PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Zhang S, Lin G, Tindel S. 2022 Two-dimensional signature of images and texture classification. *Proc. R. Soc. A* **478**: 20220346.
https://doi.org/10.1098/rspa.2022.0346

Received: 21 May 2022
Accepted: 6 September 2022

Subject Areas:
applied mathematics, statistics

Keywords:
rough paths theory, signature, image classification

Author for correspondence:
Guang Lin
e-mail: guanglin@purdue.edu

THE ROYAL SOCIETY PUBLISHING

Two-dimensional signature of images and texture classification

Sheng Zhang[1], Guang Lin[2] and Samy Tindel[3]

[1]Department of Mathematics, [2]Department of Mathematics and School of Mechanical Engineering, and [3]Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA

SZ, 0000-0003-1539-0454; GL, 0000-0002-0976-1987

We introduce a proper notion of two-dimensional signature for images. This object is inspired by the so-called rough paths theory, and it captures many essential features of a two-dimensional object such as an image. It thus serves as a low-dimensional feature for pattern classification. Here, we implement a simple procedure for texture classification. In this context, we show that a low-dimensional set of features based on signatures produces an excellent accuracy.

1. Introduction

Signatures of paths are fascinating objects which have been popularized by K-T Chen in an algebraic context [1]. The topic has then resurfaced as an essential tool in stochastic calculus, thanks to Lyons [2]. It has now been the object of intense scrutiny for the past 25 years. In this introduction, we will first put our own investigation into context, summarizing some of the contributions for signatures of paths indexed by a one-dimensional parameter. This will be the content of §1a. Then in §b, we introduce the concept of signature for paths indexed by a two-dimensional parameter. This is a natural framework for image processing. Eventually, we will summarize our main results and draw some conclusions.

(a) One-dimensional signatures

Before going further, let us mention that the *1-d* in one-dimensional signatures refers to the fact that the paths x we are considering for now are indexed by a one-dimensional parameter $t \in [0, \sigma]$ for a given time horizon $\sigma > 0$ (as opposed to the two-dimensional parameter $(s;t) \in [0,\sigma] \times [0,\tau]$ considered in the next sections). Nevertheless, the path x is generally $\mathbb{R}^d$-valued, that is $x_t \in \mathbb{R}^d$ for all $t \in [0,\sigma]$.

Signatures of paths are prominent objects in analysis and data science due to several remarkable properties: they show up naturally in fundamental computations, they enjoy suitable algebraic and analytic relations, they characterize paths, and they have been successfully applied in a data analysis context. Let us briefly review those features.

(i) *Prominence of signatures in system approximations.* One of the simplest situations in which signatures pop up in a very natural way is through basic change of variables formulae. Namely consider a regular enough C^1-path (or curve) $x:[0,\sigma]\to\mathbb{R}^d$ and a smooth function $f:\mathbb{R}^d\to\mathbb{R}$. We denote by x^i each component of x, so that $x=(x^1,\dots,x^d)$. For the sake of notation, we also write $\partial_i f$ for the partial derivative $\partial f/\partial x_i$. Then the most basic form of a change of variable formula asserts that for $0\le s<\hat{s}\le\sigma$, we have

$$f(x_{\hat{s}})-f(x_s)=\sum_{i_1=1}^{d}\int_s^{\hat{s}}\partial_{i_1}f(x_{r_1})\,\mathrm{d}x_{r_1}^{i_1}. \tag{1.1}$$

In order to get further expansions according to relation (1.1), let us introduce the first two elements of the signature of x. They are defined as iterated integrals of x as follows:

$$\mathbf{x}_{s\hat{s}}^{\mathbf{1},i_1}=x_{\hat{s}}^{i_1}-x_s^{i_1}=\int_{s<r_1<\hat{s}}\mathrm{d}x_{r_1}^{i_1}\quad\text{and}\quad\mathbf{x}_{s\hat{s}}^{\mathbf{2},i_1,i_2}=\int_{s<r_1<r_2<\hat{s}}\mathrm{d}x_{r_1}^{i_1}\,\mathrm{d}x_{r_2}^{i_2}.$$

With this notation in hand and iterating formula (1.1), we get the following approximation:

$$f(x_{\hat{s}})-f(x_s)\simeq\sum_{i_1}\partial_{i_1}f(x_s)\,\mathbf{x}_{s\hat{s}}^{\mathbf{1},i_1}+\sum_{i_1,i_2}\partial_{i_1,i_2}^2f(x_s)\,\mathbf{x}_{s\hat{s}}^{\mathbf{2},i_1,i_2}. \tag{1.2}$$

As one can see from relation (1.2), the elements $\mathbf{x}^{\mathbf{1}},\mathbf{x}^{\mathbf{2}}$ play the role of monomials in a Taylor type expansion along the path x. They can thus be thought of as building blocks for a faithful representation of the path x.Extrapolating on this kind of consideration, the signature of x summarizes all the iterated integrals of x in a single object $[S(x)]_{s\hat{s}}$ which can be written as

$$[S(x)]_{s\hat{s}}=1+\sum_{n=1}^{\infty}\int_{s<r_1<r_2<\cdots<r_n<\hat{s}}\mathrm{d}x_{r_1}\otimes\mathrm{d}x_{r_2}\otimes\cdots\otimes\mathrm{d}x_{r_n}. \tag{1.3}$$

For a given couple $s,\hat{s}$ with $s<\hat{s}$, the element $[S(x)]_{s\hat{s}}$, called the signature of x, lies in the space $T(\mathbb{R}^d)=\oplus_{n=0}^{\infty}(\mathbb{R}^d)^{\otimes n}$. It should be mentioned that $S(x)$ also appears very naturally when computing Taylor type expansions of ordinary differential equations. This fundamental property is the one that made signatures the central object in rough paths analysis. Rough paths can be seen as a new point of view on stochastic differential equations, and had a profound impact on stochastic analysis over the past two decades. Proper generalizations of rough paths are also at the heart of the celebrated regularity structure theory [3].

