



Aalborg Universitet

AALBORG UNIVERSITY  
DENMARK

## Authentication of paintings using hidden markov modelling of contourlet coefficients

Jacobsen, Christian Robert; Nielsen, Morten

*Publication date:*  
2011

*Document Version*  
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*

Jacobsen, C. R., & Nielsen, M. (2011). *Authentication of paintings using hidden markov modelling of contourlet coefficients*. Department of Mathematical Sciences, Aalborg University. Research Report Series No. R-2011-07

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

AALBORG UNIVERSITY

**Authentication of paintings  
using hidden Markov modelling  
of contourlet coefficients**

by

Robert Jakobsen and Morten Nielsen

R-2011-07

June 2011

DEPARTMENT OF MATHEMATICAL SCIENCES  
AALBORG UNIVERSITY

Fredrik Bajers Vej 7 G ■ DK-9220 Aalborg Øst ■ Denmark

Phone: +45 99 40 80 80 ■ Telefax: +45 98 15 81 29

URL: <http://www.math.aau.dk>



# AUTHENTICATION OF PAINTINGS USING HIDDEN MARKOV MODELLING OF CONTOURLET COEFFICIENTS

C. ROBERT JACOBSEN AND MORTEN NIELSEN

ABSTRACT. In this paper we present a method for determining the authenticity of a painting by modelling the contourlet transform of digital photographs of the painting. Specifically, we model the contourlet transform by a hidden Markov model and use these models for classification by various means.

We have tested our methods on different artists and the results are promising.

## 1. INTRODUCTION

The task of authenticating paintings – that is, determining whether a painting is indeed painted by the claimed artist – has traditionally been performed by art experts and connoisseurs. This is a time consuming task where the conclusion is also subjective. It is of interest to assist the authentication by automatic methods, that can provide an unbiased opinion in a fast and reliable way. In addition to the authentication, it is also possible that we gain insight into the development of the artistic style of a painter by finding a quantitative description of his/hers paintings.

A very distinctive characteristic of an artist is the brushstrokes in a painting. Brushstrokes are typically consistent throughout a painters work, independent of the contents. Forgeries are often characterized by more hesitant brushstrokes, when the forger tries to imitate the style of another painter [4].

There have been previous attempts to perform automatic authentication of paintings by training the methods to be able to detect paintings that are known forgeries. All automatic authentication methods works on digital reproductions of the paintings and the fundamental idea is to extract a limited number of features from the digitized paintings that are sufficiently expressive to distinguish the styles of different artists.

The authentication task have been tried most thoroughly on paintings by van Gogh [4, 2, 3, 19] and Pieter Bruegel the Elder [17, 15]. Several of the before mentioned methods employ multiresolution analysis to extract relevant features from the paintings. Both wavelets [4, 17, 19, 16] and curvelets [15] 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 brushstrokes between authentic paintings and forgeries.

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

---

*Date:* September 8, 2011.

*Key words and phrases.* Artist authentication, contourlet transform, hidden Markov models, classification.

C. R. Jacobsen and M. Nielsen are with the Department of Mathematical Sciences, Aalborg University, Denmark.

transform resemble brushstrokes well. Hence it is likely to obtain very sparse representations of the paintings which in turn gives better parameters for distinguishing subtle differences between paintings made by different artists.

The contourlet transform of the paintings is modelled by a hidden Markov model [18] 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 a development of the Princeton approach in [4], where complex wavelets are used in the multiscale decomposition.

The rest of the paper is organized as follows: In Section 2 we present the data used in the experiments. Section 3 describes how we apply the hidden Markov models of the contourlet transform to distinguish between authentic paintings and forgeries. 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 attributed to three different artists, as described later in this section. All paintings have been digitised into colour images, but the method for obtaining the digital images have varied. For the data we have gathered ourselves this is a deliberate choice since we want to test the robustness of our method and avoid the problems elaborated in Section 2.3.

Each colour image with Red, Green and Blue (RGB) components is converted to grayscale ( $\text{gray} = 0.2989 R + 0.5870 G + 0.1140 B$ ) as in [17, 15] in an attempt to reduce our models dependency of color variations in the images.

Before applying the contourlet transformation we divided each image into non-overlapping, square patches with a side length of 1024 pixels. The patches were taken from the largest possible central area of the image. We have also used smaller patches, but the results obtained were not nearly as good here. The number of patches varied between 2 and 6, depending on the size of the image.

**2.1. Candido Portinari.** Candido Portinari (1903 - 1962) was one of the most important Brazilian cultural personalities of the last century. In 1979 his son initiated The Portinari Project [1] to document the work of his father and provide a view of Brazilian culture during the lifetime of Candido Portinari. An important part of this project is to detect forgeries amongst the paintings attributed to Portinari.

The dataset that initiated the work for this paper was provided courtesy of The Portinari Project. The data includes both authentic Portinari paintings and known forgeries. The digital versions of the paintings have been obtained by scanning photographic negatives from an analog camera. We have not had access to information about the details in the data collection and the data is far from homogeneous, as the distance to the camera, illumination and angle of the camera varies considerably. As a consequence we have ended up deselecting this data material from our final investigations.

**2.2. Asger Jorn.** We expanded our dataset by photographing pictures of the Danish painter Asger Jorn (1914 - 1973) and of pictures made by his apprentices (that are stylistically similar to Jorn's paintings). The images by Asger Jorn were provided courtesy of Museum Jorn, Silkeborg, Denmark. We photographed the paintings of Asger Jorn with

two different cameras that had very different technical specifications to see how this influenced our method; a Canon Powershot G2 and a Nikon D90 with a AF-S Nikkor 50mm f/1.4G lens. The Nikon camera records pictures in the raw format, which we then exported to lossless tiff format. The Canon camera records pictures directly to a lossy jpeg format.

With the Canon camera all photographs were taken with the same distance to the painting. With the Nikon camera the distance to the photographs varied, but this camera records the distance to its focus point, we are able to correct the digital images to have the same number of pixels per area.

