for organizing large digital collections, for automatic artistic recommendation, and even for mobile capture and identification by consumers. Our evaluation uses variable-resolution painting data gathered across Internet sources rather than solely using professional high-resolution data. Consequently, we believe this solution better addresses the task of classifying consumer-quality digital captures than other existing approaches. We include a comparison to existing feature extraction and classification methods as well as an analysis of our own approach across classifiers and feature vectors.**

I. Introduction

Digital captures of artistic paintings are pervasive on the Internet and in personal collections. These encompass the works of the Old Masters which have been scrutinized and classified by human experts, as well as the work of current painters whose work is appreciated but not classified. The classification of paintings into their genre, or style, is often left to human experts within the discipline, but recent approaches have attempted to automate this task. An overview of image processing techniques and algorithms applied to fine art can be found in [1]. An accurate approach to automatically classify paintings by artistic genre has clear utility: museums and websites could quickly organize large digital collections, and consumer art appreciators could gain insight into a painting by automatically classifying a digital capture.

Artistic painting genres may span many different painters. While individual artists often have idiosyncrasies that make their own paintings recognizable, a genre reflects the properties shared across these artists. Consequently, this classification task may depend on different features than the task of classifying by individual artist. In this analysis, we use five genres for classification: Abstract Expressionism, Cubism, Impressionism, Pop Art, and Realism.

Art connoisseurs and professionals can easily identify a paintings genre based on the artist and year, as well as using visual cues. For instance, the motifs present in a painting, the color palette utilized and the techniques employed, such as stroke style, color mixing, edge softness, color reflection, parallel lines, and gradients often indicate the artistic genre of the painting. These properties provide a strong incentive for developing image processing techniques that attempt to extract features that roughly correspond to the perceptually salient aspects of a particular genre. However, we do not wish to address the semantics of a painting in our solution. For example, Magritte's *"Ceci n'est pas une pipe"* (Fig. 1) depicts a pipe with a caption beneath it, translating to "this is not a pipe." The content, rather than the technique, makes this a canonical work of surrealism, though the technique is similar to that of Realism. Extracting semantic content from digital media is a separate task; this paper focuses on classification using technical features.

Fig. 1. Rene Margitte, *Ceci n'est pas une pipe*

An overview of previous work is given in Section II. Feature extraction and classification problem are given in Section III. Description of the database and experiment setup are given in Section IV, and finally we present and analyze our results in Section V. Final conclusions are drawn in Section VI.

II. Background

With the development of image processing techniques and readily available data, digital processing of fine art images

MMSP'09, October 5-7, 2009, Rio de Janeiro, Brazil.
978-1-4244-4464-9/09/$25.00

has gained more attention from researchers. However, little work has been done towards classification into genres. Similar problems have been explored, however, and the intuition from these works leads to ideas for the task of classifying paintings by genre.

Firstly, one may ask if digital image processing techniques are even suitable for meaningful painting description. Genres are a higher-level semantic category, and thus the digital descriptors should, in some way, have correlation with the visual features human brains extract. One work that addresses this question is by Derefeldt et al. [2]. Here, the authors try to establish whether there is any correlation between features acquired by digital image processing and the Wölfflin's descriptors that characterize the development of art [3]. They conclude that, by using simple features, it is possible to grasp the concepts used for describing a painting, and we use this result to justify our approach.

Additional explorations into image processing of fine arts paintings include other problems, aside from classification into genres. Kammerer et al. [4] used the information acquired with infrared cameras to capture the "hidden" parts of the paintings - namely, the sketches underneath the paint. They analyzed the strokes of the sketches to determine the tool they were made from, and this is used as one of the first stages in systemizing the paintings' genres. Sablatnik et al. [5], in addition to face models, utilized similar stroke analysis technique for classifying miniature portrait paintings according to the artist. Abas et al. [6] analyzed the cracks on a painting's surface and used that as a classification feature. Lewis et al. [7] developed an art-oriented CBIR, with possibility to retrieve original paintings based on a small sub-image, a low-quality version of the original or based on crack patterns.

