ANSIG — An Analytic Signature for Permutation-Invariant Two-Dimensional Shape Representation *

José J. Rodrigues Pedro M. Q. Aguiar
João M. F. Xavier
Institute for Systems and Robotics / IST, Lisboa, Portugal
aguiar@isr.ist.utl.pt

Abstract

Many applications require a computer representation of 2D shape, usually described by a set of 2D points. The challenge of this representation is that it must not only capture the characteristics of the shape but also be invariant to relevant transformations. Invariance to geometric transformations, such as translation, rotation and scale, has received attention in the past, usually under the assumption that the points are previously labeled, i.e., that the shape is characterized by an ordered set of landmarks. However, in many practical scenarios the landmarks are obtained from an automatic process, e.g., edge/corner detection, thus without natural ordering. In this paper, we represent 2D shapes in a way that is invariant to the permutation of the landmarks. Within our framework, a shape is mapped to an analytic function on the complex plane, leading to what we call its analytic signature (ANSIG). We show that different shapes lead to different ANSIGs but that shapes that differ by a permutation of the landmarks lead to the same ANSIG, i.e., that our representation is a maximal invariant with respect to the permutation group. To store an ANSIG, it suffices to sample it along a closed contour in the complex plane. We further show how easy it is to factor out geometric transformations when comparing shapes using the ANSIG representation. We illustrate the ANSIG capabilities in shape-based image classification.

1. Introduction

Many objects are primarily recognized by their shape, rather than their color or texture. However, automatic shape-based classification has proved to be a very hard task, remaining an open problem, underlying which is the fundamental question of how to represent shape. This paper deals with the representation of two-dimensional (2D) shape. In our context, a shape is described by the coordinates of a set of points, or landmarks. We seek efficient ways to represent such sets, in particular we seek representations that are suitable to shape-based recognition. When the 2D shape is described by a set of *labeled* landmarks, there is an established theory that copes with geometric transformations and shape variations, the *statistical theory of shape* [15]. Although this theory has lead to significant results when the images to compare are characterized by feature points whose correspondences from image to image can be obtained, in many practical scenarios, such correspondences are not available. We thus focus on *unlabeled* data.

The majority of the methods that cope with unlabeled points focus on representing a "blob", *i.e.*, a shape that consists of a connected set of points. A number of techniques, usually called region-based, describe these shapes by using moment descriptors, *e.g.*, geometrical [12], Legendre [24], Zernike [24, 16], or Tchebichef [20]. Other approaches, called contour-based, represent the boundary of the shape using, *e.g.*, curvature scale space [19], wavelets [7], contour displacements [1], splines [9], or Fourier descriptors [25, 3]. Some of these representations exhibit desired invariance to geometric transformations but they are restricted to shapes well described by closed contours.

In image analysis, shape cues come primarily from image edges. Since, in general, it is hard to extract complete contours when dealing with real images, researchers developed local shape descriptors that, at each point of the shape, capture the relative distribution of the remaining points, *e.g.*, shape contexts [4] and distance multisets [10]. Although these local representations can deal with contour discontinuities, they do not deal with general shapes and geometric transformations. In this paper, we seek representations for shapes characterized by *arbitrary sets of points*.

A number of approaches to deal with shapes described by general sets of unlabeled points are motivated by the need to register the corresponding images, *i.e.*, to compute the rigid transformation that best "aligns" them. The majority of these registration methods are inspired by the fact

*To appear in CVPR'08, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, Alaska, June 2008.

that the rigid transformation is easily computed when the labels, *i.e.*, the point correspondences, are known (the Procrustes Matching problem). To cope with unlabeled points, they develop iterative algorithms that compute, in alternate steps, the rigid registration parameters and the point correspondences. One of the most widely known examples is the *Iterative Closest Point (ICP)* algorithm [5]. More recently, other researchers have proposed statistical methods that use "soft" correspondences [8, 18, 17], leading to *Expectation-Maximization (EM)*-like two-step iterative algorithms. Although these methods have succeed, even in challenging scenarios, *e.g.*, shape part decomposition [18] or nonrigid registration [8], they have the limitations of iterative algorithms, including the uncertain convergence and the sensitivity to initialization.

The relevance of being able to lead with unlabeled data in learning tasks, *i.e.*, the relevance of *permutation-invariant* representations, has been recently pointed out [13, 14, 23]. In these works, the invariance to permutation is explicitly constructed, *i.e.*, the permutation is factored out, after being computed as the solution of a convex optimization problem. However, this formulation does not deal with geometric transformations such as rotation. In this paper, rather than attempting to compute the permutation between two sets of points describing the shapes to compare, we propose a new permutation-invariant representation for sets of 2D points. We represent a 2D shape by what we call its *analytic signature* (ANSIG), an analytic function defined over the complex plane. We show that shapes that differ by a re-ordering of the set of landmarks have the same ANSIG. Thus, the ANSIG representation is permutation-invariant. Furthermore, we show that this representation enables discriminating different shapes, *i.e.*, different shapes are represented by different analytic functions.

Under our approach, shape-based classification boils down to comparing the ANSIG of a candidate shape with the ones in a database. As any analytic function, the ANSIG of a shape is completely described by the values it takes on a closed contour on the complex plane. Thus, we store an ANSIG by sampling it on the unit-circle. To compare ANSIGs, it suffices to measure the difference between the vectors collecting the corresponding samples.

The ANSIG representation is not invariant with respect to geometric transformations such as translation, rotation, and scale. However, an adequate pre-processing step factors out translation and scale. Although rotation can not be factored out, we show that the rotation that best aligns two shapes is easily obtained from their ANSIGs. In fact, the rotation that minimizes the above mentioned error measure is obtained in a computationally simple way by using FFTs.

In many situations, the points describing a shape are obtained automatically, *e.g.*, by detecting edges in an image. It is then common that two sets of points to compare, besides being noisy, have distinct cardinality (particularly when dealing with images of different sizes/resolutions). Our experiments with synthetic and real data show that the ANSIG representation is robust to these perturbations.

