DOI: 10.53375/ijecer.2022.208 INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING RESEARCH
VOL. 2, NO. 1, 2022

Classification of Pointillist paintings using colour and texture features

Kristina Georgoulaki*
Department of Informatics and Computer Engineering, University of West Attica, Athens, Greece
Email: cgeorg@uniwa.gr
** Corresponding author*

Abstract: **Fine art paintings classification based on artistic style is a field of growing interest. Pointillist style is one of the most easily recognized painting styles by humans, due to its characteristic tiny detached paintbrushes of pure colour. In this paper automatic discrimination of artworks belonging to the style of Pointillism is investigated. The opposite styles considered are Cubism, Purism, Naïve art and Impressionism. Several colour and texture features are considered and a feature selection procedure is employed to reveal the most relevant ones to pointillist movement. Binary classification is performed, both in supervised and unsupervised mode, to assess the features' discriminative ability. A small number of selected features is shown, by simulations results, to be quite powerful predictors resulting in a classification accuracy of 94% for a SVM classifier, 93.5% for a KNN classifier and 87% for a k-means classifier.**

Keywords: **Art painting styles, Feature selection, Image features, Pointillist style classification.**

I. INTRODUCTION

The widespread paintings digitization made the automatic classification of art styles a growing interest field of research. Artistic styles such as Pointillism, Cubism, Impressionism, etc. have a set of distinctive properties useful for the style discrimination task. The question of which features should be adopted to encode style information in paintings has been addressed in several studies [1-4].

This paper concerns the extraction of features from fine art paintings, in order to distinguish the style of Pointillism from other artistic styles. Pointillism is one of the most easily recognized painting styles by a human. Pointillist artworks are inferred by their granular repetitive pattern and by the representation of objects like a collection of colour points [5]. However, human ability differs from computers ability to classify artworks. To distinguish pointillist paintings from other styles the appropriate features should be used.

In this paper, many different and sometimes redundant features are utilized in an attempt to conclude the pointillist characteristics of artworks. Four styles are considered to collate with Pointillism, namely: Purism, Naïve art, Cubism and Impressionism. The adopted features are low level global features describing colour, intensity and texture characteristics of artworks and are enumerated as follows. Initially, granulometry is employed to calculate the size distribution of paintbrushes in the artworks [6]. It is observed that pointillist and non-pointillist artworks exhibit a highly different paintbrushes size distribution and this can form a discriminative feature between the two opposite styles. Other features used are, the total number of edges in the painting, the mean length of edges, the degree of color based image segmentation, the number of local minima and maxima in the grayscale image histogram, Haralick texture features [7], the gradient magnitude [8], the means and standard deviations of colors of pixels within the whole region of the image, in four color spaces, the percentage of dark colors, the color range that the peak point of the luminance histogram corresponds, the deviation of average luminance and the deviation of luminance distribution from Gauss distribution [8].

After feature extraction from the digitized paintings, the features reduction step takes place. Feature selection procedure reduces the number of features by eradicating irrelevant or redundant data. The ReliefF algorithm, a popular supervised approach for the evaluation of the discrimination capability of each individual feature is employed for the feature selection step [9].

Finally, classification of the paintings is performed using two different approaches: a) supervised classification by means of the celebrated k-nearest neighbor algorithm (KNN) algorithm and a support vector machine (SVM) classifier b) unsupervised classification with the k-means algorithm [10]. The feature selection and the supervised classification results are used for the selection of the 10 most important features. Equipped with these selected features unsupervised classification is performed and a classification accuracy of 87% is achieved. Finally, comparison of the adopted classification schemes with a SVM classifier using SURF features is also evaluated.

The paper is organized as follows. In Section 2, the image data set and the preprocessing of artworks are described. In Section 3, the employed features are presented. The next Section concerns the experiments setup, describing the procedures of features extraction, features selection and classification. In Section 5, the experimental results used to evaluate the classification performance are described and in Section 6 discussion follows. Finally, in Section 7 conclusions are drawn.