(ii) *Algebraic and analytic properties.* The fact that $S(x)$ belongs to the free algebra $T(\mathbb{R}^d)$ induces very convenient properties for algebraic manipulations. To name just a few, one can prove that if $x\sqcup y$ denotes the concatenation of two paths x and y, then (see e.g. [4], Theorem 7.11) we have

$$S(x\sqcup y)=S(x)\otimes S(y), \tag{1.4}$$

where the product in the right-hand side of (1.4) is the polynomial type product on $T(\mathbb{R}^d)$. Invariance by reparametrization also holds. More specifically, if $\phi:[0,\sigma]\to[0,\sigma]$ is

a non-decreasing surjection and if we set $x^\phi = x \circ \phi$, then for all $0 \le s < \hat{s} \le \sigma$ the following holds true:

$$[S(x)]_{\phi(s)\phi(\hat{s})} = S(x^\phi)_{s\hat{s}}. \tag{1.5}$$

On the analytic side, the essential bound on signatures asserts a factorial decay with respect to the order of the integral. Namely, if $S_n(x)$ denotes the n-th order integral in (1.3), then we have the following upper bound:

$$||S_n(x)|| \le \frac{(C_{\tau,x})^n}{n!}, \tag{1.6}$$

where the constant $C_{\tau,x}$ does not depend on n. Note that relation (1.6) yields crucial bounds on differential equations driven by x, which can be then extended to stochastic cases.

(iii) *Characterization of paths.* One of the most fundamental properties of signatures (especially with data analysis in mind) is that they characterize paths. Specifically, for two Lipschitz paths x, y, we have

$$S(x)_{01} = S(y)_{01} \quad \text{iff } x \sim y, \tag{1.7}$$

where $x \sim y$ means that x, y only differs by a tree-like path. This result is proved in [5], while [6] provides an algorithm allowing to reconstruct a path from its signature. Along the same lines, any continuous map $f : C^1([0,\sigma]) \to \mathbb{R}$ can be approximated by a linear functional of the signature. This point of view leads to more quantitative versions of (1.7), see [7].

(iv) *Signatures and data analysis.* Due to the algebraic, analytic and characterization properties recalled above, signatures have been recently used as efficient features in data analysis for paths. The literature on this topic is now abundant. Among those contributions, let us single out the very successful Chinese character recognition algorithm [8]. Other significant applications include finance time series [9], topological data analysis [10] and diagonosis prediction [11]. It is fair to claim that signatures are now accepted as an efficient set of features for classification or prediction of paths.

(b) Two-dimensional signatures

Although there are now promising steps in the analysis of signatures for fields indexed by m-dimensional rectangles, this area of research is still in its infancy. Following the structure of §1a, let us recall what has been done in case of fields indexed by $[0,\sigma] \times [0,\tau]$.

(i) *Signatures for calculus in the plane.* Let us start from the equivalent of relation (1.1) in the plane. Namely, consider a C^2-field $x : [0,\sigma] \times [0,\tau] \to \mathbb{R}^d$ and a smooth function $f : \mathbb{R}^d \to \mathbb{R}^d$. In order to write more compact equations below, we will use the following conventions:

$$\mathrm{d}_{12} x^i_{s;t} = \frac{\partial^2 x^i_{s;t}}{\partial s \partial t}\, \mathrm{d}s\mathrm{d}t \quad \text{and} \quad \mathrm{d}_{\hat{1}\hat{2}} x^{ij}_{s;t} = \frac{\partial x^i_{s;t}}{\partial s}\frac{\partial x^j_{s;t}}{\partial t}\, \mathrm{d}s\, \mathrm{d}t. \tag{1.8}$$

Also recall that we write $\partial_i f$ for the partial derivatives of f. In addition, increments like $f(x_{\hat{s}}) - f(x_s)$ in (1.1) should be replaced by rectangular increments. Specifically, for a rectangle $[s,\hat{s}] \times [t,\hat{t}]$ and a field y defined on $[0,\sigma] \times [0,\tau]$, we write

$$\square_{s\hat{s};t\hat{t}}\, y = y_{\hat{s};\hat{t}} - y_{s;\hat{t}} - y_{\hat{s};t} + y_{s;t}. \tag{1.9}$$

Then the equivalent of (1.1) in rectangles gives a change of variables formula for the rectangular increments of $y = f(x)$. It reads

$$\square_{s\hat{s};t\hat{t}} f(y) = \sum_{i,j=1}^{d} \int_{s<s_1<\hat{s}} \int_{t<t_1<\hat{t}} \partial^2_{ij} f(x_{s_1;t_1})\, \mathrm{d}_{1\hat{2}} x^{ij}_{\hat{s}_1;\hat{t}_1} + \sum_{i=1}^{d} \int_{s<s_1<\hat{s}} \int_{t<t_1<\hat{t}} \partial_i f(x_{s_1;t_1})\, \mathrm{d}_{12} x^{i}_{\hat{s}_1;\hat{t}_1}. \tag{1.10}$$

With respect to (1.1), the appearance of the term $\mathrm{d}_{1\hat{2}}x$ in (1.10) is obviously a major difference. As a result when one tries to push forward Taylor type expansions up to order 2 like in (1.2), the number of terms explodes (much faster than in the already factorial regime observed in the one-dimensional parameter case). A description of those terms is given in [12,13] for second-order calculations, but we are not even aware of extensions up to an arbitrary order. Therefore, obtaining an expression for a two-dimensional signature $S(x)$ similar to (1.3) is still a challenging open problem. Let us also mention that the stochastic calculus for plane-indexed processes does not fit into the general regularity structure framework [3], due to some boundary type singularities.

(ii) *Algebraic and analytic properties.* Since the very definition of $S(x)$ for paths indexed by (s,t) is still open, the literature on its algebra is obviously scarce. The recent preprint [14] gives an account on the algebraic structure generated by increments which can be roughly described as

$$\int_{s<s_1<\cdots<s_n<\hat{s}} \int_{t<t_1<\cdots<t_n<\hat{t}} \mathrm{d}_{1\hat{2}} x_{s_1;t_1} \otimes \cdots \otimes \mathrm{d}_{1\hat{2}} x_{s_n;t_n}, \tag{1.11}$$

together with some correcting terms involving Jacobians, and where the domain of integration is in fact based on permutations of the simplex displayed in (1.11). Some potential generalizations of (1.4)–(1.6) are also provided in [14]. In particular, the invariance property (1.5) is restricted to coordinate-wise changes of variables of the form

$$\phi(s,t) = \phi_1(s)\phi_2(t), \tag{1.12}$$

with two non-decreasing surjections ϕ_1 and ϕ_2.