Classification of traditional Chinese paintings has been explored in [8] and [9]. Li et al. [8] use wavelet decomposition of images as the basis for feature extraction, and 2-D multiresolution hidden Markov models for classification. This approach would arguably be ineffective for our task, given that it only utilizes grayscale information, due to the nature of Chinese traditional art. Lu et al. [9] use very simple color descriptors and gray-level co-occurrence matrices. Given the diversity of our database and the nature of western paintings, utilizing simple color descriptors such as mean and variance of one channel within the entire image may not be sufficient. Also, even though co-occurrence matrices have produced good results, they are not generally superior to other texture extraction methods and require much computational time [10]. Deac et al. [11] explored the significant features for assessing the authenticity of paintings. Their features included gray level co-occurrence matrices, Gabor texture features, edge features, and also color features. However, their method is tested only on two artists. The conlusion they draw is that the most discriminating feature is the texture information, which is related to brush stroke. One rather simple method is presented in [12]. It is based on DCT values and a naïve Bayes classifier. The performance was tested on 5 different artists. This method only addresses the grayscale aspect of images and, as stated earlier, would not produce good results on the diverse images included in our database. On the other hand, Lombardi et al. [13] define features as "palette" and "canvas", both addressing the colors within the image. The palette feature gives the total number of different RGB colors used in the image, and canvas features capture frequency and spatial distribution of colors. It is shown that best results are obtained by using both texture and color features, e.g., in [14].

A recent work by Shen [15] attempts to classify western paintings according to artist. It relies on both global color and texture features, and local texture features. Color features consist of a quantized HSV histogram and a color layout descriptor, while texture features are characterized by Gabor features. In addition, shape features are incorporated as a histogram of directions of edges. The classification is done for 25 classes - artists - by a RBF neural network.

The work that most closely relates to ours is done by Gunsel et al. [16]. They extract features from the luminance component of the image, however, they argue that the variations in luminance correspond to some color features, e.g., low luminance represents a dark color etc. Their classification is broken into 3 genres, and they tested various classifiers.

III. Our Approach

The paintings classification consists of two stages: feature extraction and classification. As in a conventional Content-Based Image Retrieval, our method uses both texture and color information. We can thus separate the feature extraction into gray-level processing and color processing. Classification is done jointly on the two types of features.

A. Gray-level Features

A grayscale image contains useful information about texture. Gabor filters are widely used as texture descriptors, since they are shown to exhibit properties of low-level processing in human eyes [17]. Our choice was to use Steerable Filter Decomposition since they are, like Gabor filters, inspired by biological visual processing, and also have nice properties, such as translation-invariance and rotation invariance, as claimed by Portilla and Simoncelli [18]. From each subband, the mean and variance of the absolute values of coefficients are extracted. This roughly corresponds to the energy in the given subbands, and characterizes the strokes utilized by the artist [11].

Other gray-level features are edges. Edges, i.e., their relative frequency within one painting, can be very informative, since some painting styles are characterized by smudged and subtle edges, like impressionism, while in, for example, pop art, the edges are very pronounced. Our edge feature consists of number of pixels in the image that are labeled as an edge, relative to the total number of pixels, extracted by Canny edge detector, for different sensitivity thresholds. The rationale is, strong edges would be present in the image for any threshold level, while the subtle transitions would only show up for lower thresholds.

This can be illustrated in Figure 2. In the first row is given a grayscale version of Jasper John's *Flag*, a good representative of pop art genre. We can see how the edge maps are almost the same for two extreme thresholds. In the second row we have Monet's *Sunset*, and the edge maps that differ a lot according to the threshold. Machine learning techniques should be able to capture this rule and utilize it for classification.

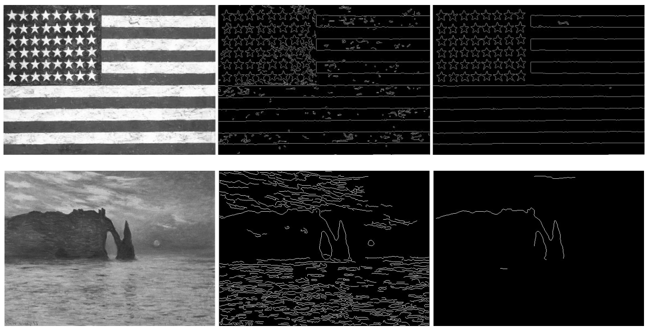

Fig. 2. Original grayscale images and edge maps for low and high thresholds

B. Color Features

The images we are using are registered in RGB colorspace. However, it is meaningful to use a colorspace that is more closely related to the way humans perceive the colors. HSV space is particularly suitable, since each component can be treated separately. H defines the hue, namely, which colors are being used; S defines the saturation of colors, or pureness; and V defines the value, or how dark the colors are. Therefore, our feature is an 8-bin histogram of each of these three values. This should characterize the color palette used (H), the color mixing (S), and also the lightness of the image (V). This results in a $8 \cdot 3 = 24$ color features. The histograms are scaled so that they represent the relative numbers, i.e., the sum of all bins for one channel equals 1.

