Handwritten Fonts Modeling
Based on Fat Lines of Variable Width

Stanislav Klimenko
Moscow Physicotechnical Institute
Moscow, Russia
klimenko@sim.ol.ru

Leonid Mestetskiy
Moscow State University
Moscow, Russia
l.mest@ru.net

Andrey Semenov
Tver State University
Tver, Russia
semenov@tversu.ru

ABSTRACT

In this paper we consider a new geometrical primitive - a curve of variable width (a fat curve). The fat curve is a trace of circle with variable radius moving along a smooth trajectory of the finite length called as an axis. For the mathematical description of such primitives cubic B-splines are used. By means of them it is possible to specify both an axis and a radius value (a thickness) of a fat curve. We solve a problem of representation of hand-written characters given as raster binary images in the form of association of fat curves. We demonstrate a possibility of modeling of "real" handwriting by entering small deformations into the form of fat curves. We consider a method of automatic kerning (visual alignment of letter spacing) of hand-written characters given as curves of variable width. We describe the way of output of arbitrary text presented by series of fat curves on raster-scan devices (printed output, displayed output). A prototype of a font editor of fat curves was developed by the authors of this paper. The presented approach can be used as embedded module (plug-in) in font editors of hand-written characters.

Keywords

B-spline fat curve, skeleton of binary bitmap, circular representation, handwritten font, personal vector cursive, automatic kerning.

1. INTRODUCTION

Attempts to use a body of mathematical tool for description of forms of alpha characters have been undertaken as long ago as in 15 century. The predecessor of all existing typographical fonts is exactly the cursive. There are three types of fonts which are used in a varying degree in personal computer: bitmapped, vector and outline. The last type is the most widespread and usable at the present time. Contours of characters are formed as straight-line segments and sections of Bézier curves of the second (TrueType) and the third (PostScript) orders. In fig. 1 the hand-written character (the letter «a») made of segments of Bézier curves of the third order is shown. Now it becomes very popular to have your personal computer cursive in a collection of fonts. Such font allows singling out the author's letter among the others and individualizing it. There are commercial firms (FontShop - Germany, ParaType - Russia, Signature Software - the USA), which offer their services in designing of full-fledged computer fonts on basis of a sample of penscript (handwriting) of any person [Kar01], in the market of software products. Such service is labor-consuming for the type designer and so it is expensive for the customer.

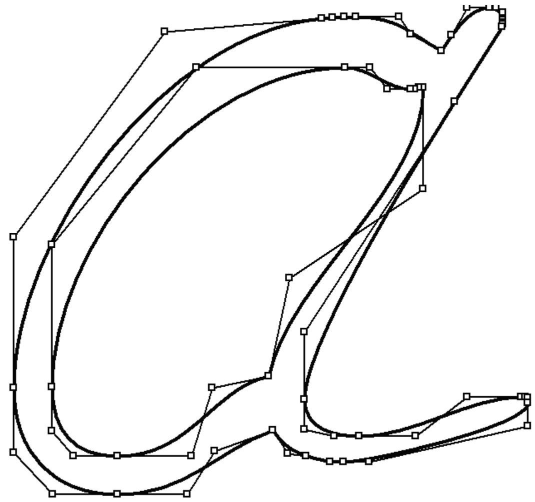


Figure 1. The hand-written character (the letter «a»), described by Bézier curves.

The important feature of a cursive is the requirement of its "revival" in the process of outputting on display devices (display or printer). Revival is understood as small variations in typeface of particular characters, inherent to real handwriting of any person. These variations are expressed in little changes of width and form of strokes for the same characters. From the point of view of such "revival" existing solutions of designing of outline handwritten fonts possess a number of disadvantages. First of all, the outline description of a font doesn't allow implementing variations of stroke width and form of pen. The outline description requires solving an intricate problem of synchronous variation of both borders of a stroke.

Secondly, during the process of stroke variations it is necessary to provide preservation of contour smoothness where it is required. Automatic maintaining of smoothness of compound Bézier curves is also nontrivial problem. Representation of cursive using a model of trace of a pen «with width» (figure 2) potentially gives more possibilities for creation of personal «computer handwriting».

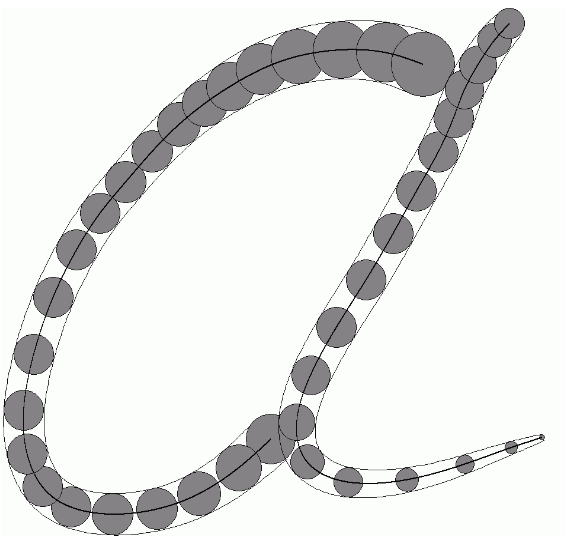


Figure 2. The hand-written character (the letter «a»), described by fat curves.

