Geometric tight frame based stylometry for art authentication of van Gogh paintings

Haixia Liu [a], Raymond H. Chan [a,*], Yuan Yao [b]

[a] *Department of Mathematics, The Chinese University of Hong Kong, Hong Kong, China*
[b] *School of Mathematical Sciences, LMAM-LMEQF-LMPR, Peking University, Beijing, 100871, China*

A R T I C L E   I N F O

A B S T R A C T

This paper is about authenticating genuine van Gogh paintings from forgeries. The paintings used in the test in this paper are provided by van Gogh Museum and Kröller-Müller Museum. The authentication process depends on two key steps: feature extraction and outlier detection. In this paper, a geometric tight frame and some simple statistics of the tight frame coefficients are used to extract features from the paintings. Then a forward stage-wise rank boosting is used to select a small set of features for more accurate classification so that van Gogh paintings are highly concentrated towards some center point while forgeries are spread out as outliers. Numerical results show that our method can achieve 86.08% classification accuracy under the leave-one-out cross-validation procedure. Our method also identifies five features that are much more predominant than other features. Using just these five features for classification, our method can give 88.61% classification accuracy which is the highest so far reported in literature. Evaluation of the five features is also performed on two hundred datasets generated by bootstrap sampling with replacement. The median and the mean are 88.61% and 87.77% respectively. Our results show that a small set of statistics of the tight frame coefficients along certain orientations can serve as discriminative features for van Gogh paintings. It is more important to look at the tail distributions of such directional coefficients than mean values and standard deviations. It reflects a highly consistent style in van Gogh's brushstroke movements, where many forgeries demonstrate a more diverse spread in these features.

1. Introduction

Art authentication is the identification of genuine paintings by famous artists and detection of forgery paintings by imitators. The traditional way in art authentication is to rely on the discerning eyes and experience of experts who are dedicated in the work and life of the artist(s). Physical means such as ultraviolet fluorescence [23], infrared reflectography [10], x-ray radiography [27], painting sampling [6], and canvas

---

* Corresponding author.
E-mail addresses: mahxliu@ust.hk (H. Liu), rchan@math.cuhk.edu.hk (R.H. Chan), yuany@math.pku.edu.cn (Y. Yao).

ARTICLE IN PRESS

YACHA:1101

2   *H. Liu et al. / Appl. Comput. Harmon. Anal. ••• (••••) •••–•••*

weave count [3] have also been used for art authentication. The term *stylometry* refers to the application of statistical or quantitative techniques for authorship and style evolution in literary arts [28]. In the past decade, research in stylometry for paintings has been benefited from the rapid progress in image data acquisition technology. By using high-resolution digital images of artists' collections, image analysis researchers and art historians have engaged in cross-disciplinary stylometric analysis of art paintings via computational techniques [1,2,17–20,24,29,30].

Although many art authentication methods were proposed and used, the authorship of many paintings is still questioned by experts, with different art scholars having different opinions. Stylometry for paintings, in particular, is still a long way from being a mature field, even for paintings from well-known artists. Here, we propose a new stylometric technique for art authentication of Vincent van Gogh paintings. Our results on 79 paintings provided by van Gogh Museum and Kröller-Müller Museum show that our method is better than existing van Gogh paintings authentication methods [18,24,29].

Stylometry is based on the assumption that there are some distinctions in styles among different artists. Each artist exhibited particular traces of natural style and habitual physical movements when painting. Therefore, characteristics reflecting these habits can be considered as features to identify the authorship of paintings. In the past two decades, various specialized features have been used in stylometric analysis, and many paintings are authenticated. An early study was given by Taylor et al. in 1999 on fractal analysis of Pollock's drip paintings [30]. They showed that the fractal dimensions increased steadily through Pollock's career and fractal analysis could be used as a quantitative and objective technique for analyzing his paintings. In a 2004 paper by Lyu et al. [24], the moment statistics of wavelet coefficients and the log error in a linear predictor are used as features to authenticate the drawings by Pieter Bruegel the Elder. In the same year, Li and Wang [19] put 2D multi-resolution Hidden Markov Model (HMM) in use to classify paintings from some China's famous artists in different dynasty periods. Later, Berezhnoy et al. [2] gave an orientation extraction technique based on circular filters for brushstroke extraction. Recently, the moment statistics of 2-D Empirical Mode Decomposition (EMD) coefficients were used by Hughes et al. [17] for stylometric analysis of drawings by Pieter Bruegel the Elder and Rembrandt van Rijn. For each forgery in their dataset, a binary classifier was trained based on this forgery together with all but one genuine drawings. Then the left-out genuine drawing was classified according to the trained classifier. However, in the paper, there is no authentication done on the forgery drawings.

In 2008, three research groups from Penn State, Princeton, and Tilburg focusing on authenticating van Gogh paintings reported their analysis of van Gogh's brushstrokes in [18]. In the work of the Penn state group, the similarity among paintings were assessed via texture and brushstroke geometry modeling. The Princeton group applied the complex wavelet and Hidden Markov Tree (HMT) for feature extraction, and then similarity distances between paintings were calculated using the first few features ranked according to their effectiveness in distinguishing van Gogh's and non-van Gogh's patches. Finally a multidimensional scaling embeds the paintings into a 3D space where the separation of genuine paintings from forgery ones was done. Binary support vector machine was used to determine the authorship by the Tilburg University group. It is based on the fact that the total energy, as calculated using the Gabor wavelet coefficients from the patches, was larger in the non-van Gogh's paintings. These studies are quite encouraging as initial works for identifying the authorship of van Gogh paintings.

In 2012, Li et al. [20] made an effort to extract those visually salient brushstrokes of van Gogh based on an integrative technique of both edge detection and clustering-based segmentation. With the extracted brushstrokes, some definitions of brushstroke features for art authentication were given in distinguishing van Gogh paintings from forgeries. In their numerical test, they compared the brushstrokes obtained manually with those extracted using their algorithm and showed that the combined brushstroke features were consistent throughout van Gogh's works during his French periods (1886–1890).