(iii) *Characterization of paths.* Here again, only partial information for two-dimensional signatures is available. However, ([14], Theorem 6) is encouraging. Indeed this theorem claims that signatures generated by (1.11) do characterize paths, albeit in a topological weak sense.

(iv) *Signature and data analysis.* An RGB image can be fairly well represented by a field $x\colon [0,\sigma]\times[0,\tau]\to\mathbb{R}^3$, where each $x_{s;t} = (x^1_{s;t}, x^2_{s;t}, x^3_{s;t})$ represents a pixel and every coordinate stands for a fundamental colour (say $1=$ red, $2=$ green, $3=$ blue). Nevertheless, to the best of our knowledge, two-dimensional signatures have not been used as features in image processing or other multiparametric data analysis problems. The need for nonlinear functionals of the field x in texture classification was acknowledged in the influential paper [15], leading to a wide variety of generalizations. One can also mention the recent contribution [16], where regularity structures-based features are used for prediction purposes. However, the fact that two-dimensional signatures are natural objects to consider for image processing is not mentioned in those two references.

(c) Outline

With the above preliminaries in mind, our objective can be summarized as follows: we wish to show empirically that two-dimensional signatures are natural and efficient low-dimensional features for image processing. We propose to achieve this by considering a simple texture identification problem. We will see that considering a 12-dimensional feature and applying

standard classification methods, one can achieve an excellent accuracy. This certainly calls for further developments, which will be highlighted in our concluding remarks.

Our paper is structured as follows: in §2, we give a mathematical description of the signatures we are using for classification purposes, together with their discretized versions. Section 3 focuses on the numerical experiment, as well as the outcome in terms of accuracy for our classification task. We finish the paper with some concluding remarks in §4.

2. Description of the two-dimensional signatures

This section is devoted to introduce the features we advocate for in this paper. In §2a, we define those objects in the (continuous) plane, while §2b focuses on the corresponding discrete objects.

(a) Features in continuous space

The integrals considered below are based on a notion of simplex in the plane. The first of these objects is simply a rectangle of the form $[s,\hat{s}]\times[t,\hat{t}]$. It is denoted below as

$$\mathcal{S}_1(s,\hat{s};t,\hat{t})=\{(s_1,t_1)\in\mathbb{R}^2, s<s_1<\hat{s}, t<t_1<\hat{t}\}. \tag{2.1}$$

Notation (2.1) can then be easily extended in order to describe the simplex used for second-order integrals. Specifically, we integrate over domains of the form

$$\mathcal{S}_2(s,\hat{s};t,\hat{t})=\{(s_1,s_2,t_1,t_2)\in\mathbb{R}^4;\ s<s_1<s_2<\hat{s}, t<t_1<t_2<\hat{t}\}. \tag{2.2}$$

With the above notation (2.1) in hand, the first-order increments used as features can be written as

$$\mathbf{x}^{(1,2);i_1}_{s\hat{s};t\hat{t}}=\int_{\mathcal{S}_1(s,\hat{s};t,\hat{t})} d_{12}x^{i_1}_{s_1;t_1} \quad\text{and}\quad \mathbf{x}^{(\hat{1},\hat{2});i_1}_{s\hat{s};t\hat{t}}=\int_{\mathcal{S}_1(s,\hat{s};t,\hat{t})} d_{\hat{1}\hat{2}}x^{i_1}_{s_1;t_1}. \tag{2.3}$$

Note that $\mathbf{x}^{(1,2);i_1}_{s\hat{s};t\hat{t}}$ above is simply the rectangular increment $\square_{s\hat{s};t\hat{t}}x$ introduced in (1.9). Taking into account the fact that the colour index i_1 lies in $\{1,2,3\}$, we get six features of order 1 defined by (2.3).

Let us now turn to the definition of second-order increments. Recalling notation (2.2), we now have several possibilities combing spatial differentials and colour indices. We get

$$\mathbf{x}^{(11,22);i_1,i_2}_{s\hat{s};t\hat{t}}=\int_{\mathcal{S}_2(s,\hat{s};t,\hat{t})} d_{12}x^{i_1}_{s_1;t_1}d_{12}x^{i_2}_{s_2;t_2} \quad\text{and}\quad \mathbf{x}^{(\hat{1}\hat{1},\hat{2}\hat{2});i_1,i_2}_{s\hat{s};t\hat{t}}=\int_{\mathcal{S}_2(s,\hat{s};t,\hat{t})} d_{\hat{1}\hat{2}}x^{i_1}_{s_1;t_1}d_{\hat{1}\hat{2}}x^{i_2}_{s_2;t_2}. \tag{2.4}$$

Remark 2.1. A wider variety of second-order increments is available, when one mixes the d_{12} and $d_{\hat{1}\hat{2}}$ differentials defined by (1.8). One can also decide to integrate in one direction only for some of the integrals. See ([13], p. 5 and p. 18) for an exhaustive first of necessary increments for a rough integration in the plane. From this long list of possible increments, we will also appeal to $\mathbf{x}^{1\hat{1};2\hat{2}}$ and $\mathbf{x}^{\hat{1}1;\hat{2}2}$, respectively, defined by

$$\mathbf{x}^{(1\hat{1},2\hat{2});i_1i_2}_{s\hat{s};t\hat{t}}=\int_{S_2(s,\hat{s};t,\hat{t})} d_{12}x^{i_1}_{s_1;t_1}d_{\hat{1}\hat{2}}x^{i_2}_{s_2;t_2} \quad\text{and}\quad \mathbf{x}^{(\hat{1}1,\hat{2}2);i_1i_2}_{s\hat{s};t\hat{t}}=\int_{S_2(s,\hat{s};t,\hat{t})} d_{\hat{1}\hat{2}}x^{i_1}_{s_1;t_1}d_{12}x^{i_2}_{s_2;t_2}. \tag{2.5}$$

Remark 2.2. In this article, we focus on low-dimensional features. We will thus only consider the increments in (2.4) for $i_1=i_2$. Specifically, the collection of first- and second-order increments used below will be

$$\{x^{(1,2);i}, x^{(\hat{1},\hat{2});i}, x^{(11,22);i,i}, x^{(\hat{1}\hat{1},\hat{2}\hat{2});i,i}, x^{(1\hat{1},2\hat{2});i,i}, x^{(\hat{1}1,\hat{2}2);i,i}; i=1,2,3\}. \tag{2.6}$$

We thus end up with an 18-dimensional feature space.

