# Stylometry of paintings using hidden Markov modelling of contourlet transforms

C. Robert Jacobsen*[1] and Morten Nielsen†[1]

[1]Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg East, Denmark

**Abstract** Visual stylometry is the task of quantifying artistic style in the visual arts. In this paper we present a method for visual stylometry of paintings from digital reproductions.

Our method is framed around modelling contourlet transforms of the digital reproductions with hidden Markov models. Using the contourlet transform in the field of classification is a new approach motivated by the contourlets' efficiency in representing piecewise smooth contours such as brushstrokes.

To test our method we have used paintings related to the Danish painter Asger Jorn and drawings related to the Flemish artist Pieter Bruegel the Elder. The paintings related to Asger Jorn are recorded in multiple digital images *and* by two different cameras. With multiple sources we are able to get insight into the robustness of our method against different means of acquisition.

Through a cross-validation of the Jorn images by one of the cameras we are able to correctly classify 39 out of 44 images; based on this classifier we can correctly classify 28 out of 36 images in the other data set.

A cross-validation of the Bruegel images correctly classifies 11 out of 13 images.

**Keywords:** Visual stylometry; authentication; contourlet transform; hidden Markov models; classification

*Email: robert@math.aau.dk; corresponding author. CRJ was supported by CSGB.
†Email: mnielsen@math.aau.dk

# 1 Introduction

The aim of visual stylometry of paintings is to quantify the artistic style of a painter and thus get an insight into the development of his style throughout his career. In the present paper we study quantification for the purpose of authenticating paintings based on digital reproductions: Is it possible to determine the authenticity of a painting based on digital photos?

Traditionally the task of authenticating paintings has been performed by art experts and connoisseurs with a profound and detailed knowledge of the artist and his contemporaries. However, this method will always involve some subjectivity and therefore the experts do not always agree. So assisting the authentication by more unbiased automatic methods may – besides being a topic of mathematical interest – also be valuable for the art community.

All automatic authentication methods work on digital reproductions of the paintings and the fundamental idea is to extract a number of features from the digitized paintings that are sufficiently expressive to distinguish the styles of different artists. Previous attempts have been made to perform automatic authentication of paintings by training models to separate paintings that are known to be authentic (i.e., painted by the claimed painter) from known forgeries.

The authentication task has been tried most thoroughly on paintings by Vincent van Gogh [3, 4, 22, 26], Pieter Bruegel the Elder [20, 24] and Jackson Pollock [2, 21, 31, 32]. But authentication from digital reproductions have also been tested on Chinese ink paintings by multiple artists [23] and unknown painters portraying the Austrian royal family [29].

Our work in this area is motivated by the theory that a forger can reveal himself by having brushstrokes that are different in style [22]. With this assumption it seems attractive to explore the fine details in paintings for the authentication purpose.

Several of the above mentioned methods employ multiresolution analysis to extract relevant features from the paintings. Both wavelets [22, 23, 24, 26] and curvelets [20] have been applied for this task. The reason for applying these signal processing tools is that they are (hopefully) able to detect subtle differences in the details between authentic paintings and forgeries at different scales in the paintings.

In this paper, we work with the contourlet transform [10] of the digital photographs. The reason for this choice of multiresolution analysis is that the atoms of the contourlet transform resemble contours well, making it likely to also represent the brushstrokes well and thereby the (subtle) differences between paintings made by different artists.

The contourlet transform of the paintings is modelled by a hidden Markov model [25] and by exploiting the differences between the hidden Markov models of different artists, we are able to predict the affiliation of new images. Our method is inspired by the Princeton approach in [22], where complex wavelets are modelled by hidden Markov models instead of contourlets.

Modelling contourlet transforms is a new contribution to the field of visual stylometry, as well as the subsequent processing of the fitted models. The main contribution of this paper is our experiments indicating that multiresolution transforms are not too sensitive

to different means of obtaining the digital reproductions.

The rest of the paper is organized as follows: In Section 2 we present the data used in the experiments. Section 3 describes the tools we utilize for our classification task. This includes the contourlet transform and how we model these with hidden Markov models to distinguish between paintings by different artists. In Section 4 we present the most important results of our experiments. Section 5 discusses the interpretation of the presented results and Section 6 contains our conclusions.

## 2 Artists used in our experiments

We have worked with paintings related to three different artists, as described in the subsections. The acquisition of the digital reproductions have varied depending on the origin of the data.

Each colour image with Red, Green and Blue (RGB) components was converted to grayscale (gray = 0.2989 R + 0.5870 G + 0.1140 B) as in [20, 24]; this preserves the brushstrokes (lines) in the image and we do not have to worry about how to combine edges from the color channels.

Each image was furthermore divided into square patches with a side length of 1024 pixels. We choose a fixed number of pixels that neighbouring patches are allowed to overlap and the patches are chosen such that they span the largest possible area of the image while not overlapping more than allowed.

### 2.1 Asger Jorn

Our primary dataset is related to the Danish painter Asger Jorn (1914 - 1973) and consists of paintings by Asger Jorn and his collaborators/apprentices. The images were provided courtesy of Museum Jorn, Silkeborg, Denmark.

We photographed the paintings with two cameras with very different technical specifications: A Canon Powershot G2 and a Nikon D90 with an AF-S Nikkor 50mm f/1.4G lens. We used two different cameras to test the robustness of our classification procedure to the means of acquisition and digital format. The Canon camera recorded images in JPEG format and the Nikon camera in raw format which we then exported to a lossless TIFF image.

The distance between the Canon camera and the paintings was consistent for all photos, however, this was not the case as to the Nikon camera. The photos taken by the Nikon camera were digitally corrected afterwards for the inconsistency in distance, i.e., for the inhomogeneous number of pixels per physical area – this was possible since the camera recorded the distance to its focus point

The paintings were photographed in their display positions on the walls in the museum where the lighting was homogeneous.

The paintings photographed are listed in Table 1.

Not all paintings could be captured in a single photograph of sufficiently high resolution per physical area unit, so these were photographed in parts – leaving us with a total of 44 digital images of the 15 paintings.

| Artist | Title | Year | Catalog no. |
|---|---|---|---|
| Asger Jorn | Automolok | 1948 | 1986/0001 |
| Asger Jorn | Euphorisme | 1970 | 1971/0274 |
| Asger Jorn | Hoved | 1935 | 1958/0043 |
| Asger Jorn | Henning. Figurstudie | 1933 | 1972/0205 |
| Asger Jorn | Portræt. Bodil | 1961 | – |
| Asger Jorn | Grand ventre - incendie | 1953 | 1958/0210 |
| Asger Jorn | Digteren Jens August Schade | 1937-44 | 2008/0001 |
| Asger Jorn | Prete Alla Spiaggia | 1957-59 | 1977/0001 |
| Asger Jorn | Trolden og fuglene | 1944 | 1962/0183 |
| Asger Jorn | Le Vent Nous Emporte | 1970 | 1988/0001 |
| Asger Jorn | Uden Titel. 1946 | 1946 | 1961/0116 |
| Asger Jorn | Uden Titel | 1956-57 | 1961/0115 |
| Jacqueline de Jong | Admiration de la reine de vert | 1961 | 1961/0119 |
| Asger Jorn/Enrico Baj | Uden titel | 1958 | 1958/0290 |
| Helmut Sturm | Uden titel | 1961 | 1961/0114 |

Table 1: The paintings related to Asger Jorn used in our experiments. The first column is the name of the artist, the second column is the original name of the painting, the third column is the year(s) of production and the fourth column is the catalogue number of the paintings at Museum Jorn. Additional information can be found at Art Index Denmark, https://www.kulturarv.dk/kid/. The painting "Portræt. Bodil" is displayed at Museum Jorn, but belongs to a private collector and therefore it has no catalog number.