More recently in 2013, Qi et al. [29] use background selection and wavelet-HMT-based Fisher information distance for authorship and dating of impressionist and post-impressionist paintings. Two novel points were

ARTICLE IN PRESS   YACHA:1101

H. Liu et al. / Appl. Comput. Harmon. Anal. ••• (••••) •••–•••   3

introduced in this work. The first point is that background information is much more reliable than the details of an intricate object which cannot represent the artist's natural style because of multiple edits and corrections. Therefore they proposed to seek out sections in the paintings that have been painted quickly without too much modification. However, in their tests the labeling of painting patches, either "background" or "detail", is done manually by a non-expert. The second point is that an artist's style should be interpreted as a probability distribution over a set of possible textures, and not just simply from the textures themselves.

For art authentication, the key point is to find the appropriate features which give a good separation between the artist's paintings and those by his imitators. In this paper we aim to find an appropriate measure so that the paintings drawn by one artist, such as van Gogh, are much more similar than those by the imitators. As in [18], we first start by analyzing the brushstrokes in the paintings by some analysis operators. Instead of using the variety of techniques such as wavelets, EMD, HMM and HMT in [17,18,24,29], here we propose to use a special tight frame, called geometric tight frame [22], to extract brushstroke information from the given paintings. Tight-frame transforms are redundant bases that can provide overcomplete but stable coding of directional variations [7]. The geometric tight frame we used has 18 filters that give the first- and second-order differences in the horizontal, vertical and diagonal directions in small neighborhoods. Therefore it can capture subtle oriented variations in the texture of the paintings.

Next to find our features, we follow the moment statistics approach explored in [17,24] and propose to use 3 simple statistics of the geometric tight frame coefficients as our features. They are the mean of the coefficients, the standard deviation of the coefficients, and the percentage of those coefficients that are more than one standard deviation away from the mean. That gives a total of 54 features for each painting. Then we select the discriminatory features by a forward stage-wise boosting procedure [13,15]. It selects features by maximizing the area under the Receiver Operating Characteristic (ROC) [13] curve such that van Gogh's paintings are highly concentrated while the forgeries are widely spread. We also used the leave-one-out cross-validation procedure to avoid overfitting while maximizing the amount of training data.

Once the features are selected, we use a simple thresholding rule to authenticate the paintings. Our test on the 79 paintings shows that we can achieve 86.08% classification accuracy and it reveals that 5 of the 54 features are predominant. By using just these 5 features, we can get a 88.61% classification accuracy which is highest so far reported in the literature [18,24,29]. Evaluation of the five features is also performed on two hundred 79-painting datasets generated by bootstrap sampling with replacement [12]. The 95% confidence interval is (78.48%, 94.94%) with median 88.61% and mean 87.77%. Our results show that a small set of statistics of the tight frame coefficients along certain orientations in small neighborhood can serve as discriminative features for van Gogh paintings. This reflects a highly consistent style in van Gogh's brushstroke movements, where many forgeries demonstrate a more diverse spread in these features.

Our proposed method, though tested only on van Gogh dataset, can easily be applied to paintings by other artists. We hope that our method may help art scholars to identify more digital evidences discriminating different artists' paintings from forgeries.

This paper is organized as follows. Section 2 describes the dataset we used for our art authentication. Section 3 introduces how we construct our features. Section 4 explains how we select the most discriminatory features among the features we constructed. Section 5 describes how we used the selected features so obtained to do the authentication. Section 6 gives the numerical results. We draw a conclusion in Section 7.

2. Dataset

Our dataset consists of 79 digitalized impressionist and post impressionist paintings provided to us by the Tilburg University group [18]. They are high-resolution color copies of paintings from the van Gogh museum and Kröller-Müller museum by professional scanners and are suitable for art research. These paintings vary in sizes, with the smallest one being 1452-by-833 pixels and the largest one 5614-by-7381 pixels. Among the 79 paintings, 64 paintings were created by van Gogh himself and the remaining 15 paintings were by

ARTICLE IN PRESS   YACHA:1101

4   *H. Liu et al. / Appl. Comput. Harmon. Anal. ••• (••••) •••–•••*

Table 1
The 6 paintings which were once wrongly attributed to van Gogh in history.

ID	Title	Date and place
f233	View of Montmartre with quarry	Paris, Late 1886
f253	Still life with a bottle, Two glasses cheese and bread	Paris, Spring 1886
f253a	A plate of rolls	Paris, first half of 1887
f278	Vase with poppies, cornflowers peonies	Paris, Summer 1886
f418	Family: Onésime Comeau/Marie-Louise Duval	Jan., 1890
f687	Reaper with sickle (after Millet)	Saint-Remy Sep., 1889

his contemporaries. The paintings by van Gogh are mainly from the Paris, Arles, and Saint-Remy periods, as well as 4 paintings from Auver-Sur-Oise days, a few months before his death. In the following, we will abbreviate them as vG (van Gogh) and nvG (non-van Gogh) paintings respectively. Sample images of vG and nvG paintings are given in Figs. 1 and 2. It should be noted that the 15 forgeries are very similar to the 64 van Gogh's artworks, with 6 of them historically attributed to van Gogh, but have been known to be forgeries now. Table 1 lists these six once-debatable paintings, which are regarded as difficult examples for stylometric analysis. Note that 9 out of the 79 paintings are stored as 8-bit unsigned integers and the remaining 70 paintings are stored as 16-bit unsigned integers. In order to unify these data, we convert all 79 images to 8-bit unsigned integers. We make a note about the boundary of the paintings here. As pointed out in Qi et al. [29], the edges of the canvas in the paintings may not be useful information for art authentication, and hence we have excluded these edges in our numerical experiments. More precisely, for each painting in the dataset, we crop off 100 pixels from its four sides, and use only the interior of the image in our numerical tests.

In the following we focus on finding a small set of features to automatically classify the 79 paintings into vG and nvG. With a small set of samples, to avoid the problem of overfitting, we adopt the leave-one-out test method [14]. Then a forward stage-wise boosting procedure [13,15] is used to construct a small set of features such that in such feature space those vG's are highly concentrated in a cluster while the nvG's are mostly spread away from such a cluster. More precisely, we consider the art authentication problem as an outlier detection problem, where vG's are the normal data and nvG's are the outliers. In the next three sections we introduce the methodology to achieve this goal and here is the outline of these three sections.