Remark 2.3. As mentioned in the introduction, the collection (2.6) of iterated integrals enjoys some complex algebraic properties. We refer to [12, Section 5.1] for an account on these relations. See also the aforementioned article [14].

Remark 2.4. Analytic properties are part of the appeal of signatures as features. Since our considerations in the current article are restricted to second-order increments, the factorial decay exhibited in relation (1.6) does not really make sense. However, assuming that x is a C^2-signal, it is readily checked that for all $0 \leq s < \hat{s} \leq \sigma$ and $0 \leq t < \hat{t} \leq \tau$, we have

$$|x^{(1,2),i}_{s\hat{s};t\hat{t}}| + |x^{(\hat{1},\hat{2}),i}_{s\hat{s};t\hat{t}}| \leq C_x|\hat{s} - s||\hat{t} - t| \tag{2.7}$$

and

$$|x^{(11,22),ii}_{s\hat{s};t\hat{t}}| + |x^{(\hat{1}\hat{1},\hat{2}\hat{2}),ii}_{s\hat{s};t\hat{t}}| + |x^{(1\hat{1},2\hat{2}),ii}_{s\hat{s};t\hat{t}}| + |x^{(\hat{1}1,\hat{2}2),ii}_{s\hat{s};t\hat{t}}| \leq C_x|\hat{s} - s|^2|\hat{t} - t|^2. \tag{2.8}$$

(b) Discretization procedure

An image is a collection of pixels. For notational sake, we will consider that it can be represented as a field indexed by $\mathbb{N}^2$ and we denote by $\tilde{x}$ the discrete quantities. Otherwise stated, we have

$$\tilde{x} = \{x^i_{k;l};\ k \in \{1, \dots, K\},\ l \in \{1, \dots, L\},\ i \in \{1, 2, 3\}\}, \tag{2.9}$$

for two constants $K, L \geq 1$. In this context, the discrete first-order increments are deduced from their continuous counterparts (2.3) as

$$\tilde{\mathbf{x}}^{(1,2);i_1}_{k\hat{k};l\hat{l}} = \sum_{k_1=k}^{\hat{k}-1}\sum_{l_1=l}^{\hat{l}-1} \square_{k_1,k_1+1;l_1,l_1+1}\, x^{i_1} \tag{2.10}$$

and

$$\tilde{\mathbf{x}}^{(\hat{1},\hat{2});i_1}_{k\hat{k};l\hat{l}} = \sum_{k_1=k}^{\hat{k}-1}\sum_{l_1=l}^{\hat{l}-1} (x^{i_1}_{k_1+1;l_1} - x^{i_1}_{k_1;l_1})(x^{i_1}_{k_1;l_1+1} - x^{i_1}_{k_1;l_1}). \tag{2.11}$$

As far as the equivalents of (2.4) are concerned, the discrete form of $\mathbf{x}^{(11,22)}$ is

$$\tilde{\mathbf{x}}^{(11,22);i_1i_2}_{k\hat{k};l\hat{l}} = \sum_{k_2=k}^{\hat{k}-1}\sum_{l_2=l}^{\hat{l}-1} \left(\sum_{k_1=k}^{k_2-1}\sum_{l_1=l}^{l_2-1} \square_{k_1,k_1+1;l_1,l_1+1}\, x^{i_1} \right) \square_{k_2,k_2+1;l_2,l_2+1}\, x^{i_2}. \tag{2.12}$$

For the sake of conciseness, the discrete expressions for the other increments in (2.6) are left to the patient reader. Let us just mention that central difference approximations will be applied when considering the first-order derivatives in $d_{\hat{1}\hat{2}}x$ (see expression (1.8)).

3. Numerical experiment

In this section, we illustrate the use of two-dimensional signatures as features for image texture classification. There are two ways to look at the set of features we are considering in this experiment:

(i) One uses classic principal component analysis (PCA) on the image pixels, then this data is enriched with the second-order increments in (2.6).
(ii) One directly uses all the increments in (2.6), and performs PCA on the first-order increments $\mathbf{x}^{(1,2);i}$ defined by (2.3).

Since the value of each pixel can be deduced from the boundary terms $\{x^i_{0;t}, x^i_{s;0}; s \in [0, \sigma], t \in [0, \tau]\}$ and the increments $\mathbf{x}^{(1,2);i}$, it is readily checked that the two approaches (i) and (ii) are equivalent.

PCA is one of the classic methods in image classification problems, successfully implemented for applications such as face classification. The technique using PCA for face classification was

named eigenfaces [17], whose basis set for all images is formed by principal components. In particular, the original training images may be represented by their projections on the principal components. Dimensionality reduction and feature extraction are achieved by keeping just the first few principal components, which have higher signal-to-noise ratio. Here, we extend this method to texture classification, enriching our data with the natural nonlinear features given by second-order signatures. We also use simple PCA features as a baseline. Let us highlight again the fact that principal components can be embedded in our signature analysis, when including both first- and second-order signatures in our set of features.