Figure 1 presents an example of a painting by Asger Jorn and a painting by one of his collaborators, Helmut Sturm.

(a) Asger Jorn, *Hoved [Head]*, 1958, oil on canvas pasted on laminate, 35.5 cm × 26 cm. © Museum Jorn, Silkeborg.

(b) Helmut Sturm, *Uden Titel [No Title]*, 1961, oil on canvas, 75.5 cm × 80.5 cm. © Museum Jorn, Silkeborg.

Figure 1: Examples of the paintings used in our experiments. Helmut Sturm was a collaborator of Asger Jorn.

## 2.2 Pieter Bruegel the Elder

In [20, 24] the authors used images of drawings by Bruegel the Elder to test their classification methods. By courtesy of Daniel Rockmore of Dartmouth College, we have been able to test our methods on the same set of Bruegel drawings.

The original Bruegel drawings stem from the Metropolitan Museum of Art in New York and in the following we refer to the drawings by their Metropolitan Museum of Art catalogue number, as it is done in [24]. The image catalogue numbers and their categories are summarized in Table 2.

The images consist of eight authenticated drawings by Bruegel and five acknowledged Bruegel imitations.

All images are 16 bit RGB color images recorded in lossless PNG format [20, 24].

## 2.3 Charlotte Caspers

The authors of [22] concluded that they could decide on the authenticity as to a series of van Gogh paintings of which some were known to be forgeries. However, as the results of [22] were found to relate more to the cameras used for acquisition, the authors continued their research [26], but this time they made sure that they had full control over the entire process of acquiring the data: They asked the conservator Charlotte Caspers to paint a series of small paintings and then to copy them herself, and afterwards the digital

| Catalog no. | Title | Artist |
|---|---|---|
| 3 | Pastoral Landscape | Bruegel |
| 4 | Mountain Landscape with Ridge and Valley | Bruegel |
| 5 | Path through a Village | Bruegel |
| 6 | Mule Caravan on Hillside | Bruegel |
| 9 | Mountain Landscape with Ridge and Travelers | Bruegel |
| 11 | Landscape with Saint Jermove | Bruegel |
| 13 | Italian Landscape | Bruegel |
| 20 | Rest on the Flight into Egypt | Bruegel |
| 7 | Mule Caravan on Hillside | — |
| 120 | Mountain Landscape with a River, Village, and Castle | — |
| 121 | Alpine Landscape | — |
| 125 | Solicitudo Rustica | — |
| 127 | Rocky Landscape with Castle and a River | Savery |

Table 2: References for the images used in our analysis and their category as authentic Bruegel drawings or forgeries. A "—" means that the artist is unknown.

reproductions were obtained with the same scanner. The digitized versions of these paintings can be found online [7].

The observation made by the authors of [22, 26] is important as an authentication method with limitations as to the acquisition and digital format is not very useful.

Another complication when quantifying the artistic styles of painters is their use of different types of paint and canvas: We have not yet investigated the influence of materials thoroughly.

## 3 Methods

In the first part of this section we explain our choice of model and how we fit the models to our painting data, i.e., training a classifier. The last part of the section is concerned with how we utilize our models in detecting forgeries.

The training part of our authentication method works in basically three steps:

1. Make a contourlet transform of patches from the digital images, Section 3.1.
2. Model the contourlet coefficients by a hidden Markov model. Hidden Markov models are very rich models that have proved to be useful for classification tasks, see e.g. [8, 25, 27], Section 3.2.
3. Construct a classifier from the hidden Markov models, Section 3.3.

The contourlet transform and hidden Markov models are both well-known tools; we include brief introductions to make the paper more self contained and to justify our choices. Furthermore, we set the notation used in later sections.

### 3.1 The contourlet transform

In this section we briefly introduce the ideas of the contourlet transform and our motivation for using it in the present work. The reader is referred to the original article [10] for a more thorough exposition.

The contourlet transform offers a multiresolution decomposition of digital images with a user specified number of directional highpass subbands.

The multiresolution decomposition is obtained by successively computing coarser versions of the given image by applying a lowpass filter in the form of a Laplacian pyramid. The highpass part that is filtered out at each level of the multiresolution is divided into directional subbands, i.e., subbands that contains high frequency content in a limited band of directions.

The contourlet transform is constructed using a multiresolution analysis and therefore, after each application of the lowpass filter, we downsample the obtained lowpass image by a factor of two in each direction. This implies that at each level of the decomposition the number of pixels in the image is a quarter of that at the former level.

The contourlet transform has a number of attractive features that have served as a motivation for using it in the present work. The *individual* features are not unique to the contourlet transform, but the fact that they are all present is what makes the contourlet transform appealing.

**Directionality** As stated in the beginning of the section, the highpass subband from each level is divided into directional subbands. Here the point is that the *number* of directional subbands is specified by the user. Having this choice is different from e.g. wavelets, where the high frequency content is distributed between a fixed set of directional subbands (vertical, horizontal and diagonal in the case of ordinary wavelets).

The directionality can also be obtained by other transforms, e.g. by Gabor wavelets or curvelets [5].

**Optimal representation** An alternative view on the contourlet transform is that it provides a frame $\{\phi_i\}_{i\in J}$ that allows us to represent an image $I$ as

$$I = \sum_{i\in J} c_i \phi_i. \tag{1}$$

When quantifying how well a transform performs at representing images, we measure how well $I$ can be represented using only a subset of the functions in (1). More formally, we let

$$I_M = \sum_{i\in J_M} c_i \phi_i,$$

where $J_M \subseteq J$ is the set of indices of the $M$ numerically largest coefficients from (1). If the representation (1) is efficient, $I$'s behavior is captured by a few $\phi$'s and $I_M \approx I$, quantified by a suitable norm $\|I - I_M\|$.

In [11] it was established, that in a continuous setting and if $I$ is sufficiently smooth, $\|I - I_M\|^2$ cannot decay faster than $M^{-2}$. The curvelet and contourlet transform almost

obtain this bound [6, 10], as

$$\|I - I_M\|^2 \leq \text{constant} \cdot M^{-2}(\log M)^3, \tag{2}$$

*provided* that the number of highpass directions double at every other scale. This requirement is known as parabolic scaling and implies that the $\phi_i$'s in (1) have elongated supports that make them excel at capturing the edges (or contours) in images.

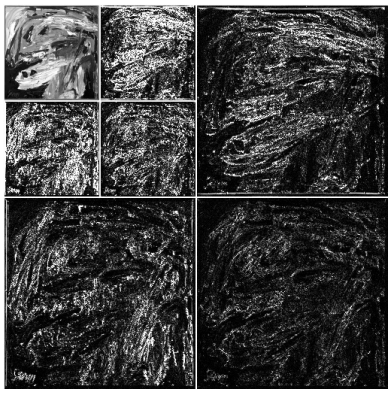
To the best of our knowledge, no transform has been found that outperforms the contourlet (and curvelet) transform in this respect.

**Discrete construction** The contourlet transform is *designed* to work on discrete data and hence the implementation for digital images is seamless.

