

ABOUT THE IDENTIFICATION OF FLAT UNCLOSED CURVES

I. M. Gostev^a, L. A. Sevastianov^b

^aMoscow State Institute of Electronics and Mathematics, Moscow

^bPeoples' Friendship University of Russia, Moscow

Questions related to process of identification of flat curves invariable to 2D affine transformations are investigated. The number and type of features, which must be used to construct the mathematical description of curves in the recognition process, are also discussed. Metrics, which allows identifying the unclosed curves, is also introduced.

В представленной статье исследуются вопросы, связанные с процессом идентификации плоских кривых, инвариантных относительно 2D аффинных преобразований. Также обсуждаются число и тип характеристик, которые нужно использовать для формулировки математического описания кривых в процессе распознавания. Введена метрика, позволяющая идентифицировать незамкнутые кривые.

PACS: 02.60.Ed; 02.60.-X

INTRODUCTION

Identification of graphical pattern essentially differs from other methods of pattern recognition. It includes several pre-processing stages of input information. On the starting stages the image is changing to a form which is convenient for next stages. To continue, one executes a process of choosing type and quantity of information properties, on which it is available to carry out pattern recognition with or without sample [1].

The most important in such a set of methods is a stage of the selection of informative properties. With the use of such properties a pattern could be received from surrounding group of objects. But what are such informative properties? How could some contour line, may be composed from separate fragments, be characterized?

Let us consider now some unclosed flat curve. How to construct a description invariant to 2D affine transformations (shifts, rotation, zoom, mirroring)?

For the random curve such features must be found from the properties of the curve.

This paper includes results of development of methodology, which describes how to form a set of properties invariant to such 2D affine transformations as rotation, shift, zoom and mirroring. We consider two metrics that could be used to identify the curves with the use of preliminary formed set of properties.

1. INTRODUCTION TO CURVE DESCRIPTION ON THE FLAT

We shall use a conception of k jets [2,3] for description of curve properties. Let us remember that k jets from k times continuously differentiable on segment $[a, b]$ and function $f(t)$ is represented by series of Taylor in which variable $(t - t_0)$ is replaced by $z \in [a, b]$

$$(J_{t_0}^k f)(t) = f(t_0) + f'(t_0)z + \dots + \frac{f^{(k)}(t_0)}{k!}z^k.$$

Under process of identification we shall understand process of comparison of two groups of properties, received from functions $f(t)$ and $g(t)$ under special designed metrics, which are considered below. For beginning let us determine a group of properties on which it must be calculated.

Definition 1.1. Let function $f(t)$ be k times continuously differentiable, then set $G = \{a = t_0 < t_1 < t_2 \dots < t_i < \dots < t_m = b\}$ is sequence of points at which some of k derivatives are equal to zero: $f'(t) = 0, f''(t) = 0, \dots, f^{(k)}(t) = 0$; these points calculated from derivatives of different orders are lying on $[a, b]$ with increasing values t . This set G will be named a zero set of k jets.

For example, for non-singular cubic parabola this set consists of three points: double points of extreme's and one of inflection. The involute of epicycloid with four petals with parameters $a = 2, b = 8$ is shown in Fig. 1, its formula is written as follows:

$$r(t) = \sqrt{\left((a + b) \cos(t) - a \cos\left(\frac{(a + b)t}{a}\right) \right)^2 + \left((a + b) \sin(t) - a \sin\left(\frac{(a + b)t}{a}\right) \right)^2}.$$

For this function the number of points of zero set on segment $[0, 2\pi]$ receiving from the first derivatives is 8, from the second derivatives — 8, from the third — 16, and so on.

Definition 1.2. Let $f(t)$ and $g(t)$ be k times continuously differentiable at point $t_0 \in [a, b]$. Then we shall say that functions are identical in k -jets manner if the next conditions $f'(t_0) = g'(t_0), f''(t_0) = g''(t_0), \dots, f^{(k)}(t_0) = g^{(k)}(t_0)$ are realized on $[a, b]$. Let this line identity at point t_0 be denoted as $\stackrel{k}{=}$.