(a) Training and testing data

The 42 textures used for this experiment are depicted in figure 1. They are from a standard dataset called CuRRET: Columbia-Utrecht Reflectance and Texture Database [18]. We randomly sample (100×100)-sized images from each texture. Ten samples from every texture are used for training. As an illustration, the training data from the texture '21-lettuce leaf' are shown in figure 2. Now, we have $10 \times 42 = 420$ different images belonging to 42 different texture categories in the training data. Since the data are easy to generate, we use a larger size of testing data to more accurately evaluate the predictive performance of the machine learning model. Specifically, one hundred images from every texture are sampled for testing. In total, this yields $100 \times 42 = 4200$ testing data.

(b) Feature construction with symmetry

Feature construction is the process of extracting useful information from raw data. If properly designed, the extracted features are of low dimension while preserving most of the important information in the data. This process can improve the performance of machine learning algorithms. As mentioned above, a traditional way of image feature construction is PCA, which projects the data onto the first few principal components. Those components are encoded in the first-order features (2.3).

For our experiment, we use the two-dimensional signatures described in §2b to construct more features for image classification. For each image, we can calculate the discretized signatures using the above formulae (2.10)–(2.12). Since textures should be symmetric, i.e. independent of the orientation of the images, we design the features as the average of the signatures under eight different orientations, the isometric group generated by rotation of 90° and reflection. As an image has three channels (red, green and blue) and we are using four second-order signatures, we have in total 12 features constructed by second-order signatures for every image. We are also adding the first $N_1 \in [0, 40]$ principal components related to the first-order increments $\tilde{\mathbf{x}}^{(1,2);i_1}_{k\hat{k};l\hat{l}}$ in (2.10). In the end, we are thus resorting to $N_2 \in [12, 52]$ features in our study. In an image processing context, this should be thought of as a low-dimensional set of features.

(c) Training

We now assume that the training data of §3a (including the images and their corresponding class labels), as well as the features constructed in §3b, are given. We then train a random forest classifier using the joint features to reconstruct the corresponding class labels of the training data. The classifier will then output a class label for any new input image. A good classifier should obviously be able to distinguish properly between different classes.

(d) Testing

For each testing image, we calculate its features using the same formula as the training data. Then, the features are fed into the trained classifier to predict the class label. Next, we compare

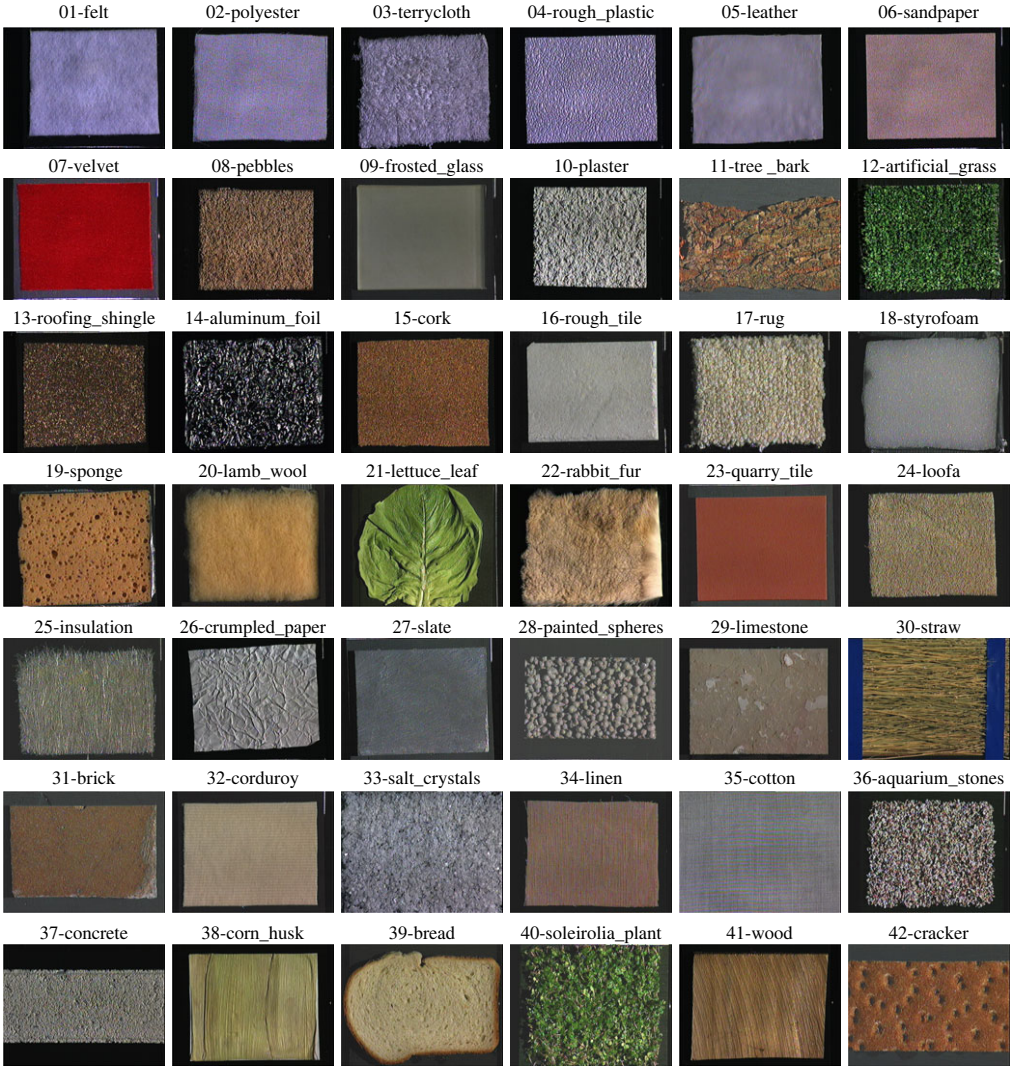

Figure 1. Forty-two different textures are used in this experiment. (Online version in colour.)