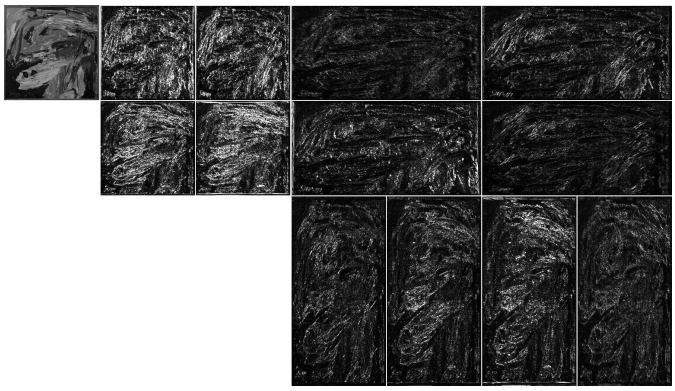
This is the major difference between the curvelet and the contourlet transform; the curvelet transform is designed to work on continuous data and is adapted to discrete data via a sampling grid.

By using a directional multiresolution transform it is our hope that we get a good representation of the brushstrokes captured in a digital image at different resolutions.

An example of a possible contourlet transformation of an image is seen in Figure 2 – along with a wavelet transformation for comparison.



(a) A two level wavelet decomposition of image.



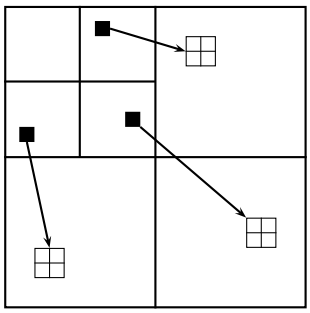
(b) A two level contourlet decomposition of image.

Figure 2: Comparison of wavelet and contourlet transformation by a two level decomposition of the grayscale version of the painting in Figure 1b. In the highpass subbands of the contourlet transformation we capture details at a variety of orientations, as opposed to the three fixed orientations of the wavelet transform.

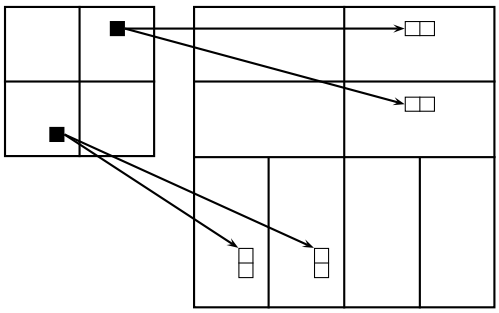
## 3.2 Hidden Markov modelling of contourlet transforms

Figure 3 illustrates the situation we want to model, i.e., how the coefficients evolve through the levels of the multiresolution decomposition. The coefficients that are con-

nected in Figure 3 are at the same spatial locations in the image, but at different levels of resolution.



(a) Coefficients in a two level wavelet transform of an image.



(b) Coefficients in a possible two level contourlet transform of an image.

Figure 3: Parent-children relationship for wavelets and contourlets. The black coefficients are parents of the white coefficients. The parents reside at the same same spatial location as their children, but at a coarser resolution.

Referring to Figure 4, they share the same major direction, as the children levels together span the same direction as their parents level.

With these common characteristics it seems plausible that the coefficients are related; if the parent coefficient is large, indicating the presence of an edge, it is likely that (at least) one of the children coefficients is large as well – and vice versa. This relationship is *local*, since we only need to know if the parent is related to an edge and not the grand parents.

From (2) we have that the contourlet transform captures an image with few large coefficients, while the majority of coefficients are insignificantly small. So the distribution of coefficients at a particular subband is highly peaked around zero and has heavy tails.

As it has been demonstrated in [25], such a distribution is well approximated by a *mixture* of normal distributions.

The hidden Markov model we utilize in our modelling was introduced for contourlets in [25] and captures the essential properties mentioned above in the following manner:

- An individual coefficient $o_i$ is a realisation of a stochastic variable $O_i$ whose distribution is a mixture of normal distributions. For each coefficient we also introduce a discrete hidden state $S_i$ indicating which normal distribution in the mixture the observation stems from.

- To model the relation between a coefficient $o_i$ and its parent $o_{\rho(i)}$, we model the relation between the associated hidden variables $S_i$ and $S_{\rho(i)}$, and this relation is Markovian, i.e., $S_i$ depends on the rest of the hidden variables only through its parent.

Each coefficient on the coarsest highpass subbands is the root of a tree and each tree is modelled by a hidden Markov model – as illustrated in Figure 5.

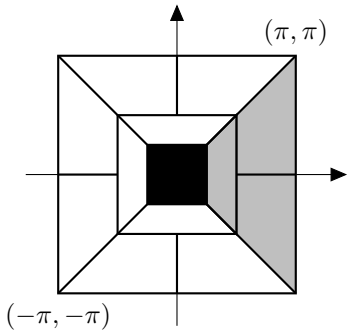



Figure 4: Ideal partitioning of the frequency plane of a contourlet transform as in Figure 2b. The outer square is the frequency plane. Each level of the contourlet transform keeps the center square of the former level. The black center region is contained in the lowpass subband and we refer to the gray shaded region as a major direction of the contourlet transform, as these are the directional frequencies covered by the coarsest highpass subband.

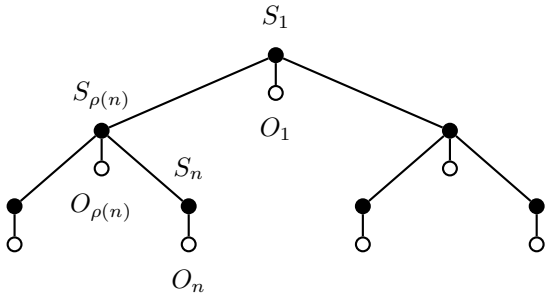



Figure 5: Illustration of a graph and the parent-children relations of a hidden Markov model. The black nodes are hidden state variables and the white nodes are observations. Index 1 is the root of the tree, $n$ is an arbitrary node with parent $\rho(n)$.

Let $\mathcal{T}$ be such a tree. If each hidden state take on values $\{1, \ldots, M\}$ (unrelated to the $M$ in (2)), we parameterize it as follows:

- The distribution of the hidden state at the root, $P(S_1 = m)$, $m = 1, \ldots, M$.
- The transition probabilities between each hidden state and its parent, $\epsilon_{i,\rho(i)}^{n,m} = P(S_i = n | S_{\rho(i)} = m)$, $m, n = 1, \ldots, M$.
- The mean $\mu_{i,m}$ and standard deviation $\sigma_{i,m}$ of each of the normal distributed variables $O_i | S_i = m$, $m = 1, \ldots, M$.

The free parameters of the tree, which we use later in the classification, remove the redundant probabilities from the list above.

To obtain robust estimates of the parameters in the model, we model the trees rooted in the same major direction (ref. Figures 3b and 4) as independent and identically distributed – this is also known as tying [8, 27].

As mentioned in Section 3.1, the number of coefficients quadruple at every level, so if for example we apply a three level contourlet transform to a square image with a side length of 1024 pixels there are $1024^2/4^3 = 128^2$ trees.

We fit hidden Markov models to each of the major directions independently.

In our application, the hidden states can each take two values since this provides an adequate description of our data with the minimum number of parameters. Furthermore, we set $\mu_{i,m} = 0$ for all $i$ and $m$, since this choice also fits data.

To verify that the hidden Markov model is indeed applicable for the images we want to model, we compare the empirical distribution of contourlet coefficients in a subband with coefficients simulated from the fitted hidden Markov model. As a by-product of the maximum-likelihood estimation we get the distribution of the hidden states in the hidden Markov model, and therefore simulation from the mixture model in a subband is straightforward. The validity of the model is verified by making a QQ-plot of the observed and simulated coefficients.

