Biomedical Signal Processing and Control 40 (2018) 163–170

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Thermodynamic analogies for the characterization of 3D human coronary arteries

C.A. Bulant[a,b,*], P.J. Blanco[a,b], A. Clausse[c], C. Bezerra[d], T.P. Lima[d], L.F.R. Ávila[d], P.A. Lemos[d], R.A. Feijóo[a,b]

[a] National Laboratory for Scientific Computing, LNCC/MCTI, Av. Getúlio Vargas 333, Quitandinha, 25651-075 Petrópolis, Brazil
[b] National Institute of Science and Technology in Medicine Assisted by Scientific Computing, INCT-MACC, Petrópolis, Brazil
[c] CNEA-CONICET and National University of Central Buenos Aires, 7000 Tandil, Argentina
[d] Heart Institute, University of São Paulo Medical School, INCOR-FM-USP, Av. Dr. Eneas de Carvalho Aguiar, 44, 3rd Floor, São Paulo, SP 05403-000, Brazil

ARTICLE INFO

Article history:
Received 3 November 2016
Received in revised form 15 June 2017
Accepted 16 September 2017

Keywords:
Thermodynamics of curves
Geometric characterization
Coronary arteries
Classification
Geometrical risk factors

ABSTRACT

The thermodynamics of three-dimensional curves is explored through numerical simulations, providing room for a broader range of applications. Such approach, which makes use of elements of information theory, enables the processing of parametric as well as non-parametric data distributed along the curves. Descriptors inspired in thermodynamic concepts are derived to characterize such three-dimensional curves. The methodology is applied to characterize a sample of 48 human coronary arterial trees and compared with standard geometric descriptors. As an application, the usefulness of the thermodynamic descriptors is tested by assessing statistical associations between arterial shape and diseases. The feature space defined by arterial descriptors is analyzed using multivariate kernel density classification methods. A two-tailed *U*-test with 95% confidence interval showed that some of the proposed thermodynamic descriptors have different mean values for healthy/diseased left anterior descending (LAD) and left circumflex (LCx) arteries. Specifically: in the LAD, the temperatures based on mean number of intersection points and curvature are larger in healthy arteries ($p < 0.05$); in the LCx, the intersection counting pressure is larger in healthy arteries ($p < 0.05$). Moreover, the shape of the right coronary artery is thoroughly characterized by these descriptors. Specifically: intersection count thermodynamics, i.e. entropy, temperature and pressure are larger in Σ-Shape RCAs, in turn curvature based entropy and pressure are larger in C-Shape RCAs ($p < 0.05$).

1. Introduction

The theory known as *thermodynamics of plane curves* was originally proposed by Mendès France [1,2]. The core idea was to characterize planar curves with classical thermodynamics quantities, e.g. entropy, temperature and pressure, preserving analogies to the corresponding physical laws. The very foundation of the theory relies on a theorem from the field of *integral geometry*, known as the Cauchy–Crofton theorem [3], which states that the expected number of intersections ($\bar{n}$) between a random line intersecting a plane curve Γ, is related to the length (ℓ) of Γ and the perimeter (C) of its convex hull. The link to thermodynamics came from an *information-theory* based analysis of the discrete probability distribution (p_n) of the intersection count function.

Over the years the ideas behind the theory of *thermodynamics of plane curves* were further explored in close relation to fractal theory, with strong theoretical flavors and a modest number of applications. For example, [4] revisited the theory for planar curves and related the entropy to the notion of dimension of curves. In [5] used the rationale behind thermodynamics analogies to define the temperature of non-random maps. In [6] adapted the concept of entropy for application in time/spatial series, showing practical examples in geological data processing. The same research group used entropy of time/spatial series to the identification of functional relationships between carbon dioxide concentration and atmospheric pressure in caves [7]. More recently, [8,9] adapted the entropy of curves, generalizing it to an arbitrary number of dimensions, with application to analysis and classification of dynamical systems. In [10] recently presented applications of the $\bar{n}$, also known as inconstancy, to numerical sequences and proposed

* Corresponding author at: National Laboratory for Scientific Computing, LNCC/MCTI, Av. Getúlio Vargas 333, Quitandinha, 25651-075 Petrópolis, Brazil.
E-mail addresses: bulant@lncc.br (C.A. Bulant), pjblanco@lncc.br (P.J. Blanco), clausse@exa.unicen.edu.ar (A. Clausse).

http://dx.doi.org/10.1016/j.bspc.2017.09.015
1746-8094/

some practical applications. More recent contributions in the area focused on the use of a variety of generalized entropy definitions, like Rényi's [11].

It is worth to remark that neither the entropy adaptations for time/spatial series [6], nor the one proposed by [8] for curves in $\mathbb{R}^n$, are linked to the expected number of intersection ($\bar{n}$) between a curve and hyperplanes, which is a cornerstone of the original theory. In fact, both works proposed a new definition of the entropy function based on the series/curve characteristics, without considering any probability distribution. This strays those contributions from the original notion: an information-theory-based entropy with analogy to statistical mechanics.

In this work, we present a natural extension of the thermodynamic descriptors to curves in three-dimensional (3D) space. In order to do that, we use directly the probability distribution p_n instead of the Cauchy–Crofton theorem. A computational approximation of p_n allows the numerical estimation of the entropy, temperature and pressure descriptors of a 3D curve. The use of probability distributions also inspired a generalization of these thermodynamic descriptors for characterizing curves using spatially distributed information, e.g. curvature, torsion.

There are several applications fields in which the characterization of 2D or 3D curves are valuable, i.e. spatio-temporal trajectory, numerical series, signal processing, handwriting and shape analysis (from contours or skeletons). To use the proposed framework in such fields, a representation of input data in the form of a curve, is required. In this work, as application examples, we use thermodynamic descriptors for the characterization of human coronary arteries, extracted from patient-specific medical images. The motivation and context for the use of such dataset are stated in Section 3.