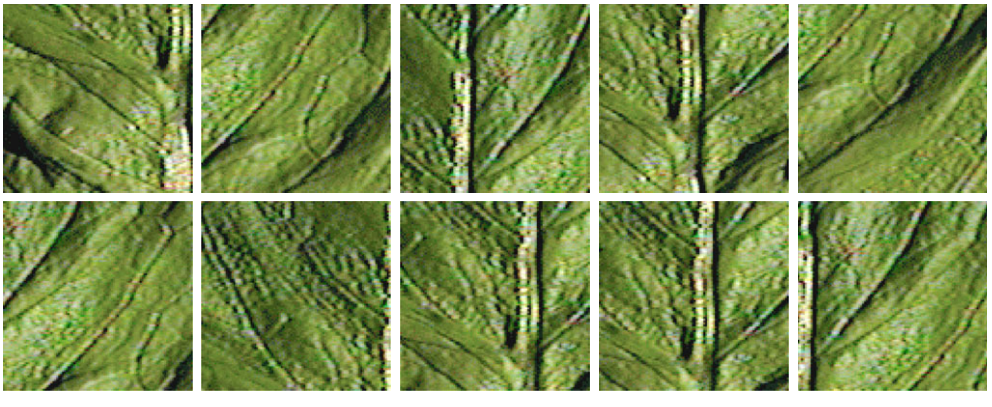

Figure 2. Ten samples from the texture '21-lettuce leaf' in figure 1. (Online version in colour.)

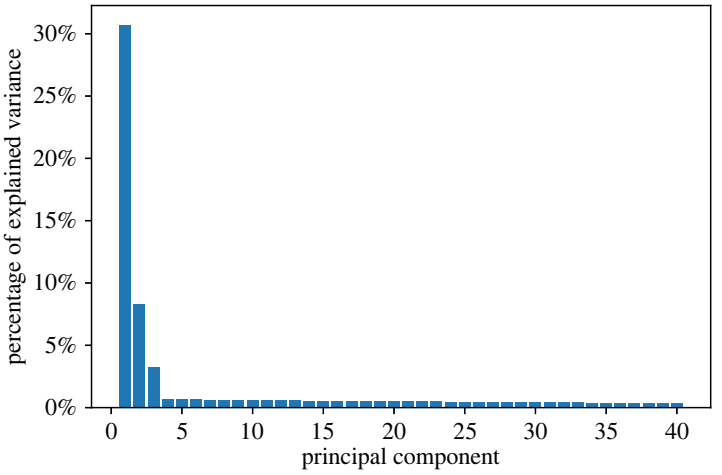

Figure 3. The percentage of explained variance of the first 40 principal components. Most of the variance is explained by the first three components. Principal component analysis is performed on the training data, where an image is reshaped into a vector. (Online version in colour.)

the predicted class labels with the true labels across all images in the testing data, and calculate the overall accuracy of the classification algorithm in identifying the true labels.

(e) Results

Our aforementioned image classification pipeline uses features generated by principal components as well as features constructed by signatures. The percentage of explained variance of the first 40 principal components is shown in figure 3. As one can see, most of the variance is explained by the first three components. Next, in order to visualize the testing data in the high-dimensional feature space, we use a nonlinear dimension reduction technique called *t*-distributed stochastic neighbour embedding (t-SNE) [19]. T-SNE embeds high-dimensional data into two-dimensional space, while trying to preserve local structures and relative distances of the data. See figure 4. The 15-dimensional feature vectors, generated by the first three principal components and 12 second-order signatures, are embedded into a two-dimensional space. Each point represents an image. The points of the same colour represent the images of the same class. As one can see in the figure, the images are well differentiated by this 15-dimensional set of features, even before the features are plugged into the classification algorithm.

The classification accuracy of the trained random forest classifiers on testing data is demonstrated in figure 5. We include accuracy under different feature sets. To be specific, we compare the accuracy with different numbers of principal components, as well as with and without features constructed by second-order signatures. We also contrast the classification task with or without symmetry in the construction of the signature features. Eventually, we compare the results under different sizes of training data, while keeping testing data the same. As one might expect, using more training data can improve the classification accuracy.

The numerical results suggest that when we use symmetric signatures and the first three principal components as features, best performance is attained. The 15-dimensional feature set yields close to 100% classification accuracy when there are just 10 training data in every class. As nonlinear features complementary to the linear features provided by PCA, second-order signatures can enhance the classification of textures. In fact, as the decay of performance when adding too many principal components suggests, not very useful features can negatively impact the machine learning classifier. On top of that, the procedure to average over eight orientations