Hidden Markov modelling of contourlet transforms was introduced in [25], which was motivated by similar modelling of wavelet transforms from [8].

Hidden Markov modelling of wavelet transforms have only limited directional separation and each direction is modelled *separately* as illustrated in Figure 3a. The modelling of the contourlet transform offers some modelling of inter-directional dependencies, since the major directions of the contourlet transform can spilt into several minor directions as seen in Figure 3b.

### 3.2.1 Parameter estimation

We estimate parameters in the hidden Markov models by maximizing the corresponding likelihood function using an Expectation Maximization (EM) algorithm; the appropriate algorithm for this type of model is described in [8]; see also [27] for precautions related to the implementation.

The EM algorithm is an iterative procedure and requires an initial estimate of the parameters as input. We initialize by simply choosing a random model obtained by simulating random values for each of the parameters.

To reduce the final model's dependency on the initial estimates, we run the EM algorithm a number of times with random initial values and choose the final model that yields the highest likelihood.

During a maximization step of the EM algorithm, it may happen that some parameters are updated to values that are approaching the computers numerical precision, which is not desirable. We therefore restrict the permissible values of the parameters to avoid numerical underflow. This probably also makes the EM algorithm less sensitive to the initial estimates [19].

## 3.3 Classification

Hidden Markov models describe our data adequately and are also widely used for classification purposes, e.g. speech recognition [27] and texture retrieval [8, 22, 25].

Classification is typically performed by computing a sensible distance between the fitted models, e.g. the Kullback-Leibler divergence. For our data this did not give satisfactory results, so we have used the fitted models in a different way.

Using the parameters of the hidden Markov models directly with a support vector machine classifier (SVM; see e.g. [18] for a good exposition) was not successful either.

A likely explanation to the failure of the SVMs is that many of the parameters in the hidden Markov models are highly correlated; we circumvent this issue by performing a feature selection prior to the classification.

We have tried two kinds of feature selection that both increased the success of the classifier. SVM can be used to perform recursive feature selection as described in [17] by recursively filtering out the features with the smallest weights (as assigned from the SVM).

However, the recursive feature selection was not as successful as using the lasso for logistic regression[1] on an embedding found by a multidimensional scaling (MDS) algorithm, see Section 3.3.1. The individual steps will be described in more details afterwards, but the workflow of our classification is as follows:

$$
\begin{aligned}
&\text{model parameters}\\
&\quad \xrightarrow{\text{lasso}} \text{select \& weigh parameters}\\
&\quad \longrightarrow \text{weighted distances between images}\\
&\quad \xrightarrow{\text{MDS}} \text{embedding in } \mathbb{R}^2\\
&\quad \xrightarrow{\text{SVM}} \text{classification}
\end{aligned}
$$

[1] Independently of our work, it was *proposed* in [16] that logistic regression could be used to weigh parameters.

### 3.3.1 Parameter selection

In the following we let $\boldsymbol{\theta}_n = (\theta_{n,1}, \ldots, \theta_{n,K})$ denote the free parameters of the hidden Markov model fitted to an image $n$. Furthermore, we let $\widetilde{\boldsymbol{\theta}}_n = (1, \theta_{n,1}, \ldots, \theta_{n,K})$.

We are interested in modelling the paintings as being either authentic or forgeries, i.e., as having two possible states. When using logistic regression to select and rank the parameters, we model this as the probabilities

$$p_n = P(\text{image } n \text{ is authentic}|\boldsymbol{\theta}_n)$$

using the logistic transformation, $x \mapsto \log \frac{x}{1-x}$. The simplest type of regression is when we assume a linear relationship

$$\log \frac{p_n}{1-p_n} = \beta_0 + \sum_{\ell=1}^{K} \beta_\ell \theta_{n,\ell} = \boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_n. \tag{3}$$

There is no a priori justification for using a linear model, except that it is the simplest model and it gives satisfactory results.

Furthermore, the model (3) has the advantage that it does not depend on the range of the pixel values of the images we classify. We typically represent the grayscale values of the pixels in a digital image either as floating point numbers in the unit interval or as 8 or 16 bit integers, i.e., integers between 0 and $2^8 - 1$ or $2^{16} - 1$, respectively. The representation range should be consistent in our dataset, but the classification method should not depend on the specific choice. That (3) ensures this will be elaborated on after the definition of the metric we introduce to measure the distance between the hidden Markov models.

Under the assumptions of a two class model and (3), the log-likelihood for $N$ images under the Bernoulli model is

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \big\{ y_i \log p_i + (1-y_i) \log(1-p_i) \big\} \\ &= \sum_{i=1}^{N} \Big\{ y_i \boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_i - \log\big(1 + \exp(\boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_i)\big) \Big\}. \end{aligned} \tag{4}$$

It is well known how to maximize this log-likelihood function, see e.g. [18].

However, using the regular logistic regression for feature selection gives rise to some problems 1) it is difficult (and sometimes impossible with the working precision) to estimate the $\beta$'s in high dimensional problems 2) many of the $\beta$'s have small numerical values, but few are exactly zero, i.e., we do not perform a good selection.

To circumvent this problem, we use the lasso logistic regression [18] instead, where we maximize (4) subject to the constraint

$$\sum_{\ell=1}^{K} |\beta_\ell| \leq t. \tag{5}$$

Equivalently, we can solve the optimization problem in the Lagrangian form

$$\max_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \left[ y_i \boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_i - \log\bigl(1 + \exp(\boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_i)\bigr) \right] - \lambda \sum_{\ell=1}^{K} |\beta_\ell| \right\}. \tag{6}$$

In [13] the authors present a fast iterative algorithm for solving (6) and we have used the implementation [14]. When using the software from [14] we get solutions to (6) for finite, decreasing sequence of $\lambda$'s. The optimal $\lambda$ is then chosen by leave-one-out cross-validation, as illustrated in Algorithm 1.

**Algorithm 1:** Cross-validation

**Input**: patches from $N$ images, where the $i$'th image consists of $N_i$ patches:
$P = \{p_{i,j} \,|\, 1 \le i \le N, 1 \le j \le N_i\}$
**Input**: decreasing, positive sequence $\Lambda = \{\lambda_k\}_{k=1}^{K}$
**Output**: cross-validation error for each $\lambda \in \Lambda$
Initialization:
$e_k = 0$, $1 \le k \le K$
**for** $i \in \{1, \ldots, N\}$ **do**
  patch set $= P \setminus \{p_{i,j} \,|\, 1 \le j \le N_i\}$
  **for** $\lambda \in \Lambda$ **do**
    train model from (6) with patch set and $\lambda$
    classify $\{p_{i,j} \,|\, 1 \le j \le N_i\}$ with model
    **if** *classification is correct* **then**
      $e_k := e_k + 1/N$
**return** $e_1, \ldots, e_K$

Note in the algorithm that when classifying a patch from an image we leave out out all other patches from that image; otherwise our classifier has an unrealistic prior.

With a set of estimated $\beta$'s we define the associated weights as

$$w_i = |\beta_i|, \quad 1 \le i \le K. \tag{7}$$

Using these weights we define the distance between two hidden Markov models with parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$, respectively, as

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) := \sum_{\ell=1}^{K} w_\ell |\theta_{i,\ell} - \theta_{j,\ell}|. \tag{8}$$

Since many of the weights are zero, the number of non-zero terms in the sum (8) is usually much smaller than $K$. The reason for using a weighted $\ell^1$-norm (as opposed to an $\ell^p$-norm with $p > 1$) is that the $\ell^1$-norm penalize small differences better than $\ell^p$-norms with larger $p$, which is necessary for the probability parameters. The classification is significantly better when using the $\ell^1$-norm instead of the $\ell^2$-norm. Using $\ell^\tau$ pseudo-norms with $0 < \tau < 1$ does not further improve the classification.