II. DATA SET COLLECTION AND PREPROCESSING

The fine art paintings employed in this research are gathered from the "Wikiart" visual art encyclopedia [11].

Except from Pointillism the opponent styles considered are Cubism, Naïve art, Purism and Impressionism. The opponent styles are considered so as to include styles which are quite different from Pointillism (Cubism, Purism) and styles very similar to Pointillism (Impressionism). Actually, Pointillism and Impressionism tend both to grasp the reality of nature in its luminous essence and by trying to capture it on the canvas through the use of colour rather than drawing [5]. The similarity of Impressionist and Pointillist artworks makes classification task more cumbersome and the performance of the classifier is deteriorate.

Paintings of sixteen artists, representing the five different schools of art were considered. The artists concerned for the representation of each style are: for Pointillism: Georges Seurat, Camille Pissarro, Henri-Edmond Cross, Albert Dubois-Pillet, Theo van Rysselberghe, Paul Signac and Marevna, for Purism: Ralston Crawford, Amedee Ozenfant, Le Corbusier and Fernand Leger, for Naïve art: Frida Kahlo, Henri Rousseau, for Cubism: Pablo Picasso, Juan Gris and for Impressionism: Claude Monet. For each painter 10 raw images are collected from Wikiart database, except from Seurat and Monet, which are represented by 20 paintings each.

Each artist was represented by different types of images (e.g., still life, landscape, etc.). However, since many painters dealt with more than one style, the paintings gathered from each artist are carefully selected to belong to the specific style of interest.

Images are pre-processed and then used for features extraction. Each image was normalized to 512 x 512 pixels. Some features are extracted from the standard colour space adopted for digital images (i.e. the RBG colour space), some others from the YCBCR colour space and some features are extracted from the gray scale space.

III. ADOPTED FEATURES

The employed features used for the discrimination of Pointillism are described in the followings:

a) Granulometric Feature. Granulometry is employed to calculate the size distribution of paintbrushes in the artwork [12]. Granulometric analysis, based on the sequence of morphological opening and closing operations and the quantification of particles of different sizes [6], is used to estimate the intensity surface area distribution of paintbrushes of the artistic image as a function of size. The minima of the first derivative of the intensity area indicate the amount of paintbrushes of a specific radius and the maximum value of the (negative) first derivative of the intensity area corresponds to the total number of most frequent sized paintbrushes. As seen in Fig. 1, in a pointillist painting, this maximum value is sizeable since the painting is dominated by small, dotted paintbrushes (of radius 2), whereas, in a non-pointillist painting the distribution of paintbrush sizes is almost uniform.

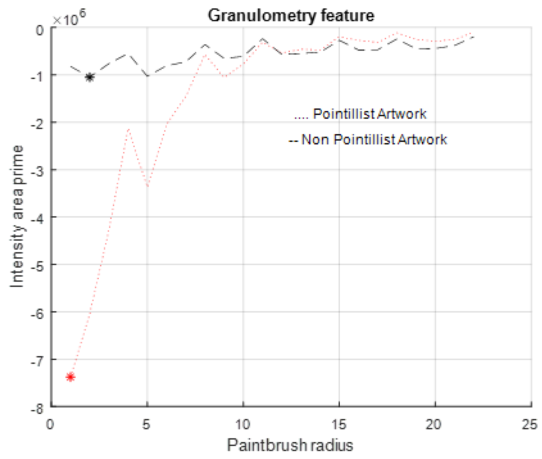

Fig. 1. The first derivative of the intensity area versus paintbrush radius for a) pointillist artwork and b) non-pointillist artwork.