Section 3 introduces our approach of extracting features from the paintings. It is based on an 18-filter geometric tight frame and 3 simple statistics so that each painting is represented by a 54-dimensional vector after feature extraction. The accuracy of our method is tested by a leave-one-out test scheme which uses one painting from the original dataset as the test painting and the remaining 78 paintings as the training data. In Section 4, we describe how the forward stage-wise boosting procedure is used to select 5 important features among the 54 features. In Section 5, we focus on the classification of the test painting. In Subsection 5.1, we describe how to construct the classification rule once the 5 features are selected and in Subsection 5.2 how to use the rule to classify the test painting. This procedure is repeated 79 times such that each painting in the dataset is tested once. The classification accuracy of our method is measured by the results of these 79 tests.

3. Feature extraction

Tight frames have been used successfully in different applications in image processing [7,8,11]. The geometric tight frame we use to analyze the brushstrokes in our paintings is proposed in [21,22] and it can capture the first- and second-order differences in the horizontal, vertical and diagonal directions in every small neighborhood of the paintings. As discovered in [17,24,29], statistical properties of quantities such as wavelet coefficients, EMD coefficients or HMT-parameters are useful in authenticating paintings by various artists. Here we combine the two ideas and propose to use some simple statistics of the geometric tight frame coefficients of each painting as our features. We will see from the numerical results in Section 6 that

ARTICLE IN PRESS   YACHA:1101

H. Liu et al. / Appl. Comput. Harmon. Anal. ••• (••••) •••–•••   5

our features can do a good job in capturing the rapid, rhythmic, and vigorous brushstroke movements of van Gogh, and hence discriminating his paintings from those forgeries of his contemporaries.

In the following, we introduce the geometric tight frame and the statistics used in this paper.

3.1. Geometric tight frame

The geometric tight frame we use has 18 filters $\tau_0, \tau_1, \ldots, \tau_{17}$:

$$
\tau_0 = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \qquad
\tau_1 = \frac{1}{16}\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, \qquad
\tau_2 = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix},
$$

$$
\tau_3 = \frac{\sqrt{2}}{16}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix}, \qquad
\tau_4 = \frac{\sqrt{2}}{16}\begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}, \qquad
\tau_5 = \frac{\sqrt{7}}{24}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix},
$$

$$
\tau_6 = \frac{1}{48}\begin{bmatrix} -1 & 2 & -1 \\ -2 & 4 & -2 \\ -1 & 2 & -1 \end{bmatrix}, \qquad
\tau_7 = \frac{1}{48}\begin{bmatrix} -1 & -2 & -1 \\ 2 & 4 & 2 \\ -1 & -2 & -1 \end{bmatrix}, \qquad
\tau_8 = \frac{1}{12}\begin{bmatrix} 0 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{bmatrix},
$$

$$
\tau_9 = \frac{1}{12}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \qquad
\tau_{10} = \frac{\sqrt{2}}{12}\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad
\tau_{11} = \frac{\sqrt{2}}{16}\begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \\ -1 & 0 & 1 \end{bmatrix},
$$

$$
\tau_{12} = \frac{\sqrt{2}}{16}\begin{bmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & -2 & 1 \end{bmatrix}, \qquad
\tau_{13} = \frac{1}{48}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}, \qquad
\tau_{14} = \frac{\sqrt{2}}{12}\begin{bmatrix} 0 & 0 & 0 \\ -1 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix},
$$

$$
\tau_{15} = \frac{\sqrt{2}}{24}\begin{bmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ -1 & 2 & -1 \end{bmatrix}, \qquad
\tau_{16} = \frac{\sqrt{2}}{12}\begin{bmatrix} 0 & -1 & 0 \\ 0 & 2 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \qquad
\tau_{17} = \frac{\sqrt{2}}{24}\begin{bmatrix} -1 & 0 & -1 \\ 2 & 0 & 2 \\ -1 & 0 & -1 \end{bmatrix}, \qquad (1)
$$

see [21,22]. We note that τ_0 is the low-pass filter. The filters τ_1, τ_2, τ_3 and τ_4 are the Sobel operators in the vertical, horizontal, $-\frac{\pi}{4}$, and $\frac{\pi}{4}$ directions, respectively whereas the filters τ_8, τ_9, τ_{14}, τ_{15}, τ_{16} and τ_{17} are the second-order difference operators in different directions.

Given the i-th color painting, $1 \leq i \leq 79$, with m_i-by-n_i pixels, we represent its grey-scale intensity by an m_i-by-n_i matrix P_i. Then we convolve P_i with each τ_j, $0 \leq j \leq 17$, to get the corresponding m_i-by-n_i tight frame coefficient matrices:

$$
A^{(i,j)} = P_i * \tau_j = \begin{pmatrix} a_{1,1}^{(i,j)} & \cdots & a_{1,n_i}^{(i,j)} \\ \vdots & & \vdots \\ a_{m_i,1}^{(i,j)} & \cdots & a_{m_i,n_i}^{(i,j)} \end{pmatrix}, \qquad 1 \leq i \leq 79, \ 0 \leq j \leq 17. \qquad (2)
$$

Therefore, there are 18 corresponding coefficient matrices for each painting after the decomposition by the geometric tight frame. We remark that we only use one level of the tight frame transform without any down-sampling. Our numerical result shows that using 2 levels of tight frame transform gives a bad classification accuracy, see Table 2. Moreover, it increases the number of features significantly.

3.2. Three statistics

Moment statistics has been used successfully to extract features in art authentication, see [17,24,25]. In [17], the moment statistics of the "outlier pixels", defined as those that are *greater than* the mean plus one standard deviation, are also considered as features. Thus here we propose to use the following three statistics as features. They are (i) the mean of the entries in the coefficient matrix, (ii) the standard deviation of the

ARTICLE IN PRESS   YACHA:1101

6   H. Liu et al. / Appl. Comput. Harmon. Anal. ••• (••••) •••–•••

entries in the coefficient matrix, and (iii) the percentage of the "tail entries" which are those entries that are *more than* one standard deviation from the mean. To be precise, given a coefficient matrix $A^{(i,j)}$ in (2) with entries $a_{l,k}^{(i,j)}$, the three statistics are defined as follows:

(i) the mean of $A^{(i,j)}$:

$$\mu^{(i,j)} = \frac{1}{m_i n_i} \sum_{l=1}^{m_i} \sum_{k=1}^{n_i} a_{l,k}^{(i,j)},$$

(ii) the standard deviation of $A^{(i,j)}$:

$$\sigma^{(i,j)} = \left(\frac{1}{m_i n_i - 1} \sum_{l=1}^{m_i} \sum_{k=1}^{n_i} \left(a_{l,k}^{(i,j)} - \mu^{(i,j)} \right)^2 \right)^{\frac{1}{2}},$$

(iii) the percentage of the tail entries $p^{(i,j)} = \#(\hat{A}^{(i,j)})/(m_i n_i)$.

