SHAPES AS EMPIRICAL DISTRIBUTIONS

Bernardo Rodrigues Pires, José M. F. Moura

Carnegie Mellon University
Department of ECE
Pittsburgh, PA
bpires@andrew.cmu.edu, moura@ece.cmu.edu

ABSTRACT

We address the problem of shape based classification. We interpret the shape of an object as a probability distribution governing the location of the points of the object. An image of the object, represented as an arbitrary set of unlabeled points, corresponds to a random drawing from the shape probability distribution and can thus be analyzed as an empirical distribution. Using this framework, classification of shapes is robust to the number of points in the image and there is no need to solve the correspondence problem when comparing two images. The framework allows us to estimate geometrical transformations between images in a statistically meaningful way using Maximum Likelihood. We formulate the decision problem associated with shape classification as a hypothesis test for which we can characterize the performance. We particularize this framework to two-dimensional shapes related by an affine transformation. Under this assumption, we develop a descriptor invariant to affine movement, permutations, and sampling density, and robust to noise, occlusion, and reasonable non-linear deformations. Experimental results demonstrate the quality of our approach.

Index Terms— Empirical shape distribution, shape classification, shape representation, shape descriptor, unlabeled data, affine-permutation invariance

1. INTRODUCTION

We consider shape-based automatic classification of objects. We represent an unknown object by a group of *unlabeled* points or landmarks. The objective is to use the set of points to infer the shape of the object and then use the shape to classify the object. How to represent the object and how to infer its shape from the observation are open problems.

The main difficulty in classifying shapes comes from the fact that certain operations preserve what a human perceives as shape but make the point configuration significantly different. These can be broadly classified as *motion* and *correspondence* distortions. *Motion* distortions arise because objects can be observed from different points of view. Specifically, an object or camera (or both) may move with relation to the setup that generated an existing database representation of the object, for example.

Correspondence distortions, on the other hand, refer to the fact that a shape should not change if: (1) we permute the order in which the points of the object are stored in memory, (2) remove a number of points from the observation (e.g., the object was partially occluded or the sample density of the points was decreased), or (3) add a number of points to the observation (e.g., the sample density was increased).

An earlier approach by Kendall, Barden, Carne, and Le [1] to shape representation provides a good framework for dealing with situations where the points of the object are *labeled* or ordered, i.e., the matching between the landmarks in the observation and the landmarks in a reference image is known and thus the *correspondence* problem is solved. Under this hypothesis, Kendall's approach developed a movement invariant framework where the shape description is invariant to translation, rotation, and scale between the observed object and its representation [1]. Even though this approach generated significant practical successes, the *correspondence* assumption is too strong for most automatic applications.

To cope with unlabeled points, many attempt to solve the *correspondence* problem, e.g., by image correlation [2] [3] [4]. Most approaches formulate correspondence as an optimization problem and can be distinguished by the optimization algorithms used: greedy algorithms, linear programming via relaxation of constraints [4], randomized search [3], dynamic programming, and convex optimization [2] have all been used with various degrees of success. One of the better known approaches, the *Iterative Closest Point (ICP)* algorithm [5], deals with one problem at a time, iteratively finding the movement and the correspondence. Along the same line, more recent approaches [6] use an Expectation Maximization-like iterative approach. These methods may be sensitive to noise and do not guarantee convergence.

Rather than attempting to solve both the *movement* and the *correspondence* problem, some approaches explore invariance in one or both of these problems. LASIC [7] is a movement-invariant method for computing the *correspondence*. Other recent methods looked for invariance to *correspondence*, e.g., [8] factors out the permutation but does not deal with movement. Reference [9], describes shape as a set of distances between pairs of points, limiting its applicability to small sets of points, where the computation of pairwise distances is possible. ANSIG [10] defines an analytical signature that is invariant to *correspondence*, scale, and translation, but is restricted to 2D shapes and does not deal with general affine movements or other more complex geometric transformations.

In this paper, we interpret the shape of an object as the probability distribution governing the location of its points. A specific view, represented as a set of unlabeled points, is a random drawing from the shape probability distribution and is analyzed as an empirical distribution. Using this framework, classification of shapes is robust to the number of points and the object representation is invariant to correspondence.