Naturally, we need to standardize the images before comparing them, but it turns out that there are several things to note about the data material. First of all we need to find the right zoom level – if the number of pixels per area is too high, we get unwanted details in the digital reproductions. For instance, the canvas might be too visible at a high resolution. If the canvas is coarse – as can be seen on the edges of Figure 1a – this can present a problem, since we might model this texture instead of characteristics of the artist. Secondly, on older paintings the paint often cracks, which again can interfere with the modelling.

If the number of pixels per area is too low we do not get enough details to distinguish the subtle differences between the paintings.

The images of Asger Jorn’s paintings have been the primary dataset used in our experiments. Figure 1 presents an example of a painting by Asger Jorn and of a painting by one of his collaborators.



(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.

**2.3. Charlotte Caspers.** In [19] the authors concluded that the primary reason for their success in [4] was because the authentic van Gogh paintings and the known forgeries

had been photographed by different cameras. This is an important observation; if an authentication method requires paintings to be digitized in exactly the same way it is not that useful.

To have access to paintings that had been produced in the same manner and photographed by the same camera, the authors of [19] asked the conservator Charlotte Caspers to paint a series of small paintings and then copy these paintings herself. The digitized versions of these paintings can be found online [7]. Since both the originals and copies are made by the same artist, it should be very challenging to separate the two categories, hence this can be thought of as a benchmark dataset.

Moreover, the authentication is also complicated by the fact that the paintings have been made with different kinds of paint and on different canvas.

### 3. METHODS

In the first part of this section we go through the motivation for our choice of model and how we fit the models to our painting data. 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 (a part of) the digital images. The contourlet transform is briefly described in 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 very useful in recognition tasks, see e.g. [20, 8, 18]. In Section 3.2 we describe how to model multiscale decompositions with hidden Markov models.
- (3) Construct a classifier from the hidden Markov models of each image.

The classification rule established can then be used to judge the category of new images.

**3.1. The contourlet transform.** In this section we briefly introduce the ideas of the contourlet transform and our motivation for working with this transform. The reader is referred to the original article [10] for a more detailed description.

For many types of one dimensional signals, an efficient tool for obtaining sparse representations is the wavelet transform. Wavelets can also be used in higher dimensions by applying the one-dimensional transform separately in each dimension. Unfortunately this approach only captures the one-dimensional singularities in higher dimensional signals which are often of limited interest. In images the significant changes in the content occurs along the contours, but the wavelet transform only detects isolated pixels along contours in three directions (horizontal, vertical and diagonal) on each level of the transform. In most images the contours are not composed of lines with only three different orientations, so the wavelet transform is suboptimal when it comes to capturing the two dimensional nature of images.

To overcome the limitations of the wavelet transform, Do & Vetterli developed the contourlet transform which enjoys the same features as wavelets, but are made for efficient representation of two-dimensional images. The contourlet transform works by finding the point discontinuities of an image and linking these discontinuities into linear structures that resemble the contours of the image, hence the name contourlets.

Attractive features of contourlets that are similar to the attractive features of wavelets is that we can obtain a multiscale decomposition of an image by recursively working on the lowpass subband from the contourlet transformation and that the basis functions of the contourlet transform are localized in both time and frequency domain. But contourlets offer more than wavelets: We can specify the number of directional highpass subbands and at every other level the number of directions can double. Additionally, the support of the basis functions of the contourlet transform are rectangular and spans a user specified number of directions and therefore contains directional information about the contours they describe. This is opposed to the square support of the wavelet basis functions that are parallel to the axes.

By providing a multiscale decomposition that captures the shapes of the contours in the image it is our hope that we get a better representation of an artists brushstrokes.

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



(A) Two level wavelet decomposition of image.

(B) Possible 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 lines at a variety of orientations – as opposed to the three fixed orientations of the wavelet transform.

The curvelet transform [6, 5] also offers a directional multiscale decomposition of images. We have chosen to work with the contourlet transform since the contourlet transform is *designed* to work on discrete data like the digital images we have access to, whereas the curvelet transform is designed to work on continuous data and is adapted to discrete data via a sampling grid.

**3.2. Hidden Markov modelling of contourlet coefficients.** Referring to Figure 4 and Figure 3 we use the following notation in connection with a contourlet transform of an image.

The coarsest level of the highpass subbands in a contourlet transform contains the main directions captured by the transform – in Figure 2b there are four main directions.

In finer subbands the number of directions may increase, but just as in the wavelet transform, the directional subbands are naturally related to the coarser scales by their affiliation to a main direction – as illustrated in Figure 4b.

Consider a contourlet coefficient  $o_i$  on scale  $J$  (which is not the finest scale). The coefficient on the same spatial position on the immediate coarser scale  $J + 1$  is called the parent of  $o_i$ , denoted  $o_{\rho(i)}$ . The coefficients on scale  $J - 1$  whose parent is  $o_i$ , are called the children of  $o_i$ . When we are working with two dimensional transforms, each coefficient has four children. Each of the coefficients on the coarsest scale of a highpass subband are roots in a tree and it is these trees we are modelling.

In [8] the authors found a good model for *wavelet* transforms by modelling such trees. Due to the excellent compression properties of the wavelet transform, the distribution of coefficients in a given subband and level is highly peaked around zero and has fairly heavy tails. In this light is not surprising that a normal distribution is not appropriate to model the many small coefficients and few large coefficients, but a *mixture* of normal distributions provides a good description of the wavelet coefficients. Furthermore, wavelet coefficients tend to cluster (i.e., if a particular wavelet coefficient is small/large it is likely that adjacent coefficients are also small/large) and be persistent across scales (i.e., if a wavelet coefficient is small/large it is likely that the children of that coefficient are also small/large). The persistency is, however, mostly local – that is, a coefficient does usually not directly influence its grandchildren.

The model proposed in [8] that captures the (local) dependencies across and within scales and the mixture of normal distributions is a Hidden Markov Tree (HMT). In a HMT we introduce a hidden state variable  $S$  for each of the observations, that determines which normal distribution in the mixture our observation stems from, that is,  $O|S = m \sim \mathcal{N}(\mu_m, \sigma_m^2)$ .

The dependencies between the hidden states is modelled by a Markov model, such that the distribution of a hidden state only depends on the rest of the hidden states through its parent coefficient. An observation  $o_i$  and the corresponding hidden state  $s_i$  are indexed in the same manner – see Figure 3.