Here $\#(\hat{A}^{(i,j)})$ is the number of nonzero entries in the tail matrix $\hat{A}^{(i,j)}$ which is defined by

$$\hat{a}_{l,k}^{(i,j)} = \begin{cases} a_{l,k}^{(i,j)}, & \text{if } |a_{l,k}^{(i,j)} - \mu^{(i,j)}| > \sigma^{(i,j)}, \\ 0, & \text{otherwise.} \end{cases}$$

Thus the feature vector of the i-th painting is represented by

$$\left[\mu^{(i,0)}, \ldots, \mu^{(i,17)}, \sigma^{(i,0)}, \ldots, \sigma^{(i,17)}, p^{(i,0)}, \ldots, p^{(i,17)} \right] \in \mathbb{R}^{54}. \tag{3}$$

In summary, we have 79 paintings and 54 features. The accuracy of our method is tested by a leave-one-out procedure where a painting, say P, in the dataset is used as the testing data and the remaining 78 paintings in the dataset are used as the training data to select a small feature subset $\mathcal{G} \subseteq \{1, 2, \ldots, 54\}$. Then $\mathcal{G}$ is used to train a classifier to test the left-out painting P. The next section describes how to select $\mathcal{G}$.

4. Forward stage-wise feature selection procedure

Considering the highly rhythmic brushstroke movements of van Gogh, it is unlikely that all the 54 features are discriminative between van Gogh and his contemporaries. If we include some noisy features in our classification task, the accuracy will deteriorate. Therefore, in this section we develop a feature selection method based on a forward stage-wise rank boosting [15] to boost the discriminating power of our feature sets. Such a feature selection plays an indispensable role in our method, which not only greatly improves our classification accuracy but also leads to interpretable models—where with only five features we can reach a classification accuracy of 88.61%. In this section, we describe our procedure to select a good feature subset from the given training dataset.

The way we discriminate van Gogh's paintings from the forgeries is based on the assumption that van Gogh exhibits highly consistent brushstroke movements in some of the texture features. Therefore under these features van Gogh's paintings will be highly concentrated toward some center points while forgeries are spread as outliers. To be more precise, let the training dataset be $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{78}\}$ and

$$X = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{78} \end{bmatrix} \in \mathbb{R}^{78 \times 54}$$

ARTICLE IN PRESS

YACHA:1101

H. Liu et al. / Appl. Comput. Harmon. Anal. ••• (••••) •••–•••

7

be the data matrix of $\mathcal{X}$ and $\tilde{X}$ be the normalization of X such that each column in $\tilde{X}$ has a unit standard deviation. Let $\{1, \ldots, 78\} = \mathcal{T}_{\text{vG}} \cup \mathcal{T}_{\text{nvG}}$ where $\mathcal{T}_{\text{vG}}$ (respectively $\mathcal{T}_{\text{nvG}}$) denotes the set of vG (respectively nvG) paintings in $\mathcal{X}$. For any feature subset $\mathcal{F}$, denote $|\mathcal{F}|$ the number of elements in the set $\mathcal{F}$ and $\mathcal{F} = \{i_1, \ldots, i_{|\mathcal{F}|}\}$. Define $\tilde{X}_{j\mathcal{F}} = (\tilde{X}_{ji_1}, \ldots, \tilde{X}_{ji_{|\mathcal{F}|}})$, i.e. $\tilde{X}_{j\mathcal{F}}$ is the j-th row of $\tilde{X}$ restricted onto the index set $\mathcal{F}$. Then we define the *vG center* w.r.t. $\mathcal{F}$ as the mean vector of vG on $\mathcal{F}$, i.e.

$$\mathbf{c}^{\mathcal{F}} = \frac{1}{|\mathcal{T}_{\text{vG}}|} \sum_{j \in \mathcal{T}_{\text{vG}}} \tilde{X}_{j\mathcal{F}}. \tag{4}$$

With this we define the distance between the j-th painting in $\mathcal{X}$ and the vG center $\mathbf{c}^{\mathcal{F}}$ by

$$d_j^{\mathcal{F}} = \|\tilde{X}_{j\mathcal{F}} - \mathbf{c}^{\mathcal{F}}\|_2, \quad 1 \le j \le 78. \tag{5}$$

For $\mathcal{F}$ to be a good feature set, $d_j^{\mathcal{F}}$ should be small for $j \in \mathcal{T}_{\text{vG}}$ and large for $j \in \mathcal{T}_{\text{nvG}}$, i.e. nvG should be far from the vG center and regarded as outliers.

To quantitatively measure any given $\mathcal{F}$, we use the theory of the ROC curve which has been widely used in literature [4,9,13,16,26,31]. Let us sort $\{d_j^{\mathcal{F}}\}_{j=1}^{78}$ in (5) in an ascending order such that $d_{j_1}^{\mathcal{F}} \le d_{j_2}^{\mathcal{F}} \le \ldots \le d_{j_{78}}^{\mathcal{F}}$. For any number ρ (smaller than $d_{j_1}^{\mathcal{F}}$, larger than $d_{j_{78}}^{\mathcal{F}}$ or in between $d_{j_1}^{\mathcal{F}}$ and $d_{j_{78}}^{\mathcal{F}}$), we can use it as a binary classifier to label all the paintings in $\mathcal{X}$. From that we can determine the true positive rate and the false positive rate w.r.t. ρ (see (6) for the definitions of the rates). By plotting the true positive rate versus the false positive rate for different ρ, we obtain the ROC curve w.r.t. $\mathcal{F}$. Then we can compute the area under the ROC curve AUC($\mathcal{F}$). Notice that the larger AUC($\mathcal{F}$) is, the better $\mathcal{F}$ is as more vG are close to the vG center and more nvG are far from the vG center [13]. In the maximal case that AUC($\mathcal{F}$) = 1, the nvG's distances are all greater than any vG's distances and there is a suitable ρ that can classify all paintings correctly.