We apply Maximum Likelihood (ML) to estimate the motion and formulate shape classification as an hypothesis test. For two-dimensional shapes related by an affine transformation, we present a descriptor that is invariant to affine movement, permutation, and sampling density and is robust to noise, occlusion, and reasonable non-linear deformations.

2. PROBABILISTIC SHAPE FRAMEWORK

Shapes as empirical distributions We define the intrinsic shape of an object as a probability distribution function (p.d.f.), which we shall refer to as the shape distribution function, $\mathfrak{P}(\boldsymbol{z})$. An observation of the object $O = \{\boldsymbol{x}_n\}_{n=1}^N$ is thus a collection of points $\boldsymbol{x}_n \in \mathbb{R}^d$ that are randomly drawn from this distribution (where d is the dimension of the shape, i.e. $d = 2$ for 2-D shapes). Unlike the work of Kendall, Barden, Carne and Le [1], ours is not a p.d.f. in the shape space (in which case a shape is a point in the shape space drawn from the shape space p.d.f.). Instead we define the shape as a p.d.f itself, and the observations as realizations of this p.d.f.

We define the empirical distribution $F(\boldsymbol{x})$ of the observation as:

$$F(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}\delta(\boldsymbol{x}_n \leq \boldsymbol{x}) \tag{1}$$

where $\delta(\cdot)$ is the indicator function that is equal to 1 if the condition is true and 0 otherwise, and the inequality corresponds to a coordinate-wise inequality.

To deal with movement, we allow for a transformation between the coordinate system of the observations, $\boldsymbol{x}$, and the coordinate system of the shape p.d.f., $\boldsymbol{z}$. We assume rigid motion and use $m(\cdot)$ to model the parametric coordinate transformation between $\boldsymbol{x}$ and $\boldsymbol{z}$, i.e., $\boldsymbol{z} = m(\boldsymbol{x}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the observation parameters.

Under our statistical shape framework, the observed *empirical* distribution approximates the shape distribution function, up to the coordinate transformation, i.e.:

$$F(\boldsymbol{x}) \;\approx\; \mathfrak{P}(m(\boldsymbol{x}, \boldsymbol{\theta})). \tag{2}$$

Observed and expected frequencies We can make equation (2) more precise if we consider a partition $\mathcal{P} = \{P_k\}_{k=1}^K$ of the domain of $\boldsymbol{x}$ (all our experiences use $K = 256$). Under this partition we define the observed frequencies $f(k)$ and the expected frequencies $\mathfrak{f}(k)$ as:

$$f(k) = \sum_{n=1}^{N}\delta(\boldsymbol{x}_n \in P_k), \quad \mathfrak{f}(k, \boldsymbol{\theta}) = N\int_{P_k}\mathfrak{p}(m(\boldsymbol{x}, \boldsymbol{\theta}))\, d\boldsymbol{x}, \tag{3}$$

where we used $\mathfrak{p}(m(\boldsymbol{x}, \boldsymbol{\theta}))$ to denote the probability density function corresponding to $\mathfrak{P}(m(\boldsymbol{x}, \boldsymbol{\theta}))$. Note that the observed frequencies $f(k)$ are simply the count of the number of points in the observation that fall into each partition, whereas the expected frequencies $\mathfrak{f}(k, \boldsymbol{\theta})$ correspond to the number of observations that would be expected to fall in each partition assuming the probability distribution $\mathfrak{P}(m(\boldsymbol{x}, \boldsymbol{\theta}))$.

Following equation (2) we must have that, if the observation comes from the known shape: $f(k) \;\approx\; \mathfrak{f}(k, \boldsymbol{\theta}) \quad \forall\, k = 1, \ldots, N$.

3. SHAPE CLASSIFICATION

Classification as an hypothesis test We formulate the problem of object recognition: Given the known shape of an object, i.e. given the expected frequencies $\mathfrak{f}(k, \boldsymbol{\theta})$, how can we decide if the a new image (characterized by the observed frequencies $f(k)$) corresponds to the object defined by $\mathfrak{f}(k, \boldsymbol{\theta})$?

We cast this problem as an hypothesis test in the following way: we want to test hypothesis $H_0 : f(k) = \mathfrak{f}(k)$ (observation comes from the know object) versus $H_1 : f(k) \neq \mathfrak{f}(k)$ (observation does not come from the known object). In the field of statistics this is called a goodness of fit test (see, for example [11]).