b) The degree of colour based image segmentation. For every pixel in the image the average distance from neighbour pixels in the 3D RGB colour space is calculated. Then a matrix is formatted which contains the calculated distances. If the mean value of the distance matrix is lower than a threshold then the image is assumed to have a single compound colour, otherwise quad tree decomposition [13] is performed recursively to define the division of image in regions with different colours [14-15]. This number is high for pointillist paintings due to the granular texture of images and smaller for styles with large compound-coloured regions, as for example in Purism where "objects are represented as powerful forms devoid of any detail" [16].

c) The number of gray scaled image histogram local minima and maxima. The histogram of the gray scaled version of the image is considered. The local maxima (peaks) and the local minima (valleys) of the histogram are found. It is observed that in pointillist paintings the number of histogram peaks and valleys is very low due to the smooth chromatic expression in the artwork. In the contrary, in paintings with abrupt chromatic changes and with a great variety in the colours pallet, like in Purism and Cubism, the number of peaks and valleys is bigger. For example in Fig. 2 is shown the histogram of the gray scaled version of artwork "Fisherman" (Fig. 3) of H. Cross (Pointillism) and in Fig. 4 the histogram of the gray scaled version of artwork "Duplicate" (Fig. 5) painted by A. Ozenfant (Purism).

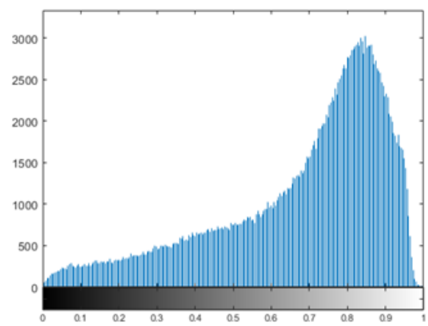

Fig. 2. Histogram of gray scaled pointillist artwork "Fisherman" by H. Cross. The number of peaks and valleys is small.

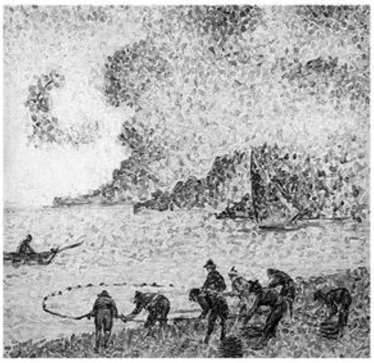

Fig. 3. "Fisherman" by H. Cross (Pointillism).

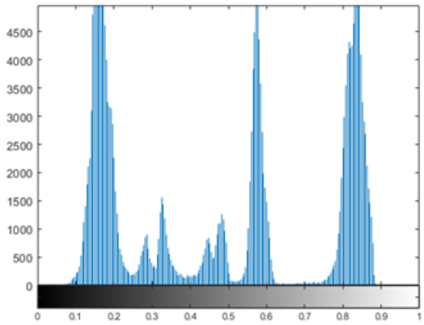

Fig. 4. Histogram of gray scaled purist artwork “Duplicate” by A. Ozenfant.The number of peaks and valleys is big.

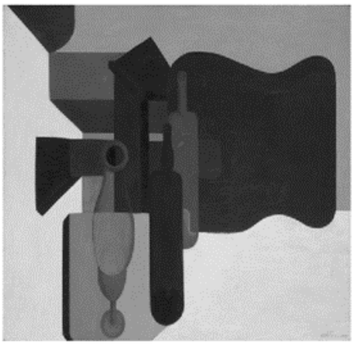

Fig. 5. “Duplicate” by A. Ozenfant (Purism).

d) Number of edges in the image and mean length of edges in the image. In a pointillist artwork the encountered edges are numerous, subtle and of small length, since, essentially, pointillist artists abandoned pronounced lines. On the contrary, the edges of artworks of many other styles like Purism, Cubism and Naïve art are explicit and clear and they are usually few and sizable [17]. For the edge detection of the grayscale image the Roberts method is utilized. Edges quantity is calculated by the normalized number of detected lines with length higher than 8 pixels divided by the total image pixels. The feature of mean edges length is described as the average number of pixels of all the encountered edges. For example, a pointillist artwork of H. Cross is shown in Fig. 6 and an illustration of the estimated edges is shown in Fig. 7. Their number is calculated as 2735 and their mean length is 1.8 pixels. From the style of Purism an artwork of Le Corbusier is shown in Fig. 8 and the corresponding detected edges are shown in Fig. 9. Their number is 218 and their mean length is 27.2 pixels.