Definition 1.3. We tell that $f(t)$ and $g(t)$ are line-identical in k -jets manner if on $[a, b]$ there exists a set of points $G = \{a = t_0 < t_1 < \dots < t_n = b\}$ such that $f'(t_i) = g'(t_i), f''(t_i) = g''(t_i), \dots, f^{(k)}(t_i) = g^{(k)}(t_i), i = \overline{0, n}$, i.e., $f(t) \stackrel{k}{=} g(t)$ are line-identical on $[a, b]$.

Theorem. Let the set F be received from function values $f(t)$ at points of zero set of k jets. Then the adjusted pair (G, F) defines all informative properties¹, necessary and sufficient for line-identification curves in k -jets manner.

Direct Proof. Under definition 1.1 on segment $[a, b]$ there exists a set of points at which extremes of k derivatives correspond to points at which $k + 1$ derivative becomes equal to zero. In this case, function $f(t)$ into intervals $(t_i, t_{i+1}), i = \overline{0, m - 1}$ will be monotone or monotone-convex (concave). Such a shape of $f(t)$ means the absence of specific behavior

¹Under informative properties we understand a formalized information, which one can receive from a function as a set of values of certain parameters.

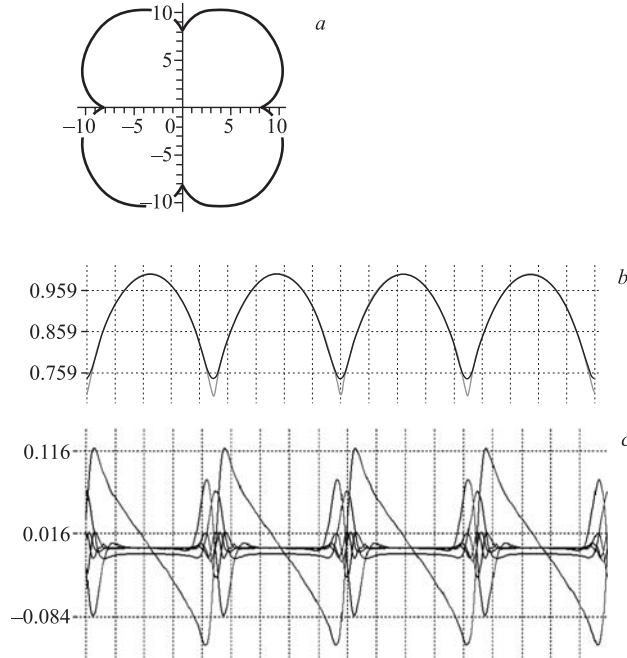


Fig. 1. Epicycloid (a), its input and smoothing involute (b), six first derivatives (c)

and, therefore, the absence of additional informative properties, which characterize function into intervals of zero set of k jets.

Reverse Proof. Suppose there is some point $t_c \notin G$, at which one of derivatives of function $f(t)$ becomes equal to zero, then appears a conflict with conditions of definition of zero set of k jets, because a set G produces just this point. Let l derivative have an extreme at point t_c , three variants are possible: if $l \leq k$, then $l + 1$ derivative must become equal to zero at this point and, therefore, it automatically is included in set G by definition 1.1; if at point t_c l derivative becomes equal to zero, then by this definition this point must be an element of set G ; if $l + 1 > k$, then the order of derivatives overcomes the bounds of consideration set of k jets, because we investigate the derivatives only until k order.

Thus, any point, at which specific behavior of curve $f(t)$ exists, is automatically included in set G . And vice versa, if any point included in G exists, then one can tell that at this point one of its derivatives becomes equal to zero.

Consequence. Cardinality of sets G and F is defined by accuracy of approximation and depends on an order of considered derivatives.