The metric (8) is between the hidden Markov models fitted to the individual patches in an image; we combine the distances between patches into a distance between images by means of the Hausdorff distance: Let $N_I$ denote the number of patches in picture $I$ and let $\boldsymbol{\theta}_{I,1}, \ldots, \boldsymbol{\theta}_{I,N_I}$ be parameters of the hidden Markov models fitted to the patches in $I$. Then the Hausdorff distance *from* image $I$ to image $J$ is

$$d_H(I,J) := \max_{1\le k\le N_I} \Big\{ \min_{1\le \ell\le N_J} \{d(\boldsymbol{\theta}_{I,k}, \boldsymbol{\theta}_{J,\ell})\}\Big\}, \tag{9}$$

and the Hausdorff distance *between* image $I$ and $J$ is

$$\max\{d_H(I,J), d_H(J,I)\}. \tag{10}$$

Before elaborating on how we use these distances for classification, we shall prove the claim from the beginning of the section, i.e., that these distances do not depend on the range of the pixel values in the images.

The probability parameters in the hidden Markov models do not depend on the range of the pixel values; hence we need only pay attention to the standard deviations. Both the contourlet transform and the standard deviations are positively homogeneous of degree 1 (when scaling the pixel values, the contourlet coefficients scale similarly; the same goes for the deviances of the fitted hidden Markov model) and the $\beta$'s in the regression are homogeneous of degree $-1$ (if a parameter $\theta_{n,\ell}$ in (3) scales with a factor $k > 0$, the corresponding weight $\beta_\ell$ scales with $1/k$). This implies that the scaling factor cancel in the expression (8): Let $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}'_i$ be the parameters of two hidden Markov models fitted to the same image, but with pixel values scaled differently. If $\sigma_{i,\ell}$ and $\sigma'_{i,\ell}$ are standard deviations in $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}'_i$, respectively, then $\sigma_{i,\ell} = k\sigma'_{i,\ell}$ for some constant $k > 0$ and the corresponding regression weights are $w_\ell = \frac{1}{k} w'_\ell$. Thus the contribution of these standard deviations to the distance (8) is

$$w_\ell |\sigma_{i,\ell} - \sigma_{j,\ell}| = \frac{w_\ell}{k} |k\sigma_{i,\ell} - k\sigma_{j,\ell}| = w'_\ell |\sigma'_{i,\ell} - \sigma'_{j,\ell}|,$$

independent of $k$.

Returning to the use of (10), we use (8)–(10) and compute pairwise distances between all the images. Using a multidimensional scaling (MDS) algorithm (see [30] for a thorough exposition on the so-called classical MDS and e.g. [12, 18] about non-classical MDS), we find a configuration of points in $\mathbb{R}^2$ whose pairwise distances are as close as possible to those computed by (10).

From the points computed by the MDS we construct a classifier based on a simple SVM with linear decision boundary.

The important part of this procedure is that we arrive at a configuration of points where the images of authentic paintings and forgeries are indeed *separable* – the subsequent classification by linear SVM might not be optimal, but it serves the point of illustrating the potential of our method.

It is worth making a couple of remarks.

- In plots where the training is based on different images, the weights from (7) are different and hence the pairwise distances between the hidden Markov models in (8) change as well. This means that it does not make sense to make a direct comparison of the subsequent plots in Figures 7, 8, 11 and 12. Consider Figures 11 and 12 as an illustration: Here the pairwise distances between the Caspers images change as different weights are used, whereas their spatial distribution does not change significantly.

- For the purpose of visualizing the pairwise distances between the models, the units in the MDS plot are not important – in this regard only the relative distances matters.

  When computing a classification rule with SVM, the points are normalized to have zero mean and unit standard deviation, as there are indications that this ensures a more robust rule [1, 15]. Hence the units are not important for the SVM classification either.

  In conclusion, we do not need units for the classification nor the display of points, but we have included the units to illustrate how the weights influence the distances.

## 4 Results

We now present the results of our experiments with the paintings related to Asger Jorn, Pieter Bruegel and Charlotte Caspers.

All numerical experiments were carried out in Matlab and the source code is distributed[2].

We used the Contourlet Toolbox by Minh Do [9]. The contourlet transform in [9] is implemented with a double filter bank, one that implements a Laplacian pyramid for separating lowpass and highpass subbands, and one that implements the partitioning into directional subbands. We therefore need to specify filters for the two filter banks, the number of levels in the Laplacian pyramid, and the number of directional subbands on each level of the Laplacian pyramid.

The classification results with the different filters were comparable and the number of levels and directions had only a small influence. The results presented here are based on the 9-7 pyramidal filter, the pkva12 directional filter and a four level contourlet transform with 8, 8, 16 and 16 directions from coarsest to finest pyramidal level.

A topic, which is not related to the transforms used, is the options considering downsampling of the images. Obviously we cannot use images of arbitrary resolution; if the resolution is too coarse the level of details is not sufficient; if the resolution is too high we may be modelling insignificant details in the canvas or paint because the brushstrokes are not part of high frequency content.

The effect of downsampling – and other preprocessing of the images – will not be investigated thoroughly here; these issues need further investigation and will be dealt with in our subsequent work. In the present experiments we have fitted hidden Markov

[2] Available online at http://www.mathworks.com/matlabcentral/fileexchange/35322

models and estimated the prediction error of the corresponding classifiers through cross-validation for a variety of downsampling factors and used the downsampling factor that yields the best result.

Our experiments showed that using an inappropriate amount of downsampling is not only visible in the final classification; it may entail that the optimization algorithm used to solve (6) is not converging and hence fitting the Bernoulli model is not feasible.

As described in Section 3.2, we verify that the HMT model is appropriate for our data by making a QQ-plot of the observed coefficients versus coefficients simulated by the fitted HMT model. An illustration of a QQ-plot for the coefficients is presented in Figure 6.

For some subbands the tails of the coefficient distributions are heavier than the fitted mixture distribution – however, the model is acceptable.

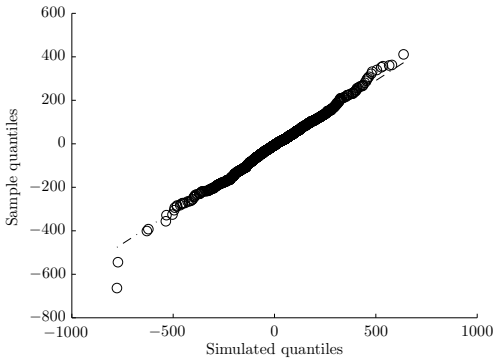



Figure 6: QQ-plot of coefficients from subband versus simulated coefficients from the fitted HMT model. The coefficients are from a patch of an Asger Jorn picture recorded with the Nikon D90 camera.