Therefore, the best feature subset $\mathcal{F}$ would be the one that maximizes AUC($\mathcal{F}$). However this is intractable due to the curse of dimensionality with an exponential blow-up of computational complexity. Thus we adopt the following forward stage-wise approach (see [15]) to maximize AUC($\mathcal{F}$). We start from the empty set $\mathcal{F}^{(0)} = \emptyset$ and iterate. Suppose at the j-th iteration, we already find $\mathcal{F}^{(j)}$. Then in the $(j+1)$-th iteration, we greedily select the next feature by

$$l_{j+1} = \arg \max_{l \notin \mathcal{F}^{(j)}} \text{AUC}\left(\mathcal{F}^{(j)} \cup \{l\}\right),$$

and update $\mathcal{F}^{(j+1)} = \mathcal{F}^{(j)} \cup \{l_{j+1}\}$. In our numerical experiment, we stop at the fifth iteration. Thus the resulting feature set for $\mathcal{X}$ is $\mathcal{G} = \mathcal{F}^{(5)}$, and it has 5 features.

Note that for a dataset of n paintings with f features (ours has $f = 54$ and $n = 78$) such a forward procedure has a computational cost of $O(njf(f - j))$ at the j-th iteration.

5. Classification and validation

In the following, we give the classification rule and how to use the classification rule to determine whether the left-out painting is genuine or fake.

5.1. Classification rule

Given the selected features in $\mathcal{G}$, we already have (see (4) and (5)) the vG center $\mathbf{c}^{\mathcal{G}}$ corresponding to $\mathcal{G}$ and the distance $d_i^{\mathcal{G}}$ of the i-th painting in $\mathcal{X}$ to $\mathbf{c}^{\mathcal{G}}$, for $1 \le i \le 78$. If $\mathcal{G}$ is a good feature set, we expect $d_i^{\mathcal{G}}$

to be small for vG and large for nvG. Therefore our classifier is based on a simple threshold δ, such that paintings with $d_i^{\mathcal{G}} < \delta$ will be classified into vGs, or into nvGs if otherwise.

To determine δ, a natural choice will be to maximize the classification accuracy (see (7) for the definition). To be precise, let us sort $\{d_i^{\mathcal{G}}\}_{i=1}^{78}$ as $d_{i_1}^{\mathcal{G}} \leq d_{i_2}^{\mathcal{G}} \leq \cdots \leq d_{i_{78}}^{\mathcal{G}}$ and define $(e_1, \ldots, e_{78}) = (d_{i_1}^{\mathcal{G}}, \ldots, d_{i_{78}}^{\mathcal{G}})$. Let $(b_1, \ldots, b_{78})$ be the labels of the paintings in $\mathcal{X}$, i.e. $b_j = 1$ if the i_j-th painting is vG and -1 otherwise. For any threshold inside the interval $(d_{i_{j-1}}^{\mathcal{G}}, d_{i_j}^{\mathcal{G}})$, its accuracy is

$$\epsilon_j = \frac{\sum_{l=1}^{j-1} |\{l : b_l = 1\}| + \sum_{l=j}^{78} |\{l : b_l = -1\}|}{78}, \quad j = 1, \ldots, 79.$$

We should therefore choose j to be the one that maximizes ϵ_j. But as such j may not be unique, we choose $j^* = \max\{\arg\max_j \epsilon_j\}$, and then the classification threshold is defined to be $\delta = \frac{e_{j^*-1} + e_{j^*}}{2}$.

5.2. Classifying the left-out painting

With the classification threshold δ defined, now we are ready to classify the left-out painting P. Let $\mathbf{z} \in \mathbb{R}^5$ be the feature vector extracted from P according to the feature set $\mathcal{G}$. Then we normalize $\mathbf{z}$ to get $\tilde{\mathbf{z}}$, i.e. we divide each entry in $\mathbf{z}$ by the corresponding column standard deviation of X. Then the distance between the test painting P and the vG center $\mathbf{c}^{\mathcal{G}}$ is $d = \|\tilde{\mathbf{z}} - \mathbf{c}^{\mathcal{G}}\|_2$. We now classify P as vG if $d < \delta$, or as nvG if otherwise.

Since we have 79 paintings in the dataset, the leave-one-out cross-validation procedure described in Sections 3–5 is repeated 79 times such that each painting in the dataset is tested as a left-out painting once. In particular, our method is used 79 times to authenticate each of the 79 paintings in our dataset once. The classification accuracy of the method is defined to be the percentage of correct classifications (either genuine or forgery) in these 79 tests. We will carry out these tests in Section 6.

6. Experimental results

This section gives the experimental results. In Subsection 6.1, we give the result for our method and compare it with those of others. Subsection 6.2 identifies the most discriminatory features obtained from our methods. In Subsection 6.3, we statistically evaluate our method and also the most discriminatory features selected by our method.

6.1. Results comparison

True positive (TP) is defined as the number of correct detection in the vG test cases and true negative (TN) is the number of correct detection in the nvG test cases. Then the true positive rate (TPR), true negative rate (TNR), and the classification accuracy are defined as

$$\text{TPR} = \frac{\text{TP}}{\text{vG number}}, \quad \text{TNR} = \frac{\text{TN}}{\text{nvG number}}, \tag{6}$$

$$\text{classification accuracy} = \frac{\text{TP} + \text{TN}}{\text{total number}}. \tag{7}$$

We have performed our method on the 79 paintings in our dataset. Recall that there are 64 vG and 15 nvG paintings. In our experiment, 60 out of the 64 van Gogh paintings are detected correctly by our method as genuine (i.e. TP = 60) and 8 out of the 15 imitations are detected correctly as forgery (i.e. TN = 8). Therefore $(60 + 8)/79 = 86.08\%$ of these paintings are classified correctly by our method. Table 2

ARTICLE IN PRESS

YACHA:1101

H. Liu et al. / Appl. Comput. Harmon. Anal. • • • (• • • •) • • • – • • •

9

Table 2
Comparison of the classification results.