2. METRICS FOR IDENTIFICATION OF CURVES

In Sec. 1 we described definition which allows introducing informative properties for line identification of curves. Now we must consider methods for comparing curves on the base of these properties. For realization of these methods we shall construct two metrics.

Definition 2.1. Let a zero set of k jets of a sample curve consist of $n + 1$ -points of $G_s = \{a = x_0 < x_1 < \dots < x_n = b\}$, and for identifying curve $G_o = \{c = x_0 < x_1 < \dots < x_m = d\}$ consists of $m + 1$ -points. Let $n > m$, that is, suppose that identifying curve is a fragment of sample¹. Let us suppose that considering functions are pseudo-periodical, then we must redefine a set G_s by adding to it a number of points produced by cardinality of G_o in the following manner $G_s^* = \{x_0, x_1, \dots, x_n, x_{n+1} = x_n + x_1 - x_0, \dots, x_{n+m} = x_n + x_m - x_0\}$. Further to simplify notations, independently of does or does not pseudo-periodicity of these functions exist, we shall denote a zero set of sample as G_s .

To realize invariability from zooming we normalize functions f_s and f_o relative to their max and min values on segment $[a, b]$, so that their max value means 1, and min value — 0. Let us introduce difference function defined on zero sets of k jets of samples and identifying objects

$$\eta_{i,j} = f_s(t_{i+j}) - f_o(t_i), \quad i = \overline{0, n - m}, \quad j = \overline{0, m},$$

where f_s and f_o — are normalized values of functions (if pseudo-periodicity $i = \overline{0, n}$ exists). Let us define a function of absolute error of identification

$$\delta_i = \frac{1}{(m + 1)} \sum_{j=0}^m |\eta_{i,j}|, \quad i = \overline{0, n - m},$$

and a function of relative error of identification

$$\sigma_i = \frac{1}{(m + 1)} \sum_{j=0}^m |\delta_i - \eta_{i,j}|, \quad i = \overline{0, n - m}.$$

Definition 2.2. For a zero set of k jets $G = \{a = x_0 < x_1 < \dots < x_n = b\}$, let us introduce *mirroring set* $G_z = \{b = y_0 = x_n > y_1 = x_{n-1} > \dots > y_n = x_0 = a\}$.

Definition 2.3. Let us determine a function of recognition for identification of unclosed curves based on line correlations #1 (LK1) as

$$\lambda_{LK1} = \begin{cases} 1, & (\rho_{LK1} < \varepsilon_{LK1}) \vee (\rho_{LK1}^z < \varepsilon_{LK1}), \\ 0, & (\rho_{LK1} \geq \varepsilon_{LK1}) \wedge (\rho_{LK1}^z \geq \varepsilon_{LK1}), \end{cases} \quad (1)$$

where $\rho_{LK1} = \min_i \delta_i$ is a metric of LK1; ρ_{LK1}^z is a metric of LK1, calculated on mirroring set G_z , and ε_{LK1} is a classification tolerance of LK1. Equality $\lambda_{LK1} = 1$ will mean successful identification of curves, and value of i — shift of objects from beginning sample curve, where min of metric was accessed.

Definition 2.4. Let us determine a function of recognition for identification of unclosed curves based on line correlations #2 (LK2) as

$$\lambda_{LK2} = \begin{cases} 1, & (\rho_{LK2} < \varepsilon_{LK2}) \vee (\rho_{LK2}^z < \varepsilon_{LK2}), \\ 0, & (\rho_{LK2} \geq \varepsilon_{LK2}) \wedge (\rho_{LK2}^z \geq \varepsilon_{LK2}), \end{cases} \quad (2)$$

where $\rho_{LK2} = \min_i \sigma(i)$ is a metric of LK2; ρ_{LK2}^z is a metric of LK2, calculated on mirroring set G_z , and ε_{LK2} is classification tolerance of LK2. Equality $\lambda_{LK2} = 1$ will mean successful

¹Otherwise sample and identifying curves change there places.

identification of curves, and value of i — shift of objects from beginning sample curve, where min of metric was accessed.

