Methods and Applications of Analysis.
Vol. 26, No. 3, pp. 269–280, September 2019

004

CURVATURE BASED AUTHENTICATION OF VAN GOGH PAINTINGS*

HAIXIA LIU†‡ AND XUE-CHENG TAI§

Abstract. Art authentication is the identification of genuine paintings by famous artists from the forgeries. In this paper, we introduce a novel curvature-based method to authenticate van Gogh paintings. We use curvature images to capture the shape information in the paintings. For each painting, we convert it from RGB to HSI color space. The features we propose are two simple statistics of the three parts: including (i) the H, S, I color information, (ii) their corresponding first order derivatives in x, y directions, and (iii) the corresponding 2D curvature images. In order to select the appropriate features for art authentication, we use a forward stage-wise feature selection method such that van Gogh paintings are highly concentrated and forgeries are spread around as outliers. Numerical results show that our method gives the 88.61% classification accuracy, which outperforms the state-of-the-art methods for art authentication so far.

Key words. Art authentication, curvature, HSI color space, van Gogh, moment statistics.

Mathematics Subject Classification. 68U10.

1. Introduction.

Art forgery was born after people started paying money for their favourite paintings. Art authentication is the identification of the fake paintings from the genuine ones by famous artists. The traditional methods of art authentication include: (i) the discerning eyes and experience of a few experts who are dedicated in the work and life of the artist(s), (ii) comparisons of a variety of technical data, for example, ultraviolet fluorescence, infrared reflectography, x-radiography, painting sampling, and/or canvas weave count. Although many methods are used, the authorship of many paintings is still questioned by experts, with different art scholars having different opinions.

As image-acquisition technology has advanced, the museums provide the high-resolution digital images of some artists' collections, which makes it possible to analyze the stylometry of paintings via computerized methods [23, 19, 2, 13, 17, 12, 8, 9, 20, 15]. For art authentication, the key point is to extract the features which can do a good separation between the artist's paintings and those by his contemporaries.

In the past two decades, various specialized features have been extracted to authenticate paintings, including Pollock's drip paintings, the drawings by Pieter Bruegel the Elder, paintings from some of China's famous artists in different dynasty periods and van Gogh's paintings and so forth. As early as 1999, Taylor *et al.* [23] proposed fractal analysis of Pollock's drip paintings. 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 2009, Berezhnoy *et al.* [2] introduced brushstroke extraction, which is an orientation-extraction technique based on circular filters. In 2012, Li *et al.* [13] designed a method to extract those visually salient brushstrokes of van Gogh based on an integrative technique of

*Received December 21, 2017; accepted for publication July 30, 2019.
†School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China.
‡Hubei Key Laboratory of Engineering Modeling and Scientific Computing, Huazhong University of Science and Technology, Wuhan 430074, China (liuhaixia@hust.edu.cn).
§Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong (xuechengtai@hkbu.edu.hk).

both edge detection and clustering-based segmentation. 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). Lyu *et al.* [17] in 2004 used the moment statistics of wavelet coefficients and the log error in a linear predictor as features to authenticate the drawings by Pieter Bruegel the Elder. At the same year, Li and Wang [12] based on a 2D multi-resolution Hidden Markov Model (HMM) to classify paintings from some of China's famous artists in different dynasty periods. In 2011, Hughes *et al.* [8] used the moment statistics of 2-D Empirical Mode Decomposition (EMD) coefficients to authenticate the drawings by Pieter Bruegel the Elder and Rembrandt van Rijn.

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 [9]. In the work of the Penn State group, the similarities 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 embedded the paintings into a 3D space where the separation of genuine paintings from those by his contemporaries 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. As an extension of Princeton group's work for authentication of van Gogh's paintings, Qi. *et al.* [20] used background selection and wavelet-HMT-based Fisher information distance for authorship and dating of impressionist and post-impressionist paintings.