Fig. 6. “A Venetian Canal” by H. Cross (Pointillism)

Fig. 7. Edges of the artwork “A Venetian Canal” by H.Cross (Pointillism).

Fig. 8. “Still Life” by Le Corbusier (Purism).

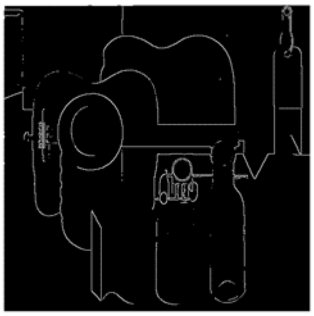

Fig. 9. Edges of the artwork “Still Life” by Le Corbusier (Purism).

e) Haralick features. Haralick features use Gray Level Co-occurrence Matrices (GLCM) to capture relationships between pairs of pixels separated by a certain distance and in a given angle [18]. Fig. 10 shows the GLCM of a non-pointillist painting indicating that repetitions are restricted mainly to the same colour (diagonal) and Fig. 11 illustrates the GLCM of a pointillist painting where patterned colours repetitions create numerous gray tone spatial dependencies. Fourteen Haralick features are utilized which correspond to the measures proposed by [18] calculated on a distance equal to one (d=1) and for the mean value of four directions ($\theta = 0^0, 45^0, 90^0, 135^0$).

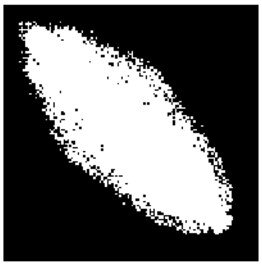

Fig. 10. GLCM of non-pointillist artwork. Repetitions are restricted mainly to the same colour (diagonal).

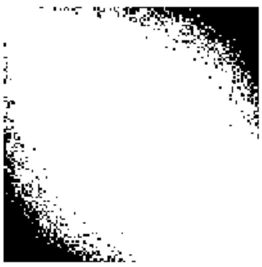

Fig. 11. GLCM of pointillist artwork. Repetitions are numerous.

f) The means and standard deviations of intensity of pixels within the whole region of the image, in four colour spaces, namely, RGB, CIEYxy, CIELUV and CIELAB, are computed giving rise to 22 features [19].

g) The percentage of dark colours, defined as the number of pixels whose luminance value corresponds to [0,64] range divided by the total number of pixels [8].

h) The colour range that the peak point of the luminance histogram corresponds [8].

i) The artwork is divided into nine identical blocks and then the deviation of average luminance within each block from the average luminance of the entire image is calculated [8].

j) The deviation of luminance component distribution from Gauss distribution [8].

k) The gradient magnitude defined as the normalized square root of the sum of the squares of the two directional gradients of the gray scaled image.

IV. EXPERIMENTS SET UP

Two particular experimental sets up were evaluated considering classification of the paintings a) in supervised and b) in unsupervised mode. Both experiments consider classification of pointillist paintings versus cubist, naïve art, purist and impressionist paintings.