2. Permutation invariance: the analytic signature of a shape

We consider that a 2D shape is a set of n unlabeled points in the complex plane $\mathbb{C}$, thus described by a complex vector

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} x_1 + jy_1 \\ x_2 + jy_2 \\ \vdots \\ x_n + jy_n \end{bmatrix} \in \mathbb{C}^n. \tag{1}$$

However, as the ordering of the landmarks is irrelevant, the choice in (1) is not unique: the same shape is equivalently represented by any vector in the set

$$\{\boldsymbol{\Pi z} \,:\, \boldsymbol{\Pi} \in \Pi(n)\}, \tag{2}$$

where $\Pi(n)$ denotes the set of $n \times n$ permutation matrices.

Using the language of group theory, a shape is seen as a point in the quotient space $\mathbb{C}^n/\Pi(n)$. Here, $\Pi(n)$ is a group and matrix multiplication is the group operation:

$$\Pi(n) \times \mathbb{C}^n \ni (\boldsymbol{\Pi}, \boldsymbol{z}) \mapsto \boldsymbol{\Pi z} \in \mathbb{C}^n. \tag{3}$$

This group action induces a partition of $\mathbb{C}^n$ into disjoint orbits, where each orbit collects all the possible ways of representing the same shape vector, *i.e.*, it is the set in (2). Each shape corresponds to an orbit and the quotient space $\mathbb{C}^n/\Pi(n)$ is the set of orbits, *i.e.*, the space of shapes. The canonical map, also referred by the quotient map,

$$\pi \,:\, \mathbb{C}^n \to \mathbb{C}^n/\Pi(n)\,, \qquad \boldsymbol{z} \mapsto \pi(\boldsymbol{z}), \tag{4}$$

maps each point to its orbit. In our case, π maps each vector $\boldsymbol{z} \in \mathbb{C}^n$ to the shape it represents, $\pi(\boldsymbol{z}) \in \mathbb{C}^n/\Pi(n)$.

From the definition above, two vectors $\boldsymbol{z}$, $\boldsymbol{w} \in \mathbb{C}^n$ that are not related by a permutation, will be mapped by π to distinct points in the quotient space, *i.e.*, $\pi(\boldsymbol{z}) \neq \pi(\boldsymbol{w})$. This is usually referred to as a *maximal invariance* property, meaning that $\pi(\boldsymbol{z}) = \pi(\boldsymbol{w})$ if and only if $\boldsymbol{z}$ and $\boldsymbol{w}$ represent the same shape. Although this is a property we look for (the map π detects whether or not $\boldsymbol{z}$ and $\boldsymbol{w}$ represent the same shape), this mechanism is hardly implementable, due to the rather abstract nature of the quotient space and map.

Definition of analytic signature (ANSIG). We now develop a version of the objects introduced above, which is suitable for use in practice. We propose to replace the quotient map $\pi \,:\, \mathbb{C}^n \to \mathbb{C}^n/\Pi(n)$ by a map

$$a \,:\, \mathbb{C}^n \to \mathcal{A}, \quad \mathcal{A} := \{f \,:\, \mathbb{C} \to \mathbb{C} \,:\, f \text{ is analytic}\}. \tag{5}$$

As the quotient map π, our surrogate map a, which maps shape vectors in $\mathbb{C}^n$ to analytic functions on the complex plane, will exhibit maximal invariance with respect to the permutation group $\Pi(n)$. This means that a maps $\boldsymbol{z}$ and $\boldsymbol{w}$ to the same analytic function if and only if $\boldsymbol{z} = \boldsymbol{\Pi}\boldsymbol{w}$, for some permutation matrix $\boldsymbol{\Pi}$, *i.e.*, if and only if $\boldsymbol{z}$ and $\boldsymbol{w}$ describe the same shape. We refer to the analytic function $a(\boldsymbol{z}, \cdot) : \mathbb{C} \to \mathbb{C}$ as the *analytic signature* (ANSIG) of the shape described by $\boldsymbol{z}$.

In particular, we propose the ANSIG map $a : \mathbb{C}^n \to \mathcal{A}$, $\boldsymbol{z} \mapsto a(\boldsymbol{z}, \cdot)$, given by

$$a(\boldsymbol{z}, \xi) := \frac{1}{n} \sum_{m=1}^{n} e^{z_m \xi}, \tag{6}$$

where ξ is a dummy complex variable. We will see that the ANSIG just defined, exhibits properties that make it adequate for shape representation, in particular, the maximal invariance with respect to the permutation group.

Maximal invariance of the ANSIG. It is clear that invariance of a with respect to the permutation group holds. In fact, from the definition of the ANSIG map a in (6), it follows that $a(\boldsymbol{z}, \cdot) = a(\boldsymbol{w}, \cdot)$ whenever $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$ are related by a permutation, *i.e.*, whenever $\boldsymbol{z} = \boldsymbol{\Pi}(n)\boldsymbol{w}$.