2. Thermodynamics of curves in 3D space

Consider a curve Γ in 3D space. We define p_n as the probability of a random plane intersecting Γ at n points. Therefore, the mean number of intersection points of a random plane is $\bar{n} = \sum np_n$. Then, we introduce the Shannon's measure of entropy from information theory [12],

$$H = -\sum_{n=1}^{\infty} p_n \log \quad p_n. \tag{1}$$

In physics, finding the probability distribution p that maximizes H, is the basis of the so called MaxEnt thermodynamics principle [13,14], which explains statistical mechanics and equilibrium thermodynamics as inference processes. Maximization of H subjected to a restriction on the mean value was first tackled by [15]. The classical solution, known as Gibbs algorithm, makes use of Lagrange multipliers to find the roots of the functional

$$\mathcal{L}(p) = -\sum_{n=1}^{\infty} p_n \log \quad p_n - \beta\left(\bar{n} - \sum_{n=1}^{\infty} np_n\right) - \lambda\left(1 - \sum_{n=1}^{\infty} p_n\right). \tag{2}$$

Solving the equation $\mathcal{L}' = 0$ yields

$$p_n = (\mathrm{e}^{\beta} - 1)\mathrm{e}^{-\beta n},\ \beta = \log \quad \left(\frac{\bar{n}}{\bar{n}-1}\right),\ \mathrm{e}^{-1-\lambda} = \mathrm{e}^{\beta} - 1, \tag{3}$$

then, the maximum entropy corresponds to a curve in "thermodynamic equilibrium", and can be written in terms of $\bar{n}$ as

$$H_{\max} = \log \quad (\bar{n}) + \frac{\beta}{\mathrm{e}^{\beta} - 1} = \bar{n} \log \quad (\bar{n}) - (\bar{n} - 1) \log \quad (\bar{n} - 1). \tag{4}$$

In quantum thermodynamics, p_n usually represents the probability that a system of particles (e.g., atoms or molecules) is in the discrete energy level E_n. Furthermore, the classical definition of temperature results $T = k\beta^{-1}$, where k is the Boltzmann constant (hereafter taken equal to one). In the present context the temperature of a curve can be defined using this analogy, that is

$$T = \frac{1}{\beta} = \left[\log\left(\frac{\bar{n}}{\bar{n}-1}\right)\right]^{-1}. \tag{5}$$

When the temperature vanishes ($T = 0$), the curve freezes to a straight line $\bar{n} = 1$. Furthermore, the entropy also vanishes ($H = H_{max} = 0$), which agrees with classical thermodynamics, for which, at zero temperature the entropy of the system vanishes.

In an attempt to push further the analogy with physics, we define the pressure (P) of a curve in terms of its entropy, Eq. (1), and its temperature, Eq. (5), in analogy with the thermodynamics of ideal gases, that is:

$$H = \frac{\gamma}{1-\gamma} \log \quad T + \log \quad P, \tag{6}$$

where γ is the ratio of specific heats, and the universal gas constant is set to unity. The motivation of Eq. (6) is that the spatial configuration of random intersections are analogous to spatial configurations of point particles. P and γ will be taken as parameters that, hopefully, remain the same for a given class of curves (e.g., with similar shape). Note that $\gamma(1-\gamma)^{-1}$ is the slope of the linear approximation in a ($\log T$ vs. H) plot.

Particularly, the original theory of *thermodynamics of plane curves*, as presented by Mendès France [1], relies on the Cauchy–Crofton theorem [3], which states that, for a planar curve Γ, the expected number of intersections between Γ and a random line, is given by

$$\bar{n} = \sum_{n=1}^{\infty} np_n = \frac{2\ell}{C}, \tag{7}$$

where ℓ is the length of Γ and C is the perimeter of the convex hull of Γ. Expression (7) allows analytical computation of H_{max} and T for a given planar curve Γ. Unfortunately, the lack of an extension of the Cauchy–Crofton theorem to higher dimensions has limited the theory to the plane. Nonetheless, observe that given p_n, for example obtained from numerical simulations, the thermodynamics can be defined for any curve in any dimension, which is the matter of this work.

2.1. An extended framework for the thermodynamics of curves

The most abstract setting of the curve thermodynamics framework depends merely on a discrete probability distribution function (DPDF). In Section 2, the DPDF accounts for the number of intersections between a 3D curve and a random plane. The procedure to obtain a generalized thermodynamic characterization for a given curve Γ is as follows:

i. Choose a random variable (X) associated to the geometry of the curve, e.g. the number of intersection points of Γ with random planes. Note that in the context of the thermodynamics analogy, $X \in \mathcal{G} \subset \mathbb{R}$ represents the "energy levels" of the curve. Here, $\mathcal{G}$ represents the subset of admissible energy levels.
ii. Compute the probability distribution, $p(X, \Gamma) = p$, for the given curve Γ.
iii. Calculate curve descriptors based on the probability function, for example
   i. Statistical moments of p, such as the mean.
   ii. Entropy (H) of p. In this work Shannon's entropy is used, but other definitions like [16] or [17] entropies may be used as well.
   iii. Using the mean, thermodynamic descriptors can be calculated through equations (1), (5) and (6).

As said, the generalization of the thermodynamic framework is done by defining X as the intersection count between the curve Γ and random planes. This generalization allows direct use of thermodynamics descriptors defined in Eqs. (1), (5) and (6), since the probability distribution p and the mean value of X, hereafter $\bar{n}$, are known. Nonetheless, estimation of p requires a numerical simulation because no analytical result is available. The following methodology generalizes the computation of p for any parametric function, not just planes.

i. Choose a parametric function $\{\mathcal{F}(\mathbf{m}) \in \mathcal{F}^N \mid \mathcal{F}: \mathbb{R}^N \to \mathbb{R}^3\}$, where $\mathbf{m} \in \mathbb{R}^N$ is the vector of parameters. And $\mathcal{F}^N$ is the space of all parametric functions with N parameters. Particularly, for planes $N = 3$.
ii. Define a functional operator $\{\mathcal{O}(\Gamma, \mathcal{F}) \in \mathcal{O} \mid \mathcal{O}: \mathcal{F}^1 \times \mathcal{F}^N \to \mathcal{G}\}$. Where $\Gamma \in \mathcal{F}^1$ is a curve parameterized by arc length; $\mathcal{F} \in \mathcal{F}^N$ is the test parametric function; $\mathcal{O}$ is the functional operator which retrieves an admissible energy level $X \in \mathcal{G}$.
iii. Define an exploration set of parameters $S = \{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_J\}$. In theory, S would be an infinite set. In practice, the cardinality of S is finite and equals J. The selection of S must ensure an homogeneous spatial distribution of the parametric function. This means that no spatial region of $\mathbb{R}^3$ should be privileged with a higher density of functions covering the region.
iv. Compute a histogram for the random variable X, hereafter called hist(X). This is performed by applying the functional operator to the complete set of parametric functions spanned by S. That is, $X_j = \mathcal{O}(\Gamma, \mathcal{F}(\mathbf{m}_j))$, $j = 1, 2, \ldots, J$.
v. Compute p from hist(X).