Estimation of the movement We use maximum likelihood (ML) to estimate the value of the observation parameters $\boldsymbol{\theta}$ under the hypothesis H_0, i.e. under the hypothesis that the observation was drawn from the shape distribution. Under this hypothesis, we write the likelihood of the observation, $\mathbf{Prob}\{\{\boldsymbol{x}_n\}_{n=1}^N|\boldsymbol{\theta}\}$ in terms of the probability of finding (for all k sections of the partition) $f(k)$ observed points in section k. Since, for each section k, we have probability $\mathfrak{f}(k, \boldsymbol{\theta})/N$ of finding a point in it, the ML estimate of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$, can be found by solving the problem:

$$\max_{\boldsymbol{\theta}}\prod_{k=1}^{K}\left[\frac{\mathfrak{f}(k, \boldsymbol{\theta})}{N}\right]^{f(k)} = \max_{\boldsymbol{\theta}}\sum_{k=1}^{K} f(k)\log\left(\mathfrak{f}(k, \boldsymbol{\theta})\right). \tag{4}$$

Goodness of Fit Test Using the ML estimate of the parameters, the Pearson chi-square test statistic is [11]:

$$X^2 = \sum_{k=1}^{K}\frac{\left(f(k) - \mathfrak{f}(k, \widehat{\boldsymbol{\theta}})\right)^2}{\mathfrak{f}(k, \widehat{\boldsymbol{\theta}})}, \tag{5}$$

which leads to the test:

$$\begin{cases} X^2 \leq \mathcal{X}_0^2 & \Rightarrow H_0 \text{ is true} \\ X^2 > \mathcal{X}_0^2 & \Rightarrow H_1 \text{ is true} \end{cases} \tag{6}$$

where $\mathcal{X}_0^2$ is the threshold of the test.

It can be shown [11] that, under H_0, the test statistic X^2 has a chi-square distribution with $K - \dim(\boldsymbol{\theta}) - 1$ degrees of freedom (where $\dim(\boldsymbol{\theta})$ is the number of unknown observation parameters). For this test we can easily compute the desired probability of false alarm (size of the test), the probability of detection (power of the test), and the threshold $\mathcal{X}_0^2$.

4. SHAPE UNDER AFFINE MOVEMENT

Movement Model We assume that we are working with 2-D shapes, i.e. $\boldsymbol{x} \in \mathbb{R}^2$. For the affine movement model, we have:

$$m(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{\delta}, \tag{7}$$

where $\boldsymbol{A}$ is a 2×2 matrix and $\boldsymbol{\delta}$ is a 2×1 vector. There are six movement parameters for the transformation, $\boldsymbol{\theta} = \{\boldsymbol{A}, \boldsymbol{\delta}\}$.

Shape Compacting It is possible to use the results in the previous section to estimate the six movement parameters and to classify the shape. However, we can greatly reduce the complexity of the problem by applying two simple pre-processing steps [12]. Let $\boldsymbol{x}_1$ be the $1 \times N$ vector that concatenates all the x entries in $\{\boldsymbol{x}_n\}_{n=1}^N$. Let $\boldsymbol{x}_2$ be the corresponding vector of y entries.

We consider the preprocessing steps:

1. Centering: The shape is centered at the origin of the coordinate system by removing the global translation. This step zeroes the mean of the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ by ensuring that $\mathbf{1}^T\boldsymbol{x}_1 = \mathbf{1}^T\boldsymbol{x}_2 = 0$.
2. Normalization: The shape is transformed so that the vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthonormal, i.e., $\|\boldsymbol{x}_1\|_2 = \|\boldsymbol{x}_2\|_2 = 1$ and $\boldsymbol{x}_1 \perp \boldsymbol{x}_2$.

These two steps, centering and normalization, are together referred to as compacting or shape normalization in the pattern recognition literature. It can be shown that the compacting process results in a shape that is continuous, easy to compute, and robust to digitalization errors. Readers are referred to [12] for details.