In [18] Po & Do demonstrated that contourlet transforms can be modelled by HMT's that are similar to those introduced in [8]. The difference is that the HMT's used to model contourlet transforms have branches that span multiple directional subbands. However, this is a major advantage of modelling the contourlet transformations instead of wavelet transformations, because we can model inter-direction dependencies between contourlet coefficients.

Consider a HMT  $\mathcal{T}$  in a main direction of a contourlet transform. Following the notation in Figure 3, the state/observation at the root is indexed by 1. If each hidden state has  $N$  possible values, we have the following parameters for  $\mathcal{T}$ :

- $p_{S_1}(n)$ ,  $1 \leq n \leq N$ : The distribution of the hidden state at the root of the tree.
- $\epsilon_{i,\rho(i)}^{n,m} = P(S_i = n | S_{\rho(i)} = m)$ ,  $i \in \mathcal{T} \setminus \{1\}$ ,  $1 \leq n, m \leq N$ : The transition probabilities between every hidden state and its parent.
- $\mu_{i,m} = E(O_i | S_i = m)$ ,  $i \in \mathcal{T}$ ,  $1 \leq m \leq N$ : The mean of the normally distributed variable  $O_i | S_i = m$ .
- $\sigma_{i,m}^2 = \text{Var}(O_i | S_i = m)$ ,  $i \in \mathcal{T}$ ,  $1 \leq m \leq N$ : The variance of the normally distributed variable  $O_i | S_i = m$ .



Each of the coefficients in the directional subbands on the coarsest levels is the root of a tree modelled by a HMT. To obtain robust estimates of the parameters in the model, we model all trees in one main direction as independent and identically distributed – this is also known as tying [8, 20]. Each level of the contourlet transform contains a quarter of the pixels from the previous level. So if we apply e.g. a four level contourlet transform to our patches of side length 1024 pixels, we have  $128^2$  trees for estimating parameters in the HMT model.

HMT's are fitted independently to each of the main directions.

We have only used two values for each hidden state, since this provides an adequate description of our data with the minimum number of parameters. This also allows the interpretation that one state is associated with edges in the image (characterized by a large variance) and one state is associated with homogeneous areas in the image (characterized by a small variance). Furthermore, we set  $\mu_{i,m} = 0$  for all  $i$  and  $m$ , since this is also in correspondence with data.

With the HMT where each hidden state variable can take two possible values and a variance characterizing the normal distributions we have the following free parameters,

$$(1) \quad \theta = \{p_{S_1(1)}\} \cup \{\epsilon_{i,\rho(i)}^{1,m} : i \in \mathcal{T} \setminus \{1\}, m = 1, 2\} \cup \{\sigma_{i,m} : i \in \mathcal{T}, m = 1, 2\}.$$

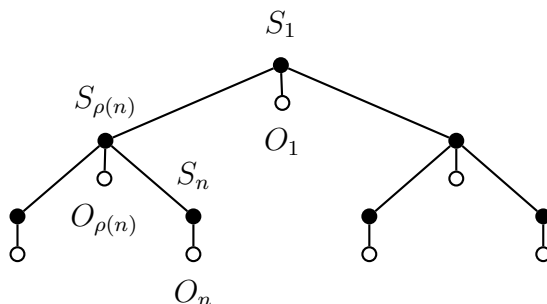


FIGURE 3. Illustration of a graph and the parent children relations of a hidden Markov tree. The black nodes are hidden state variables and the white nodes are observations.

To verify that the HMT 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 HMT model. As a by-product of the maximum-likelihood estimation we get the distribution of the hidden states in the HMT, so 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.

**3.2.1. Estimation of parameters.** Finding a Maximum Likelihood Estimate (MLE) of the parameters in the HMT's modelling the contourlet transform is not an easy task: There are many parameters and since we do not observe the values of the hidden states, we can not estimate the distribution of the hidden variables directly. Hence, it is not feasible to find MLE analytically. To find a (local) MLE we employ an Expectation-Maximization (EM) algorithm. The specific steps in an EM algorithm are problem dependent and for the HMT's we consider here, the appropriate EM algorithm is described in [8].

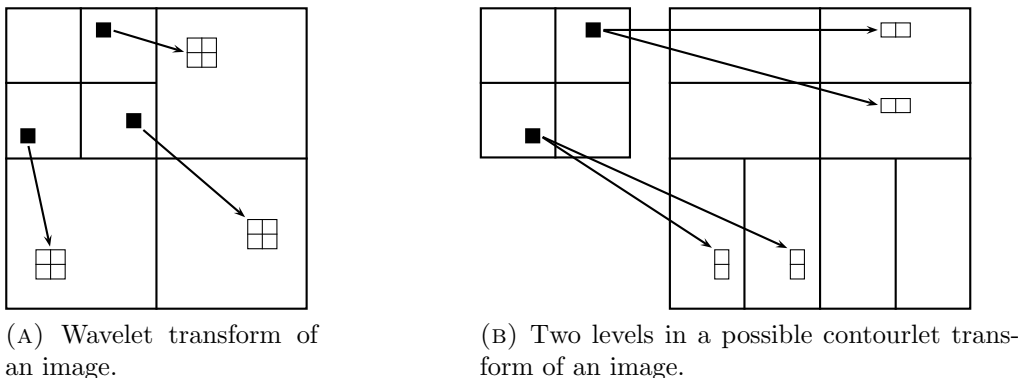


FIGURE 4. Parent-children relationship for wavelets and contourlets. The black coefficients are parents of the white coefficients.

There are some practical issues with numerical precision that must be taken into account when applying the EM algorithm for the problem at hand.

The EM algorithm is an iterative procedure and expect as such an initial estimate of the parameters. When performing maximum likelihood estimation we run the EM algorithm a number of times with different initial values and pick the final estimate with the greatest likelihood.

The initial values for the probabilities in (1) are simulated from a uniform distribution on the interval  $[\epsilon, 1]$  for some small  $\epsilon > 0$ . The initial values for the standard deviations are chosen uniformly in the interval  $[\epsilon, 2s]$ , where  $s$  is the observed variance in the relevant subband.

During the iterations of the EM algorithm we also have to take precautions – both in the expectation and maximization steps.