More recently, Liu *et al.* [15] introduced geometric tight frame based method for authentication of van Gogh paintings. Instead of using the variety of techniques such as wavelets, EMD, HMM and HMT in [17, 8, 9, 20], they proposed to use a special tight frame, called geometric tight frame [14], to extract brushstroke information from the given paintings to capture subtle oriented variations in the texture of the paintings. They used 3 simple statistics of the 18 geometric tight frame coefficients as the features for each painting. Then they selected the discriminatory features with the assumption of van Gogh's paintings are highly concentrated while the forgeries are widely spread. They maximize the area under the Receiver Operating Characteristic (ROC) [5] curve, which is based on the distance between each painting and the mean of the van Gogh paintings.

In this paper we introduce a curvature based method for authentication of van Gogh paintings. Curvature is a useful tool to capture the shape information in the painting. In order to extract brushstrokes in paintings, we calculate the curvature images. Instead of the grayscale images used in [9, 20, 15], we convert from RGB to HSI color space. The features we extract are two simple moment statistics of following three parts: (i) 3 components of HSI, (ii) the corresponding first order derivatives in x and y directions and (iii) the corresponding curvature image. Here we use mean and percentage of the tail entries. In order to exclude those features which may deteriorate the accuracy, we use a forward stage-wise rank boosting to select features.

We use the same technique as that in [15] for authentication of painting. We select the discriminatory features by maximizing the area under the ROC curve. Once the features are selected, we use a simple thresholding rule to authenticate the paintings. Our test is on the dataset with 79 paintings. Numerical results show that we can achieve 88.61% classification accuracy, which is the best result reported for van Gogh paintings authentication so far.

This paper is organized as follows. Section 2 introduces the van Gogh dataset we used for art authentication. Section 3 introduces the features we extract, which are based on two simple statistics of the HSI information and their corresponding first order derivatives and curvature images. Sections 4 shows how to select the most discriminatory features among the features we constructed and how to use them for authentication. Numerical results are shown in Section 5. We draw a conclusion in Section 6.

2. Dataset. In the dataset, there are 79 digitalized impressionist and post impressionist paintings with different sizes. The smallest one is 1452-by-833 pixels and the largest one is 5614-by-7381 pixels. Among the 79 paintings, 64 paintings belong to van Gogh himself and the remaining 15 paintings are by his contemporaries. The paintings by van Gogh are mainly from different periods of van Gogh, including 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. The 15 paintings by his contemporaries 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 fakes now. Table 1 lists these six once-debatable paintings, which are regarded as difficult examples for stylometric analysis.

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

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

All images are scaled to $[0, 1]$ with double precision. As noted in [20, 15], 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 adaptively. We use the Prewitt method [21] to detect edges by the Prewitt approximation to the derivative. Figure 1 visualize the van Gogh painting $f607$ in Subfigure 1(a) and the fake $f418$ in Subfigure 1(b). The displayed paintings was downsampled due to the privacy.

In the following we focus on art authentication method to automatically classify the 79 paintings into vG and nvG.

3. Feature extraction. As the first step of art authentication, it is of great importance to extract efficient features. In image processing, the HSI color space is a very important, attractive color model. it represents colors similarly how the human

(a) $f607$

(b) $f418$

FIG. 1. *The van Gogh painting $f607$ and the fake $f418$.*

eye senses colors [11]. Instead of the grayscale used in [9, 20, 15], we convert them to HSI color space for each painting. At the same time, image gradients and curvatures are very useful in image processing and machine learning, such as [18, 1, 25]. Image gradients (the first order derivative of image in x and y directions) are used to detect edges and curvature images are to describe a shape. In this paper, we convert each painting from RGB color space to HSI space. For each component of H, S and I, we compute the corresponding gradient images in x and y and curvature images [22]. As discovered before, moment statistics [15, 8, 17] are very successful for art authentication. Here we calculate their H, S, I information and the corresponding first order derivatives and curvature images together and use some simple statistics of them as our features. In the following, we will review how to calculate the 2D curvature image in Subsection 3.1 and introduce the features we extracted in Subsection 3.2.