The thermodynamics of a curve, as defined in this work, depends on the DPDF of a random variable X. When X stands for the number of intersections between a planar curve and a random line, the thermodynamics proposed by Mendés is recovered. In this section we exploit spatially distributed information which can be associated to the curves, e.g. curvature, to construct a probability distribution and use it to compute thermodynamic descriptors. In this work, and for simplicity, we assume that a curve is described by a polyline (a set of points linked through line segments). Such DPDF is computed in four steps:

i. Define a spatial partition of the curve into M buckets, called B_i, $i = 1, \ldots, M$.
ii. List the buckets in ascending order according to the natural order of the curve.
iii. For a given spatially distributed feature f, e.g. local curvature, compute the cumulative value (F_i) of f for each bucket B_i, $i = 1, \ldots, M$, namely

$$F_i = \sum_{j=1}^{L_i} f_j,$$

for a given bucket B_i, containing L_i points that describe, discretely, the curve.
iv. Compute the discrete pseudo-probability distribution

$$p_i = F_i \left(\sum_{j=1}^{M} F_j \right)^{-1}. \tag{8}$$

Note that in this case, the discrete random variable, $X = B_m$, represents the bucket index with expected value $\bar{B}_m = \sum i p_i$. Given p_i and $\bar{B}_m$, the entropy, temperature and pressure as defined in Eqs. (1), (5) and (6) can be calculated.

Table 1
Statistical description of lesions among the arterial samples. Lesions are characterized according to [28].

Lesion characteristic	RCA ($n = 28$)	LAD ($n = 39$)	LCx ($n = 15$)
Stenosis grade			
Minimal, %	39.2	15.4	33.3
Mild, %	42.9	56.4	53.3
Moderate, %	14.3	17.9	6.7
Severe, %	3.6	10.3	0
Occluded, %	0	0	6.7
Position			
Proximal, %	60.7	53.8	53.3
Middle, %	21.4	33.4	33.4
Distal, %	17.9	12.8	13.3
Tissue type			
Soft tissue, %	21.4	10.3	6.7
Calcification, %	35.7	46.1	40.0
Mixed, %	42.9	43.6	53.3

3. Application case: human coronary arteries

Coronary artery disease (CAD) is the leading cause of dead worldwide, over the years, several risk factors were identified, namely: (i) Systemic factors; (ii) Biomechanical predictors and (iii) Geometrical risk factors. Particularly, the later type of risk suggests that the geometric variability of the human vasculature contributes to the development of atherosclerosis [18]. The amount of CAD explained by systemic risk factors is still controversial, mainly because they do not explain the localization and distribution of atherosclerosis. These medical observations are the motivation behind the research conducted on biomechanical and geometrical risk factors.

Up to date, the shape of the coronary arteries is not to be a known risk factor for CAD. In turn, there is evidence that the shape of the RCA is influential [19–22]. Such studies relay on invasive angiography images and manual measurements to classify the RCA shape.

The proposed methodology is applied in this section to characterize human coronary arteries. Then classification examples are presented targeting the discrimination of healthy/disease arteries and also the classification of the shape of the right coronary artery, i.e. C or Σ according to [20]. Potential applications in clinical practice include the development of objective indexes for CAD geometric risk factors assessment. In such scenario, given the arterial model of a patient, an automatic characterization of the geometry in terms of descriptors is performed and a risk stratification is derived based on the classification techniques presented here, or in similar clusterization strategies.

Note that geometric analysis of vascular structures is an active research field for the generation of atlas-like morphometrics, and the study of genesis and progression of cardiovascular diseases, e.g. the analysis of cerebral aneurysms [23–26].

3.1. Description of the dataset

A detailed description of the patient sample can be found in [27]. The sample consist of 48 patients (24 sibling pairs), all of which have at least one major systemic risks factor, i.e. smoker, hypertension, diabetes or dyslipidemia. This study pay particular attention to the main coronary arteries: left anterior descending (LAD), left circumflex (LCx) and right coronary artery (RCA).

The presence of arterial lesions is specified by physicians as described in [27]. From the 48 patients, 12 LCx, 19 RCA and 27 LAD vessels were found to be diseased. Table 1 summarizes the type, location and severity of all lesions for each vessel.

A specialized cardiologist classified the RCA shape using the criterion defined in [22], namely, Σ-Shape arteries have an indentation of at least the width of the artery in at least one view angle. If an RCA did not comply with such definition, it is classified as C-Shape. Two healthy RCA vessels that could not be classified by the expert were excluded from the study in Section 3.2.2. The classification performed by the expert resulted in 16 Σ- and 30 C-Shape RCAs.

All arterial models were generated from Coronary Computed Tomography Angiography (CCTA) images. The acquisition protocols and the processing methodology were described in [27]. Arteries in 3D space are modeled by the corresponding centerlines, which are computed following [29]. In this work, centerline is a synonymous for the arterial skeleton. Centerlines are curves in 3D space represented by polylines and, for the present context, contain spatially distributed information (e.g. lumen radius, curvature and torsion). For each arterial centerline, six thermodynamic descriptors are calculated, three computed from intersection counting and three using point-wise curvature.