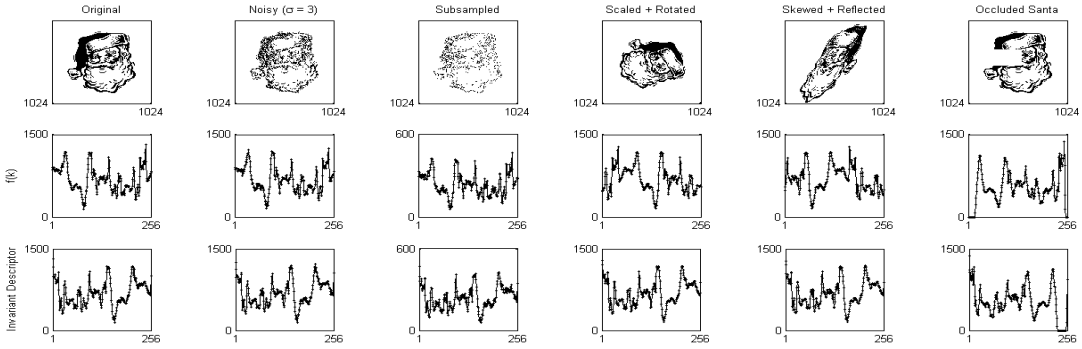

Fig. 1. Shape distribution and invariant descriptor under a several distortions

Rotation-Reflection Model Two shapes related by the affine model in (7) will, after compacting, be related by a simple rotation-reflection model of the form (see, for example [12]):

$$m(\boldsymbol{x}, \boldsymbol{\theta}) = \begin{bmatrix} \gamma\cos(\theta) & -\gamma\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \boldsymbol{x}, \tag{8}$$

where $\theta \in [0, 2\pi]$ is the rotation (in the positive or counter-clockwise direction) and $\gamma = \{-1, 1\}$ is the reflection parameter. Thus compacting reduces the number of parameters in the affine model from six to two.

Rotation-specific partition We quantize the rotation parameter, i.e., we make $\theta = 2\pi\, t/K$, $t = 0, 1, \ldots, K-1$ and use the following partition of $\mathbb{R}^2$:

$$P_k = \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \frac{(k-1)\,2\pi}{K} \leq \text{angle}(\boldsymbol{x}) < \frac{k\,2\pi}{K} \right\}, \tag{9}$$

where $k = 1, \ldots, K$ and angle$(\boldsymbol{x})$ denotes the angle that the vector from the origin to the point $\boldsymbol{x}$ (the vector $\overrightarrow{\mathbf{0}\boldsymbol{x}}$) makes with a pre-specified vector, arbitrarily chosen as one of the axis of the coordinate system.

While the definitions of $f(k)$ and $\mathfrak{f}\,(k, \boldsymbol{\theta})$ remain as in (3), the model in (8) and the partition in (9) allow us to write (using simple trigonometry) the shape distribution as:

$$\mathfrak{f}\,(k, \boldsymbol{\theta}) = \mathfrak{f}\,(k, \theta, \gamma) = \mathfrak{f}\,(\,\gamma\,(k + t)). \tag{10}$$

Parameter estimation Particularizing equation (4) for the rotation-reflection, we obtain using equations (10) and (4):

$$\{\widehat{\gamma}, \hat{t}\} = \arg\max_{\gamma,\, t} \sum_{k=1}^{K} f(k)\, \log\left[\,\mathfrak{f}\,(\gamma\,(k + t))\right]. \tag{11}$$

Since the rotation has period 2π, we have that, for a specific value of $\gamma = \pm 1$, the $\mathfrak{f}$ function will be periodic with period K. This allows us to write the ML estimators as:

$$\{\widehat{\gamma}, \hat{t}\} = \arg\max_{\gamma,\, t} \sum_{k=1}^{K} f(k) \cdot (\,\log\left[\,\mathfrak{f}\,(\gamma\, k)\right]\ \mathsf{cir}\ \gamma\, t\,), \tag{12}$$

where cir represents the circular-shift operator. Equation (12) shows that the ML estimators for γ and t maximize the correlations between the functions $f(k)$ and $\log[\mathfrak{f}\,(k)]$, where the parameter t shifts the $\log[\mathfrak{f}\,(k)]$ function while γ inverts it about the k axis.

