A Folding Index of 2D Curves

Yi-King Choi[1], Dongming Yan[2] and Wenping Wang[3]

[1]The University of Hong Kong, ykchoi@csis.hku.hk
[2]Tsinghua University, yandm@cg.cs.tsinghua.edu.cn
[3]The University of Hong Kong, wenping@csis.hku.hk

ABSTRACT

A new index is proposed for describing the degree of folding of a planar curve segment. Based on a basic concept from integral geometry, the Curve Folding Index (CFI) of a curve segment is defined to be the expected number of intersections that a random line has with the curve. The CFI provides a simple measure to characterize planar curves in terms of the degree of folding, and is shown to be invariant under the similitude transformations. We show by experiments that the CFI conforms to the intuitive perception of folding complexity.

Keywords: curve folding, curve matching, integral geometry

1. INTRODUCTION

Planar curves have extensively been studied in geometry. Many geometric properties of a curve, either local or global, e.g. curvature and length, are quantifiable and computable. Yet there are characteristics of a curve that is not easy to quantify. For example, it is of interest to characterize the degree of folding of a curve. We consider in this paper a new index to describe the degree of folding of a planar curve segment. For brevity, a curve segment will be called a curve.

Intuitively, there are many turns and twists in a highly folded curve. In contrast, an unfolded or mildly folded curve is one that is straightened out and is close to a straight line. Human cognition can exploit this conceptual information to differentiate and identify a curve from another. In applications such as curve matching, the degree of folding may serve as a good starting point to quickly eliminate some curves from considerations. There are also applications of folding complexity in 3D space, such as in the studies of protein folding [1]. In this case, we may model the backbone chain of a protein by a space curve.

Probability distributions obtained from geometric properties of objects have been used for 3D shape matching [4]. For a curve, the distribution of the number of intersections of randomly distributed lines intersecting the curve can be extracted. We define the Curve Folding Index (CFI) of a 2D curve by the expected number of intersection points that a random line has with the curve. While the distribution encapsulates a more comprehensive view on the degree of folding of a curve, the CFI gives a handy single-number index to represent the distribution.

2. PRELIMINARIES

We first review some basic facts from integral geometry [5] about lines in 2D plane. Integral geometry is the study of the measure of a set of geometric figures and is closely related to combinatorial geometry, convex geometry, and geometric probability. A basic definition, called *measure*, gives a description for a set of geometric entities such as points, lines, chords of curve, etc., which is invariant under the group of rigid motions.

Consider a set of straight lines S in E^2. Each straight line G can be represented by its normal coordinates (p,ψ) and the equation of G is $x\cos\psi + y\sin\psi - p = 0$. The *density* of S is given by

$$dG = dp \wedge d\psi \ ,$$

where $\wedge$ stands for the exterior product of two differential forms. The *measure* of S is defined by

$$m(\mathrm{S}) = \int_{\mathrm{S}} dG \tag{1}$$

which is invariant under Euclidean transformations in E^2 up to a constant factor.

The density of a set of straight lines intersecting a curve is shown [5] to be equal to

$$dG = |\sin\theta|\, ds \wedge d\theta,$$

where θ is the angle between a line G and the tangent at a point p, and ds is the differential arc element at p on the curve. For any rectifiable curve C, i.e. a curve of finite length, there is

$$\int n\, dG = 2l\,, \tag{2}$$

where n is the number of intersection points each straight line has with the curve C and l is the length of C. This is the Cauchy-Crofton formula [2]. Since a straight line must either intersect a convex closed curve K at exactly two contact points or it does not intersect K at all, we have

$$\int_{G\cap K\neq\varnothing} dG = l \tag{3}$$

where K is a convex closed curve of length l. This means that the measure of the set of straight lines which intersect a convex closed curve equals to the length of the curve. Detailed derivation of the above formulas can be found in [5, 6].