To establish the maximal invariance, consider two vectors $\boldsymbol{z} = [\, z_1 \, z_2 \, \cdots \, z_n \,]^T$ and $\boldsymbol{w} = [\, w_1 \, w_2 \, \cdots \, w_n \,]^T$ that have equal ANSIGs, $a(\boldsymbol{z}, \cdot) = a(\boldsymbol{w}, \cdot)$. We will show that $\boldsymbol{z}$ and $\boldsymbol{w}$ represent the same shape, *i.e.*, that they differ only by a permutation of their entries. Since the analytic functions $a(\boldsymbol{z}, \cdot)$ and $a(\boldsymbol{w}, \cdot)$ are equal, their first n-th order derivatives at the origin also coincide:

$$\frac{d^k}{d\xi^k} a(\boldsymbol{z}, \xi)\bigg|_{\xi=0} = \frac{d^k}{d\xi^k} a(\boldsymbol{w}, \xi)\bigg|_{\xi=0}, \quad k = 1, 2, \ldots, n.$$

Using the definition of the ANSIG a in (6), this system of equations is written in terms of the entries of $\boldsymbol{z}$ and $\boldsymbol{w}$ as

$$\begin{cases} z_1 + z_2 + \cdots + z_n &= w_1 + w_2 + \cdots + w_n \\ z_1^2 + z_2^2 + \cdots + z_n^2 &= w_1^2 + w_2^2 + \cdots + w_n^2 \\ & \vdots \\ z_1^n + z_2^n + \cdots + z_n^n &= w_1^n + w_2^n + \cdots + w_n^n \,. \end{cases}$$

This set of equalities implies that the polynomials $p(t) = (t - z_1)(t - z_2)\cdots(t - z_n)$ and $q(t) = (t - w_1)(t - w_2)\cdots(t - w_n)$ are identical [11]. In particular, $p(t)$ and $q(t)$ share the same system of roots (including their multiplicities), thus the (multi-)sets $\{z_1, z_2, \ldots, z_n\}$ and $\{w_1, w_2, \ldots, w_n\}$ are equal. This way we conclude that the vectors $\boldsymbol{z}$ and $\boldsymbol{w}$ are equal, up to a permutation, thus proving the maximal invariance of our ANSIG map a in (6).

Storing the ANSIG. The ANSIG map $a : \mathbb{C}^n \to \mathcal{A}$ in (6) maps shape vectors to an analytic functions on the complex plane. As the space of analytic functions $\mathcal{A}$ is infinite-dimensional, our ANSIG may seems inadequate to a computer implementation. However, a major consequence of Cauchy's integral formula is that any analytic function f is unambiguously determined by the values it takes on a simple closed contour, see, *e.g.*, [2]. Choosing this contour to be the unit-circle $\mathbb{S}^1$ (the circle of radius 1 centered at the origin), the analytic function f is uniquely determined by $\{f(e^{j\varphi}) : \varphi \in [0, 2\pi]\}$. In practice, this is approximated by sampling, *i.e.*, by considering the values of f on a finite set of K points in the unit-circle, say $\{1, W_K, W_K^2, \ldots, W_K^{K-1}\}$, where

$$W_K := e^{j\frac{2\pi}{K}}. \tag{7}$$

In summary, we approximate the ANSIG map $a : \mathbb{C}^n \to \mathcal{A}$ in (6) by its discrete counterpart $\boldsymbol{a}_K : \mathbb{C}^n \to \mathbb{C}^K$, $\boldsymbol{z} \mapsto \boldsymbol{a}_K(\boldsymbol{z})$, given by

$$\boldsymbol{a}_K(\boldsymbol{z}) := \begin{bmatrix} a(\boldsymbol{z}, 1) \\ a(\boldsymbol{z}, W_K) \\ a(\boldsymbol{z}, W_K^2) \\ \cdots \\ a\left(\boldsymbol{z}, W_K^{K-1}\right) \end{bmatrix} \in \mathbb{C}^K. \tag{8}$$

3. Quotienting out shape-preserving geometric transformations

We now address how the ANSIG representation handles shape-preserving geometric transformations, such as translation, rotation and scale. As introduced in the previous section, a shape is an orbit generated by a group G of pertinent transformations. We started by considering the permutation group $G = \Pi(n)$ and provided a maximal invariant with respect to this group action, the ANSIG a in (6). In fact, the ANSIG map a is such that $a(\boldsymbol{z}, \cdot) = a(\boldsymbol{w}, \cdot)$ if and only if $\boldsymbol{z}$ and $\boldsymbol{w}$ differ by a permutation. Building on this result, we now enlarge the group G to also accommodate shape-preserving geometric transformations.

Two vectors $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$ represent the same shape if they are equal, up to, not only a permutation of their entries, but also a translation, rotation, and scale factor, affecting the set of n points they describe in the plane. To take this into account, we consider the group of transformations $G = \Pi(n) \times \mathbb{R}^+ \times \mathbb{C} \times \mathbb{S}^1$, defining the action of G on $\mathbb{C}^n$ as the map $G \times \mathbb{C}^n \to \mathbb{C}^n$,

$$\left(\boldsymbol{\Pi}, \lambda, v, e^{j\theta}\right) \cdot \boldsymbol{z} := \boldsymbol{\Pi}\lambda e^{j\theta}\boldsymbol{z} + v\mathbf{1}_n \,, \tag{9}$$

where $\mathbf{1}_n := [\, 1 \, 1 \, \cdots \, 1 \,]^T$ is the n-dimensional vector with all entries equal to 1. It is clear from (9) that the action of one element of the group G on a shape vector $\boldsymbol{z}$ corresponds

to a permutation ($\mathbf{\Pi}$), translation (v), rotation (θ), and scaling (λ), applied to the shape described by $\boldsymbol{z}$.

Obviously, deciding if two given vectors $\boldsymbol{z}$, $\boldsymbol{w}$ represent the same shape, *i.e.*, if they are in the same orbit, corresponds to checking if the value of the optimization problem

$$\min_{(\mathbf{\Pi},\lambda,v,e^{j\theta})\in G} \left\|\boldsymbol{z}-(\mathbf{\Pi},\lambda,v,e^{j\theta})\cdot\boldsymbol{w}\right\| \tag{10}$$

is zero. However, to the best of our knowledge, solving (10) requires an exhaustive search over the group $\Pi(n)$, which has cardinality $n!$. Clearly, this is not feasible, even for moderate values of the number of landmarks, say $n = 100$ ($100! \simeq 10^{158}$). In contrast, our approach is to use the ANSIG representation to devise a scheme that circumvents the combinatorial search. This way, in our experiments, we were able to process shapes described by very large sets of points, *e.g.*, with n up to 40000.