The temperature (T_n), entropy (H_n) and pressure (P_n) are derived from intersection counting. Recalling Section 2.1: (i) the parametric functions are intersecting planes, and so $N = 3$. (ii) The functional operator is the operation through which the number of intersections (n) between a plane and the curve is counted. The number of intersections is in this case the chosen random variable ($X = n$). (iii) The set of parameters S, in this work, has cardinality $J = 440{,}000$ and was defined from a regular grid containing all possible curves. A series of rotations of the grid in azimuthal and polar angles is performed and for each grid configuration, the normal vectors to the planes are also rotated in both directions. This procedure ensures an homogeneous density of planes in the region where the curves are defined.

On the other hand, a different triad of temperature (T_κ), entropy (H_κ) and pressure (P_κ) is derived from the point-wise distribution of curvature. Recalling Section 2.1: (i) the centerline is divided into $N = 20$ equally sized buckets; (ii) the bucket order starts at the arterial ostium (inlet) and increases following the flow direction.

In addition to the six thermodynamic descriptors, arterial centerlines can be characterized by several geometric descriptors available in the literature, here called conventional features. In this work, for any given artery A, a set of 19 descriptors is used: branch count (Υ), length (ℓ), tortuosity (χ), mean radius ($\bar{r}$), mean ($\bar{\kappa}$) and total (κ_T) curvature, mean ($\bar{\tau}$) and total (τ_T) torsion, mean ($\bar{\zeta}$) and total (ζ_T) combined curvature, aspect ratio (Λ_r), curvature ratio (Λ_κ), torsion ratio (Λ_τ), curvature energy (ξ_κ), torsion energy (ξ_τ), fractal dimension (ϱ), rising angle (α), offspring mean raising angle (β), average distal curvature ($\bar{\kappa}_d$). The reader is referred to [27,32] for more details.

For each arterial type (LAD, LCx and RCA), features were normalized to have a sample-wide unit variance and zero mean.

3.2. Statistical analysis

To assess the suitability of the proposed thermodynamic descriptors, the following statistical analysis was performed:

i. Perform a Mann–Whitney U-test to detect a set of geometric descriptors with statistically significant difference in the mean value of groups. This test produces the first assessment of those descriptors which better classify clinical groups.
ii. Use multivariate kernel density estimation for posterior probability-based classification, parameterized with zero mean and 0.75 standard-deviation. Kernel density estimation is an unsupervised learning procedure, which leads naturally to non-parametric classification. A good review of these methods can be found in [31], Chapter 6.6. Furthermore, such classification strategy intrinsically provides a risk stratification mechanism based on the classification probability. Also, it out-performed the standard linear classification (results not reported here).

Taking into account that the size of the available data sample is small and that trained classification algorithms often produce overoptimistic assessments, the resulting classifications are evaluated by means of a leave-one-out cross validation (LOOCV) test.

The groups, or classes, are the following: healthy/diseased for the applications presented in Section 3.2.3, and C/Σ shapes for the RCA shape classification presented in Section 3.2.2.

The classification of the different arterial samples (RCA, LAD and LCx) using posterior probabilities computed with multivariate kernel density estimation, requires the definition of a feature set ($\mathcal{S}$). Choosing this set such that maximizes the classification score is known as the feature selection problem. We tested standard sequential (forward and backward) feature-selection strategies, also the features sets containing the complete set (25 features), only geometric (19 features), only thermodynamic (6 features), and the one specified by the U-test for each artery. All of these sets resulted in suboptimal classification scores when compared with a "brute force" search of a 6-feature set that maximizes the average of the LOOCV classification scores. We call such feature set $\mathcal{S}_{B6}$ and, for the sake of clarity, we only present results using such set. In our tests, the maximum LOOCV score presented here was reached by several sets of 6 features, we randomly selected one set for each artery to present here.

Convergence tests were performed to ensure that the computational methodology presented in Section 2.1 does not introduce errors in the statistical analysis of the thermodynamics variables. For the parameters used in this work, it was found that the mean relative difference in all thermodynamic variables for all the 144 curves was significantly higher (30–60 times) than the mean of relative convergence difference for successive planes of intersection. These results ensures that errors due to the iterative procedure used in the numerical computation of thermodynamic variables are negligible compared to the differences among different curves.

3.2.1. Association among curve descriptors

The associations between the complete feature set can be assessed by identifying significant correlation between descriptors in each arterial sample (LAD, LCx and RCA). Fig. 1 presents significant ($p < 0.01$) feature correlations using arc plots. The Spearman correlation coefficient was used for continuous and ordinal variables. In such plots, red and blue links stand for positive and negative correlations.

The following association patterns can be recognized in the arc plots of Fig. 1:

- Thermodynamics descriptors are not correlated neither to the arterial rising angle (α), the mean offspring rising angle (β) nor the number of visible branches (Υ). The mean radius ($\bar{r}$) only correlates with one thermodynamic feature (P_κ) in the RCA. The thermodynamics descriptors are more correlated to curvature-derived features than to torsion-derived ones.
- All features related to the curvature showed significant correlation.
- No association between lesion number (η) and any geometric feature can be observed, for the chosen level of significance.
- H_κ and P_κ showed negative correlation.

Most of the feature associations are reasonable because of the underlying definition based on similar point-wise features, e.g. $\bar{\kappa}$, κ_T, $\bar{\zeta}$, indicating that a single geometric aspects is characterized by several descriptors. Besides, features of the RCA are more cor-

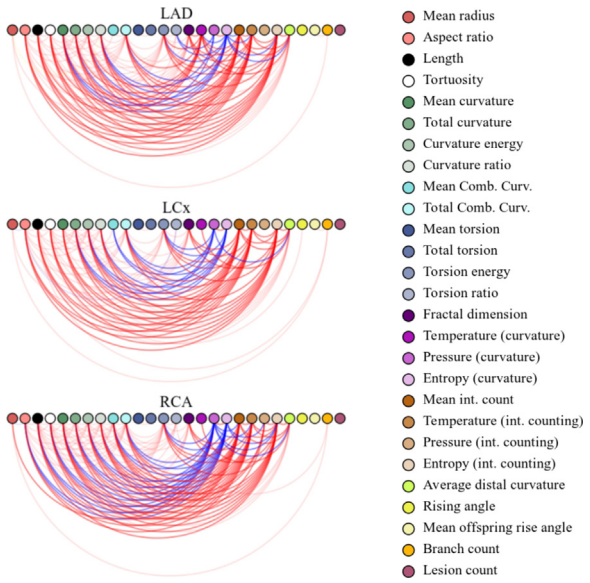

Fig. 1. Arc plots showing significant Spearman's correlations (red for positive and blue for negative correlation coefficients). Correlations involving thermodynamic descriptors are highlighted. On top the LAD artery, in the middle the LCx and at the bottom the RCA artery. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

related than those of the LAD and LCx, the latter presenting fewer correlations. Nevertheless, there are some relationships that are characteristics of each artery, namely:

- Curvature-derived thermodynamics present different association patterns in the three arteries, i.e. few correlations in the LCx, more positive correlations involving T_κ in the LAD, and more negative correlations involving P_κ in the RCA.
- The rising angle (α) only correlates in the LCx.
- In the LCx and LAD, the number of visible (large) branches Υ is positively correlated to the mean rising angle of the offspring (β). Although Υ increases with $\bar{r}$ in the three arteries, only in the LCx a strong correlation exists between Υ and ℓ.
- An expected negative correlation is present between $\bar{r}$ and Λ_r in the RCA and LCx, but it is not present at the LAD with a significant level of statistical confidence.
- The average distal curvature $\bar{\kappa}_d$ presents more associations with other features in the LAD than in the other arteries.

Association patterns given by correlation analysis suggest redundancy in the feature characterization. A principal component analysis (PCA) was performed for each arterial set (LAD, LCx and RCA). Regarding the LAD artery, it was found that the first two principal components explained 50.0% of the variability, and 96.7% is explained by the first 12 principal components. Analogously, the first two principal components for the LCx dataset explained 49.8% of the variability, and 95.7% is explained by the first 11 principal components. Finally, for the RCA artery, the first two principal components explained 56.6% of the variability, while 96.2% is explained by the first 11 principal components.

Generally, the $T_n, H_n, \ell, \chi, \kappa_T, \zeta_T, \bar{\zeta}, \xi_\kappa, \xi_\tau$ and ϱ contribute to the first two principal components in at least one artery. Such contribution is assessed by significant correlation (Spearman's $\rho > 0.8$, $p < 0.01$) between the principal component and the features.

3.2.2. Association of descriptors to the shape of the RCA

In this section, the association between RCA shape and thermodynamic conventional features is explored. The degree of association of thermodynamic and geometric descriptors to RCA shape was tested by means of the standard Mann–Whitney U-test. Fig. 2 presents a box-plot of thermodynamic features with statistically significant difference in the mean values between C-Shape and Σ-Shape RCAs ($p < 0.05$). See caption of Fig. 2 for a complete list of features whose mean values are significantly different. Feature values have been normalized to have zero mean and unitary standard deviation. The following observations can be made:

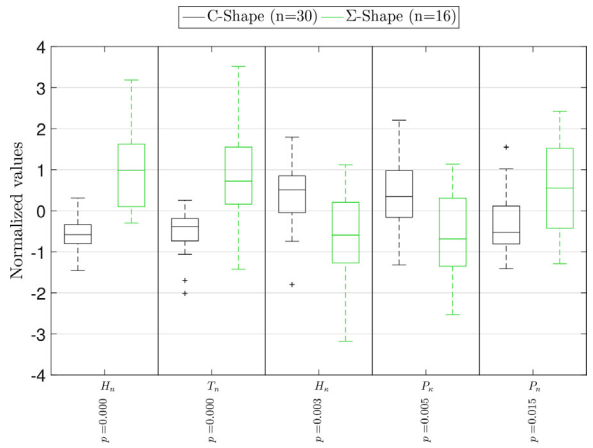

Fig. 2. Subset of thermodynamic features for which difference in mean values for the C and Σ shaped RCA is statistically significant ($p < 0.05$). Features are sorted from left to right in ascending order of the associated p-value. The following set of conventional features also presented significant differences between the shape classes: $\bar{\kappa}, \kappa_T, \Lambda_\kappa, \xi_\kappa, \bar{\zeta}, \zeta_T, \xi_\tau, \ell, \chi, \varrho, \Lambda_\tau, \bar{\kappa}_d$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Multivariate kernel density classification scores for LOOCV. The prevalence indicated in parentheses, of the Σ-shape is 34.8%.

	$\mathcal{S}_{B6}$ (34.8%)
Acc.	0.96
Sen.	0.88
Spe.	1.00
PPV	1.00
NPV	0.94

- Regarding the thermodynamic descriptors introduced in this work, it was found that all, except for curvature-derived temperature (T_κ), are associated to the RCA shape. Particularly, the entropy associated to the number of intersections between the curve and planes (H_n) obtained the second smallest p-value.
- With the exception of the entropy derived from curvature (H_κ) and the associated pressure (P_κ), all the other feature values are greater for the Σ-Shape group.
- Although the mean radius ($\bar{r}$) is not directly associated to the RCA shape, the curvature and torsion ratios (Λ_κ, Λ_τ), which combine spatially distributed radii with curvature and torsion respectively, are different between groups.
- All the features associated with the curvature are strongly associated to the RCA-shape.

Table 2 presents the performance of the multivariate kernel density classification using LOOCV. The brute-force search yields $\mathcal{S}_{B6} = \{\bar{\kappa}, \zeta_T, \Lambda_r, \Lambda_\tau, T_n, H_n\}$.