3.1. Curvature image. Let us consider a two-dimensional image which is given by a continuous bounded mapping $f : \mathbb{R}^2 \to \mathbb{R}$. That is, $f(x, y)$ is the value at position (x, y). Suppose $f_x(x, y)$ and $f_y(x, y)$ are the image gradients in the x and y directions respectively, then the contour normal to the image gradients is

$$\mathbf{N}(x,y) = \frac{1}{\sqrt{f_x^2 + f_y^2}} \begin{pmatrix} f_x \\ f_y \end{pmatrix}.$$

The image curvature [22] is defined as the negative divergence of normal concour. That is,

$$\begin{aligned} -\kappa(x,y) &= \operatorname{div} \cdot \mathbf{N} \\ &= \frac{\partial}{\partial x} \frac{f_x}{\sqrt{f_x^2 + f_y^2}} + \frac{\partial}{\partial y} \frac{f_y}{\sqrt{f_x^2 + f_y^2}} \\ &= \frac{f_{xx} f_y^2 + f_{yy} f_x^2 - 2 f_{xy} f_x f_y}{\left(f_x^2 + f_y^2\right)^{3/2}}. \end{aligned}$$

Therefore,

$$\kappa(x,y) = -\frac{f_{xx} f_y^2 + f_{yy} f_x^2 - 2 f_{xy}^2 f_x f_y}{\left(f_x^2 + f_y^2\right)^{3/2}}.$$

In general, curvature is sensitive to noises [7]. There is a necessity to smooth the image first. One of the most widely used methods [24] for smoothing f is to regard it as the initial state of a homogeneous linear diffusion process:

$$\begin{array}{c} \partial_t u = \triangle u \\ u(x,y,0) = f(x,y) \end{array} . \tag{1}$$

The solution of (1) is

$$u(x,y,t) = \begin{cases} f(x,y), & t = 0, \\ (\tau_{\sqrt{2t}} * f)(x,y), & t > 0, \end{cases} \tag{2}$$

where $\tau_\sigma = \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$. Notice that the equation in (2) when $t > 0$, the solution of (1) is a convolution with Gaussian filtering.

3.2. Features. The HSI color space is a very important and attractive color model for image processing application, because it represents colors similarly to how the human eye senses them [11]. In HSI color space, we can consider each component (hue, saturation, and lightness (or luminosity)) as a two-dimensional matrix (image). For the given i-th ($1 \le i \le 79$) color painting with size $r_i \times s_i \times 3$, let $M^{(i,p,H)}$, $M^{(i,p,S)}$, $M^{(i,p,I)}$ be the hue, saturation and lightness of the i-th painting, respectively and the corresponding first order derivatives in x and y be $M^{(i,x,H)}$, $M^{(i,x,S)}$, $M^{(i,x,I)}$ and $M^{(i,y,H)}$, $M^{(i,y,S)}$, $M^{(i,y,I)}$. As stated before, it is necessary to smooth the image before computing the curvature images because of the sensitivity to noise [7]. We choose a linear diffusion process, which corresponds to a Gaussian filtering [24]. Let $\tilde{M}^{(i,p,H)}$, $\tilde{M}^{(i,p,S)}$, $\tilde{M}^{(i,p,I)}$ be the corresponding matrices of the three components after smoothing. That is, the smoothed image of $M^{(i,p,T)}$ with $T = H,\ S$ or I is

$$\tilde{M}^{(i,p,T)} = \tau * M^{(i,p,T)},$$

where τ is a Gaussian filter. Then we calculate the curvature images of $\tilde{M}^{(i,p,H)}$, $\tilde{M}^{(i,p,S)}$, $\tilde{M}^{(i,p,I)}$. The corresponding curvature image is

$$M^{(i,c,T)} = \mathrm{div} \cdot \left(\mathbf{N}(\tilde{M}^{(i,x,T)}, \tilde{M}^{(i,y,T)})\right),$$