Translation and scale. We start by quotienting out all the transformations, except the rotation, through a map ϕ. Then, we show that ϕ is equivariant with respect to the rotations, what will enable a computationally simple scheme to detect equality of orbits. To factor out translation and scale, we make the simple pre-processing step of centering and normalizing the shape. This corresponds to considering the map $\phi : \mathbb{C}^n \to \mathcal{A}$, defined by

$$\phi(\boldsymbol{z},\cdot) := a\left(\sqrt{n}\frac{\boldsymbol{z}-\overline{\boldsymbol{z}}}{\|\boldsymbol{z}-\overline{\boldsymbol{z}}\|},\cdot\right), \tag{11}$$

where a is the ANSIG (6), $\overline{\boldsymbol{z}} := \frac{1}{n}\mathbf{1}_n^T\boldsymbol{z}\mathbf{1}_n$ is a vector with all entries equal to the mean value of $\boldsymbol{z}$, and $\|\cdot\|$ denotes the 2-norm, $\|\boldsymbol{v}\| := \boldsymbol{v}^H\boldsymbol{v}$, where $\boldsymbol{v}^H$ is the Hermitian (*i.e.*, conjugate) transpose of $\boldsymbol{v}$. The reader may wonder why the factor $\sqrt{n}$ in (11). In fact, this factor has nothing to do with factoring out translation or scale — it is constant for shapes described by the a fixed number of landmarks, n. The motivation for the factor $\sqrt{n}$ in (11) is precisely to enable dealing well with shapes described by different numbers of landmarks, providing robustness to over/under sampling the shapes.

Using the maximal invariance of the ANSIG a with respect to permutation, it is easy to show that the map ϕ in (11) is a maximal invariant with respect to permutation, translation, and scale, *i.e.*, that $\phi(\boldsymbol{z},\cdot) = \phi(\boldsymbol{w},\cdot)$ if and only if $\boldsymbol{z} = (\mathbf{\Pi},\lambda,v,1)\cdot\boldsymbol{w}$, for some $(\mathbf{\Pi},\lambda,v,1)\in G$. Thus, the map ϕ in (11) quotients out all transformations in G, except the rotation.

Rotation. We now show that, although the map ϕ is not invariant to rotations, it is equivariant. Let us begin by looking at how the rotation of a shape affects its ANSIG. From the definition of the ANSIG a in (6), it follows that

$$\begin{aligned} a(e^{j\theta}\boldsymbol{z},\xi) &= \frac{1}{n}\sum_{k=1}^{n} e^{e^{j\theta}z_k\xi} \\ &= \frac{1}{n}\sum_{k=1}^{n} e^{z_k\left(e^{j\theta}\xi\right)} \\ &= a(\boldsymbol{z},e^{j\theta}\xi)\,. \end{aligned} \tag{12}$$

Thus, the ANSIG of the rotated shape is a rotated version (in the complex plane) of the ANSIG of the original shape.

This property enables the derivation of the desired equivariance of ϕ, through the following chain of equalities:

$$\begin{aligned} \phi\left((\mathbf{\Pi},\lambda,v,e^{j\theta})\cdot\boldsymbol{z},\xi\right) &= \phi\left(\mathbf{\Pi}\lambda e^{j\theta}\boldsymbol{z}+v\mathbf{1}_n,\xi\right) && (13)\\ &= a\left(\sqrt{n}\frac{\mathbf{\Pi}e^{j\theta}(\boldsymbol{z}-\overline{\boldsymbol{z}})}{\|\boldsymbol{z}-\overline{\boldsymbol{z}}\|},\xi\right) && (14)\\ &= a\left(\sqrt{n}\frac{e^{j\theta}(\boldsymbol{z}-\overline{\boldsymbol{z}})}{\|\boldsymbol{z}-\overline{\boldsymbol{z}}\|},\xi\right) && (15)\\ &= a\left(\sqrt{n}\frac{\boldsymbol{z}-\overline{\boldsymbol{z}}}{\|\boldsymbol{z}-\overline{\boldsymbol{z}}\|},e^{j\theta}\xi\right) && (16)\\ &= \phi(\boldsymbol{z},e^{j\theta}\xi)\,, && (17) \end{aligned}$$

where: (13) comes from the definition of the group action (9); (14) uses the definition of the map ϕ (11); (15) results from the permutation invariance of the ANSIG; (16) comes from property (12); and (17) uses again the definition of ϕ.

Since the analytic function $\phi(\boldsymbol{z},\cdot)$ is univocally determined by the values it takes on a closed contour, *i.e.*, by its the restriction to the unit-circle, $\phi_{\mathbb{S}^1} : \mathbb{C}^n\times\mathbb{S}^1 \to \mathbb{C}$, $\phi_{\mathbb{S}^1}(\boldsymbol{z},\cdot) := \phi(\boldsymbol{z},\cdot)$, it will be useful to state the equivariance (17) in terms of this restriction:

$$\phi_{\mathbb{S}^1}\left((\mathbf{\Pi},\lambda,v,e^{j\theta})\cdot\boldsymbol{z},\xi\right) = \phi_{\mathbb{S}^1}\left(\boldsymbol{z},e^{j\theta}\xi\right). \tag{18}$$

This leads to the following short summary of our results: vectors $\boldsymbol{z}$ and $\boldsymbol{w}$ describe the same shape, *i.e.*, they differ by permutation, translation, rotation, and scale, if and only if

$$\phi_{\mathbb{S}^1}(\boldsymbol{z},\cdot) = \phi_{\mathbb{S}^1}\left(\boldsymbol{w},e^{j\theta}\,\cdot\right), \quad \text{for some } \theta\in[0,2\pi]. \tag{19}$$

4. Implementation

As referred in Section 2, a practical way to store the ANSIG $a : \mathbb{C}^n \to \mathcal{A}$ is through its unit-circle sampled version, $\boldsymbol{a}_K : \mathbb{C}^n \to \mathbb{C}^K$, introduced in (8). Thus, in practice, the map $\phi_{\mathbb{S}^1}$ is approximated by $\boldsymbol{\phi}_K : \mathbb{C}^n \to \mathbb{C}^K$, defined as

$$\boldsymbol{\phi}_K(\boldsymbol{z}) := a_K\left(\sqrt{n}\frac{\boldsymbol{z}-\overline{\boldsymbol{z}}}{\|\boldsymbol{z}-\overline{\boldsymbol{z}}\|}\right), \tag{20}$$

where K, the number of samples in the unit-circle, is chosen by the user. In all our experiments, we used $K = 512$.