Dataset	Methods	TPR (TP)	TNR (TN)	Accuracy (TP + TN)
Our dataset	Our method	**93.75%** (60)	53.33% (8)	86.08% (68)
	All 54 features used	89.06% (57)	0.00% (0)	72.15% (57)
(64 vG, 15 nvG)	2-level tight frame	84.36% (54)	26.67% (4)	73.42% (58)
IP4AI1	WHMT_FI	92.59% (50)	63.64% (7)	**87.69%** (57)
	HMT with MDS	90.74% (49)	36.36% (4)	81.54% (53)
(54 vG, 11 nvG)	LP with MDS	96.30% (52)	36.36% (4)	86.15% (56)
IP4AI2	WHMT_FI	87.69% (57)	**73.33%** (11)	85.00% (68)
	HMT with MDS	86.15% (56)	26.67% (4)	75.00% (60)
(65 vG, 15 nvG)	LP with MDS	84.62% (55)	40.00% (6)	76.25% (61)

(a) f249   (b) f371   (c) f522   (d) f752
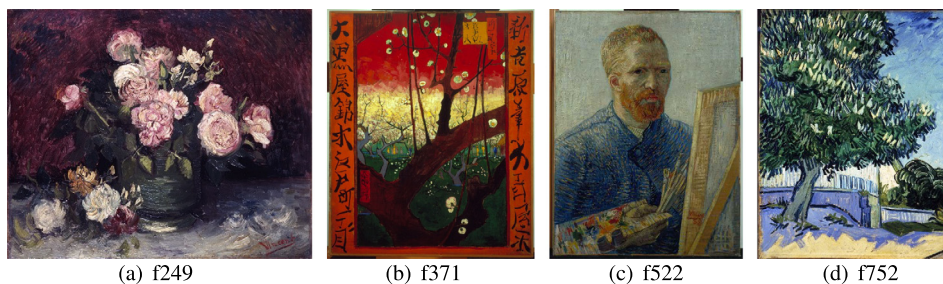
Fig. 1. The four van Gogh's paintings detected wrongly by our method.

gives the classification results by our method as compared with some previous methods. We emphasize that there are 3 datasets: our dataset (64 vG, 15 nvG), IP4AI1 (54 vG, 11 nvG) and IP4AI2 (65 vG, 15 nvG). The numbers reported on IP4AI1 and IP4AI2 are all computed by Qi et al. [29].

We see from Table 2 that our method gives the highest true positive rate and a rather good classification rate. In the second row of Table 2, we list the result if no feature selection is done and all 54 features are used in our method. We see that the result is bad and it is indeed necessary to perform feature selection to exclude those features which are irrelevant or noisy and hence may obscure accurate classification. In the third row of Table 2, we give the results when the 2-level geometric tight frame is used instead of 1-level. We see that the classification accuracy is also bad, indicating features with bigger neighborhood are not good in discriminating van Gogh's paintings. Note that, in the 2-level case, we have 105 features to begin with instead of 54.

In Figs. 1 and 2, we give the misclassified paintings together with their ID numbers. In particular, there are 3 forgeries (f253a, f418, and f687) which were once wrongly regarded as van Gogh's paintings and are indeed highly similar in such a stylometric analysis, see Table 1. They successfully cheat both experts and our method. On the other hand, van Gogh's paintings f249, f371 and f752 exhibit so many unusually diverse movements of brushstroke that they look different from van Gogh's other paintings.

Comparing the results in Table 2, the TPR of our method is quite high while its TNR is a bit low. In Section 6.2, we show that we can significantly improve the TNR while keeping the TPR by carefully selecting the features.

6.2. Feature analysis

Recall that in our method for each test painting a set of five features $\mathcal{G}$ is selected, see Section 4. In order to identify the most discriminatory features which are useful in accurate classification, we gather the features sets $\mathcal{G}$ in all 79 tests and count the frequency of each feature that occurs in these 79 $\mathcal{G}$'s. It turns out that only 11 out of the 54 features occur in these 79 feature sets and they are listed in Table 3.
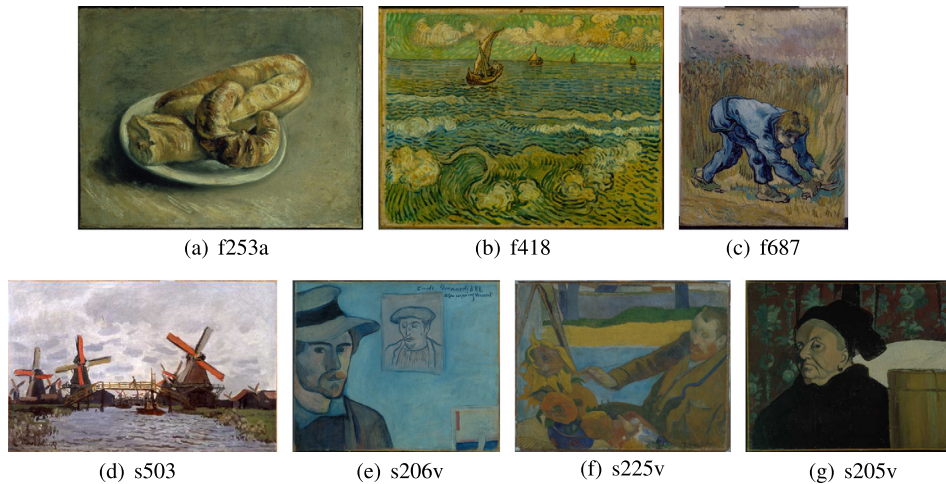
ARTICLE IN PRESS   YACHA:1101

10   *H. Liu et al. / Appl. Comput. Harmon. Anal. • • • (• • • •) • • • – • • •*

(a) f253a   (b) f418   (c) f687

(d) s503   (e) s206v   (f) s225v   (g) s205v

Fig. 2. The seven forgeries detected as van Gogh's painting by our method.

Table 3
The eleven features (with their filter and statistics) and their frequencies of occurrence.