A. Supervised classification

In the first set of experiments style discrimination is treated as a supervised classification task. The steps for the classification are summarized below:

a) Raw image data preprocessing: As described in Section 2.

b) Features extraction: The 47 features described in the previous Section are considered, as they offer insight into artworks style classification.

c) Feature normalization: Usually, the raw numbers produced by the feature extraction step are not scaled consistently. Normalization of features ensures the equal importance of all features [2]. Features are standardized by means of mean subtraction and standard deviation division.

d) Feature set partition: The set of features is randomly partitioned into two disjoint subsets, namely, the training set and the testing set. By a rule of thumb, the training set contains about 80% of the total features, whereas, the remaining features are used for testing.

e) Feature selection: The predictors are ranked using the ReliefF algorithm with k nearest neighbors [9]. Feature selection step can give us insight into the semantic relation of the ranked features and the pictorial characteristics of Pointillism and consequently their influence on the classification process.

f) Classification: Pointillism discrimination from others styles is treated as a binary classification task, with the first class corresponding to Pointillism and the second class to the opposite styles. The KNN classifier and the SVM classifier are used for classification of the paintings within the testing set.

g) Validation: As a measure of performance, the number of errors (false classification) detected in the classification of the testing set of features is adopted. Usually, the sequence of steps *(d)-(g)* is repeated several times.

B. Unsupervised classification

In this set of experiments, classification of the paintings is performed without the use of training data. However, the results obtained from the supervised feature selection step are utilized to restrict the features used for the unsupervised classification to the most important ones.

Unsupervised clustering is performed by means of the well-known k-means algorithm [20].

The k-means algorithm faces the optimization problem of finding the observations partition to k clusters by minimizing the within clusters sum-of-squares

$$\min_{(S_j)_{j=1}^k} \sum_{j=1}^{k} \sum_{i \in S_j} \left\| \boldsymbol{x_i} - \boldsymbol{c_j} \right\|^2$$

where $(S_j)_{j=1}^k$ is the requested partition, $\boldsymbol{x_i}$ the observations vector and $\boldsymbol{c_j}$ the means of the clusters.

V. EXPERIMENTAL RESULTS

A. Supervised classification

In order to reduce the number of considered features and to find the features that best highlight the style of Pointillism, feature selection procedure is employed. The most important features were found and ranked by means of the ReliefF algorithm. After experimentation, the parameter k of the ReliefF is set to 5. The resulting 10 most important features are listed below:

- Degree of color based image segmentation.
- Granulometric feature.
- Mean length of edges and their total number.
- Number of gray scaled image histogram local maxima.
- Haralick features, namely, inverse Difference or Homogeneity, that is, a spatial autocorrelation measure measuring distribution around to the GLCM diagonal, Entropy and Difference Entropy, which are related to the randomness of intensity and Information measure of correlation.
- Gradient magnitude.

The less informative features are mostly features related to the colour of paintings, that is, the mean and standard deviation of colour intensity in the colour spaces.

As the next step, classification of artworks is performed by means of a KNN and a SVM classifier. After experimentation the parameter k of KNN is set to 5 and the SVM kernel function adopted is the Gaussian kernel.

The separability properties of the descriptors and the effect of the number of used features in the classification accuracy are illustrated in Fig. 12. The ranking of the features is dictated by the results of the Relieff algorithm. The classification error count in terms of the number of used features is achieved as a result of the mean of 20 iterations of steps *(d)-(g)* of the classification procedure.

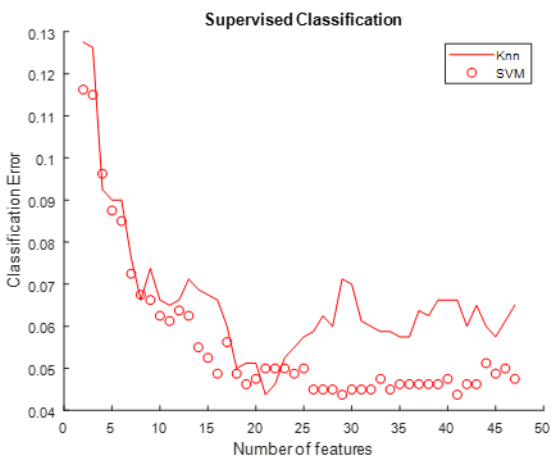

Fig. 12. The classification error count in terms of the number of used features for a KNN and an SVM classifier.