When the rotation angle matches one of those of the samples in the unit-circle, *i.e.*, when $e^{j\theta} = W_K^k$, for some k, the equivariance of $\phi_{\mathbb{S}^1}$ in (18) gracefully transfers to $\boldsymbol{\phi}_K$:

$$\boldsymbol{\phi}_K\left(\left(\boldsymbol{\Pi}, \lambda, v, e^{i\theta}\right) \cdot \boldsymbol{z}\right) = \boldsymbol{\phi}_K(\boldsymbol{z}) \bmod k\,, \qquad (21)$$

where $\bmod k$ denotes a k-step cyclic shift. Thus, our test for deciding if two shape vectors $\boldsymbol{z}$ and $\boldsymbol{w}$ correspond to the same shape, expressed in (19), boils down to checking if $\boldsymbol{\phi}_K(\boldsymbol{w})$ is a cyclic-shifted version of $\boldsymbol{\phi}_K(\boldsymbol{z})$. Naturally, with K sufficiently large, $e^{j\theta} \simeq W_K^k$, for some k, and (21) holds for practical purposes. The shape similarity test can then be carried out by checking if the following error is below a small threshold:

$$\min_{k=0,1,\ldots,K-1} \left\|\boldsymbol{\phi}_K(\boldsymbol{z}) - \boldsymbol{\phi}_K(\boldsymbol{w}) \bmod k\right\|^2 . \qquad (22)$$

To evaluate (22), we must find the cilic-shift k^* that best "aligns" $\boldsymbol{\phi}_K(\boldsymbol{z})$ and $\boldsymbol{\phi}_K(\boldsymbol{w})$. Solving this by exhaustive search leads to an algorithm of complexity $\mathcal{O}(K^2)$. We reduce the complexity by using FFTs. Denote the Discrete Fourier Transform (DFT) of $\boldsymbol{v} \in \mathbb{C}^K$ by $\widehat{\boldsymbol{v}} \in \mathbb{C}^K$,

$$\widehat{\boldsymbol{v}} := \boldsymbol{D}_K^H \boldsymbol{v}\,, \qquad (23)$$

where $\boldsymbol{D}_K$ is the $K \times K$ DFT matrix, see, *e.g.*, [21]. Using the facts that the DFT is an unitary operator ($\boldsymbol{D}_K^H \boldsymbol{D}_K = \boldsymbol{I}_K$) and that the DFT of a k-cyclic-shifted vector equals the DFT of the original vector, multiplied by $\sqrt{K}\boldsymbol{d}_k$, where $\boldsymbol{d}_k$ is the k-th column of $\boldsymbol{D}_K$[21], the minimizer of (22) is written in the frequency domain as

$$k^* = \arg\min_k \left\| \widehat{\boldsymbol{\phi}_K(\boldsymbol{z})} - \widehat{\boldsymbol{\phi}_K(\boldsymbol{w})} \odot \sqrt{K}\,\boldsymbol{d}_k \right\|^2 , \qquad (24)$$

where $\odot$ is the Schur-Hadammard (*i.e.*, elementwise) product. Expressing the square norm in (24) and removing the terms that do not depend on k, we get (Re is the real part):

$$k^* = \arg\max_k \mathrm{Re}\left\{ \boldsymbol{d}_k^H \left(\widehat{\boldsymbol{\phi}_K(\boldsymbol{z})} \odot \overline{\widehat{\boldsymbol{\phi}_K(\boldsymbol{w})}} \right) \right\} . \qquad (25)$$

Expression (25) encodes a computationally simple procedure to "align" $\boldsymbol{\phi}_K(\boldsymbol{z})$ and $\boldsymbol{\phi}_K(\boldsymbol{w})$: i) compute their DFTs; ii) compute the DFT of the elementwise product of these DFTs; iii) locate the entry with largest real part. Since the complexity of computing a DFT using the Fast Fourier Transform (FFT) algorithm is $\mathcal{O}(K \log K)$, the overall complexity of our procedure is $\mathcal{O}(K + 3K \log K)$.

Naturally, the appropriate measure of ANSIG similarity need not be the one in (22). In fact, our shape-based image classification experiments have shown that a better choice is the (cosine of) the angle between the ANSIG vectors,

$$\text{similarity}\,(\boldsymbol{z}, \boldsymbol{w}) := \frac{\left|\boldsymbol{\phi}_K^H(\boldsymbol{z})\, \boldsymbol{\phi}_K(\boldsymbol{w}) \bmod k^*\right|}{\|\boldsymbol{\phi}_K(\boldsymbol{z})\|\, \|\boldsymbol{\phi}_K(\boldsymbol{w})\|} . \qquad (26)$$

5. Experiments

We start by describing experiments that illustrate the behavior of the ANSIG representation, emphasizing its invariance and its capability to deal with shapes described by sets of points of distinct cardinality. Then, we report the use of the ANSIG representation in shape-based image classification, emphasizing its robustness to noise.

Invariance of the ANSIG. We use the shapes represented in Fig. 1, where the shape on the right plot is obtained by translating, rotating, and resizing the one on the left. The vectors containing the coordinates of the points of these shapes also differ by a random permutation of their entries (obviously, the plots in Fig. 1 do not capture this difference).