3.2.3. Association between geometry and lesions

Fig. 3 depicts the DPDF of the number of intersections between planes and centerlines (3b), and of the bucket index for spatially distributed curvature of centerlines (3c). This visual example compares a healthy LAD (blue) with a diseased LAD (red), for which the corresponding centerlines are presented in Fig. 3a. It can be seen

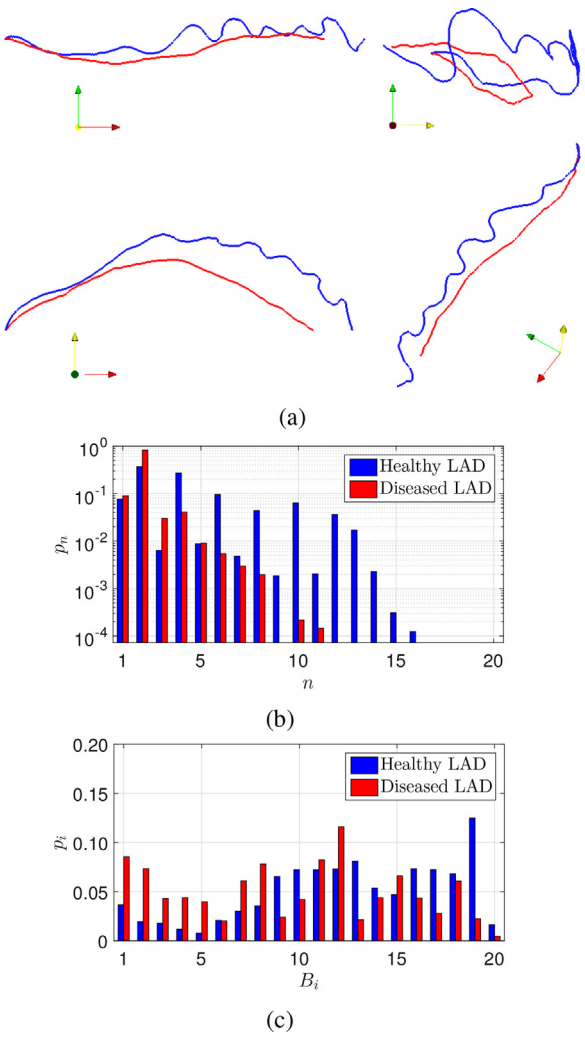

Fig. 3. Centerlines of a healthy (blue) and a disease (red) LAD from different perspectives. The plots show the discrete probability distributions for the number of intersection points, in log scale (b) and for the bucket index partition of curvature, B_i (c). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that both types of DPDF's reflect qualitative differences of the vessel shape in different ways. On the other hand, Fig. 4 presents scatter plots for log(T) vs. H, with the associated linear approximation (black) used to compute the pressure using Eq. (6).

Association between lesion presence and descriptors can be detected with the Mann–Whitney U-test. The LAD artery shows significant association of disease with descriptors. Fig. 5 presents box-plots of descriptors with significant ($p < 0.05$) difference in the mean value between healthy and diseased LAD according to the U-test. Feature values have been normalized to have sample-wide zero mean and unit standard deviation. The results indicate that:

- Regarding thermodynamic descriptors, the curvature-based temperature (T_κ) and the intersection counting temperature (T_n) are larger in healthy vessels than in diseased ones.
- The average distal curvature ($\bar{\kappa}_d$), scores the best results in the U-test, [32].
- A few conventional features were found to be associated to lesion presence: curvature energy (ξ_κ), branch count (Υ), fractal dimension (ϱ), arterial length (ℓ), total curvature (κ_T) and offspring mean rising angle (β).
- All the associated features present smaller mean values for disease LAD vessels than for healthy ones.

Regarding the RCA, no statistically significant (95% confident interval) association was found between lesion presence and thermodynamic/geometric features. On the other hand, the LCx presented association between disease and two features: intersection counting pressure (P_n) is larger in healthy vessels and torsion ratio (Λ_τ) is larger in diseased vessels.

Table 3 presents the performance results of the kernel density classifiers, when using LOOCV. The performance is assessed in terms of the accuracy (Acc.), sensitivity (Sen.), specificity (Spe.), positive and negative predictive values (PPV and NPV). As explained in Section 3.2, a brute force search for the 6 features maximizing classification output (summation of performance indicators), was carried out. Such search resulted in: (i) $\mathcal{S}_{B6}^{\text{RCA}} = \{\zeta_T, \Lambda_\kappa, \xi_\tau, \beta, H_\kappa, H_n\}$; (ii) $\mathcal{S}_{B6}^{\text{LAD}} = \{\kappa_T, \bar{\zeta}, \Lambda_r, \Lambda_\tau, \xi_\kappa, \beta\}$; (iii) $\mathcal{S}_{B6}^{\text{LCx}} = \{\ell, \bar{\zeta}, \Lambda_\tau, \Upsilon, P_n, H_n\}$.

3.2.4. Discussion

The large number of significantly correlated descriptors indicates some degree of redundancy in the characterization. Nevertheless, not all correlated features are useful in the discrimination of classes (see U-test results for both classification examples). Such result strengthens the value of the new thermodynamic features, suggesting possible applications of dimensionality reduction techniques. Particularly, PCA showed that two principal components are needed to explain ~50% of the variability and ~11 components are needed to explain ~95% in every artery. Furthermore, the T_n and

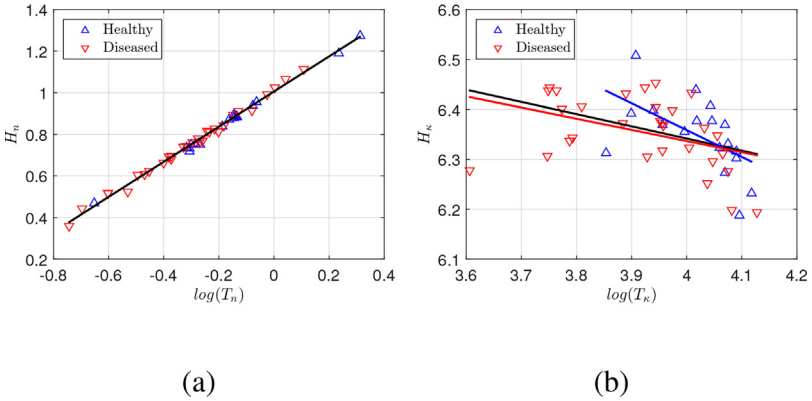

Fig. 4. Entropy (H) as a function of the logarithm of temperature (T), for thermodynamic indexes derived from intersection count (a) and from bucket partition of curvature (b). Scatter points represents healthy (blue) and diseased (red) LAD. Linear approximations for the entire sample are shown in black, and in blue/red for healthy and diseased sub-samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

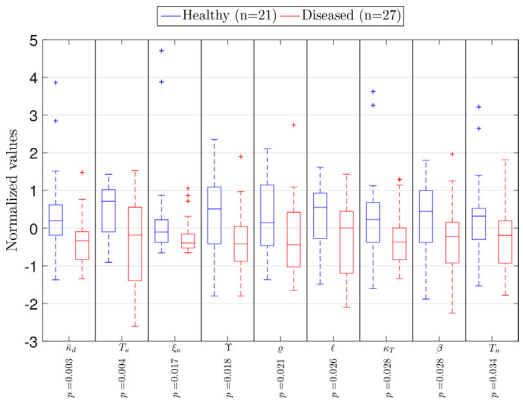

Fig. 5. Subset of features for which difference in mean values for the healthy and diseased LAD arteries are statistically significant ($p < 0.05$). Features are sorted from left to right in ascending order of the associated p-value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Multivariate kernel density estimation classification scores for LOOCV. The prevalence of the disease is indicated in parentheses for each type of artery.