In the expectation step this is necessary since some of the intermediate results decays exponentially fast as information is progressed through the coefficient tree. This is a common problem circumvented by scaling the intermediate results [20].

In the maximization step we might obtain estimates of the parameters that are too small and we therefore set a lower bound for values of the parameters – just as in the initialization. If the variances are too small the likelihood of observing large coefficients is numerically zero, which is not desirable if it happens for a large number of coefficients. If the probabilities are too small we run into trouble in the next iteration of the EM algorithm; the states with small probabilities will not be visited very often and hence the state probability will again be estimated as being small – in this way we will start a self-reinforcing effect [20].

A positive effect of restricting the admissible values of the parameters is that our EM algorithm supposedly is less sensitive to the choice of initial values [14].

**3.3. Weighing of parameters.** To determine which parameters in the HMT are most important in distinguishing between the authentic paintings and the known forgeries, we perform a suitable classification.

In the case where we have only two categories for a patch – authentic or fake – we model this category by a Bernoulli distribution. In a more general case, where we want

to determine the affiliation to one of  $K$  different artists, we would use the multinomial distribution with  $K$  classes.

Let  $N$  denote the number of patches in our data set. For  $1 \leq i \leq N$  we let  $\boldsymbol{\theta}_i$  denote the free variables for the HMT fitted to patch  $i$ . We also use the notation  $\tilde{\boldsymbol{\theta}}_i = (1, \theta_{i,1}, \dots, \theta_{i,K})$ .

Furthermore, we introduce the categorical variables  $y_i$  by

$$y_i = \begin{cases} 1 & \text{if patch } i \text{ is authentic} \\ 0 & \text{if patch } i \text{ is a forgery} \end{cases}$$

and  $p_i = P(y_i = 1 | \boldsymbol{\theta}_i, \boldsymbol{\beta})$ , where  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)$  is a vector of scalars described below.

We model the probabilities by a logistic model:

$$(2) \quad \log \frac{p_i}{1 - p_i} = \beta_0 + \sum_{\ell=1}^K \beta_\ell \theta_{i,\ell} = \boldsymbol{\beta}^\top \tilde{\boldsymbol{\theta}}_i.$$

A priori it is not given that the relationship in (2) should be linear, but it is a simple model which ensures that the final classification is scale independent and gives satisfactory results. The issue with independence of scale will be explained in the end of Section 3.4.

With the assumption (2), the log-likelihood for  $N$  patches under the Bernoulli model is

$$(3) \quad \begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^N \{y_i \log p_i + (1 - y_i) \log(1 - p_i)\} \\ &= \sum_{i=1}^N \{y_i \boldsymbol{\beta}^\top \tilde{\boldsymbol{\theta}}_i - \log(1 + \exp(\boldsymbol{\beta}^\top \tilde{\boldsymbol{\theta}}_i))\} \end{aligned}$$

It is well known how to maximize this log-likelihood function, see e.g. [13]. When presented with a patch from an image whose authenticity we would like to determine, one could be inclined to predict the probability that the patch is authentic by the logistic model:

$$(4) \quad p_i = \frac{\exp(\boldsymbol{\beta}^\top \tilde{\boldsymbol{\theta}}_i)}{1 + \exp(\boldsymbol{\beta}^\top \tilde{\boldsymbol{\theta}}_i)}.$$

However, as discussed in Section 5, there are some potential issues with this approach. The other option we have used is presented in the Section 3.4.

With the standard logistic model (3) it is difficult (and sometimes impossible with the working precision) to obtain estimates of the  $\boldsymbol{\beta}$ 's in high dimensional problems. To circumvent this problem, we use the lasso logistic regression [13] instead, where we maximize (3) subject to the constraint

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

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

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

There are two noticeable effects of using the lasso regression. First of all we are able to obtain estimates of the  $\beta$ 's even in high dimensional problems. Secondly, when we impose an  $\ell^1$  constraint on the parameters, we force many of the parameters to be exactly zero [13]. This is opposed to the usual regression, where many of the parameters have small numerical values, but few parameters are exactly zero. The second effect of the lasso regression is desirable since it greatly reduces the number of parameters in the HMT model that actually influence the decision about the affiliation of a painting, when the training material is sparse.

In [12] the authors present a fast iterative algorithm for solving (6) and we have used the implementation made available through [11]. When using the software from [12] we get solutions for finite, decreasing sequence of  $\lambda$ 's in (6). The optimal  $\lambda$  is then chosen by cross-validation, performed in the following way (based on the guidelines described in [13]): We obtain  $\boldsymbol{\beta}$  for a sequence of  $\lambda$ 's based on all patches except those from a particular image. Then we predict the authenticity of the patches in the image that was not used in the training and use a suitable loss function to determine if the prediction is successful. This procedure is repeated for all images and in this way we obtain an estimate of the prediction error for each  $\lambda$  in the sequence.

As mentioned in the beginning of this section, we can also extend this method to deal with multiple artists. Now let

$$p_i = P(\text{patch } i \text{ is made by artist } k \mid \boldsymbol{\theta}_i, \boldsymbol{\beta})$$

where  $1 \leq k \leq K$ . Instead of the binomial logistic model (2), we model these probabilities by the symmetric multinomial logistic model

$$(7) \quad p_i = \frac{\exp(\boldsymbol{\beta}_i^\top \tilde{\boldsymbol{\theta}})}{\sum_{k=1}^K \exp(\boldsymbol{\beta}_k^\top \tilde{\boldsymbol{\theta}})}, \quad 1 \leq i \leq K.$$

With this model we get a  $\boldsymbol{\beta}$  vector for each of the  $K$  artists. To find the MLE of (7) subject to the constraint (5) we can also use the software presented in [12].

**3.4. Multidimensional scaling.** Here we present a classification alternative to (4) that resolves some of the issues with the prediction procedure described above. This method has also been our primary method.

It is desirable to present the patches as points in a euclidean space, i.e. assign meaningful coordinates to a patch. After representing patches in this way we are able to use a wide range of classification tools. The important part here is the way we obtain the euclidean representation – the tools we have used in the subsequent classification might not be optimal – but they serve more as an illustration of the *possibility* of obtaining good classification schemes.