From the figure is seen that for both classifiers the artworks separability is high. The SVM classifier outperforms the KNN classifier, however, as seen from Fig. 12, some additional features may degrade classifiers performance [3] and especially, for the KNN algorithm, the evolution of the error count in terms of the number of used features is worsening due to the curse of dimensionality.

B. Unsupervised classification

In this experiment unsupervised classification is considered. In this case, no training data are needed and thus the discriminative abilities of the features are tested in a bigger number of testing data.

In this case, the 10 most important features, as dictated by feature selection step, are used to feed the k-means algorithm employed for the unsupervised classification task. The confusion matrix resulting from the unsupervised classification of artworks using 10 features appears in Fig. 13. Accuracy is 87%, Precision 79% and Recall 100%, revealing that all pointillist artworks are classified correctly.

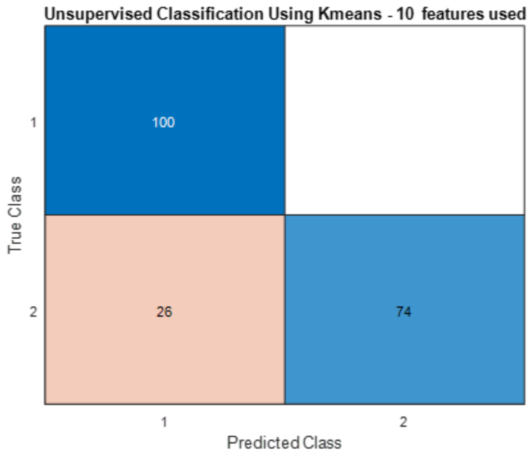

Fig. 13. Confusion matrix from unsupervised classification using k-means and 10 proposed features.

Finally, a last experiment is performed focusing on the favorable situation where Impressionism is excluded from the opponent styles. Pointillism is an artistic style branching from Impressionism and thus these two styles share many common characteristics. Subtracting Impressionism from the set of opponent styles, makes classification process much more straightforward. This is highlighted in the scatter diagram of Fig. 14 illustrating the cluster of pointillist artworks (red dots), the cluster of non-pointillist (excluding Impressionism) artworks (cyan circles) and the cluster of impressionist artworks (black stars). The two features encountered for the formation of the scatter diagram are the Granulometric feature and the Mean length of edges. From the figure is observed that the cluster of pointillist features is almost linearly separable from the cluster of non-pointillist (excluding impressionist) features.

For a fair comparison of the two classes, 20 pointillist artworks are also removed, leading to 160 artworks used as observations. The confusion matrix corresponding to the case of unsupervised classification is seen in Fig. 15. Accuracy is 96.2%, Precision 93% and Recall 100%. Result analysis shows that, in the case of supervised classification a 100% accuracy is achieved.

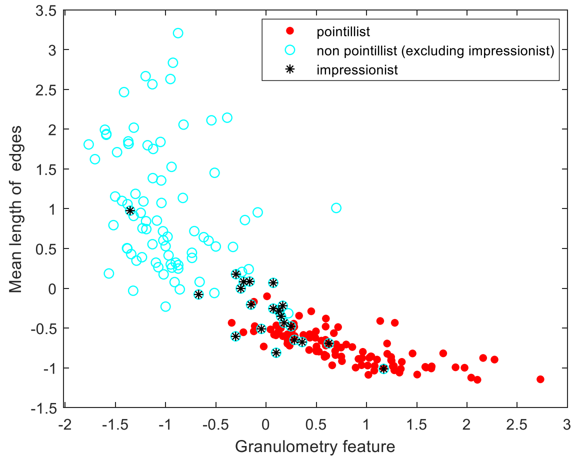

Fig. 14. Scatter diagram including the cluster of pointillist features, the cluster of non-pointillist (excluding impressionist) features and the cluster of impressionist features.