Filter	Statistics	Frequency
τ_3	Mean	78
τ_{16}	Percentage of the tail entries	77
τ_0	Percentage of the tail entries	77
τ_1	Percentage of the tail entries	76
τ_8	Percentage of the tail entries	75
τ_3	Percentage of the tail entries	5
τ_1	Mean	3
τ_4	Mean	1
τ_9	Mean	1
τ_{17}	Mean	1
τ_2	Mean	1

From the table, we see that the first five features occurs with an average of 96.96% frequency (total 383 occurrences out of 5×79). Thus they are the most discriminatory features. Notice that the features on standard derivation are not selected at all. Thus in hindsight, we could start our method with only 2 statistics, i.e. the mean $\mu^{(i,j)}$ and the percentage $p^{(i,j)}$ (see (3)), and end up with the same result.

To test the discriminatory power of the top features in Table 3, we use the top four and five features to do the classification under the leave-one-out cross-validation. They give the accuracies of 86.08% and 88.61% (see Table 4), respectively. Indeed using the top five features, we can further identify Fig. 1 (f) and (g) to be forgery, thus improving the TNR. The accuracy of our method using only the top five features (i.e. 88.61%) is better than the best ones (87.69% for IP4AI1 and 85.0% for IP4AI2) obtained by others so far or by us (86.08%), see Table 2.

The success of this small set of five features reflects a highly consistent style in van Gogh's brushstroke movements, where many forgeries demonstrate a more diverse spread in these features. Our method also leads to an interpretable model: e.g. the first feature in Table 3 is related to filter τ_3 (see (1)) indicating that the $-\pi/4$ direction is an important direction in discriminating van Gogh's paintings.

6.3. Method evaluation

In this section, we statistically evaluate our method and the top 5 features we selected in Section 6.2. Since we only have one dataset of 79 paintings, we generate 200 similar datasets by bootstrap sampling with replacement [12, p.12]. More precisely, each of these 200 datasets are generated by randomly choosing 64

ARTICLE IN PRESS YACHA:1101

H. Liu et al. / Appl. Comput. Harmon. Anal. • • • (• • • •) • • • – • • • 11

Table 4
Classification results on different sets of classifiers.

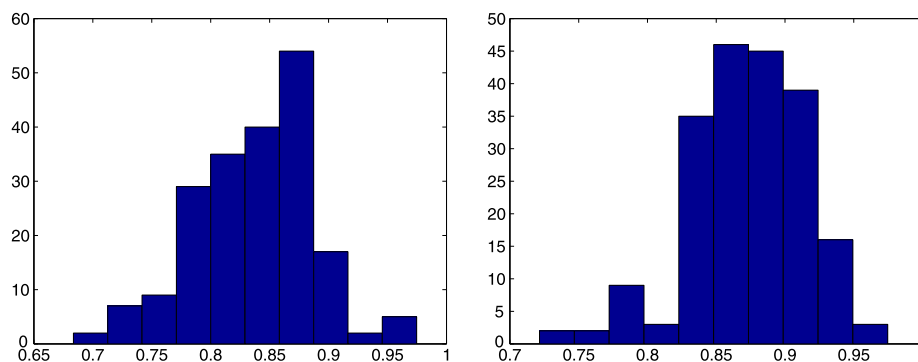|	TPR (TP)	TNR (TN)	Classification accuracy (TP + TN)
Our method	93.75% (60)	53.33% (8)	86.08% (68)
4-feature classifier	92.19% (59)	60.00% (9)	86.08% (68)
5-feature classifier	93.75% (60)	66.67% (10)	88.61% (70)

Fig. 3. Histograms of the accuracy values for (i) our methods (left) and (ii) for the top 5 features (right).

samples from the 64 vG paintings with replacement and 15 samples from 15 nvG paintings with replacement. We will compute the accuracy of our method and also the accuracy of the top 5 features on these 200 datasets. Suppose the accuracy of the i-th dataset is y_i, $i = 1, \ldots, 200$. The 95% confidence interval we give below is defined to be $(y_{i_6}, y_{i_{195}})$ with $y_{i_1} \leq \cdots \leq y_{i_{200}}$, see [5].

6.3.1. Evaluation of our method

As mentioned above, there are 200 randomly chosen datasets for evaluation of our method, with each dataset having 79 samples and 54 features. For each such dataset, the accuracy of our method is again tested by using the leave-one-out cross-validation described in Sections 3–5, where a painting in the dataset is used as a testing data and the remaining 78 samples are the training data to select the five important features. Then a classifier is trained for authentication of the left-out painting. This procedure is repeated 79 times until each painting in the 79 samples is tested once and a classification accuracy value is obtained from these 79 tests. Fig. 3 (left) is the histogram of the 200 accuracy values for these 200 datasets. The mean, median and standard deviation are 83.73%, 83.54% and 0.0507 respectively. The 95% confidence interval is (73.42%, 92.41%).

6.3.2. Evaluation of the top five features

Here we evaluate the top five features given in Table 3. Similar to Section 6.3.1, for each of the 200 datasets, the accuracy is tested under the leave-one-out cross-validation, except that everything is done w.r.t the top five features only. Fig. 3 (right) is the histogram of the 200 accuracy values. The mean, median and standard deviation are 87.77%, 88.61% and 0.0435 respectively. The 95% confidence interval is (78.48%, 94.94%). Using only the best five out of the 54 features, the classification accuracy is quite good and the results reflect the consistency of van Gogh's habitual brushstroke movements.

7. Conclusion

We have proposed a geometric tight frame based visual stylometry method to discriminate paintings by van Gogh from those by imitators. The methodology consists of some simple statistics of the geometric tight frame coefficients, as well as a boosting procedure for feature selection. Our methodology has been

ARTICLE IN PRESS   YACHA:1101

12   *H. Liu et al. / Appl. Comput. Harmon. Anal.* • • • (• • • •) • • • – • • •

tested on a data set of 79 paintings provided by the van Gogh museum and Kröller-Muller museum. The classification accuracy of our method is 86.08%. The high classification accuracy shows that our features are appropriate in identifying the authorship of van Gogh's paintings. In particular, our method identifies five robust features such that van Gogh's paintings show a higher degree of similarity in that feature space while forgeries exhibit a wider spread tendency as outliers. The accuracy using these 5 features is 88.61% which is the best one compared with the existing methods so far (see Tables 2 and 4). The success of this small set of features reflects the consistency of van Gogh's habitual brushstroke movements. From our results, we see that the "statistical outliers" of certain tight frame coefficients are not noise, but important signals to distinguish van Gogh's painting from his contemporaries. Such "outliers" and their tail distributions may due to the intrinsic creativity of the maestro expressed through his brushstroke styles. We hope these features may help art scholars to find new digital evidences in van Gogh's art authentication. In this paper, we did not discuss how the digitization or quantization may affect our results. We expect higher order filters are more sensitive to small changes in pixel values resulting from digitization or quantization. In our future study, we will consider how robust these filters are with respect to these or other possible errors. Our methodology can easily be generalized to authenticate paintings for other artists and that will also be our future research directions.