where $\tilde{M}^{(i,x,T)}$ and $\tilde{M}^{(i,y,T)}$ are the first derivatives in x and y directions, respectively. Therefore, there are 12 corresponding matrices $M^{(i,p,H)}$, $M^{(i,p,S)}$, $M^{(i,p,I)}$, $M^{(i,x,H)}$, $M^{(i,x,S)}$, $M^{(i,x,I)}$, $M^{(i,y,H)}$, $M^{(i,y,S)}$, $M^{(i,y,I)}$, $M^{(i,c,H)}$, $M^{(i,c,S)}$, $M^{(i,c,I)}$ corresponding to the i-th painting, $1 \le i \le 79$.

Moment statistics are used successfully to extract features in art authentication [15, 8, 17]. Following [15], we use the statistics of the above 12 matrices for each painting. Instead of all the 3 simple statistics used in [15], we focus on two of them: (i) the mean of the entries in the coefficient matrix, (ii) the percentage of the "tail entries" which are those entries that are more than one standard deviation from the mean. More precisely, suppose the matrix

$$M^{(i,t,T)} = \begin{pmatrix} m_{11}^{(i,t,T)} & m_{12}^{(i,t,T)} & \cdots & m_{1s_i}^{(i,t,T)} \\ m_{21}^{(i,t,T)} & m_{22}^{(i,t,T)} & \cdots & m_{2s_i}^{(i,t,T)} \\ \vdots & \vdots & \ddots & \vdots \\ m_{r_i1}^{(i,t,T)} & m_{r_i2}^{(i,t,T)} & \cdots & m_{r_is_i}^{(i,t,T)} \end{pmatrix}$$

with $t = p, x, y$ or c and $T = H,\ S$ or I.
1. The mean of the entries in the coefficient matrix $M^{(i,t,T)}$ is

$$\mu^{(i,t,T)} = \frac{1}{r_i s_i} \sum_{l=1}^{r_i} \sum_{k=1}^{s_i} m_{lk}^{(i,t,T)}.$$

2. The percentage of the tail entries is

$$\eta^{(i,t,T)} = \#\left(\hat{M}^{(i,t,T)}\right)/(r_i s_i),$$

where $\#\left(\hat{M}^{(i,t,T)}\right)$ is the number of nonzero entries in the tail matrix $\hat{M}^{(i,t,T)}$ with the (l,k)-th entry

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

and $\sigma^{(i,t,T)}$ is the standard deviation of $M^{(i,t,T)}$

$$\sigma^{(i,t,T)} = \sqrt{\frac{1}{r_i s_i - 1} \sum_{l=1}^{r_i} \sum_{k=1}^{s_i} \left(m_{lk}^{(i,t,T)}\right)^2}.$$

Thus, the feature we extracted for the i-th painting is a 24-dimensional vector.

4. Authentication. In this Section, we will give the art authentication method based on the features extracted above. As stated in [15], it is unlikely that all of the features are discriminative in art authentication. It is necessary to exclude those noisy features which may deteriorate the accuracy. We use the same technique as that in [15], which is a supervised learning technique. For the sake of completeness, we will give the feature selection method based on a forward stage-wise rank boosting to select features in Subsection 4.1, which is followed by the classification rule and how to use this rule to give one authentication result for testing data in Subsection 4.2. For more details about this method, we refer the reader to [15].

4.1. Forward stage-wise feature selection procedure. In this section, the forward stage-wise feature selection procedure [6, 10, 16] is described to select a good feature subset by the given training dataset.