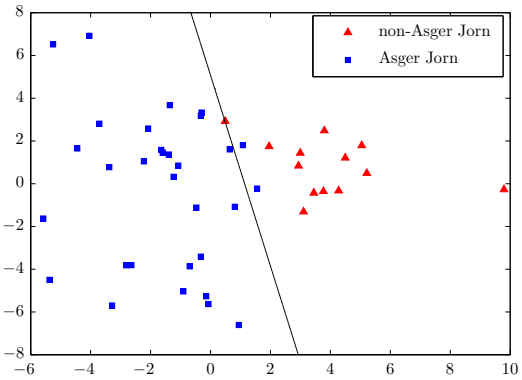
## 4.1 Results for Asger Jorn

As mentioned in Section 2.1 some of the paintings from Museum Jorn were photographed in multiple images and we divided our data into two groups, one for classification based on the digital images, and one for classification based on the paintings, i.e., by concatenating the images of the same painting. This means that when classifying the digital images we used (9) as it stands, but when classifying paintings we calculated the pairwise distances between the paintings $P$ and $P'$ as

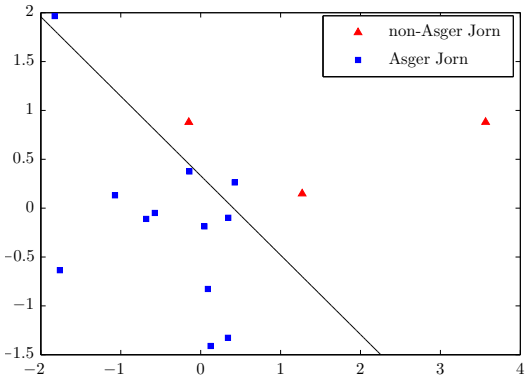
$$d_H(P, P') := \max_{1 \le k \le N_P} \Big\{ \min_{1 \le \ell \le N_{P'}} \{ d(\boldsymbol{\theta}_{P,k}, \boldsymbol{\theta}_{P',\ell}) \} \Big\}, \tag{11}$$

where $N_P$ is the number of hidden Markov models fitted to the patches in *all* images from $P$.

The result of the multidimensional scaling and the following classification with a linear SVM is seen in Figure 7 and summarized in Table 3.




(a) Classification of the Asger Jorn images.




(b) Classification of the Asger Jorn paintings. The pairwise distances are smaller than in Figure 7a due to the second application of the Hausdorff distance (9).

Figure 7: Classification of Asger Jorn's images using a linear SVM on the embedding in $\mathbb{R}^2$ computed with a multidimensional scaling algorithm. In Figure 7b the classification is based on the paintings – each of which is captured in several images. Figure 7a shows the classification of the individual images.

| | AJ | not AJ | total |
|---|---|---|---|
| images | 28/31 | 11/13 | 39/44 |
| paintings | 9/12 | 2/3 | 11/15 |

Table 3: Classification results of the images related to Asger Jorn. As discussed in the paper some paintings are captured in multiple images. The first number in a category is the number of correctly classified images/paintings; the second number is the total number of images/paintings in that category.

The $\lambda$'s in the regression model (6) that give the best classification results rely on between 10 and 30 parameters from the HMTs – which is a considerable reduction from the full parameter set of 184 parameters.

Furthermore, it is notably difficult to classify the images of the painting in Figure 1b by Helmut Sturm; when performing cross-validation, the painting by Helmut Sturm is misclassified for all $\lambda$ values.

As mentioned in Section 2.3 other researchers have reported difficulties with classification independent of the camera, i.e., classifying digital reproductions of paintings related to the same artist, but acquired by different cameras/scanners [26].

It is of interest to see how our method performs in such a situation and as mentioned in Section 2.1 we photographed the paintings at Museum Jorn with two different cameras. Comparing the two sets of images should yield good results since they are digital reproductions of the same paintings. However, since the two cameras produce digital reproductions that differ significantly in resolution and DPI, we need to downsample[3] the high resolution images to a level where the details are of the same size as in the low resolution images. If the high resolution images are not downsampled, the objects in the two sets of images are not comparable in any way. In our situation we can determine the appropriate level of downsampling by comparing the size of a painting's characteristic elements in different images of that painting.

As it is seen in Figure 8 our method can perform cross-camera classification if we choose the right level of downsampling; it is also evident that a comparison of images from the two cameras is meaningless if the level of downsampling is inappropriate.

With Figure 8b we correctly classified 28 out of 36 images from the Canon camera.

We mentioned in the introduction that our original inspiration was hidden Markov modelling of complex wavelets. To support our theoretical motivation for using contourlets instead of wavelets, we have performed our experiments with complex wavelets instead of contourlets, i.e., the only thing changed is the multiresolution transform. Hidden Markov modelling of complex wavelets was introduced in [28] and the code for computing the complex wavelet transforms and fitting hidden Markov models was kindly provided by Justin Romberg.

This is of course not a precise replication of the experiments with complex wavelets in [22] as we have different data sets and code, but it serves to illustrate our point with contourlets.

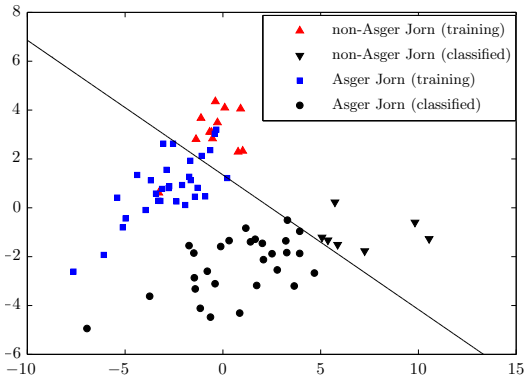
The complex wavelet counterpart to Figure 7 is seen in Figure 9 where there is a more significant overlap between the two classes.

## 4.2 Results for Pieter Bruegel

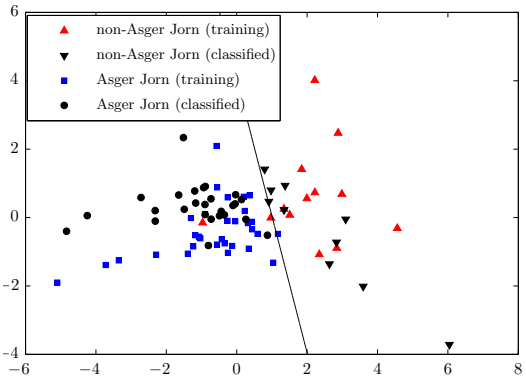
When performing cross-validation on the images related to Pieter Bruegel the Elder, we can correctly classify 11 out of 13 drawings; the two drawings that are not correctly classified are 13 and 125 from Table 2.

When computing the embedding, it appears as in Figure 10.

[3] Performed with standard tools and settings in Matlab

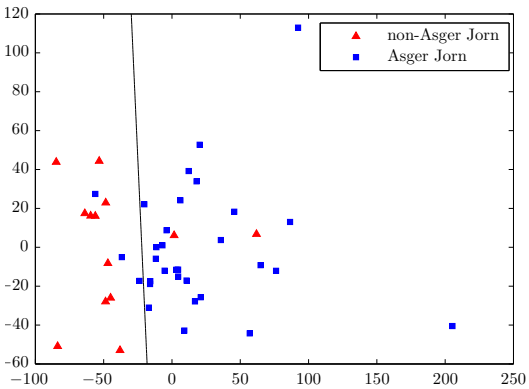



(a) The zoom level of the training and test images are very different.

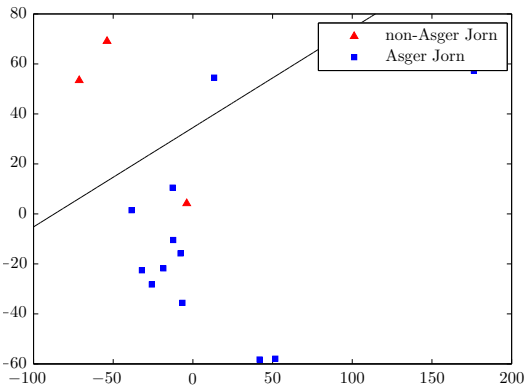



(b) The zoom level of the training and test images are comparable.