This problem can be solved in a straightforward manner by exhaustive search in t for the two possible values of γ. This leads to an algorithm of quadratic complexity in the number of partitions, $\mathcal{O}(2K^2)$, which is clearly superior to the factorial complexity of considering all possible permutations of the N points (note that $N \gg K$). Using a method similar to [10], it is possible to use the Fast Fourier Transform to solve the problem in $\mathcal{O}(K \log K)$ time.

Affine Shape Classification Using the ML estimate for the movement parameters $\{\widehat{\gamma}, \hat{t}\}$ found by solving (12), we write the test statistic for the affine shape with compacting as:

$$X^2 = \sum_{k=1}^{K} \frac{\left(f(k) - \mathfrak{f}\,(\gamma\,(k + t))\right)^2}{\mathfrak{f}\,(\gamma\,(k + t))}. \tag{13}$$

The test is as in (6). The threshold $\mathcal{X}_0^2$ is found by specifying a value for the probability of false alarm and using the fact that, under hypothesis H_0, X^2 has a chi-squared distribution with $K-2$ degrees of freedom.

5. AFFINE-PERMUTATION INVARIANT DESCRIPTOR

Invariant Descriptor As discussed in the previous section, equation (12) shows that the reflection-rotation parameters are estimated so as to "align" the shape distribution function $\mathfrak{f}\,(k, \boldsymbol{\theta})$ with the observation empirical distribution $f(k)$.

But there are other ways of aligning the two functions. Instead of estimating the rotation-reflection parameters, we create a descriptor that is invariant to them. This way we have a 2D shape descriptor that is fully invariant to affine and permutation distortions. Our experiments will focus on this descriptor.

Algorithm To create such a descriptor we apply the following deterministic algorithm to both $\mathfrak{f}\,(k, \boldsymbol{\theta})$ and $f(k)$ (let $g(k)$ be either one of these functions):

1.a) Circularly rotate $g(k)$ so that its maximum is in the first position, i.e. $g(1) \geq g(k)$, $\forall k$.

1.b) If $g(k)$ has multiple maximums $\{k_j^*\}_{j=1}^{J}$, break ties by maximizing $g(k_j^* - 1) + g(k_j^* + 1)$.

1.c) If tie persists consider points further away from maximum until tie is broken.

2.a) Consider the K/m points closest to $k = 1$. If necessary flip $g(k)$ about $k = 1$ so that the $K/2m$ points to its right sum more that the points to its left, i.e. ensure that:
$$\textstyle\sum_{j=2}^{K/2m+1} g(j) \;>\; \sum_{j=K-K/2m}^{K} g(j)$$

2.b) If there is a tie in the previous step, include more points in the summation (increase the value of m) until the tie is broken

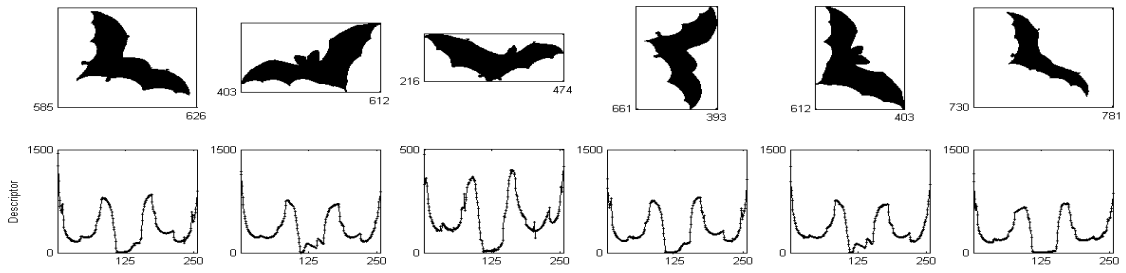

Fig. 2. Performance of the invariant descriptor when the shape is deformed. Images from the bat category of the MPEG-7 database.

6. EXPERIMENTS

Distortion invariance Figure 1 plots the observed frequency and the invariant descriptor under a series of synthetically introduced distortions. Note that, although the shape of the observed frequencies changes, none of the severe distortions introduced causes a significant change in the descriptor. The first two distortions show that the descriptor is robust to severe noise and to drastic changes in the number of points used to describe the shape (approximately 65% of the points in the shape were removed in the subsampled experiment). As expected the descriptor is invariant to scale and rotation but note that it is also invariant to skew and reflection – two distortions that can not be modeled by scale-translation-rotation methods. Finally note that the descriptor is also able to handle significant occlusion (approximately 12.5% of the shape area was removed to simulate occlusion). Additionally, the proposed hypothesis test (correctly) classifies all distorted images as representing the same shape as the original image (when using a probability of false alarm of 2%). This attests to the robustness of the classification scheme proposed.