In general, each artist has his (her) own style when painting. The brushstroke movement for one artist will be highly consistent. The good feature are those that can separate van Gogh's paintings from those by his contemporaries. Therefore we will select those features which make van Gogh's paintings be highly concentrated toward some center points while forgeries are spread as outliers. To be unification of the training data, we do normalization to get X such that each column of X has a unit standard deviation. Let $\mathcal{T}_{\mathrm{vG}}$ (respectively $\mathcal{T}_{\mathrm{nvG}}$) be the set of vG (respectively nvG) paintings in training data. For any feature subset $\mathcal{F} = \{i_1, \cdots, i_{|\mathcal{F}|}\}$with $|\mathcal{F}|$ the number of elements in the set $\mathcal{F}$. Define $X_{j\mathcal{F}} = (X_{ji_1}, \cdots, X_{ji_{|\mathcal{F}|}})$, i.e. $X_{j\mathcal{F}}$ is the j-th row of X restricted onto the index set $\mathcal{F}$. Then the *vG center* w.r.t. $\mathcal{F}$ is the mean vector of vG on $\mathcal{F}$, i.e.

$$\mathbf{c}^{\mathcal{F}} = \frac{1}{|\mathcal{T}_{\mathrm{vG}}|} \sum_{j \in \mathcal{T}_{\mathrm{vG}}} X_{j\mathcal{F}} \tag{3}$$

and the distance between the j-th painting and the vG center $\mathbf{c}^{\mathcal{F}}$ is

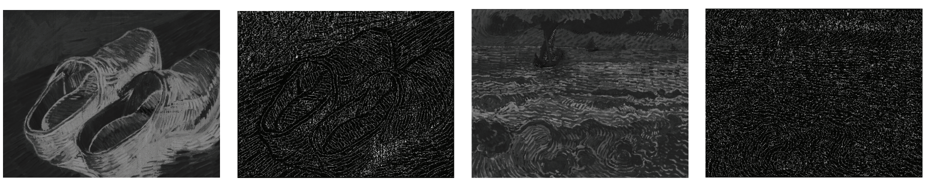
$$d_j^{\mathcal{F}} = \|X_{j\mathcal{F}} - \mathbf{c}^{\mathcal{F}}\|_2. \tag{4}$$

For $\mathcal{F}$ to be a good feature set, $d_j^{\mathcal{F}}$ should be small for $j \in \mathcal{T}_{\mathrm{vG}}$ and large for $j \in \mathcal{T}_{\mathrm{nvG}}$, i.e. nvG should be far from the vG center and regarded as outliers. To quantitatively measure any given $\mathcal{F}$, we base on the theory of the ROC curve [26, 5, 4, 3]. We use a binary classifier to label all the paintings by a simple threshold. That is, for any number ρ, the painting will be labeled as van Gogh if the distance in (4) is smaller than ρ, otherwise nvG. Thus, we can calculate the true positive rate and false negative rate with respect to ρ. We obtain the ROC curve by plotting true positive rate versus false negative rate for different ρ. The larger of the area under the ROC curve (AUC), the better of the feature set $\mathcal{F}$. The ideal case is that AUC equals one. That is, all the nVG painting are much far away from the vG center. Therefore, the best feature set $\mathcal{F}$ should be the one maximizing the AUC on $\mathcal{F}$. Unfortunately, it is with an exponential computational complexity due to the course of dimensionality.

We use forward stage-wise approach (see [6]) to maximize AUC($\mathcal{F}$), where AUC($\mathcal{F}$) is the area under the ROC curve respect to the feature set $\mathcal{F}$. We start from the empty set $\mathcal{F}^{(0)} = \emptyset$ and then iterate. Suppose at the j-th iteration, we already find $\mathcal{F}^{(j)}$. In the $(j+1)$-th iteration, we greedily select the next feature by

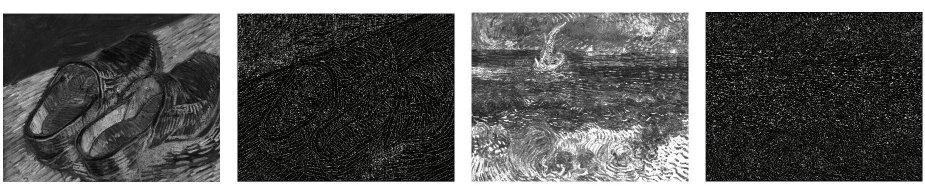
$$l_{j+1} = \arg\max_{l \notin \mathcal{F}^{(j)}} \mathrm{AUC}\left(\mathcal{F}^{(j)} \cup \{l\}\right),$$

and update $\mathcal{F}^{(j+1)} = \mathcal{F}^{(j)} \cup \{l_{j+1}\}$. We stop after k iterations.