The way we have achieved this is by defining a suitable dissimilarity<sup>1</sup> between individual patches and then employed a Multi-Dimensional Scaling (MDS) algorithm to find a configuration of points in  $\mathbb{R}^2$ , whose pairwise distances resembles those of the original data.

When comparing HMT's it is common to use the Kullback-Leibler divergence [20, 18]. We have tried to use the Kullback-Leibler divergence as a dissimilarity measure between patches, but this did not give satisfactory results.

Instead we have used a dissimilarity between paintings that utilises the regression described in Section 3.3.

After performing the logistic regression to obtain  $\beta$  we get an ordering of the  $\theta_i$ 's – the larger the numerical value of  $\beta_i$ , the more  $\theta_i$  influences the authenticity.

We use this to define a norm between the patches as a weighted  $\ell^1$ -norm between the  $\theta$ 's of different patches. Let

$$(8) \quad w_i = |\beta_i|, \quad 1 \leq i \leq K.$$

The distance between two patches  $P_i$  and  $P_j$  with HMT parameters  $\theta_i$  and  $\theta_j$ , respectively, is then

$$(9) \quad d_P(P_i, P_j) := \sum_{\ell=1}^K w_\ell |\theta_{i,\ell} - \theta_{j,\ell}|.$$

Since many of the regression coefficients are zero the number of nonzero terms in the sum (9) is usually much smaller than  $K$ .

We compute the patch-to-patch distance between every pair of patches and combine these into image-to-image distances by computing the Hausdorff distance between the patches belonging to each image. So if image  $I_i$  contains the patches  $P_{i,1}, \dots, P_{i,M_i}$ , the distance from image  $I_i$  and  $I_j$  is

$$(10) \quad d_I(I_i, I_j) := \max_{1 \leq k \leq M_i} \left\{ \min_{1 \leq \ell \leq M_j} \{d_P(P_{i,k}, P_{j,\ell})\} \right\}$$

and the Hausdorff distance between image  $I_i$  and  $I_j$  is

$$(11) \quad d_H(I_i, I_j) := \max\{d_I(I_i, I_j), d_I(I_j, I_i)\}.$$

We use a MDS algorithm to find a configuration of points in  $\mathbb{R}^2$  whose pairwise distances are as close as possible to the distances in (11).

The MDS algorithm that provide the best results is the classical MDS (see e.g. [21]). The two dimensional representation obtained from the classical MDS accounts for as much of the variation in data as is possible in two dimensions. It is as such roughly equivalent to a principal component analysis [13].<sup>2</sup>

<sup>1</sup>A dissimilarity between patches  $r$  and  $s$  is a map  $\delta$  from the set of patches to the real numbers that satisfies  $\delta(r, s) \geq 0$ ,  $\delta(r, r) = 0$  and  $\delta(r, s) = \delta(s, r)$ .

<sup>2</sup>The classical MDS aims at finding a configuration of points in some  $\mathbb{R}^n$  – possibly very high dimensional – whose pairwise euclidean distances are the same as the specified pairwise distances. The coordinates of the euclidean configuration are ordered in terms of explaining variance in data – the lower entries in a coordinate vector explains more of the variance than the higher entries. Hence, we get a good two dimensional representation by choosing the first two entries in the configuration obtained with the classical MDS.

We also tried using a weighted  $\ell^2$ -norm in (9), but this gave significantly worse results. The reason for this is that the most significant parameters of the HMT is the transition probabilities and the  $\ell^1$ -norm penalizes small deviations better than the  $\ell^2$ -norm. The multidimensional scaling algorithms only require that the  $d_I(I_i, I_j)$  in (10) are dissimilarities, so in (8) we also tried  $\ell^\tau$  pseudo-norms with  $\tau < 1$ . However, this approach did not outperform the  $\ell^1$ -norm.

If we include additional images in the pairwise distance matrix, the configuration of coordinates changes. So when using this method for classification we proceed as follows. We calculate the weights (8) using the patches from a training set and perform a MDS for *all* the available patches, that is, for both the training images and the images to be tested. From this MDS we calculate a suitable decision boundary in the *training* set and classify the test images with this classifier.

In Section 4 we have used this method in a leave-one-out cross-validation with the decision boundary calculated with a simple support vector machine classifier.

3.4.1. *A note on independence of representation.* We can represent a digital image in multiple equivalent ways by scaling the pixel values, but the final classification should not depend on the representation.

When calculating distances between models as in (9), we are ensured that the distances are independent of the image scale, as is seen by the following argument.

The contourlet transform and the standard deviations in the HMT's are both positively homogeneous of degree 1 and the weights in the regression are homogeneous of degree  $-1$ . This implies that scaling factors cancel in the expression (9): We only have to consider the standard deviations in a fitted HMT model and these are homogeneous of degree 1. Let  $\sigma_{i,\ell}$  and  $\sigma'_{i,\ell}$  be the two of the standard deviations in an HMT model of a patch  $P_i$  with equivalent representations. That is,  $\sigma_{i,\ell} = k\sigma'_{i,\ell}$  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 (9) 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 the scale.

## 4. RESULTS

We now present the results obtained by our methods. In our experiments we used the Contourlet Toolbox by Minh Do [9]. In the Contourlet Toolbox we have to specify the pyramidal and directional filter as well as the number of directional subbands on each pyramidal level. In our experiments we tried a number of different combinations of filters, but most filters performed comparably. In the results presented here we have used the 9-7 pyramidal filter and the pkva12 directional filter.

Besides the number of directions on the pyramidal level, we can also choose the zoom level – that is, how much digital zoom should be applied before fitting a HMT model. It turns out that the zoom level is influential – for the high resolution Asger Jorn images

---

If the classical MDS is successful at finding a *perfect* euclidean configuration, this dimension reduction technique is the same as applying a principal component analysis to the euclidean configuration.

obtained with the Nikon camera, the best results are obtained when the images are zoomed to be 75% of their original size.

The effect of choosing the right zoom level is not only visible in the final classification; with the wrong zoom levels the regression algorithm [11] did not converge – hence fitting the binomial model is not feasible.

The effect of zoom levels – and other preprocessing of the images – needs further investigation and will not be explored in the present work.

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 5.