	$\mathcal{S}_{B6}^{\text{RCA}}$ (41.3%)	$\mathcal{S}_{B6}^{\text{LAD}}$ (56.3%)	$\mathcal{S}_{B6}^{\text{LCx}}$ (26.1%)
Acc.	0.78	0.83	0.88
Sen.	0.74	0.89	0.50
Spe.	0.81	0.76	1.00
PPV	0.74	0.83	1.00
NPV	0.81	0.84	0.86

H_n are strongly correlated with the first two principal components in the three arteries.

Albeit the modest size of the sample calls for further studies to confirm the present findings, some preliminary conclusions can be drawn from the results. It is worth noting that the sets that maximize classification for diseased LCx and RCA arteries include thermodynamic descriptors, namely H_κ and H_n for LCx and P_n and H_n for the RCA. The analysis of RCA shape descriptors identified 17 features with significant difference in the mean value among classes, five of which are thermodynamic descriptors. In turn, classification scores show 96% of correctly identified shapes (accuracy), 88% of correctly classified Σ-shaped RCAs (sensitivity), 100% of correctly identified C-shaped RCAs (specificity), none incorrectly classified C-shape (100% of PPV), and a low number of Σ-shape classified as C-shape (94% of NPV). Furthermore, two thermodynamics descriptors (T_n, H_n) are included in the $\mathcal{S}_{B6}$. Another interesting preliminary result is that H_n showed significant performance to differentiate arteries with lesions.

4. Study limitations and future directions

It should be remarked that in the present study the patient sample is small to perform conclusive statistical analysis, but at the same time, it is sufficiently large to serve as proof-of-concept for the suitability of combined thermodynamic and conventional descriptors. Also, since the sample consists of siblings, bias is a possibility due to heritable coronary geometry [30]. However, despite the familial relationship present in the dataset, it was shown in [27] that the patient sample is representative in terms of anatomical description of the coronary vasculature, i.e. circulation dominance, vessel radii and arterial occurrence/distribution.

Class imbalance could be considered as a source of low sensitivity for the diseased LCx classification, for which the prevalence of disease is ∼26%. The KDE classifier may be biased towards the major class, i.e. healthy arteries. Nevertheless, for the RCA shape classification, the ∼35% prevalence of Σ-shape is not a problem. Generally, problems associated with class imbalance are approached by specialized classification algorithms, feature selection or data pre-processing. The first two approaches require specialized algorithms that maximize the classification accuracy of the imbalanced class, but such study is beyond the scope of this work. Data pre-processing aims to balance the classes by random under-sampling the major classes or by synthetically over-sampling the minority classes. Due to the small sample size and complexity of generating artificial healthy/diseased arteries, such approaches are not an option.

In future studies with larger samples, dimensionality reduction techniques, e.g. PCA, could be used to explore classification in the transformed spaces. Incorporation of patient descriptors such as systemic risk factors could also help to improve classification of healthy/diseased arteries. The ultimate goal for this kind of analysis is to provide risk stratification indexes to develop CAD using patient specific coronary arterial models. As for the thermodynamic framework, it is general enough to be used in other fields of applications such as computer vision and shape classification, spatio-temporal trajectories, numerical series, handwriting and shape analysis and signal processing.

5. Conclusions

A generalization of the theory of *thermodynamics of plane curves* to curves in 3D space, as well as an adaptation for using spatially distributed information attached to the curves, was presented. The method was tested to characterize a set of coronary arterial models. It was found that some thermodynamic descriptors have different mean values ($p < 0.05$) for healthy/diseased LAD and LCx, as well as for Σ/C-shaped RCA arteries. Some of the thermodynamic descriptors contributed to the optimal feature sets. Particularly, identification of the shape of the RCA showed excellent results, and identification of geometric risk factors for CAD are promising. The new set of thermodynamic descriptors presented here proved to be a valuable complement to the conventional set of geometric descriptors in the context of characterizing human coronary arteries. Also, the fact that T_n and H_n strongly contribute to the two principal components of the PCA, indicates that these quantities contain valuable information to explain variability in the geometric characterization of the coronary arteries. The authors hope that the tools proposed here provide a suitable approach to further extend the range of applications of the thermodynamics of curves.

Acknowledgements

The support of Brazilian agencies CNPq and FAPERJ is gratefully acknowledged.

References

[1] M. Mendès France, Les courbes chaotiques, Courr. Centre Natl. Recer. Sci. 51 (1983) 5–9.
[2] Y. Dupain, T. Kamae, M. Mendés, Can one measure the temperature of a curve? Arch. Ration. Mech. Anal. 94 (2) (1986) 155–163.
[3] M.W. Crofton, On the theory of local probability, applied to straight lines drawn at random in a plane; the methods used being also extended to the proof of certain new theorems in the integral calculus, Philos. Trans. R. Soc. Lond. 158 (0) (1868) 181–199.
[4] R.R. Moore, A.J. van der Poorten, et al., On the thermodynamics of curves and other curlicues, in: Miniconference on Geometry and Physics, Centre for Mathematics and its Applications, Mathematical Sciences Institute, The Australian National University, 1989, pp. 82–109.
[5] G. Jumarie, A new information theoretic approach to the entropy of nonrandom discrete maps relation to fractal dimension and temperature of curves, Chaos Solitons Fractals 8 (6) (1997) 953–970.
[6] A. Denis, F. Crémoux, Using the entropy of curves to segment a time or spatial series, Math. Geol. 34 (8) (2002) 899–914.