Suppose a curve C of length l_C is enclosed by a convex closed curve K of length l_K. If we consider all lines intersecting K, the expected number of intersection points of these lines with C would be given by

$$\bar{n} = \frac{\int n\, dG}{\int dG} = \frac{2l_C}{l_K}\,. \tag{4}$$

Consider a set L of N lines that are randomly sampled from the set of lines intersecting a convex closed curve K that encloses C. Suppose that the length l_K of the curve K is known. Let n_K and n_C be the total number of intersection points of L with K and C, respectively. By Eq. (2), with integration approximation, we have

$$\frac{n_K}{N} \approx 2wl_K \quad \text{and} \quad \frac{n_C}{N} \approx 2wl_C$$

where w is a constant of proportionality. Therefore,

$$l_C \approx \frac{n_C}{n_K} l_K\,.$$

This gives us a combinatorial way of computing the length of a curve C. The extension to this idea to computing surface area is presented in [3].

3. THE 2D CURVE FOLDING INDEX

We give the definition for the 2D Curve Folding Index, which provides a quantitative description of the degree of folding for a 2D curve.

Definition: *Given a curve C in E^2, the Curve Folding Index (CFI) of C, denoted by $fi(C)$, is defined as the expected number of intersections with C of all straight lines intersecting C, i.e.*

$$fi(C) = \frac{\int n\, dG}{\int_{G\cap C\neq\varnothing} dG}\,.$$

The CFI of a 2D curve is invariant under Euclidean transformations due to the same property possessed by the measure of a set of straight lines. We also have the following

Proposition 1 *The CFI of a curve is invariant under uniform scaling.*

PROOF. Let C be a curve of length l_C. Let C' be a scaled copy of C obtained by uniform scaling of factor k such that its length is $l_{C'} = k\, l_C$, where $k > 0$ is a constant. Let H_C of length l_{H_C} be the closed boundary curve of the convex hull of C and let $H_{C'}$ of length $l_{H_{C'}}$ be the closed boundary curve of the convex hull of C'. The measure of lines intersecting C is equal to the measure of lines intersecting H_C, and therefore

$$\int_{G\cap C\neq\varnothing} dG = \int_{G\cap H_C\neq\varnothing} dG = l_{H_C}\,,$$

where the second equality follows from Eq. (3). Similarly, we have $\int_{G\cap C'\neq\varnothing} dG = l_{H_{C'}}$. Also, by Eq. (2), we have $\int n_C\, dG = 2\,l_C$ and $\int n_{C'}\, dG = 2\,l_{C'}$. Hence,

$$fi(C') = \frac{\int n_{C'}\, dG}{\int_{G\cap C'\neq\varnothing} dG} = \frac{2\,l_{C'}}{l_{H_{C'}}} = \frac{2\,k\,l_C}{k\,l_{H_C}}$$
$$= \frac{2\,l_C}{l_{H_C}} = \frac{\int n_C\, dG}{\int_{G\cap C\neq\varnothing} dG} = fi(C),$$

i.e. the CFI of a curve is invariant under uniform scaling.

Hence, the CFI of a curve is invariant under the similitude transformations, i.e. Euclidean transformations together with uniform scaling.

Since $n = 0$ for any line G without an intersection with C, the CFI is different from $\bar{n}$ of Eq. (4) in that the index ignores those straight lines not intersecting the curve under consideration; otherwise, its value would only depend on the curve length.

Intuitively, the CFI of a curve C gives the expected number of intersections that a straight line segment may have with C by considering only those lines that are in contact with C. In general, the CFI of a highly tangled curve would be greater than the CFI of a less tangled curve; and in particular, the CFI of a straight line is 1 and that of a convex closed curve is 2.