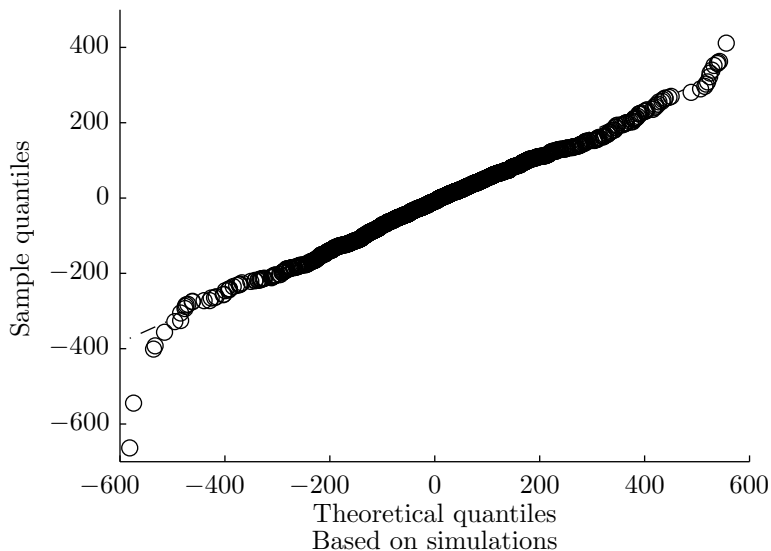


FIGURE 5. QQ-plot of coefficients from subband versus simulated coefficients from the fitted HMT model. The coefficients are from a patch from an Asger Jorn picture recorded with the Nikon D90 camera.

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

**4.1. Results for Asger Jorn.** When deciding which model(s) to work with, we perform cross-validation as described in Section 3.3 either directly with the logistic model or by classifying the MDS configuration.

Since we have fixed the filters used in the contourlet transforms, the different transforms we have worked with are characterized by the number of directions on each pyramidal level. For the different contourlet transforms we have performed leave-one-out cross-validation to get an estimate of the error rate when prediction the category of an image.

The results are presented in Figure 6.

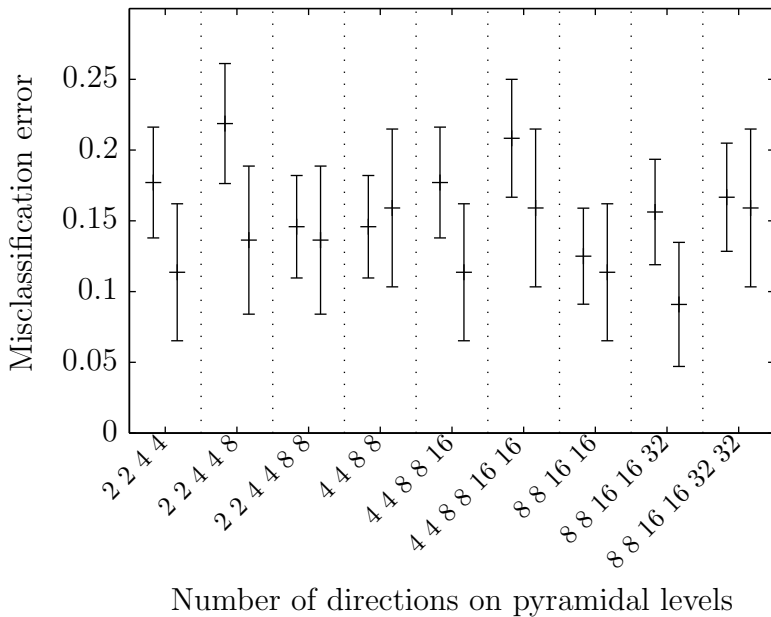


FIGURE 6. Results from leave-one-out cross-validation for the different settings used in the training. The length of a bar is twice the standard deviation of the estimated misclassification error. For each choice of directions we have estimated the misclassification error with (4) (left bar) and by the method described in Section 3.4 (right bar).

The reason that the left bars are shorter than the right is that with (4) we determine the authenticity of each patch *individually* whereas for the other method we determine the authenticity one image at a time.

With the method from Section 3.4 the number of misclassified images varies between 6 and 11 out of 44.

It is seen from Figure 6 that the contourlet transform with 8, 8, 16, 16 directions on the pyramidal levels performs well with both classification methods – the rest of the results presented is based on this transform. The contourlet transform with 2 major orientations directions is best with MDS based classification. We have chosen to work with the transform that has 8, 8, 16, 16 directions none the less, since the separation for *all* images is much better for this transform.

The  $\lambda$ 's in the regression model that give the best classification results rely on between 10 and 30 nonzero parameters from the HMT's – which in most cases is a considerable reduction from the full parameter set, cf. Table 1.

Regarding the misclassification probabilities with (4) in Figure 6 the following observations should be noted: For the best model we noted the misclassification probability for both the authentic images and forgeries separately. It turns out that for most  $\lambda$  values used in the lasso regression (6) all the authentic patches were classified correctly, whereas for the forgeries the minimum misclassification error is around 20 % and this is only achieved for a few  $\lambda$  values.



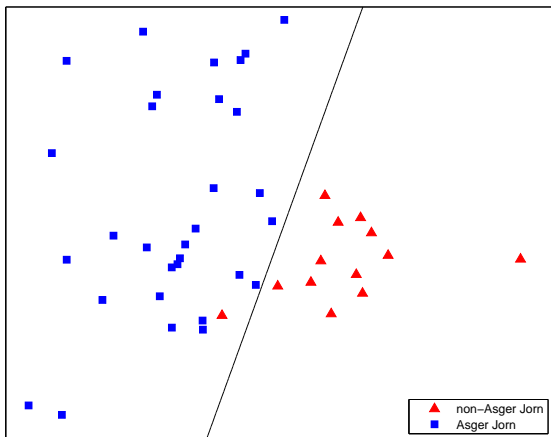
Directions	Free parameters
2 2 4 4 8 8	110
2 2 4 4 8	78
2 2 4 4	46
4 4 8 8 16 16	220
4 4 8 8 16	156
4 4 8 8	92
8 8 16 16 32 32	440
8 8 16 16 32	312
8 8 16 16	184

TABLE 1. The number of free parameters for the HMT model (1) with the specified number of directions on each pyramidal level.