[7] A. Denis, Identification of functional relationships between atmospheric pressure and CO_2 in the cave of Lascaux using the concept of entropy of curves, Geophys. Res. Lett. 32 (5) (2005).
[8] A. Balestrino, A. Caiti, E. Crisostomi, A classification of nonlinear systems: an entropy based approach, Chem. Eng. Trans. 11 (2007) 119–124.
[9] A. Balestrino, A. Caiti, E. Crisostomi, Generalised entropy of curves for the analysis and classification of dynamical systems, Entropy 11 (2) (2009) 249–270.
[10] J.-P. Allouche, L. Maillard-Teyssier, Inconstancy of finite and infinite sequences, Theor. Comput. Sci. 412 (22) (2011) 2268–2281.
[11] M.M. Dodson, M.M. France, M. Mendes, On the entropy of curves, J. Integer Sequences 16 (2) (2013) 3.
[12] C. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948).
[13] E.T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (4) (1957) 620–630.
[14] E.T. Jaynes, Information theory and statistical mechanics. II, Phys. Rev. 108 (2) (1957) 171–190.
[15] J.W. Gibbs, Elementary Principles in Statistical Mechanics Developed with Especial Reference to the Rational Foundation of Thermodynamics, C. Scribner, 1902.
[16] C. Tsallis, Possible generalization of Boltzmann–Gibbs statistics, J. Stat. Phys. 52 (1) (1988) 479–487.
[17] A. Rènyi, On measures of entropy and information, in: Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1961, pp. 547–561.
[18] M.H. Friedman, O.J. Deters, F.F. Mark, C.B. Bargeron, G.M. Hutchins, Arterial geometry affects hemodynamics. A potential risk factor for atherosclerosis, Atherosclerosis 46 (2) (1983) 225–231, ISSN 0021-9150.
[19] D. Dvir, R. Kornowski, T. Ben-Gal, M. Berman, B. Vidne, D. Aravot, Relation of amounts of narrowing to the length of the right coronary artery, Am. J. Cardiol. 90 (1) (2002) 46–48.
[20] D. Dvir, R. Kornowski, J. Gurevich, B. Orlov, D. Aravot, Degrees of severe stenoses in sigma-shaped versus C-shaped right coronary arteries, Am. J. Cardiol. 92 (3) (2003) 294–298, ISSN 0002-9149.
[21] R. Demirbag, R. Yilmaz, Effects of the shape of coronary arteries on the presence, extent, and severity of their disease, Heart Vessels 20 (5) (2005) 224–229, ISSN 0910-8327, 1615-2573.
[22] Y. Arbel, D. Dvir, M.S. Feinberg, R. Beigel, M. Shechter, The association between right coronary artery morphology and endothelial function, Int. J. Cardiol. 115 (1) (2007) 19–23, ISSN 01675273.
[23] M. Fisher, S. Fieman, Geometric factors of the bifurcation in carotid atherogenesis, Stroke 21 (2) (1990) 267–271, ISSN 0039-2499.
[24] H. Bogunovíc, J.M. Pozo, R. Cárdenes, M.C. Villa-Uriol, R. Blanc, M. Piotin, A.F. Frangi, Automated landmarking and geometric characterization of the carotid siphon, Med. Image Anal. 16 (4) (2012) 889–903, ISSN 13618415.
[25] T. Passerini, L.M. Sangalli, S. Vantini, M. Piccinelli, S. Bacigaluppi, L. Antiga, E. Boccardi, P. Secchi, A. Veneziani, An integrated statistical investigation of internal carotid arteries of patients affected by cerebral aneurysms, Cardiovasc. Eng. Technol. 3 (1) (2012) 26–40.
[26] S.N. Wright, P. Kochunov, F. Mut, M. Bergamino, K.M. Brown, J.C. Mazziotta, A.W. Toga, J.R. Cebral, G.A. Ascoli, Digital reconstruction and morphometric analysis of human brain arterial vasculature from magnetic resonance angiography, NeuroImage 82 (2013) 170–181, ISSN 1095-9572.
[27] C.A. Bulant, P.J. Blanco, T.P. Lima, A. Assunçãao, G. Liberato, J.R. Parga, L.F.R. Ávila, A.C. Pereira, R.A. Feijóo, P.A. Lemos, A computational framework to characterize and compare the geometry of coronary networks, Int. J. Numer. Methods Biomed. Eng. 33 (3) (2017) e02800.
[28] G.L. Raff, A. Abidov, S. Achenbach, D.S. Berman, L.M. Boxt, M.J. Budoff, V. Cheng, T. DeFrance, J.C. Hellinger, R.P. Karlsberg, SCCT guidelines for the interpretation and reporting of coronary computed tomographic angiography, J. Cardiovasc. Comput. Tomogr. 3 (2) (2009) 122–136, ISSN 19345925.
[29] L. Antiga, B. Ene-Iordache, A. Remuzzi, Computational geometry for patient specific reconstruction and meshing of blood vessels from MR and CT angiography, IEEE Trans. Med. Imaging 22 (5) (2003) 674–684.
[30] C. Bulant, P. Blanco, A. Clausse, A. Assunção, T. Lima, L. Ávila, R. Feijóo, P. Lemos, Association between three-dimensional vessel geometry and the presence of atherosclerotic plaques in the left anterior descending coronary artery of high-risk patients, Biomed. Signal Process. Control 31 (2017) 569–575.
[31] J. Friedman, T. Hastie, R. Tibshirani, The Elements of Statistical Learning, vol. 1, Springer Series in Statistics Springer, Berlin, 2001.
[32] C.A. Bulant, P.J. Blanco, A. Pereira, T.P. Lima, A.N. Assunção, G. Liberato, C.G. Bezerra, J.R. Parga, L.F. Ávila, R.A. Feijóo, P.A. Lemos, On the search of arterial geometry heritability, Int. J. Cardiol. 221 (2016) 1013–1021.