Remark 2.1. If possibility of identification mirroring curves is absent, then a part of formulas with mirroring function ρ^z in (1) and (2) is excluded.

3. DISCUSSION

Introduction of notion of k jets and consideration of process of identification of curves on base of zero set of k jets discover the way to great number of new potentialities for operation on function.

Firstly, this approach makes it possible to construct functions description, oriented to their identification invariant to 2D affine transforming (shift, rotation, zoom, mirroring). If readers want to get to know more about these transforms, they can see works of L. Euleri [4]. Using of metrics of line correlations permits one to realize process of identification of functions — shape recognition of unclosed curves.

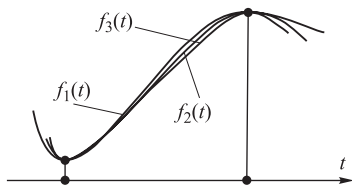


Fig. 2. Variability of behaviour of functions between points of a set G

Secondly, in proposed methods of identification there is some variability leading to the fact, that two curves of different orders, but with identical zero set of k jets, will be arranged to the same class (see Fig. 2). That means that there exists an ability not only to recognize shape of curves, but also to differentiate functions on classes according to their behaviour, i.e., number of points of G and their location. Description of a class may be found a priori, theoretically, during calculations of zero set of k jets, which fully determine a class of functions under consideration. For example, knowing parameters of some particle and characteristics of mass-spectrometer one may predict its trace in this device. Having a location of zero set of k jets one may approve trace behaviour in these points using identification of tracks. This is possible on the base of comparing functions using introduced metrics.

Thirdly, proposed methods have varying accuracy to identification of curves relative to the order of using derivatives. Obviously, while identification of cubic parabola all derivatives higher than the third will be equal to zero. Nevertheless, we may seek for fragments of this parabola as a part of curve of higher orders (on three points), and there may be more than one such fragments. This technique permits permanent checking on appearance of fragments of predefined form in inspected signals. Accuracy of this inspection will be determined by orders of using derivatives.

On the other hand, there is a possibility to cut out a three-dotted fragment from curve of higher order and to use it for a search of a place at which it is congruent to cubic parabola fragment. This situation is possible in the case when a part of function is lost, and with the help of conserved fragments it is not possible to determine its order. More details of this situation will be considered in following papers.

Fourthly, proposed methods of identification of curves are very simple in calculative aspects and may work in on-line, i.e., at the same time with carrying experiments on registration of particles.

At the same time, proposed methodology has some imperfections. Among them complexity and instability in calculation of derivatives must be marked. In process of computational experiments for compensation of these effects we have used function's smoothing and calculation of derivatives with interpolation formulas of Newton and Gauss.

CONCLUSION

Proposed methodology has a wide area of applications. It may be used in analysis of tracks of charged particles, identification of form of signals, recognition of parts of graphical images (when integrity of shapes of objects is lost) and in many other fields of science and technology. Advantage of this approach is in invariance to 2D affine transforms and easiness of metrics calculation. All of this allows one to work with these methods in real time.

Using for identification of curves the methodology of k jets permits presentation and recognition not only flat curves, but allows identification of curves in n -dimensional space. It makes this methodology still more relevant and perspective in scientific aspects and practical applications.

REFERENCES

1. *Gostev I.M.* Recognition of Graphic Patterns. Part 1 // *Izv. Ross. Akad. Nauk. Teor. Sist. Upr.* 2004. No. 1; *Comp. Syst. Sci.* 2004. V. 43, No. 1. P. 129.
2. *Goursat E.* Course of Mathematical Analysis. M., 1933. V. 1. P. 1.
3. *Arnold V.I.* Geometric Methods in Theory of Ordinary Differential Equations. Springer, 1988.
4. *Euleri L.* Introduction in Analysis Infinities. 2. M.: State Publ. of Phys. and Math. Lit. 1961. V. 1.