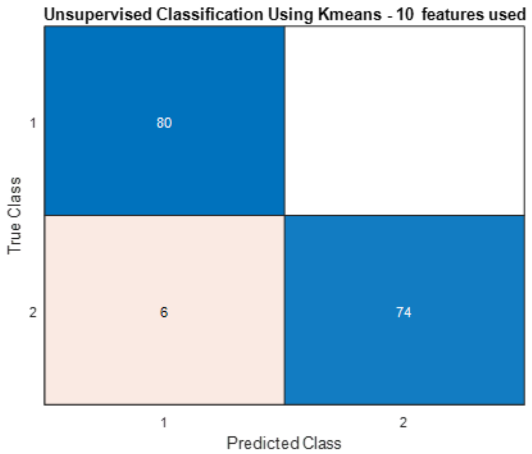

Fig. 15. Confusion matrix from unsupervised classification using k-means and 10 proposed features. Impressionism excluded from the opponent styles.

C. COMPARISON WITH SURF DESCRIPTORS

For comparison reasons, classification of the artworks is also performed by means of a SVM classifier using Speeded-up Robust Features (SURF) descriptors [21]. A Bag of Features (BOF) is constructed from the SURF descriptors using k-means for a vocabulary of 500 visual words.

The accuracy achieved by the SVM with the SURF descriptors as a result of the mean of 10 iterations is 94% as is depicted in Table I. The table illustrates also the accuracy achieved when the 10 proposed features are used with a) the SVM classifier b) the KNN classifier c) k means classifier.

TABLE I. ACCURACY COMPARISON.

SVM 10 proposed features	KNN 10 proposed features	k-means 10 proposed features	SVM SURF features
94%	93.5%	87%	94%

From the table is seen that the SVM- SURF scheme gives results close to the results achieved by the proposed simple scheme, of 10 features and KNN classifier. Even in the unsupervised case the proposed features are strong enough to give a high and comparable accuracy.

VI. DISCUSSION AND FUTURE WORK

Pointillism is characterized by a strong technical manner - the combination of many small dots of pure colour - which makes it one of the most easily recognizable artistic movements by a human. This makes it interesting to research for the automatic recognition of Pointillism among other artistic movements, using machine learning techniques. Articles on the automatic recognition of artistic movements seldom include Pointillism, usually including the style of Post-impressionism, of which Pointillism is a subset [22-23]. In the present research, I focus on finding features adapted to the recognition of Pointillism. Through a variety of features, a search is made for those that lead to the recognition of pointillist artworks with great accuracy. The simulation results lead to a small number of features capable of distinguishing pointillist artworks with both supervised and unsupervised classification.
Research on the distinction of Pointillism from other artistic movements could in the future consider the inclusion of other movements in rival styles. Also, while colour characteristics do not appear to play a decisive role in distinguishing Pointillism, further research should be done on the role of colour combination, contrast, and harmony that may exist in pointillist paintings.

VII. CONCLUSIONS

Several features are proposed for the discrimination of pointillist artworks from purist, cubist, naive art and impressionist artworks. Texture features are evaluated, by the ReliefF algorithm, as the most important features.
Among them the Granulometric texture feature, the degree of color based image segmentation, the mean length of edges and their total number and Haralick features contribute the most to successful classification reflecting the characteristics of pointillist paintings, i.e., the granular nature of paintings, the numerous unblended small rounded brushstrokes. On the contrary, the feature selection procedure showed up the most colour related features to be irrelative and even misleading. The discriminant ability of the described features is evaluated in the case of supervised and unsupervised classification. A simple supervised classification scheme formed by a KNN classifier and the 10 most prevailed of the proposed features can give almost the same accuracy with an SVM classifier using SURF features and a Bag Of Word histogram constructed using k-means vocabulary of 500 words. Simulations results, show that, although simple to compute, the 10 most informative features are quite powerful predictors to discriminate pointillist artworks even in an unsupervised classification scheme.

REFERENCES

[1] L. Shamir, T. Macura, N. Orlov, D. Mark Eckley and I. G. Goldberg, “Impressionism, expressionism, surrealism: automated recognition of painters and schools of art,” *ACM Trans. Appl. Percept*. vol. 7, no. 2, pp. 1-17, February 2010.