Acknowledgments

References

[1] I.E. Berezhnoy, E.O. Postma, H.J. van den Herik, Authentic: computerized brushstroke analysis, in: IEEE International Conference on Multimedia and Expo, IEEE, 2005, pp. 1586–1588.
[2] I.E. Berezhnoy, E.O. Postma, H.J. van den Herik, Automatic extraction of brushstroke orientation from paintings, Mach. Vis. Appl. 20 (2009) 1–9.
[3] G.A. Berger, Weave interference in vacuum lining of pictures, Stud. Conserv. 11 (1966) 170–180.
[4] W.M. Briggs, R. Zaretzki, The skill plot: a graphical technique for evaluating continuous diagnostic tests, Biometrics 64 (2008) 250–256.
[5] S.T. Buckland, Monte Carlo confidence intervals, Biometrics 40 (1984) 811–817.
[6] B. Caddy, Forensic Examination of Glass and Paint: Analysis and Interpretation, CRC Press, 2002.
[7] J.-F. Cai, B. Dong, S. Osher, Z. Shen, Image restoration: total variation, wavelet frames, and beyond, J. Amer. Math. Soc. 25 (2012) 1033–1089.
[8] R.H. Chan, S.D. Riemenschneider, L. Shen, Z. Shen, Tight frame: an efficient way for high-resolution image reconstruction, Appl. Comput. Harmon. Anal. 17 (2004) 91–115.
[9] J. Davis, M. Goadrich, The relationship between precision–recall and ROC curves, in: Proceedings of the 23rd International Conference on Machine Learning, ACM, 2006, pp. 233–240.
[10] J.V.A. De Boer, A note on the use of an improved infrared Vidicon for reflectography of paintings, Stud. Conserv. 19 (1974) 97–99.
[11] B. Dong, Z. Shen, MRA Based Wavelet Frames and Applications, IAS Lecture Notes Series, Park City Mathematics Institute, 2010, Summer Program on "The Mathematics of Image Processing".
[12] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, vol. 57, 1994.
[13] T. Fawcett, ROC graphs: notes and practical considerations for researchers, Mach. Learn. 31 (2004) 1–38.
[14] S. Geisser, Predictive Inference, vol. 55, CRC Press, 1993.
[15] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, vol. 2, Springer, 2009.
[16] Y. Huang, M. Sullivan Pepe, Z. Feng, Evaluating the predictiveness of a continuous marker, Biometrics 63 (2007) 1181–1188.
[17] J.M. Hughes, D. Mao, D.N. Rockmore, Y. Wang, Q. Wu, Empirical mode decomposition analysis for visual stylometry, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 2147–2157.

ARTICLE IN PRESS

YACHA:1101

H. Liu et al. / Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • • 13

[18] C.R. Johnson Jr., E. Hendriks, I.J. Berezhnoy, E. Brevdo, S.M. Hughes, I. Daubechies, J. Li, E. Postma, J.Z. Wang, Image processing for artist identification, IEEE Signal Process. Mag. 25 (2008) 37–48.

[19] J. Li, J.Z. Wang, Studying digital imagery of ancient paintings by mixtures of stochastic models, IEEE Trans. Image Process. 13 (2004) 340–353.

[20] J. Li, L. Yao, E. Hendriks, J.Z. Wang, Rhythmic brushstrokes distinguish van Gogh from his contemporaries: findings via automated brushstroke extraction, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 1159–1176.

[21] Y.-R. Li, D.-Q. Dai, L. Shen, Multiframe super-resolution reconstruction using sparse directional regularization, IEEE Trans. Circuits Syst. Video Technol. 20 (2010) 945–956.

[22] Y.-R. Li, L. Shen, D.-Q. Dai, B.W. Suter, Framelet algorithms for de-blurring images corrupted by impulse plus Gaussian noise, IEEE Trans. Image Process. 20 (2011) 1822–1837.

[23] L. H. Liang, Apparatus and methods for authentication using partially fluorescent graphic images and OCR characters, US patent 6,373,965, 16 Apr. 2002.

[24] S. Lyu, D. Rockmore, H. Farid, A digital technique for art authentication, Proc. Natl. Acad. Sci. USA 101 (2004) 17006–17010.

[25] D. Mao, Y. Wang, Q. Wu, A new approach for analyzing physiological time series, preprint, 2010.

[26] C.S. Moskowitz, M.S. Pepe, Quantifying and comparing the predictive accuracy of continuous prognostic factors for binary outcomes, Biostatistics 5 (2004) 113–127.

[27] R. Newman, Applications of x rays in art authentication: radiography, x-ray diffraction, and x-ray fluorescence, in: Photonics West'98 Electronic Imaging, International Society for Optics and Photonics, 1998, pp. 31–41.

[28] A. Pawłowski, A. Pacewicz, Wincenty Lutosławski (1863–1954): philosophe, helléniste ou fondateur sous-estimé de la stylométrie?, Historiogr. Linguist. 31 (2004) 423–447.

[29] H. Qi, A. Taeb, S.M. Hughes, Visual stylometry using background selection and wavelet-HMT-based Fisher information distances for attribution and dating of impressionist paintings, Signal Process. 93 (2013) 541–553.

[30] R.P. Taylor, A.P. Micolich, D. Jonas, Fractal analysis of Pollock's drip paintings, Nature 399 (1999) 422.

[31] M.H. Zweig, G. Campbell, Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, Clin. Chem. 39 (1993) 561–577.