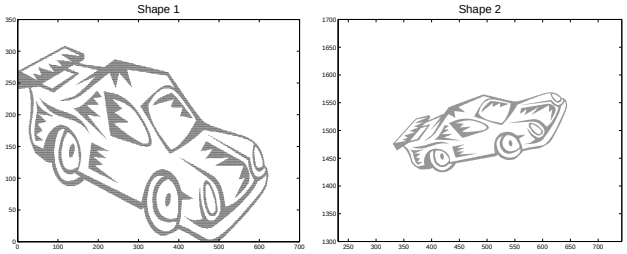

Figure 1. Two shapes that differ by shape-preserving geometric transformations. The ordering of the points describing the shapes is also distinct; thus, although the left and right images represent the same shape, this is not trivially inferred from the stored data.

In Fig. 2, we represent the magnitude (left plot) and the phase (right plot) of the discretized ANSIGs ($\boldsymbol{\phi}_K$ in (20)) of the two shapes in Fig. 1. Note that both the magnitude and the phase plots of the two shapes only differ by a (circular) translation. This is in agreement with our derivations, see expressions (18) and (21): the rotation of a shape induces the same rotation on its ANSIG and this rotation is seen as a (circular) shift of the vector collecting the ANSIG samples on the unit-circle. Since aligning ANSIGs is very simple, as described in Section 4, see (25), our representation fully captures the similarity of the shapes in Fig. 1, factoring out geometric transformations and point labeling.

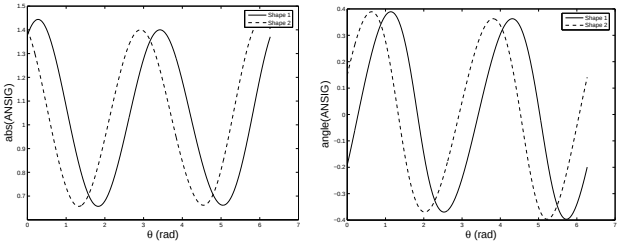

Figure 2. Magnitude and phase of the ANSIGs of the two shapes in Fig. 1. Note that, although these shapes differ by position, rotation, scale, and ordering of the points, their ANSIGs only differ by a (circular) translation that is easily computed.

The experiment just described used shapes described by sets of points of the same cardinality. We now show that ANSIG also deals well with shapes described by sets of

(very) distinct cardinality. This characteristic is important in practice, to enable comparing shapes obtained from images, by an automatic process (due do the pixelization, the same shape often leads to point sets of different cardinality, when seen at different scales). We use the shapes represented in Fig. 3: both shapes represent the same object but the shape on the right is described by just 30% of the points of the one on the left. In Fig. 4, we represent the ANSIGs of both shapes. Clearly, the ANSIGs of the differently sampled shapes almost coincide, as desired. This robustness to distinct sampling density was anticipated in Section 3: it motivated the factor $\sqrt{n}$ in the normalization (11).

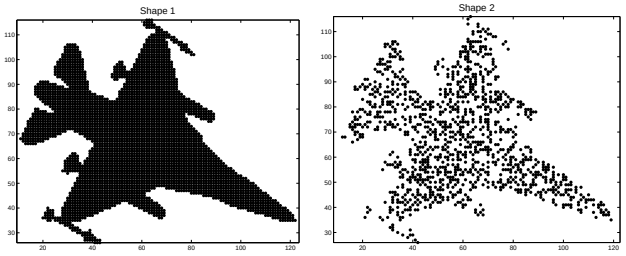

Figure 3. Two shapes that differ in sampling density.

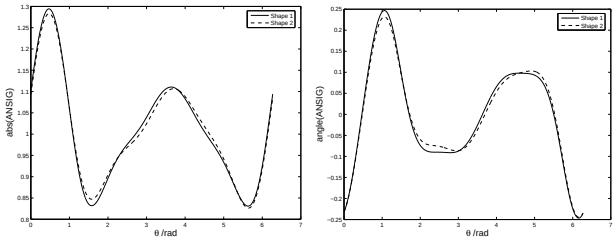

Figure 4. Magnitude and phase of the ANSIGs of the two shapes in Fig. 3. Although the number of points describing these shapes differ significantly, their ANSIGs are similar.

Shape-based classification. We now use the ANSIG representation in automatic classification. The shape database is a collection of prototype shapes (one ANSIG per prototype). The classification is simply made by comparing the ANSIG of a test shape with the ANSIGs in the database and selecting the most similar one, according to (26).

To test the robustness to noise, we used a particularly challenging database of four geometric shapes — a circumference, an hexagon, a square, and a triangle — that are difficult to distinguish when in presence of noise. In Fig. 5, we show some of the test shapes of our experiments. They are noisy (besides translated, rotated, scaled, and randomly point re-ordered) versions of the four geometric shapes. Note in particular how the circle and the hexagon become very similar when the noise level is high.

We tested the classifier with 2000 tests for each of the four geometric shapes, at each noise level. The results are summarized in Table 1. We obtained 100% correct classifications in all tests with SNR above 28dB. Shapes with this level of noise are displayed in the plots of the middle

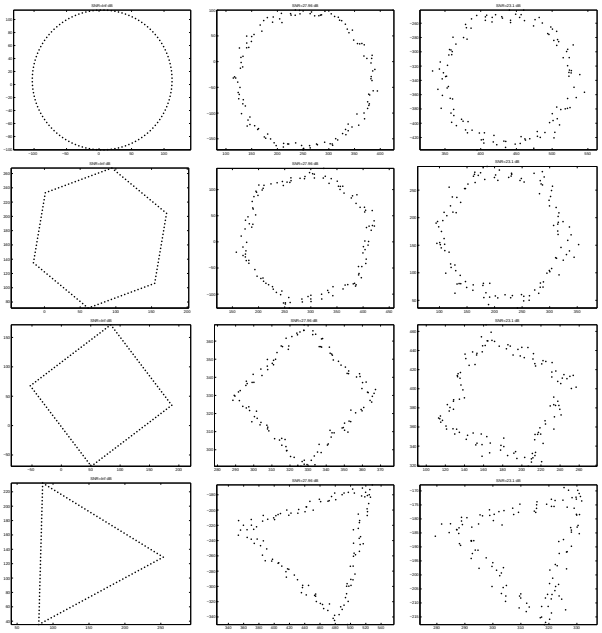

Figure 5. Noisy test shapes for classification. From top to bottom, circle, hexagon, square, and triangle. Noise level increases from the left to the right. The various shapes exhibit distinct translations, rotations, and scales (besides point ordering). Note how the shapes become difficult to distinguish, when the noise increases.