4. COMPUTING THE INDEX

To compute the CFI of a 2D curve, we use a similar method to that for computing the curve length as described in section 2. Given a 2D curve C, the algorithm for computing $fi(C)$ is as follows:

1. Determine a bounding circle B that encloses C.
   The role of B is to assist in generating a set of lines which intersect the curve C.
2. Generate a set L of N random lines that intersect the bounding circle B.
   The lines in L sample the set $\bar{\mathrm{L}}$ of all lines intersecting B.
3. Compute the number of lines (n') in L that intersect C and the total number of intersections (n) of these lines with C.
4. Compute the CFI of the curve C. By integration approximation, the CFI of the curve C is given by

   $$fi(C) \approx \frac{n}{n'}.$$

It is crucial that the set L of N lines is a good sampling of the set $\bar{\mathrm{L}}$ of all lines intersecting the bounding circle B, so as to ensure that the computed CFI is invariant under the similitude transformations. The approximation error introduced in the last step of the algorithm depends not only on the number of lines intersecting C (and therefore the number of lines generated in L), but also on whether the lines in L are evenly distributed. In our testing, we employ the *chord model* [7] in which a random line is defined by its two end points which are uniformly distributed points on a circle B. Therefore, all these lines are guaranteed to intersect the circle B and they are shown to be uniformly distributed [7].

5. DISCUSSIONS

In this section, we show how the CFI and the intersection distribution are used to describe the degree of folding of various curves.

Fig. 1 shows four curves with different degrees of tangling. The CFI and the intersection distributions are also given alongside with the corresponding curves. In computing the CFI of each curve presented here, a total of 100,000 random and uniformly distributed chords of a bounding circle are generated. From the figures, we see that the CFI generally reflects the degree of folding of a curve: the CFI of a highly folded curve is greater than the CFI of a mildly folded one. However, it should be noted that although the CFI is a useful indicator to characterize the distribution of intersection points, there are other statistical characteristics about the distribution, e.g. variance, that a single index may not represent. For example, consider two curves C and C' of the same length. Then by Eq. (2), the total number of intersections that they have with all the lines in the plane would be the same. Now, if the perimeters of the convex hulls of the two curves are also of the same length, it is easy to show by Eq. (3) that the measures of the lines intersecting the two curves are the same as well. In this case, the two curves have the same CFI, no matter how different the degrees of folding they may possess within their convex hulls (Fig. 2). This example shows that while the CFI allows us to have a glance at the degree of the curve folding, a better understanding can be gained by a detailed analysis of the intersection distribution.

More properties of a curve are revealed by its intersection distribution. In Fig. 3(a) and (b), we have two similar curves with the only difference that one is a close curve and the other one is open. The intersection distributions of the two curves are quite different in that the frequencies of the odd number of intersections are all zero for the close curve. It indeed reveals the fact that a