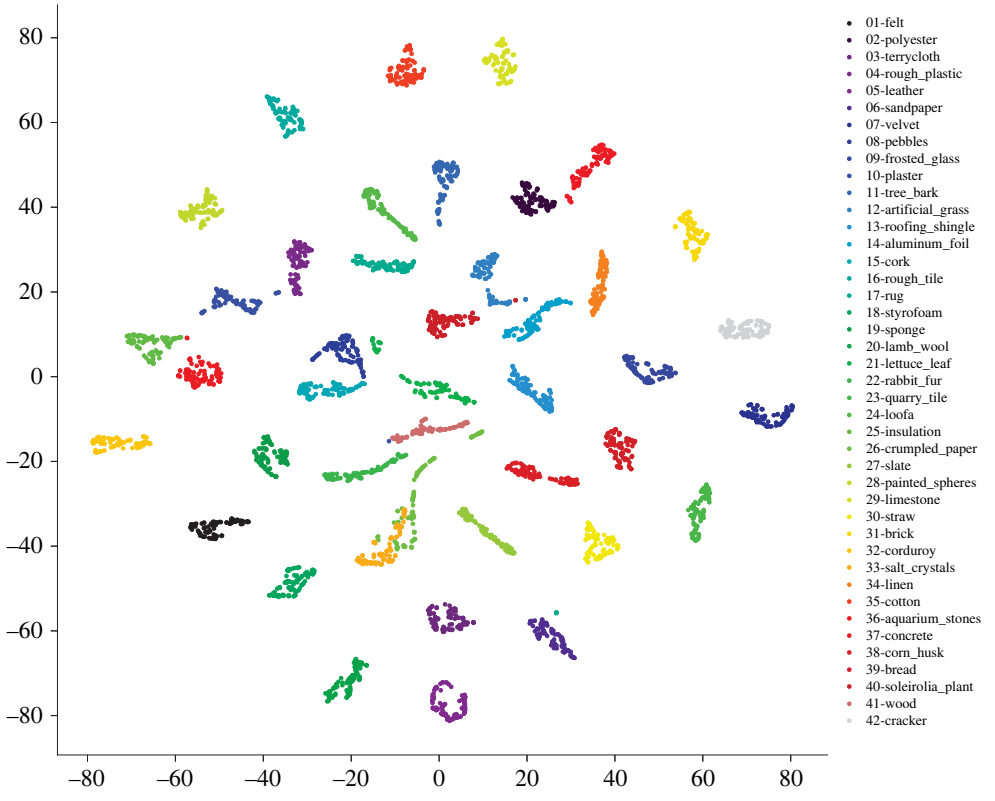

Figure 4. Two-dimensional *t*-distributed stochastic neighbour embedding of the 15-dimensional feature vectors of testing data. The features are generated by the first three principal components and 12 second-order signatures. Each point represents an image. The points of the same colour represent the images of the same class. The images are well differentiated by this 15-dimensional set of features, even before the features are plugged into the classification algorithm. (Online version in colour.)

when constructing features from signatures is helpful in improving the classification accuracy, and this is consistent under different sizes of training data.

(f) Discussions

The numerical experiment demonstrates that signatures are useful features for texture classification. The 15-dimensional feature vector constructed by first- and second-order signatures contains most of the texture information of an image (of $100 \times 100 \times 3 = 30\,000$ dimensions). PCA and random forest classifier are used here, but signatures as features can also be plugged into other machine learning pipelines such as neural networks. A complete comparison with benchmarks (including filter banks among others), as well as the strengths and weaknesses of signatures, is a research line to explore in the future.

All results in this paper take a total of a few minutes to run, coded in Python and using one core of Intel i7-6700HQ CPU. Time complexity and memory complexity of the signature calculation step are linear to the number of data (images), while the classification step might scale differently depending on what machine learning pipeline is used.

4. Conclusion

We have produced a low-dimensional set of features, based on first- and second-order signatures of two-dimensional indexed fields. Those objects stem naturally from elementary calculus in the

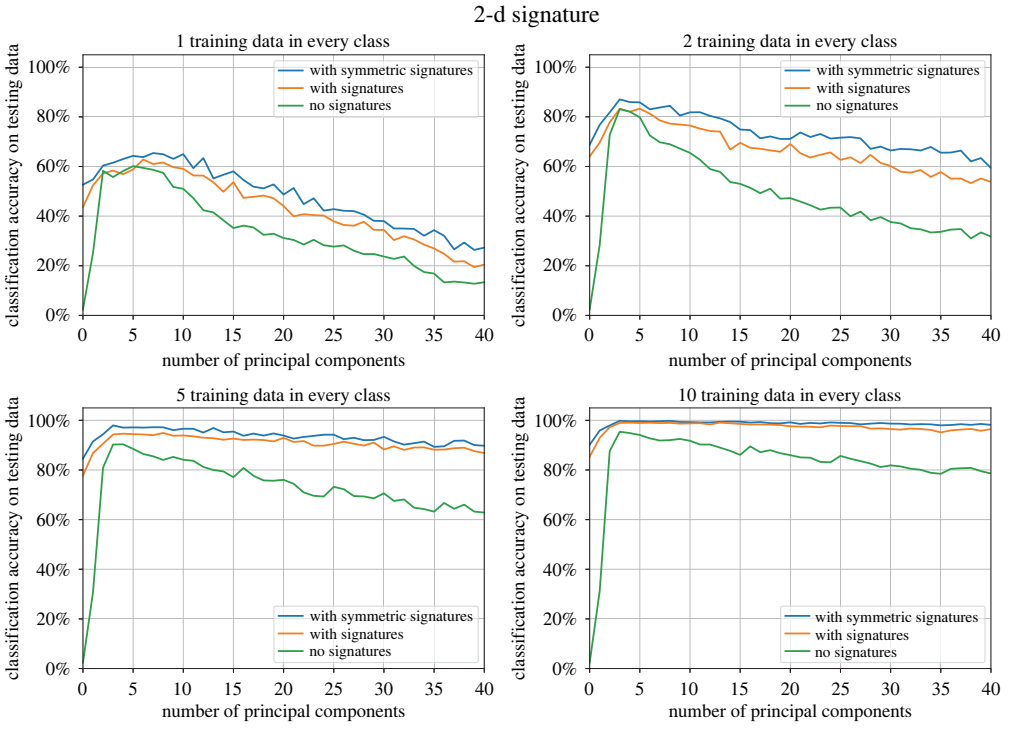

Figure 5. Classification accuracy of the random forest classifiers. When there is no signature or principal component, the accuracy is fixed at 1/42 since there is no feature and we are predicting 1 out of 42 classes. (Online version in colour.)

plane considerations. Therefore, they provide a very convenient set of parameters, expressed as nonlinear functionals of a field indexed by $\mathbb{R}^2$. In addition, we have shown that two-dimensional signatures yield excellent performances for texture classification in a concrete example.

In view of the above observations, two-dimensional and more generally higher dimensional signatures certainly deserve further investigation. Below is a non-exhaustive list of research lines we wish to explore:

(i) Increase the number of experiments, progressively asserting the validity of two-dimensional signatures as a meaningful set of characteristics for different categories of fields. Specific domains of application should include material science as well as civil engineering, where texture type features play a prominent role. We should also look at inserting our signature-based features into broader neural network procedures.

(ii) Preprocessing of signals often enhances the performance of classifiers. Natural explorations in this direction would be additions of lags or of components yielding parameterization invariance [20]. The introduction of kernels [21] could also lead to improvements.

(iii) Get a better grasp on the underlying algebraic structures related to two-dimensional signatures. This study should go beyond the investigation lead in [14] (restricted to the $\mathrm{d}_{\hat{1}\hat{2}}x$ differentials introduced in (1.8)), and will probably involve advanced structures such as Hopf algebras. One of the main objectives in this direction is to construct all signatures (up to a given order) in a systematic way. Invariances of signatures might also be encoded by group actions such as the ones exhibited in [14].