Figure 8: Comparison of the effect of zooming the images. In both cases we have tried to classify the images from the Canon camera (with low resolution) from a Support Vector Machine decision rule calculated from images obtained with the Nikon camera (with different resolutions). In Figure 8a we used the original images for the training and in Figure 8b we scaled the training images. The right zoom level was determined manually by comparing fixed objects in images of the same painting with different zoom levels.




(a) Classification of the Asger Jorn images.




(b) Classification of the Asger Jorn paintings.

Figure 9: Counterpart to Figure 7: Here complex wavelets are used instead of contourlets as basis for the classification; notice that the separation is worse than in the contourlet case.

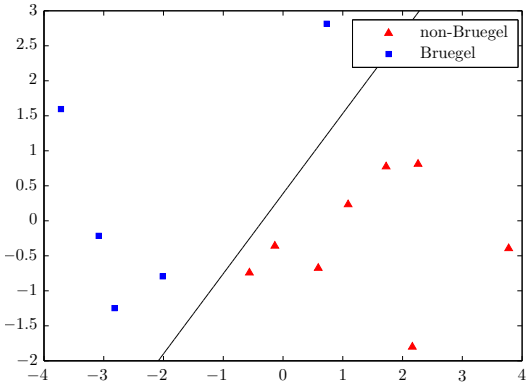



Figure 10: Classification of images related to Pieter Bruegel the Elder with a linear SVM on an embedding in $\mathbb{R}^2$ found with an MDS.

### 4.3 Results for Charlotte Caspers

Our results regarding the images by Charlotte Caspers are far from satisfactory in terms of the number of correct classifications, but they are nonetheless interesting.

In Figure 11 is the result of a MDS for all images by Caspers and here it is worth noticing that the copies and originals are clustered pairwise and separated from the other pairs.

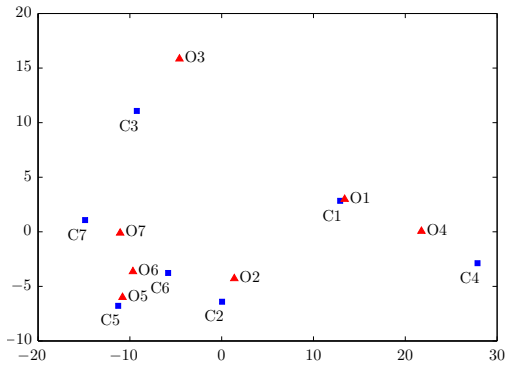



Figure 11: Embedding in $\mathbb{R}^2$ of the models fitted to paintings by Charlotte Caspers from a multidimensional scaling algorithm, as described in Section 3.3. The originals and copies are denoted by O# and C#, respectively, with $\# = 1, \ldots, 7$.

Also, for many $\lambda$'s in (6), the regression algorithm did not converge and in all cases the number of non-zero parameters were very high compared to the experiments with Asger Jorn. This indicates that the Bernoulli model is not suitable, as more than two classes are needed.

It is interesting to study the behaviour of our method when classifying paintings that are totally unrelated to the training material, e.g. the paintings by Charlotte Caspers

using a classification rule obtained from the paintings related to Asger Jorn. The result is presented in Figure 12; as expected the results related to Charlotte Caspers' paintings do not overlap with those of Asger Jorn.

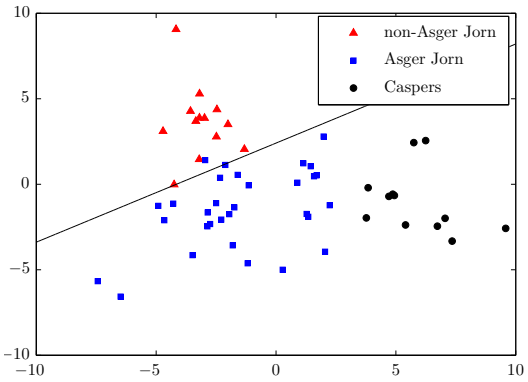



Figure 12: Classification of the paintings by Charlotte Caspers using the paintings related to Asger Jorn. As can be seen, the paintings by Charlotte Caspers do not overlap the paintings related to Asger Jorn.

## 5 Discussion

The classification method presented yields good results when separating paintings by Asger Jorn from those of his collaborators; however, there are at least two issues that are worth mentioning 1) our method *does* contain black-box elements 2) the poor classification of the paintings by Charlotte Caspers in Figure 11.

Regarding 1) it would of course be desirable to know exactly what we are capturing in the paintings using the hidden Markov models – is it indeed the brushstrokes? On the other hand, all present methods for digital authentication contain some kind of black-box element; there is no guarantee that our interpretation of the features is exhaustive.

Regarding 2) it is disappointing that the method cannot handle this data set. On the other hand, the "copies" of Charlotte Caspers' paintings are made by herself and thus the styles can be expected to have a high degree of similarity – as can be seen online [7]. Hence it may be that the reason that our method indicates the pairwise cohesion is because it is in fact so. This conjecture is supported by the observation from Section 4.3 of the Bernoulli model being inappropriate for modelling the hidden Markov models: The originals and copies are not two homogeneous classes.

This can also be considered part of a more general problem. When training a classifier to distinguish between the works of a given artist and "the rest", it is not realistic to model all of "the rest", since "the rest" potentially contains every image not made by the given artist, which is definitely not homogeneous.

In the situation where we have multiple imitations that could not be modelled as representations of the same class, one could imagine training multiple classification rules

– one for each of the homogeneous subclasses – and deem a painting authentic if it passed all/a majority of these classification rules.

Finally it is interesting that our classification methods depend on the zoom level of the images. The effect of data collection and preprocessing of the digital images is an issue that will be addressed further in later work.

## 6 Conclusion

In this paper, we have constructed methods for classifying paintings. The contourlet transform of digital images of paintings is able to efficiently capture the contours of the image – including the brushstrokes of a painter, which is believed to be characteristic. The contourlet transform is well described by a hidden Markov model, which captures both the coefficient distribution on the individual resolutions and the dependency structure between the resolutions.

By applying a lasso regression we perform variable selection and weighing among the features of the hidden Markov models. Using such a selection to define a metric between the trained hidden Markov models, we find an embedding of points in $\mathbb{R}^2$ that resembles the pairwise distances. Once this embedding is found, we train a classifier to distinguish between the classes of paintings.

With this procedure we have performed a leave-one-out cross-validation where we have successfully classified 39 out of 44 images related to Asger Jorn and 11 out of 15 images related to Pieter Bruegel.

A promising aspect of our method is that we can determine the authenticity of images digitized differently than the training data – from the model found through the cross-validation mentioned above, 28 out of 36 images from the other test set was classified correctly.

To experimentially motivate contourlets over other multiresolution transforms we have also performed our experiments with complex wavelets, that have been used in similar experiments [22, 26]. The results with contourlets are better than those with complex wavelets.

## Acknowledgement

The authors would like to thank Arne Jensen for suggesting this topic and for providing contact to the Portinari Project, that initiated our interest in this subject.

Furthermore, we thank Museum Jorn, Silkeborg, Denmark, and curator Teresa Østergaard Pedersen for allowing us to use paintings of Asger Jorn. We thank Daniel Rockmore, Dartmouth College, for providing us with drawings related to Pieter Bruegel the Elder.

We thank Justin Romberg for sharing his code on hidden Markov modelling of complex wavelets.

Finally, we thank P. Svante Eriksen for helpful discussions and valuable comments on this manuscript as well as the anonymous reviewers for their valueable feedback.