(a) Hue for $f607$ (b) Curvature of Hue for $f607$ (c) Hue for $f418$ (d) Curvature of Hue for $f418$

FIG. 2. *The visualization of the hue.*



(a) Saturation (b) Curvature of Saturation (c) Saturation (d) Curvature of Saturation

FIG. 3. *The visualization of Saturation.*

4.2. Classification and Validation. In the following, we give the classification rule and how to use it to determine whether the left-out painting (testing painting) is genuine or fake.

Suppose $\mathcal{G}$ is the selected features, we already have (see (3) and (4)) the vG center $\mathbf{c}^{\mathcal{G}}$ corresponding to $\mathcal{G}$ and the distance $d_i^{\mathcal{G}}$ of the i-th painting in the training data to $\mathbf{c}^{\mathcal{G}}$. 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 by maximazing the classification accuracy, such that paintings with $d_i^{\mathcal{G}} < \delta$ will be classified into vGs, or into nvGs otherwise.

With the classification threshold δ defined, now we are ready to classify the left-out painting P. Let $\mathbf{z} \in \mathbb{R}^k$ be the feature vector extracted from P restricted 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.

5. Numerical results. In this section, we present some numerical results. In Subsection 5.1, we visualize the HSI information and the corresponding curvature images. In Subsection 5.2, we list the classification results with different numbers of features selected and make a comparison with [15], which gave the best result for van Gogh dataset classification reported so far.

5.1. Visualization. Recall that the features we use are some statistics of the H, S, I of the paintings themselves in HSI color space, the corresponding first order derivatives in x, y directions and curvature images. In the following, we will visualize HSI information of the genuine van Gogh painting $f607$ and the fake painting $f418$ in our dataset, as well as their corresponding curvature images. Figure 2 is the visualization of the hue corresponding to $f607$ and $f418$, where Subfigures 2(a) and 2(b) are the hue for the genuine painting $f607$ and its curvature image and Subfigures 2(c) and 2(d) are the hue for the fake painting $f418$ and its curvature image. Similarly, Figures 3 and 4 are the visualizations of the saturation and lightness corresponding to $f607$ and $f418$.

(a) Lightness (b) Curvature of Lightness (c) Lightness (d) Curvature of Lightness

FIG. 4. *The visualization of lightness.*

5.2. Comparison. In this subsection, we try to give the yes-or-no answers to evaluate the quality of separations between vGs and nvGs. To evaluate the quality of the separation, we use supervised learning method based on the assumption that van Gogh's paintings will be highly concentrated towards some center point while forgeries are spread as outliers. We test our method on 79 paintings (64 vG and 15 nvG) in our dataset by leave-one-out cross-validation procedure. That is, we use one painting as the testing data and the remaining 78 paintings as the training data. We repeat the procedure 79 times such that each painting is used once as the testing data. Table 2 gives the true positive (TP), true positive rate (TPR), true negative (TN), true negative rate (TNR) and the accuracy with the number of selected features set to $k = 1, \cdots, 8$. In the cases when $k = 5$ or $k = 7$, 62 out of 64 van Gogh paintings are identified as genuine van Gogh painting and 8 out of 15 copies are identified as forgeries. Here the classification accuracy is (62+8)/79=88.61%, which outperforms the results reported in [15]. Table 3 gives the comparison results, where GTF the geometric tight frame based method introduced by [15].