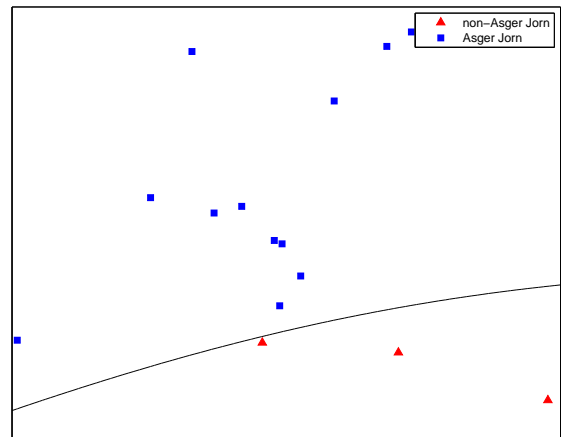
Furthermore, it is notoriously difficult to classify the images of the painting in Figure 1b by Helmut Sturm. When performing cross-validation, the painting by Helmut Sturm is misclassified for all the transform tested in Figure 6.

With the HMT model based on the contourlet transform with 8, 8, 16, 16 directions on the pyramidal levels we have visualized the decision boundary calculated from *all* of our available data, as described in Section 3.4.

The result for the Asger Jorn paintings is seen in Figure 7. Here we have presented separation for the individual photographs and for the individual paintings. We see that the authentic Asger Jorn paintings are well separated from the non-Asger Jorn paintings.



(A) Classification of the Asger Jorn images.

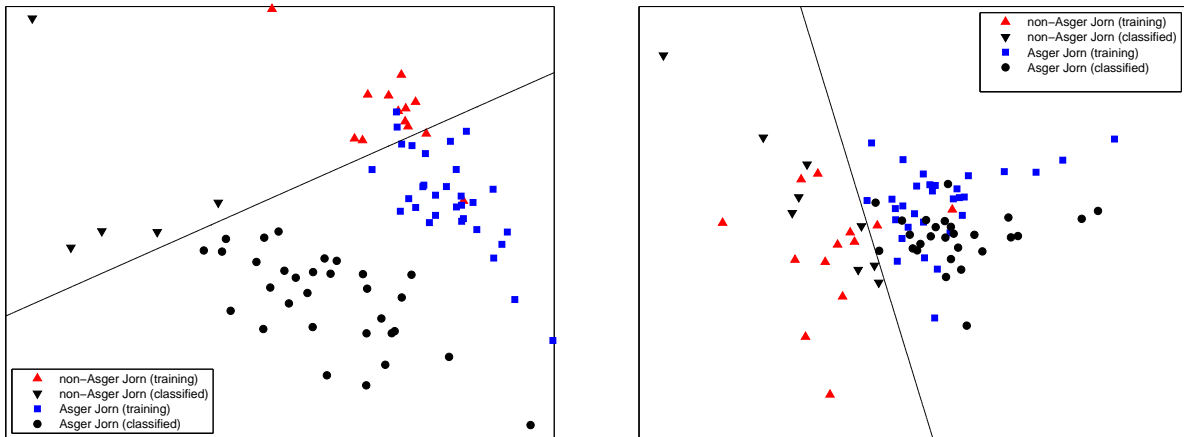


(B) Classification of the Asger Jorn paintings.

FIGURE 7. Classification of Asger Jorn's images from the multidimensional scaling configuration described in Section 3.4. One classification is based on the paintings – each of which is captured in several images. The other classification is of the individual images.

As mentioned in Section 2.3 other multiscale methods does not work well across different cameras. To test if our method is robust enough to classify images of the same paintings obtained with a different camera, we have used the images from our Nikon camera to classify the images obtained with the Canon camera. As it is seen in Figure 8 the models trained from images obtained with different cameras are comparable – if we have chosen the right zoom level.

With Figure 8b we correctly classified 32 out of 36 images from the Canon camera.



(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 (a) we used the original images for the training and in (a) 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.

**4.2. Results for Charlotte Caspers.** Our results for the images by Charlotte Caspers are far from satisfactory and except for one curiosity it is not worth elaborating the details here. In Figure 9 is the result of a MDS for all images by Caspers and the thing worth noticing here is that the copies and originals are clustered together and separated from the rest. The clustering tendency is more dominant in an interactive 3 dimensional MDS, but the main point is still distinct in the 2 dimensional MDS.

Also, the number of nonzero coefficients in the regression is very high for the Caspers images.

## 5. DISCUSSION

We have shown that the HMT provides an adequate description of the contourlet coefficients from the pictures we work with of paintings. By modelling the class of a

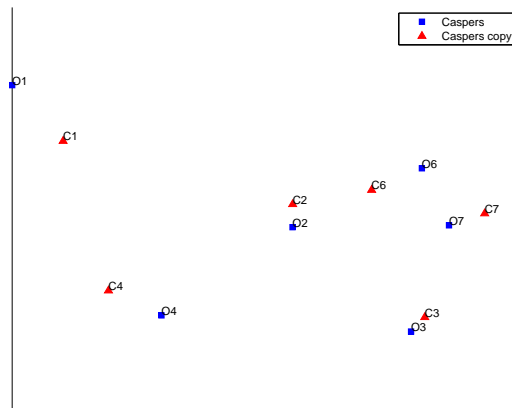


FIGURE 9. Visualization of the images by Charlotte Caspers as explained in Section 3.4. The originals and copies are denoted by  $O\#$  and  $C\#$ , respectively, with  $\# = 1, \dots, 7$ .

painting by a logistic model we obtain two ways of determining the affiliation of a new painting, namely (1) estimation of the probability that a patch belongs to a certain class via (4), and (2) applying a classification algorithm on the MDS computed from the dissimilarities in (11).

As for the method 1), we can obtain fairly good overall performance, as presented in Figure 6. However, as demonstrated by the separate misclassification probabilities for the authentic Asger Jorn paintings and the images made by his apprentices, this method does have its drawbacks. Correct classification of the authentic Asger Jorn paintings seems to be an easy task for the model, but correct classification of the non-Asger Jorn paintings is more challenging. The reason for this is that the non-Asger Jorn paintings is not a well-defined class, since it contains every painting not made by Asger Jorn. When determining which class is more likely for a new painting to belong to, the model compares the features of the painting with the features of the different classes. It is not certain that a new non-Asger Jorn paintings shares the features of the so far observed non-Asger Jorn paintings and it is also unrealistic to model this class.