## References


[1] Shawkat Ali and Kate Smith-Miles. "Improved Support Vector Machine Generalization Using Normalized Input Space". In: *AI 2006: Advances in Artificial Intelligence*. Ed. by Abdul Sattar and Byeong-ho Kang. Vol. 4304. Lecture Notes in Computer Science. Springer Berlin/Heidelberg, 2006, pp. 362–371. ISBN: 978-3-540-49787-5. DOI: 10.1007/11941439_40.

[2] Mahmoud Al-Ayyoub, Mohammad T. Irfan, and David G. Stork. "Boosting multi-feature visual texture classifiers for the authentication of Jackson Pollock's drip paintings". In: *Proceedings of SPIE*. Ed. by David G. Stork, Jim Coddington, and Anna Bentkowska-Kafel. Vol. 7869. 2011.

[3] Igor E. Berezhnoy, Eric O. Postma, and H. Jaap van den Herik. "Automatic extraction of brushstroke orientation from paintings". In: *Machine Vision and Applications* 20.1 (Jan. 2009), pp. 1–9. DOI: 10.1007/s00138-007-0098-7.

[4] Igor Berezhnoy, Eric Postma, and H. Jaap van den Herik. "Computer analysis of Van Gogh's complementary colours". In: *Pattern Recognition Letters* 28.6 (2007), pp. 703–709. DOI: 10.1016/j.patrec.2006.08.002.

[5] Emmanuel J. Candès and David L. Donoho. "Continuous curvelet transform - II. Discretization and frames". In: *Applied and Computational Mathematics* 19 (Sept. 2005), pp. 198–222. DOI: 10.1016/j.acha.2005.02.004.

[6] Emmanuel J. Candès and David L. Donoho. "New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise $C^2$ Singularities". In: *Communications on Pure and Applied Mathematics* 57 (Feb. 2004), pp. 219–266. DOI: 10.1002/cpa.10116.

[7] Charlotte Caspers. *Caspers Data Set*. http://www.math.princeton.edu/ipai/datasets.html.

[8] Matthew S. Crouse, Robert D. Nowak, and Richard G. Baraniuk. "Wavelet-Based Statistical Signal Processing Using Hidden Markov Models". In: *IEEE Transactions on Signal Processing* 46.4 (Apr. 1998), pp. 886–902. DOI: 10.1109/78.668544.

[9] Minh N. Do. *Contourlet Toolbox*. http://www.ifp.illinois.edu/~minhdo/software.

[10] Minh N. Do and Martin Vetterli. "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation". In: *IEEE Transactions on Image Processing* 14.12 (Dec. 2005), pp. 2091–2106. DOI: 10.1109/TIP.2005.859376.

[11] David L. Donoho. "Sparse Components of Images and Optimal Atomic Decompositions". In: *Constructive Approximation* 17.3 (2001), pp. 353–382. DOI: 10.1007/s003650010032.

[12] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification, second edition*. Wiley, 2001.

[13] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1 (Feb. 2010), pp. 1–22. URL: http://www.jstatsoft.org/v33/i01.

[14] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *GLMnet for Matlab*. http://www-stat.stanford.edu/~tibs/glmnet-matlab.

[15] Arnulf B. A. Graf, Alexander J. Smola, and Silvio Borer. "Classification in a normalized feature space using support vector machines". In: *IEEE Transactions on Neural Networks* 14.3 (May 2003), pp. 597–605. DOI: 10.1109/TNN.2003.811708.

[16] Daniel J. Graham, James M. Hughes, Helmut Leder, and Daniel N. Rockmore. "Statistics, vision, and the analysis of artistic style". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 4.2 (Mar. 2012), pp. 115–123. DOI: 10.1002/wics.197.

[17] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. "Gene Selection for Cancer Classification using Support Vector Machines". In: *Machine Learning* 46 (1 2002), pp. 389–422. DOI: 10.1023/A:1012487302797.

[18] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning, 2. edition*. Springer, 2008.

[19] Richard J. Hathaway. "Constrained maximum-likelihood estimation for a mixture of $M$ univariate normal distributions". PhD thesis. Rice University, 1983.

[20] James M. Hughes, Daniel J. Graham, and Daniel N. Rockmore. "Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder". In: *Proceedings of the National Academy of Sciences* 107 (2009), pp. 1279–1283. DOI: 10.1073/pnas.0910530107.

[21] Mohammad Irfan and David G. Stork. "Multiple visual features for the computer authentication of Jackson Pollock's drip paintings: Beyond box-counting and fractals". In: *Proceedings of SPIE*. Ed. by Kurt S. Niel and David Fofi. Vol. 7251. 2009. DOI: 10.1117/12.806245.

[22] C. Richard Johnson, Jr, Ella Hendriks, Igor J. Berezhnoy, Eugene Brevdo, Shannon M. Hughes, Ingrid Daubechies, Jia Li, Eric Postma, and James Z. Wang. "Image Processing for Artist Identification". In: *IEEE Signal Processing Magazine* 25.4 (2008), pp. 37–48. DOI: 0.1109/MSP.2008.923513.

[23] Jia Li and James Z. Wang. "Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models". In: *IEEE Transactions on Image Processing* 13 (2004), pp. 338–351. DOI: 10.1109/TIP.2003.821349.

[24] Siwei Lyu, Daniel Rockmore, and Hany Farid. "A digital technique for art authentication". In: *Proceedings of the National Academy of Sciences* 101.49 (Dec. 2004), pp. 17006–17010. DOI: 10.1073/pnas.0406398101.

[25] Duncan D.-Y. Po and Minh N. Do. "Directional Multiscale Modeling of Images using the Contourlet Transform". In: *IEEE Transactions on Image Processing* 15.6 (June 2006), pp. 1610–1620. DOI: 10.1109/TIP.2006.873450.

[26] Güngör Polatkan, Sina Jafarpour, Andrei Brasoveanu, Shannon Hughes, and Ingrid Daubechies. "Detection of Forgery in Paintings Using Supervised Learning". In: *IEEE International Conference on Image Processing*. 2009. DOI: 10.1109/ICIP.2009.5413338.

[27] Lawrence R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition". In: *Proceedings of the IEEE* 77 (Feb. 1989), pp. 257–286. DOI: 10.1109/5.18626.

[28] Justin Romberg, Hyeokho Choi, Richard Baraniuk, and Nick Kingsbury. "Hidden Markov tree modeling of complex wavelet transforms". In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2000, pp. 133–136. DOI: 10.1109/ICASSP.2000.861889.

[29] Robert Sablatnig, Paul Kammerer, and Ernestine Zolda. "Hierarchical classification of paintings using face- and brush stroke models". In: *Proceedings of the Fourteenth International Conference on Pattern Recognition*. Vol. 1. 1998, pp. 172–174. DOI: 10.1109/ICPR.1998.711107.

[30] G. A. F. Seber. *Multivariate Observations*. Wiley, New York, 1984.

[31] Richard P. Taylor, Adam P. Micolich, and David Jonas. "Fractal analysis of Pollock's drip paintings". In: *Nature* 399 (June 1999), p. 422. DOI: 10.1038/20833.

[32] R.P. Taylor, R. Guzman, T.P. Martin, G.D.R. Hall, A.P. Micolich, D. Jonas, B.C. Scannell, M.S. Fairbanks, and C.A. Marlow. "Authenticating Pollock paintings using fractal geometry". In: *Pattern Recognition Letters* 28.6 (Apr. 2007), pp. 695–702. DOI: 10.1016/j.patrec.2006.08.012.