Performance under deformation In figure 2 we plot the proposed descriptor for several images in the bat category of the MPEG-7 database. Note that the images were deformed in a way that cannot be modeled by affine transformations, but the descriptor is similar in every case. In fact, when we take any pair of bats in the figure and conduct the hypothesis test proposed (with a probability of false alarm of 2%), the test concludes (correctly) that the shapes are the same. This shows that both the descriptor and the hypothesis test are robust to challenging non-affine deformations.

Performance on the MPEG-7 Database As a final test, we used the proposed descriptor and hypothesis test to classify a set of 18 objects in the MPEG-7 database. For each object, 20 views are contained in the database, for a total of 360 images in the database. In our experiment, we added two realistic levels of salt and pepper noise to each of the images in the database and attempted to classify the image into one of the 18 object categories. Table 7 shows the overall percentage of correct classifications and compares our results with ART, the MPEG-7 standard region-based shape descriptor [13]. The high rate of success of our descriptor shows that it is well suited for practical applications.

7. CONCLUSION

We introduced a new framework for shape-based classification based on interpreting images as empirical distributions. We developed this framework for general shapes and solved the problem of object classification using an hypothesis test with known performance. Assuming parametric motion, we developed a Maximum Likelihood estimator for the movement parameters.

We particularized our framework for 2D shapes under affine motion. Under this assumption we developed a descriptor that is fully invariant to affine movement and permutation. Experiments show that this descriptor is robust to full affine movement (including skew and reflection), severe noise, extreme under-sampling, significant occlusion and reasonable non-affine deformation.

Noise Probability	0.005	0.01
Proposed Method	$98.6\,\%$	$92.2\,\%$
ART algorithm [13]	$70.6\,\%$	$65,8\,\%$

Table 1. Summary of results for MPEG-7 database

8. REFERENCES

[1] D. G. Kendall, D. Barden, T. K. Carne, and H. Le, *Shape and Shape Theory*, Wiley, first edition, 1999.

[2] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of the 7th Int. Joint Conf. on Artificial Intelegence*, 1981, pp. 674–679.

[3] P. Torr and D. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *Intl. Journal of Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.

[4] J. Maciel and J. P. Costeira, "A global solution to sparse correspondence problems," *Trans. PAMI*, vol. 25, no. 2, pp. 187–199, 2003.

[5] P. Besl and N. McKay, "A method for registration of 3-d shapes," *Trans. PAMI*, vol. 14, no. 2, pp. 239–256, 1992.

[6] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching," in *Proc. CVPR*, SC,USA, 2000, pp. 44–51.

[7] B. R. Pires, J. M. F. Moura, and J. Xavier, "Lasic: A model invariant framework for correspondence," in *Proc. ICIP*, San Diego CA, October 2008, pp. 2356–2359.

[8] T. Jebara, "Images as bags of pixels," in *CVPR*, Nice, 2003, pp. 265–272.

[9] M. Boutin and M. Comer, "Faithful shape representation for 2d gaussian mixtures," in *Proc. ICIP*, San Antoinio TX, 2007, vol. VI, pp. 369–372.

[10] J. J. Rodrigues, P. M. Q. Aguiar, and J. M. F. Xavier, "Classification of unlabeled point sets using ansig," in *Proc. ICIP*, San Diego CA, 2008, pp. 2360–2363.

[11] M. G. Kendall, *The Advanced Theory of Statistics*, vol. 2A, Hodder Arnold Publication, 1994.

[12] V. Ha and J. M. F. Moura, "Affine-permutation invariance of 2d shapes," *Trans. on Image Processing*, vol. 14, no. 11, pp. 1687–1700, 2005.

[13] S.-K. Hwang and W.-Y. Kim, "Fast and efficient method for computing art," *Trans. on Image Processing*, vol. 15, no. 1, pp. 112–117, 2006.