C. Classification

For classification, we used the open source program Weka [19], that allows for choosing between many different types of classifiers. We tested the following: naïve Bayes classifier, Knn with 1 nearest neighbor, Knn with 10 nearest neighbors, Support Vector Machine, Artificial Neural Network with 4 hidden layers, and AdaBoosted [20] J48 decision tree.

IV. Experiment Setup

A. Database

Our database consists of 353 different paintings, belonging to 5 artistic genres: Abstract Impressionism, Cubism, Impressionism, Pop Art and Realism. The paintings have been acquired through various websites, and thus the size and quality of images vary. This works both towards our advantage and disadvantage: we may get worse classification results due to these inconsistencies, however, this also may prove our system to be robust to such changes, therefore making it suitable for large databases that have not been necessarily acquired under the same conditions.

The source of the original images and distribution over genres are given in Table I.

TABLE I
Database sources and distribution

Genre	Source	#
Abstract Expressionism	Artlex, Google	59
Cubism	Artlex, Google	60
Impressionism	CARLI Digital Collections	96
Pop Art	Artlex, Google	58
Realism	CARLI Digital Collections	80

B. Feature Extraction

All feature extraction is done using Matlab, with Steerable Pyramid implementation taken from the website [21].

In this work, the 3-scales 4-orientations decomposition is used. Note that, apart from 12 steerable subband filters, we also have a low-pass and a high-pass subband.

Edges are found with Canny edge detector, with the following thresholds: 0.2, 0.3, 0.4 and 0.6.

H, S and V components are divided into 8 uniformly spaced bins between $[0, 1]$, yielding a total of three 8-bins histograms.

V. Results

The analysis of results is done in two directions: firstly, we compare classification results for various classifiers using proposed features. Secondly, we compare our work to the work in [16], and also to the modified version of [2].

A. Comparison of Proposed Features vs. Classifiers

For this part, we compared how single features perform over different classifiers, and also their joint performance. The results are given in Table II. Label "Pyramid" or "P" denotes the features obtained through pyramid decomposition; "Edge" or "E" are all edge features, for 4 different thresholds; and "HSV" refers to the HSV quantized histograms. "Knn" stands for the K nearest neighbors method, "ANN" is artificial neural network with 4 hidden layers, "SVM" is linear kernel support vector machine, and "AdaBoost" is the AdaBoosted J48 decision tree with 100 iterations. the classification accuracy is computed using 10 fold cross validation and averaged across folds, which is suitable for small datasets like ours.

TABLE II
Proposed Features vs. Classifiers (accuracy in %)

	Pyramid	**Edge**	**HSV**	**P+E**	**P+E+HSV**
Naïve Bayes	40.8	35.4	53.8	42.8	48.7
Knn1	54.4	36.8	39.1	53.8	47.6
Knn10	50.1	40.2	46.2	53.8	57.5
ANN	59.2	38.5	50.4	58.9	64
SVM	62	44.2	47.9	59.5	57.8
AdaBoost	62	39.4	57.5	63.2	**68.3**

Based on these results, several comments can be made. As expected, the use of multiple features leads to better results overall. In fact, for some classifiers the additional features degrade their performance, which can be due to the augmented dimensionality, noise in the newly added feature etc. Also, there is no "best classifier" - for each set of features, there is a classifier that works best, but that is not necessarily the same classification method for all the different features or their combinations.

The best performance overall can be achieved by increasing the number of iteration to 1000 in AdaBoosted J48 tree algorithm. The best overall performance is 69.1%. The number of iterations drastically increases the training time, however, this is not meant to be a real-time application and therefore training time should not be regard as a performance measure.

B. Comparison of Proposed Features vs. Related Work

In this subsection, we compared the performance of our best set of features (that is, all features combined), versus the features proposed in [16]. Also, inspired by work of Derefeldt et al. [2], in place of HSV 8-bin histograms, we computed for each channel following statistical measures: mean, variance, skewness and kurtosis, and evaluated performance of such features vs. using quantized histograms. The comparison is given in Table III. "HSVs" denote the statistical measures for HSV channels.