k	TPR (TP)	TNR (TN)	accuracy (TP+TN)
1	90.62% (58)	46.67% (7)	82.28% (65)
2	93.75% (60)	40.00% (6)	83.54% (66)
3	92.19% (59)	40.00% (6)	82.28% (65)
4	92.19% (59)	40.00% (6)	82.28% (65)
5	**96.88% (62)**	53.33% (8)	**88.61% (70)**
6	95.31% (61)	46.67% (7)	86.08% (68)
7	**96.88% (62)**	53.33% (8)	**88.61% (70)**
8	92.19% (59)	**60.00% (9)**	86.08% (68)

TABLE 2
The classification results with k features selected.

methods	TPR (TP)	TNR (TN)	accuracy (TP+TN)
Our method	**96.88% (62)**	**53.33% (8)**	**88.61% (70)**
GTF	93.75% (60)	**53.33% (8)**	86.08% (68)

TABLE 3
Comparison of the classification results.

Besides the two statistics used in this paper (see Subsection 3.2, we use 24-dimentional vector to represent a painting), the standard deviation is also used in [15]. Table 4 is the result provided by the features with the standard deviation added. That is, each painting is represented as a 36-dimensional vector. It can be seen that performance is not good as that in Table 2 for art authentication with the standard deviation features added. Without those features from the standard deviation, our method can give a better result. Thus, exclusion of standard deviation gives better results. Moreover in [15], the authors pointed out the features on standard deviation are not selected at all when selecting the effective features for art authentication.

6. Conclusion. In this paper, we have proposed a curvature based method for art authentication of van Gogh paintings. In general, curvature is a useful tool to describe the shape information. There are many applications in image processing and machine learning for curvature images. We use it here to capture the brushstroke in the paintings. Instead of grayscale images, we convert from RGB to HSI color space for each painting. In order to extract features, we use two simple statistics of the three parts: (i) H,S, I information, (ii) the corresponding first order derivatives in x and y directions and (iii) the corresponding curvature images. As stated before, it is expected that not all of the extracted features are discriminative for art authentication of paintings (such as van Gogh paintings). In order to exclude those features which may deteriorate the accuracy, we select some features which are highly concentrated

k	TPR (TP)	TNR (TN)	accuracy (TP+TN)
1	95.31% (61)	13.33% (2)	79.75% (63)
2	93.75% (60)	40.00% (6)	83.54% (66)
3	90.62% (58)	53.33% (8)	83.54% (66)
4	93.75% (60)	40.00% (6)	83.54% (65)
5	**96.88% (62)**	46.67% (7)	**87.34% (69)**
6	92.19% (59)	46.67% (7)	83.54% (66)
7	92.19% (59)	46.67% (7)	83.54% (66)
8	90.62% (58)	**53.33% (8)**	83.54% (66)

TABLE 4
The classification results with k features selected with the standard deviation features added.

toward some center points for van Gogh paintings while forgeries are spread as outliers. Our method can give a 88.61% accuracy, which is better than the state-of-the-art art authentication method reported so far.

For the future directions, one is to use some new techniques, such as deep learning, to extract new effective features and then do classification. Another one will be some applications in image processing, giving a visualization to guide the art experts.

Acknowledgement. The authors would like to thank Profs. Haixiang Lin, Eric Postma and Raymond H. Chan for providing the 79 van Gogh paintings for use. The authors also would like to thank the Isaac Newton Institute for Mathematical Sciences for its hospitality during the programme 'variational methods and effective algorithms for imaging and vision' which was support by EPSRC Grant Number EP/K032208/1. Haixia Liu would like to thank Norwegian Research Council project through ISP-Matematikk (Project no. 239033/F20) to support her stay.

REFERENCES

[1] E. Bae, J. Shi, and X.-C. Tai, *Graph cuts for curvature based image denoising*, IEEE Transactions on Image Processing, 20 (2011), pp. 1199–1210.

[2] I. E. Berezhnoy, E. O. Postma, and H. J. van den Herik, *Automatic extraction of brush-stroke orientation from paintings*, Machine Vision and Applications, 20 (2009), pp. 1–9.