(iv) Generalizations of rough paths notions to fields indexed by $\mathbb{R}^d$ have recently given rise to breakthroughs in the definition of singular PDEs. Those advances have been achieved either in the landmark of regularity structures [3,22] or paracontrolled calculus

[23], and they all rely on convolution type iterated integrals. Since PDEs are known to be related to various fundamental problems in image processing (see [24]), it is likely that those iterated convolution integrals are also meaningful for image classification purposes. In particular, the paracontrolled approach hinges on Fourier modes, which could complement the direct modes feature introduced in (2.6).

As the reader can see, the signature method for image processing is a promising research direction, which deserves further investigations and improvements. Those questions will be addressed in subsequent publications.

Data accessibility. This article has no additional data.

Authors' contributions. S.Z.: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review and editing; G.L.: funding acquisition, investigation, project administration, resources, supervision, writing—review and editing; S.T.: conceptualization, formal analysis, funding acquisition, investigation, project administration, resources, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. S.Z. and S.T. are partially supported by National Science Foundation under grant DMS-1952966. S.Z. and G.L. are partially supported by National Science Foundation (DMS-1555072, DMS-1736364, DMS-2053746 and DMS-2134209), Brookhaven National Laboratory Subcontract (382247) and Department of Energy Office of Science Advanced Scientific Computing Research program (DE-SC0021142).

Acknowledgements. S.Z. and S.T. would like to thank Joscha Diehl, Leonard Schmitz and Fabian Harang for some valuable conversation about 2-d signatures and related objects. Our acknowledgements also go to Harald Oberhauser and Darrick Lee, for suggestions about future lines of investigation.

References

1. Chen KT. 1957 Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. Math.* **65**, 163–178. (doi:10.2307/1969671)
2. Lyons TJ. 1998 Differential equations driven by rough signals. *Rev. Math. Iberoam.* **14**, 215–310. (doi:10.4171/RMI/240)
3. Hairer M. 2014 A theory of regularity structures. *Inventiones Math.* **198**, 269–504. (doi:10.1007/s00222-014-0505-4)
4. Friz PK, Victoir NB. 2010 *Multidimensional stochastic processes as rough paths: theory and applications*, vol. 120. Cambridge studies in advanced mathematics, Cambridge, UK: Cambridge University Press. (doi:10.1017/CBO9780511845079)
5. Hambly B, Lyons T. 2010 Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. Math.* **171**, 109–167. (doi:10.4007/annals.2010.171.109)
6. Geng X. 2017 Reconstruction for the signature of a rough path. *Proc. Lond. Math. Soc.* **114**, 495–526. (doi:10.1112/plms.12013)
7. Kalsi J, Lyons T, Arribas IP. 2020 Optimal execution with rough path signatures. *SIAM J. Financ. Math.* **11**, 470–493. (doi:10.1137/19M1259778)
8. Graham B. 2013 Sparse arrays of signatures for online character recognition. (http://www.arxiv.org/abs/1308.0371)
9. Lyons T, Ni H, Oberhauser H. 2014 A feature set for streams and an application to high-frequency financial tick data. In *Proc. of the 2014 Int. Conf. on Big Data Science and Computing - BigDataScience'14*, pp. 1–8. New York, NY: ACM Press. (doi:10.1145/2640087.2644157)
10. Chevyrev I, Nanda V, Oberhauser H. 2020 Persistence paths and signature features in topological data analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 192–202. (doi:10.1109/TPAMI.2018.2885516)
11. Moore PJ, Lyons TJ, Gallacher J. 2019 Using path signatures to predict a diagnosis of Alzheimer's disease. *PLoS ONE* **14**, e0222212. (doi:10.1371/journal.pone.0222212)
12. Chouk K, Gubinelli M. 2014 Rough sheets. (http://www.arxiv.org/abs/1406.7748)
13. Chouk K, Tindel S. 2015 Skorohod and Stratonovich integration in the plane. *Electron. J. Probab.* **20**, 1–39. (doi:10.1214/EJP.v20-3041)

14. Giusti C, Lee D, Nanda V, Oberhauser H. 2022 A topological approach to mapping space signatures. (http://www.arxiv.org/abs/2202.00491)
15. SIfre L, Mallat S. 2014 Rigid-motion scattering for texture classification. (http://www.arxiv.org/abs/1403.1687)
16. Chevyrev I, Gerasimovics A, Weber H. 2021 Feature engineering with regularity structures. (http://www.arxiv.org/abs/2108.05879)
17. Turk MA, Pentland AP. 1991 Face recognition using eigenfaces. In *Proc. 1991 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 586–591. IEEE Comput. Sco. Press. (doi:10.1109/CVPR.1991.139758)
18. CUReT: Columbia-Utrecht Reflectance and Texture Database. Available from: www1.cs.columbia.edu/CAVE/software/curet/ (visited on 03 January 2022).
19. Van der Maaten L, Hinton G. 2008 Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605.
20. Toth C, Oberhauser H. 2020 Bayesian learning from sequential data using gaussian processes with signature covariances. In *Int. Conf. on Machine Learning*, pp. 9548–9560. PMLR.
21. Kiraly FJ, Oberhauser H. 2019 Kernels for sequentially ordered data. *J. Mach. Learn. Res.* **20**, 1–45.
22. Chen X, Deya A, Ouyang C, Tindel S. 2021 Moment estimates for some renormalized parabolic Anderson models. *Ann. Probab.* **49**, 2599–2636. (doi:10.1214/21-AOP1517)
23. Gubinelli M, Imkeller P, Perkowski N. 2015 Paracontrolled distributions and singular PDEs. In *Forum of Mathematics, Pi*, vol. 3, p. e6. Cambridge, UK: Cambridge University Press.
24. Aubert G, Kornprobst P. 2006 Mathematical problems in image processing: partial differential equations and the calculus of variations, with a foreword by Olivier Faugeras. vol. 147 of Applied Mathematical Sciences, 2nd edn. New York, NY: Springer. (doi:10.1007/978-0-387-44588-5)