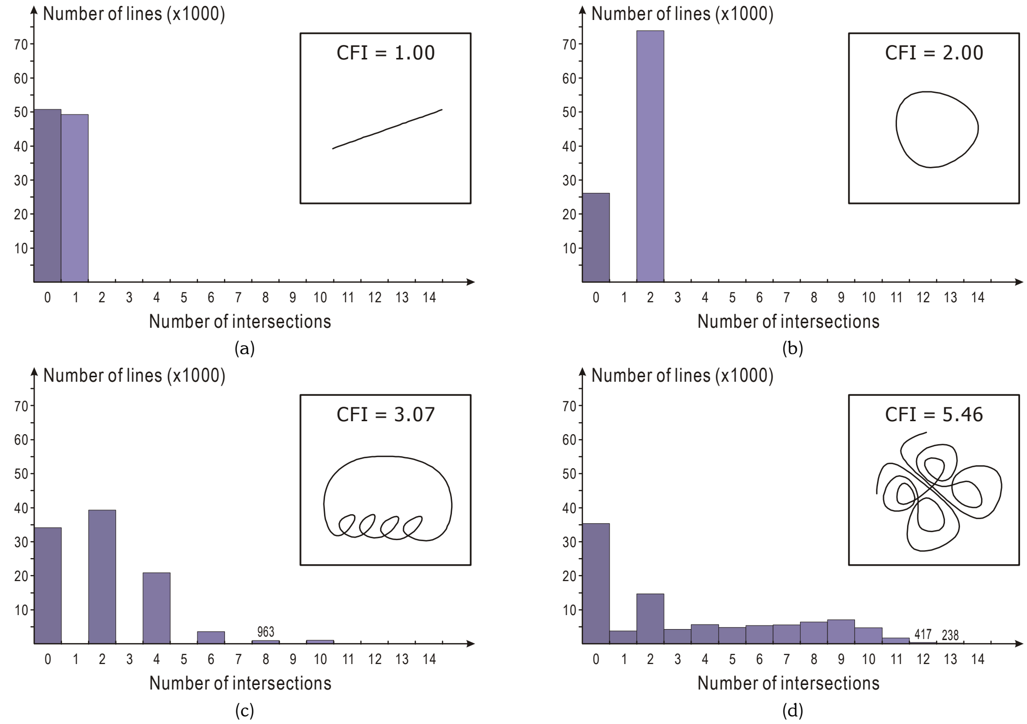

Fig. 1. Intersection distributions and CFI of different curves. The CFI in general reflects the degree of curve folding.

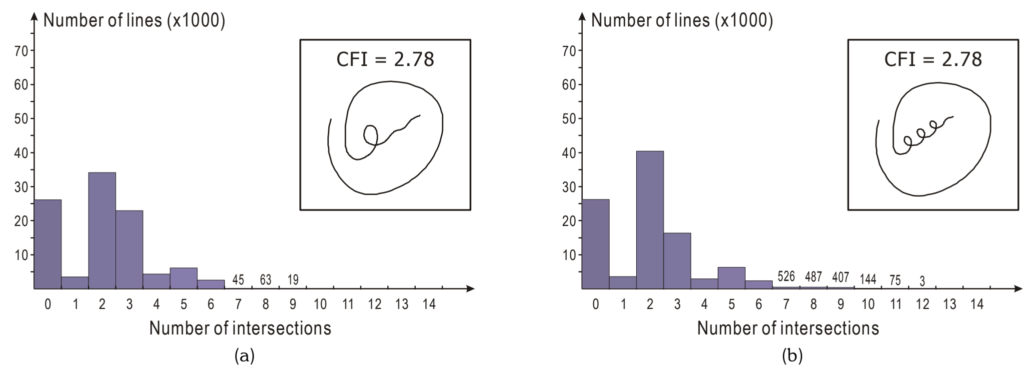

Fig. 2. The two curves in (a) and (b) are of the same length and the perimeters of their convex hulls are also of the same length. The CFI of the two curves are theoretically the same.

line must intersect a close curve at an even number of contact points. Moreover, the percentage of lines with only one intersection suggests that there is a significant portion of the curve that is open or stays loose from the rest of the curve (Fig. 3(c) and (d)). Also, the maximum number of intersections that a line can have with a curve may also tell the complexity of the curve folding. Therefore, the CFI and the intersection distribution can be used together to effectively characterize the degree of folding of a curve, depending on what level of detail one would like to attain in describing the folding complexity.

6. CONCLUSION

In this paper, we have presented a novel method to describe the degree of folding of a given planar curve quantitatively. The 2D Curve Folding Index (CFI) is based on the theory of integral geometry and is shown to be invariant under similitude transformations, i.e.