TABLE III
PROPOSED FEATURES VS. RELATED WORK (ACCURACY IN %)

	Gunsel	**HSVs**	**HSV**	**P+E+Hs**	**P+E+H**
Naïve Bayes	49	42.8	53.8	45.6	48.7
Knn1	38.8	33.4	39.1	50.4	47.6
Knn10	42.2	45.6	46.2	56.1	57.5
ANN	46.4	42.8	50.4	64	64
SVM	49.8	46.7	47.9	64.3	57.8
AdaBoost	47	45.6	57.5	67.1	**68.3**

These results show clear advantage of the proposed method vs. the one by Gunsel et al. However, in comparison, using HSV statistics in place of quantized histograms proves to be worse both as stand-alone feature and also combined with the rest, but in that case the performance is degraded only by a little (and not even for all classifiers). This leads to the conclusion that it may be possible to use 4 in place of 8 descriptors for each channel with no significant impact on the performance, but having 12 features less may or may not prove to be significant in computational time.

C. Analysis of Results

Here, we were present the results obtained with AdaBoosted J48 decision tree with 1000 iterations, as it proved to be the best classification method for the given dataset and features.

Accuracy over different genres and overall is given in Figure 3. We can see that Abstract Impressionism is the most difficult one to correctly classify. This could be explained by the fact that this genre resembles cubism and pop-art, in the image processing sense. The presence of strong lines and corners can easily fool the classifier into erroneously categorizing an abstract impressionism painting as, say, a cubist's.

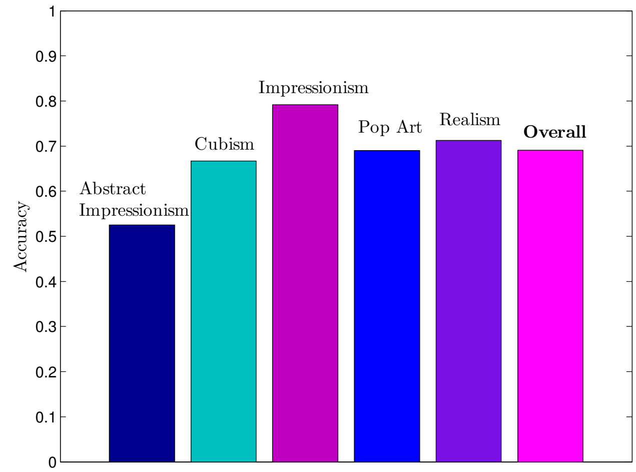

Fig. 3. Accuracy over genres and total accuracy

More in-depth analysis can be derived from the confusion matrix, given in Table IV. It shows that there are problems with misclassifying abstract impressionism paintings as cubism or pop art, which we already explained. Also, there is a considerable misclassification of realism paintings into impressionism category. This may best be explained by an example. In Figure 4, we see a correctly classified impressionistic painting (Pissarro, *Corner of the Hermitage, Pontoise*) and a misclassified realism painting (Thomas, *Max Schmitt in a single scull*). These two images that have very similar color composition; texture in the realism painting is being pronounced from the cracks of the paint and from the water reflections; most of the edges are produces by the change in color, not lighting, and they have similar percentage of edges for all thresholds. These may be the reasons that the classifiers mistake the realism painting as impressionism painting.

TABLE IV
CONFUSION MATRIX

	Classified as:				
True Class ↓	Abs.Imp.	Cubism	Imp.	PopArt	Realism
Abs.Imp.	31	12	4	11	1
Cubism	4	40	6	9	1
Imp.	3	3	76	0	14
PopArt	9	8	0	40	1
Realism	1	1	21	0	57

VI. CONCLUSIONS

We presented a simple and efficient way of feature extraction for the purpose of classification of paintings into genres. Our proposed features address the salient aspects of a painting - color, texture and edges - and are proven to work well in the real world scenario. Since our database was not uniform in quality or image size, we would expect better performance with the ideal, laboratory conditions, as would be museum databases. One possible direction for future research is separating edges from the texture. Subband decompositions inevitably incorporate edges, since they contribute to the

(a) (b)

Fig. 4. Two paintings belonging to realism (4(a)) and impressionism (4(b))

overall spectral content of the images, and it may be beneficial to develop a method that would separate edge and texture analysis. Our method outperforms the previous works, however, there is still a lot of space for improvement.