[2] T. E. Lombardi, “The classification of style in fine-art painting”, Ph.D. dissertation,School of Computer Sci. Inf.n Sys., Pace University , NY, 2005.

[3] J. Zujovic, L. Gandy, S. Friedman, B. Pardo and T. N. Pappas, “Classifying paintings by artistic genre: an analysis of features & classifiers,” *IEEE International Workshop on Multimedia Signal Processing*, Rio De Janeiro, pp. 1-5, 2009.

[4] R. S. Arora and A. Elgammal, "Towards automated classification of fine-art painting style: a comparative study", *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Tsukuba, pp. 3541-3544, 2012.

[5] L. Fichner-Rathus, *Foundations of art and design*, 2nd ed. Clark Baxter.

[6] A. Haas, G. Matheron, and J. Serra, "Morphologie mathématique, géostatistique et stéréologie : contributions à l'élaboration de modèles", *Annele des Mines*, vol. 12, pp. 768-782, 1967.

[7] R. M. Haralick, "Statistical and structural approaches to texture", Proceedings *of the IEEE*, vol. 67, no. 5, pp. 786-804, 1979. doi: 10.1109/PROC.1979.11328.

[8] O. Icoglu, B. Gunsel and S. Sariel, "Classification and indexing of paintings based on art movements", *12th European Signal Processing Conference*, pp. 749-752, 2004.

[9] M. Robnik-Šikonja, I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, pp. 23-69, 2003.

[10] S. Theodoridis and K. Koutroumbas, *Pattern recognition*, 4th. ed., Academic Press Inc., USA, 2008.

[11] WikiArt, Visual art encyclopedia, https://www.wikiart.org.

[12] M. Khatun, A. Gray and S. Marshall, "Classification of Ordered Texture Images Using Regression Modelling and Granulometric Features," *Proceedings - 2011 Irish Machine Vision and Image Processing Conference*, pp. 64-69, 2011.

[13] J. R. Parker, *Algorithms for image processing and computer vision*, 1st. ed., John Wiley & Sons, Inc., USA, 1996.

[14] M. Lai, P. MacAlpine, G. Printy, J. Navarro and S. Har-Noy, Le Art Spy, http://www.owlnet.rice.edu/~elec301/Projects02/artSpy/

[15] G. J. Awcock and R. Thomas, *Applied image processing*, McGraw-Hill, Inc., USA, 1995.

[16] J. Morra, "Purism and the object-type: tradition and modernity, Art and Society", Master Thesis, CUNY Academic Works, 2017.

[17] Y. Bar, L. Noga and L. Wolf, "Classification of artistic styles using binarized features derived from a deep neural network" *European Conference on Computer Vision*, pp. 71-84, Cham, 2014.

[18] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification", *IEEE Transactions on Systems*, vol. 3, no. 6, pp. 610-621, 1973.

[19] G. Lu, Z. Gao, D. Qin, X. Zhao and M. Liu, "Content-Based Identifying and Classifying Traditional Chinese Painting Images" *2008 Congress on Image and Signal Processing*, pp. 570-574, 2008. doi: 10.1109/CISP.2008.477.

[20] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281-297, 1967.

[21] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "Speeded-up robust features (SURF)", *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346-359, 2008.

[22] M. Spehr, C. Wallraven, R. Fleming, O. Deussen, D. Fellner and N. Dodgson, "Image statistics for clustering paintings according to their visual appearance". *Computational Aesthetics in Graphics, Visualization, and Imaging*, pp. 57-64, 2009 .

[23] S.G. Lee and E.Y. Cha, "Style classification and visualization of art painting's genre using self-organizing maps". *Hum. Cent. Comput. Inf. Sci.* vol. 6, no. 7, 2016. doi:10.1186/s13673-016-0063-4.