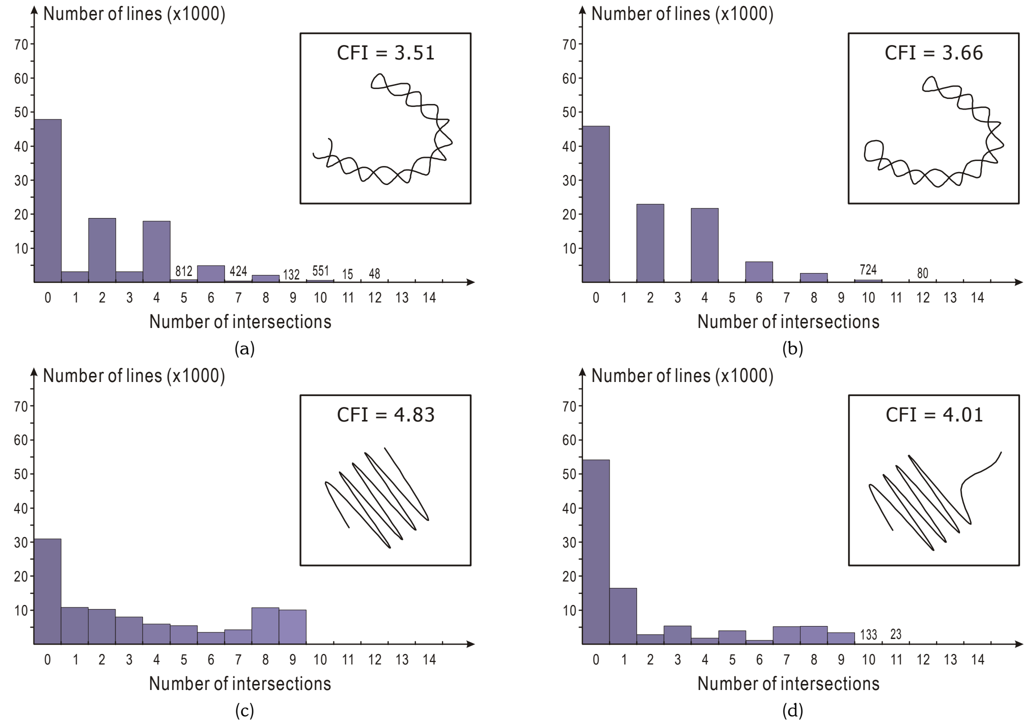

Fig. 3. Curve properties revealed by the intersection distributions; (a) & (b): open vs. closed curves. For a closed curve, there is no line with odd numbers of intersections with the curve; (c) & (d): open portion of the curve in (d) is indicated by having more lines intersecting the curve at only one point in the distribution.

Euclidean transformations as well as uniform scaling. This provides with us a simple and convenient tool to characterize a curve based on its degree of folding.

The CFI of a curve is defined to be the expected number of intersections that a random line has with the curve. It may be computed by generating random chords of a bounding circle of the curve and calculating the expected number of intersections that the curve may have with those lines intersecting the curve. It is shown that the intersection distribution of the curve obtained by the above method demonstrates distinguishable properties of the curve's folding complexity.

There are more problems about the description and analysis of the degree of curve folding in the three dimensional space, e.g. in protein folding problems by representing the backbone chain with a 3D space curve. Therefore, an extension of the CFI to three dimensional space is a problem for further research. Also, it would be interesting to study the relationship of the CFI with the integral of the curvature of a planar curve.

7. REFERENCES

[1] Creighton, T. E., *Protein Folding*, W.H. Freeman and Co., New York, 1992.

[2] Do Carmo, M. P., *Differential Geometry of Curves and Surfaces*, Prentice Hall, New Jersey, 1976.

[3] Li, X.Q., Wang, W., Martin, R., and Bowyer, A., Using low-discrepancy sequences and the Crofton formula to compute surface areas of geometric models, *Computer-Aided Design*, vol. 35, no. 9, 2003, pp. 771–782.

[4] Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D., Shape distributions, *ACM Transactions on Graphics*, 21(4), October, 2002, pp. 807–832.

[5] Santaló, L. A., *Introduction to Integral Geometry*, Hermann, Paris, 1953.

[6] Santaló, L. A., Integral geometry, In *Studies in Global Geometry and Analysis*, Chern, S. S., editor, The Press of The Mathematical Association of America, Washington DC, 1967, pp. 147–195.

[7] Solomon, H., *Geometric Probability*, SIAM, Philadelphia, 1978.