[3] R. Bettinger, *Cost-sensitive classifier selection using the ROC convex hull method*, SAS Institute, (2003).

[4] J. Davis and 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.

[5] T. Fawcett, *ROC graphs: Notes and practical considerations for researchers*, Machine Learning, 31 (2004), pp. 1–38.

[6] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1, Springer series in statistics Springer, Berlin, 2001.

[7] T. D. Gatzke and C. M. Grimm, *Estimating curvature on triangular meshes*, International journal of shape modeling, 12 (2006), pp. 1–28.

[8] J. M. Hughes, D. Mao, D. N. Rockmore, Y. Wang, and Q. Wu, *Empirical mode decomposition analysis for visual stylometry*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (2012), pp. 2147–2157.

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

[10] A. Kalousis, J. Prados, and M. Hilario, *Stability of feature selection algorithms: a study on high-dimensional spaces*, Knowledge and information systems, 12 (2007), pp. 95–116.

[11] T.-h. Kim, H. Adeli, W. I. Grosky, N. Pissinou, T. K. Shih, E. J. Rothwell, B.-H. Kang, and S.-J. Shin, *Multimedia, Computer Graphics and Broadcasting, Part I: International Conference, MulGraB 2011, Held as Part of the Future Generation Information Technology Conference, FGIT 2011, in Conjunction with GDC 2011, Jeju Island, Korea, December 8-10, 2011. Proceedings*, vol. 262, Springer, 2011.

[12] J. Li and J. Z. Wang, *Studying digital imagery of ancient paintings by mixtures of stochastic models*, IEEE Transactions on Image Processing, 13 (2004), pp. 340–353.

[13] J. Li, L. Yao, E. Hendriks, and J. Z. Wang, *Rhythmic brushstrokes distinguish van Gogh from his contemporaries: findings via automated brushstroke extraction*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (2012), pp. 1159–1176.

[14] Y.-R. Li, L. Shen, D.-Q. Dai, and B. W. Suter, *Framelet algorithms for de-blurring images corrupted by impulse plus Gaussian noise*, IEEE Transactions on Image Processing, 20 (2011), pp. 1822–1837.

[15] H. Liu, R. H. Chan, and Y. Yao, *Geometric tight frame based stylometry for art authentication of van Gogh paintings*, Applied and Computational Harmonic Analysis, 41 (2016), pp. 590–602.

[16] H. Liu and L. Yu, *Toward integrating feature selection algorithms for classification and clustering*, IEEE Transactions on Knowledge and Data Engineering, 17 (2005), pp. 491–502.

[17] S. Lyu, D. Rockmore, and H. Farid, *A digital technique for art authentication*, Proceedings of the National Academy of Sciences of the United States of America, 101 (2004), pp. 17006–17010.

[18] S. Osher and J. A. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations*, Journal of computational physics, 79 (1988), pp. 12–49.

[19] G. Polatkan, S. Jafarpour, A. Brasoveanu, S. Hughes, and I. Daubechies, *Detection of forgery in paintings using supervised learning*, in Image Processing (ICIP), 2009 16th IEEE International Conference on, IEEE, 2009, pp. 2921–2924.

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

[21] N. Senthilkumaran and R. Rajesh, *Edge detection techniques for image segmentation–a survey of soft computing approaches*, International journal of recent trends in engineering, 1 (2009), pp. 250–254.

[22] X.-C. Tai, *Fast numerical schemes related to curvature minimization: a brief and elementary review*, Courbure discrète: théorie et applications, (2013), pp. 17–30.

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

[24] J. Weickert, *A review of nonlinear diffusion filtering*, Scale-space theory in computer vision, (1997), pp. 1–28.

[25] W. Zhu, X.-C. Tai, and T. Chan, *Augmented lagrangian method for a mean curvature based image denoising model*, Inverse problems and imaging, 7 (2013), pp. 1409–1432.

[26] M. H. Zweig and G. Campbell, *Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine*, Clinical Chemistry, 39 (1993), pp. 561–577.