column of Fig. 5; they are far from being visually "clean". From Table 1, we see that we only get a few classifications errors for higher levels of noise and when dealing with circles or hexagons, which, as noted before, for such levels of noise, are in fact difficult to distinguish, even for humans, see the rightmost plots of the first two lines of Fig. 5.

SNR[dB]	≥ 27.96	26.02	24.44	23.1	21.94
○	100	99.95	99.9	99.45	97.3
⬡	100	100	99.9	99.85	99.25
◇	100	100	100	100	100
▷	100	100	100	100	100

Table 1. Percentages of correct classifications the shapes illustrated in Fig. 5 (2000 tests per shape and per noise level).

MPEG-7 shape database. To test the ANSIG representation with a larger database, we used 216 shapes of the MPEG-7 database [22]. Since this database is divided into 18 categories, with 12 (similar) shapes in each category, see the tiny images in Table 2, besides evaluating the classifier performance over the entire database (just as in the experiment with the four geometric shapes), we also evaluated its performance in terms of classification into shape categories. We performed 200 tests for each of the 216 shapes and for each of the two noise levels that are illustrated in Fig. 6.

In what respects to classification in shape categories, all the shapes were correctly classified within the correspond-

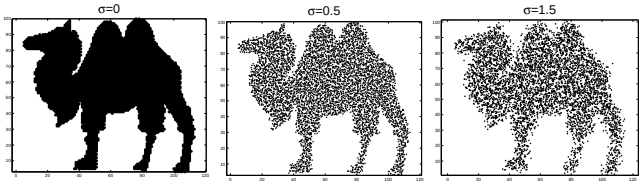

Figure 6. A shape from the MPEG-7 database (left image) and noisy versions for classification (middle and right images).

ing category in 100% of the tests, for both noise levels.

When we treat each of the 216 shapes as a different prototype, we obtain the classification performance detailed in Table 2. We clearly notice two distinct situations: while for shapes belonging to the majority of the categories, the generality of the experiments lead to close to 100% correct classifications, even for the higher level of noise, for shapes belonging to some other categories, the classification errors are much more frequent. What distinguishe these last categories is that the shapes they contain are very similar to each other. In some cases, they are even visually indistinguishable, after factoring out the geometric transformations, see the examples in Fig. 7. Naturally, we can not expect the ANSIG (or any other) representation to be able to distinguish between (noisy versions of) these shapes. All shape classification errors in Table 2 are due to this intra-category similarity, thus, we recall, they disappear when classifying into shape categories.

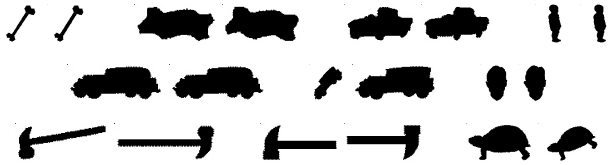

Figure 7. Pairs of shapes in the MPEG-7 database that are visually indistinguishable. When examples like these are considered as a prototypes of the same shape category, we get 100% correct classifications for both noise levels illustrated in Fig. 6.

Automatic trademark retrieval. We now describe an experiment using real images of trademarks. We got, from the internet, logos that are well described by their shape. They are shown in Fig. 8, but not at their real scale: they range from 80×80 to 500×500 pixels. We processed each of these images with the Canny edge detector [6] and stored the ANSIGs of the corresponding edge maps in our database.

Figure 8. Trademark images obtained from the internet.

σ	0.5	3	σ	0.5	3	σ	0.5	3	σ	0.5	3
	100	99.5		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	89.0	97.0		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	87.0	80.0		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	100	98.0		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	32.9	24.2		100	100		100	100		100	100
	24.8	1.0		100	100		100	100		100	100
	58.0	40.8		100	100		100	100		58.0	62.5
	65.0	65.5		100	100		100	100		98.5	88.5
	56.0	68.5		100	100		100	100		100	100
	20.6	19.4		100	100		100	100		62.0	54.0
	24.8	24.8		100	100		100	100		100	100
	44.8	37.5		100	100		100	98.5		100	100
	29.6	27.0		100	100		100	100		46.0	64.5
	47.4	36.3		100	100		100	100		100	100
	57.0	41.9		100	100		100	100		100	100
	61.5	53.0		100	100		100	100		100	100
	33.3	24.8		100	100		100	80.0		100	100
	55.5	87.5		100	100		100	100		100	100
	47.6	20.6		100	100		100	100		96.5	73.5
	100	100		100	100		100	99.0		100	100
	100	100		100	100		100	100		100	100
	89.0	93.5		94.0	72.0		100	100		100	100
	100	100		100	100		100	100		100	100
	65.5	55.0		96.5	98.0		100	100		100	100
	100	100		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	100	100		100	100		100	100		100	100
	40.5	39.0		100	100		100	99.0		100	100
	100	100		55.5	60.0		100	100		100	100
	100	100		49.8	41.9		100	100		100	100
	100	100		37.1	3.9		57.0	57.0		100	100
	100	100		55.5	81.5		100	100		100	100
	100	100		50.0	28.6		100	100		100	100
	100	100		45.7	46.2		100	100		100	100
	100	100		51.0	52.0		100	100		100	100
	100	100		40.1	43.8		100	100		100	100
	100	100		29.6	19.4		100	100		100	100
	100	100		67.0	62.5		100	100		100	100
	100	100		76.0	78.0		51.0	44.1		100	100
	100	100		70.0	79.5		100	100		100	100
	79.0	71.0		100	100						
	100	100		100	100						
	78.5	60.5		100	100						
	77.0	75.5		100	100						
	100	100		100	100						
	100	98.0		100	100						
	100	100		100	100						
	100	100		100	100						
	100	59.0		100	100						
	100	94.0		100	100						
	100	100		100	100						
	100	100		100	100						

Table 2. Percentages of correct classifications for the MPEG-7 database. When classifying into shape categories, we got 100%.