References

[1] F. Bartolini, V. Cappellini, A. D. Mastio, and A. Piva, "Applications of image processing technologies to fine arts," in *Optical Metrology for Arts and Multimedia*, R. Salimbeni, Ed., vol. 5146, no. 1. SPIE, 2003, pp. 12–23.

[2] G. Derefeldt, S. Nyberg, J. Alfredson, and H. Allberg, "Is wolfflin's system for characterizing art possible to validate by methods used in cognitive-based image-retrieval (cbir)?" in *Proceedings of SPIE-The International Society for Optical Engineering*, vol. 6057. International Society for Optical Engineering, 2006, p. 605710.

[3] H. Wölfflin, *Principles of Art History. The Problem of the Development of Style in Later Art*. New York, USA: Dover Publication, Inc., 1932.

[4] P. Kammerer, M. Lettner, E. Zolda, and R. Sablatnig, "Identification of drawing tools by classification of textural and boundary features of strokes," *Pattern Recognition Letters*, vol. 28, no. 6, pp. 710 – 718, 2007.

[5] R. Sablatnig, P. Kammerer, and E. Zolda, "Hierarchical classification of paintings using face- and brush stroke models," in *in Proc. of 14th Int. Conf. on Pattern Recognition*, 1998, pp. 172–174.

[6] F. S. Abas and K. Martinez, "Classification of painting cracks for content-based analysis," in *Machine Vision Applications in Industrial Inspection XI*, M. A. Hunt and J. R. Price, Eds., vol. 5011, no. 1. SPIE, 2003, pp. 149–160.

[7] P. H. Lewis, K. Martinez, F. S. Abas, M. F. A. Fauzi, S. C. Y. Chan, M. J. Addis, M. J. Boniface, P. Grimwood, A. Stevenson, C. Lahanier, and J. Stevenson, "An integrated content and metadata based retrieval system for art," *Image Processing, IEEE Transactions on*, vol. 13, no. 3, pp. 302–313, 2004.

[8] J. Li and J. Wang, "Studying digital imagery of ancient paintings by mixtures of stochastic models," *Image Processing, IEEE Transactions on*, vol. 13, no. 3, pp. 340–353, March 2004.

[9] G. Lu, Z. Gao, D. Qin, X. Zhao, and M. Liu, "Content-based identifying and classifying traditional chinese painting images," in *Image and Signal Processing, 2008. CISP '08. Congress on*, vol. 4, May 2008, pp. 570–574.

[10] T. Randen and J. Husoy, "Filtering for texture classification: a comparative study," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 4, pp. 291–310, Apr 1999.

[11] A. Deac, J. van der Lubbe, and E. Backer, "Feature selection for paintings classification by optimal tree pruning," in *Multimedia Content Representation, Classification and Security*, 2006, pp. 354–361.

[12] D. Keren, "Painter identification using local features and naive bayes," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 2, 2002, pp. 474–477.

[13] T. Lombardi, S.-H. Cha, and C. Tappert, "A graphical user interface for a fine-art painting image retrieval system," in *MIR '04: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, 2004, pp. 107–112.

[14] S. Saha, A. Das, and B. Chanda, "Cbir using perception based texture and colour measures," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2, Aug. 2004, pp. 985–988.

[15] J. Shen, "Stochastic modeling western paintings for effective classification," *Pattern Recognition*, vol. 42, no. 2, pp. 293 – 301, 2009, learning Semantics from Multimedia Content.

[16] B. Gunsel, S. Sariel, and O. Icoglu, "Content-based access to art paintings," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2, Sept. 2005, pp. 558–561.

[17] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex." *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, December 1987.

[18] J. Portilla and E. P. Simoncelli, "Texture modeling and synthesis using joint statistics of complex wavelet coefficients," in *In IEEE Workshop on Statistical and Computational Theories of Vision, Fort Collins*, 1999.

[19] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd ed. San Francisco, USA: Morgan Kaufmann, 2005.

[20] Y. Freund and R. E. Schapire, "A Short Introduction to Boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, no. 5, pp. 771–780, 1999.

[21] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack, "Image and video quality assessment research at live." [Online]. Available: http://live.ece.utexas.edu/research/quality/