One drawback of method 2) is that it is necessary to find a new configuration of points and compute a new decision boundary every time a new painting needs to be tested.

Method 2) is better than method 1) in the sense that it aims at clustering the points that share dominant features with authentic training images instead of distributing probability mass between two (or more) classes.

To illustrate this we have classified the peppers image seen in Figure 10 with both methods. The peppers image should not resemble Asger Jorn's image nor the ones made by apprentices.

Using (4) to predict the category of the peppers image, we get that it is fake with probability 0.82. But as it is seen from Figure 11, the peppers image is an outlier of the authentic Asger Jorn images.

A potential problem with our method(s) is seen in Figure 9, where the images cluster according to the content of the image and/or the canvas rather than according to the



FIGURE 10. The peppers image is used to demonstrate how our classifications work on a picture completely different from the training images.

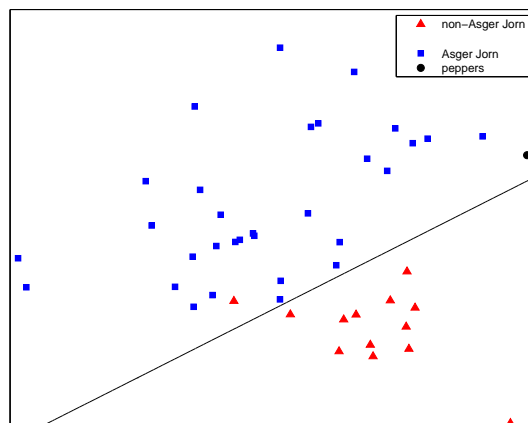


FIGURE 11. The peppers image is used to demonstrate how our classifications work on a picture completely different from the training images.

class. A reasonable explanation to this behavior is that the materials induce systematic patterns that the contourlet transform captures.

An interesting observation is that our classification methods depends on the zoom level of the images. The effect of data collection and preprocessing of the digital images is an issue that should be addressed further.

## 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 painter, which is believed to be very characteristic. The multiscale contourlet transformation is well described by a hidden Markov tree, which captures both the coefficient distribution on the individual scales and the dependency structure between the scales.

By applying a lasso regression we are able to link the parameters of the hidden Markov trees with the class of the each painting. We have utilized this in two classification methods. We can either get the probability of affiliation for each class based on the likelihood between the test image and the known representatives of each class. Alternatively we use the regression coefficients to define a metric between patches as a weighted euclidean distance and collect the distances between the patches of an image to a distance between images. When we know the pairwise distances between every image, we can find a configuration of points in a low dimensional space whose pairwise distances resembles the distances between the images. If the different classes from the training data in such a configuration are separated, we can use a suitable classification method to calculate decision rules that can be used on test data.

A promising aspect of our method is that we – with reasonable success – can determine the authenticity of images digitized differently than the training data.

#### ACKNOWLEDGEMENT

The authors would like to thank Arne Jensen (Department of Mathematical Sciences, Aalborg University) for suggesting this topic and for providing contact to the Portinari Project. Furthermore, we thank the Portinari Project for allowing us to use their paintings and Museum Jorn, Silkeborg, Denmark, for allowing us to use paintings of Asger Jorn.

Furthermore, we would like to thank P. Svante Eriksen (Department of Mathematical Sciences, Aalborg University) for helpful discussions and valuable comments on this manuscript.

#### REFERENCES

- [1] The Portinari Project. <http://www.portinari.org.br>.
- [2] Igor Bereznoy, Eric Postma, and H. Jaap van den Herik. Computer analysis of Van Gogh's complementary colours. *Mach Vision Appl*, 20:1–8, 2009.
- [3] Igor E. Bereznoy, Eric O. Postma, and H. Jaap van den Herik. Automatic extraction of brushstroke orientation from paintings. *Mach Vision Appl*, 20:1–9, JAN 2009.
- [4] Jr. C. Richard Johnson, Ella Hendriks, Igor J. Bereznoy, Eugene Brevdo, Shannon M. Hughes, Ingrid Daubechies, Jia Li, Eric Postma, and James Z. Wang. Image processing for artist identification. *IEEE Signal Process Magazine*, 25(4):37–48, 2008.
- [5] Emmanuel Candès, Laurent Demanet, David Donoho, and Lexing Ying. Fast discrete curvelet transform. *Multiscale Modeling & Simulation*, 5:861–899, 2006.
- [6] Emmanuel J. Candès and David L. Donoho. Continuous curvelet transform - II. Discretization and frames. *Applied and Computational Mathematics*, 19:198–222, 2005.
- [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. *IEEE Trans. Signal Process.*, 46(4):886–902, April 1998.
- [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. *IEEE Trans. Image Process.*, 14(12):2091–2106, DEC 2005.
- [11] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. GLMnet for Matlab. <http://www-stat.stanford.edu/~tibs/glmnet-matlab>.
- [12] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*, 33, 2010.

- [13] Trevor Hastie, Rob Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning, 2. edition*. Springer, 2008.
- [14] Richard J. Hathaway. *Constrained maximum-likelihood estimation for a mixture of  $M$  univariate normal distributions*. PhD thesis, Rice University, 1983.
- [15] 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. *Proc. Natl. Acad. Sci.*, 107:1279–1283, 2009.
- [16] Jia Li and James Z. Wang. Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Trans. Image Process.*, 13:338–351, 2004.
- [17] Siwei Lyu, Daniel N. Rockmore, and Hany Farid. A digital technique for art authentication. *Proc. Natl. Acad. Sci.*, 101(49):17006–17010, December 2004.
- [18] Duncan D.-Y. Po and Minh N. Do. Directional multiscale modeling of images using the contourlet transform. *IEEE Trans. Image Process.*, 15(6):1610–1620, JUN 2006.
- [19] Gngr 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.
- [20] Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.
- [21] George Arthur Frederick Seber. *Multivariate Observations*. Wiley, New York, 1984.