We printed the trademarks in Fig. 8 and photographed their paper version with a low quality digital camera, with several paper-camera positions, orientations, and distances. This way we got a set of challenging test images, where the candidate logos appear at different sizes and positions, see some examples in Fig. 9. Finally, we computed the edges of these test images, without any particular care to tune detection parameters, obtaining a set of test shapes.

Figure 9. Examples of trademark images to be classified.

Being edge maps of real photos, our test shapes exhibit distortions (relative to their versions in the database) that are not modeled explicitly by the ANSIG, like: missing points (when edges are not detected, *e.g.*, due to inaccurate focusing); spurious points (when too much edges are detected, *e.g.*, due to paper texture noticed at large zoom); or perspective distortion (when the camera image plane was not parallel to the paper). We classified these test shapes by proceeding as in the experiments above. Except in cases where the distortions just mentioned are dramatic (*e.g.*, out-of-focus images), the classification was correct.

6. Conclusion

We proposed a new method to represent 2D shapes, described by a set of points, or landmarks, in the plane. Our method is based on what we call the analytic signature (ANSIG) of the shape, whose most distinctive characteristic is its invariance to the way the landmarks are labeled. This makes ANSIG particularly suited to cope with shapes described by large sets of edge points in images. We illustrated its performance in shape-based classification tasks.

We envisage paths for future research based on the ANSIG representation. For example, while in this paper we store ANSIGs by sampling them on the unit-circle, a topic that deserves further study is the adoption of different sampling schemes, *e.g.*, the use of two or more circles, for robustness. Also, the derivations in this paper are targeted to the representation and comparison of complete shapes. However, in many practical scenarios, it is also necessary to recognize a set of points as being a part, *i.e.*, a subset of a given shape. A good challenge is then to adapt the ANSIG representation to deal with incomplete shapes.

References

[1] T. Adamek and N. O'Connor. A multiscale representation method for nonrigid shapes with a single closed contour. *IEEE T-PAMI*, 14(5):742–753, 2004. 1

[2] L. Ahlfors. *Complex Analysis*. McGraw-Hill, USA, 1978. 3

[3] I. Bartolini, P. Ciaccia, and M. Patella. Warp: Accurate retrieval of shapes using phase of fourier descriptors and time warping distance. *IEEE T-PAMI*, 27(1):142–147, 2005. 1

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE T-PAMI*, 24(24):509–522, 2002. 1

[5] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE T-PAMI*, 14(2), 1992. 2

[6] J. Canny. A computational approach to edge detection. *IEEE T-PAMI*, 8:679–714, 1986. 7

[7] G. Chauang and C. Kuo. Wavelet descriptor of planar curves: Theory and applications. *IEEE T-IP*, 5:56–70, 1996. 1

[8] H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *IEEE CVPR*, Hilton Head Island, South Carolina, USA, 2000. 2

[9] P. Dierckx. *Curve and Surface Fitting with Splines*. Oxford Science Publications, 1995. 1

[10] C. Grigorescu and N. Petkov. Distance sets for shape filters and shape recognition. *IEEE T-IP*, 12(10), 2003. 1

[11] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985. 3

[12] M. Hu. Visual pattern recognition by moment invariants. *IRE Trans. on Information Theory*, 8:179–187, 1962. 1

[13] T. Jebara. Images as bags of pixels. In *Proc. of the IEEE Int. Conf. on Computer Vision*, Nice, France, 2003. 2

[14] T. Jebara. Kernelizing sorting, permutation and alignment for minimum volume PCA. In *Proc. of Annual Conf. on Learning Theory*, Banff, Alberta, Canada, 2004. 2

[15] D. Kendall, D. Barden, T. Carne, and H. Le. *Shape and Shape Theory*. John Wiley and Sons, 1999. 1

[16] A. Khotanzad and Y. Hong. Invariant image recognition by Zernike moments. *IEEE T-PAMI*, 12:489–497, 1990. 1

[17] G. McNeill and S. Vijayakumar. Hierarchical procrustes matching for shape retrieval. In *IEEE CVPR*, NY, 2006. 2

[18] G. McNeill and S. Vijayakumar. Part-based probabilistic point matching using equivalence constraints. In *Proc. of NIPS*, Vancouver BC, Canada, 2006. 2

[19] F. Mokhtarian and A. Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE T-PAMI*, 8:34–43, 1986. 1

[20] R. Mukundan, S. Ong, and P. Lee. Image analysis by Tchebichef moments. *IEEE T-IP*, 10:1357–1364, 2001. 1

[21] A. Oppenheim, R. Schafer, and J. Buck. *Discrete-Time Signal Processing*. Prentice Hall, 1999. 5

[22] T. Sebastian, P. Klein, and B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE T-PAMI*, 26(5), 2004. 6

[23] P. Shivaswamy and T. Jebara. Permutation invariant SVMs. In *Int. Conf. on Machine Learning*, Pittsburgh PA, 2006. 2

[24] M. Teague. Visual pattern recognition by moment invariants. *Journal of the Optical Society of America*, 70, 1980. 1

[25] D. Zhang and G. Lu. Study and evaluation of different fourier methods for image retrieval. *IJCV*, 1234:33–49, 2005. 1