Such model is used in various graphic editors of hieroglyphs [Kla01], and also in Knut's system *METAFONT* [Knu79,Knu01]. This model makes possible to vary trajectory and pen width, i.e. it is much better adapted for modeling real variations of sample font during writing. However practical implementation of such approach for creation of cursives faces a problem of representation of font characters in the form of curves with width on the initial scanned sample of text. The method presented in this work describes solution of this problem. The source data are the handwritten text given on a paper, for example, the alphabet (fig. 3). And the result of the solution is the description of characters in the form of curves with width.

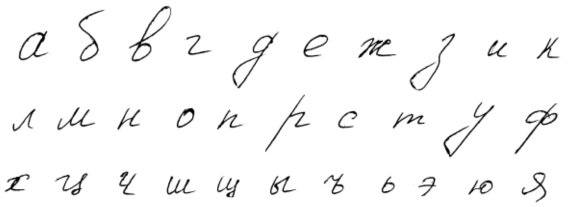


Figure 3. The hand-written Russian alphabet.

On basis of presented method is the idea of representation of the initial raster hand-written text in the form of set of continuous graphic primitives - fat curves [Yao91] (figure 4). A fat curve describes as a trace of circle with variable radius moving along a smooth trajectory called as an axis of this fat curve. B-splines of the third order are used as a mathematical apparatus for description of fat curves.

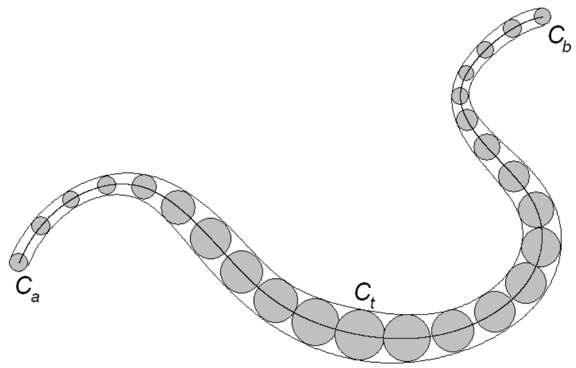


Figure 4. Fat curve.

2. FAT CURVES

Elementary B-spline fat curve of the third order [Mes00] is specified by the following vector equation

$$C(t) = (x(t), y(t), r(t)) = \frac{1}{6}\sum_{i=0}^{3} B_i(t) H_i \text{ ,}$$

where $t \in [0,1]$ - a curve parameter, $H_i = \{H_{ix}, H_{iy}, H_{ir}\}$ - a set of control circles centered in points (H_{ix}, H_{iy}), with radiuses H_{ir}, $B_i(t)$ - base functions of the third order B-spline satisfying the following conditions [Boo72]:

$$B_0(t) = (1-t)^3$$

$$B_1(t) = 3t^3 - 6t^2 + 4$$

$$B_2(t) = -3t^3 + 3t^2 + 3t + 1$$

$$B_3(t) = t^3$$

 ISBN 978-86943-15-2

Each control circle H_i can be considered as three-dimensional vector. The example of elementary B-spline fat curve with control circles $\{H_0,..., H_3\}$ is shown in fig. 5. The fat curve (as a trace of moving circle) is presented by grey color, the axis of the fat curve - by black color, control circles - by dotted line.

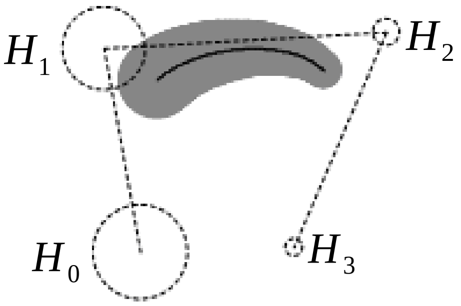


Figure 5. The elementary B-spline fat curve.

Elementary curves can unite into more complex graphic primitives - compound fat curves. Let a curve γ, presented as an association of elementary cubic B-spline fat curves $\gamma^{(1)},...,\gamma^{(m-2)}$, be a compound cubic B-spline curve with common set of control circles $H = (H_0,..., H_m)$, where $H_{k-1}, H_k, H_{k+1}, H_{k+2}$ - control circles of curve $\gamma^{(k)}$. There is the compound B-spline fat curve constructed of 6 elementary sections and having 9 control circles $H_0,..., H_8$ is presented in fig. 6.

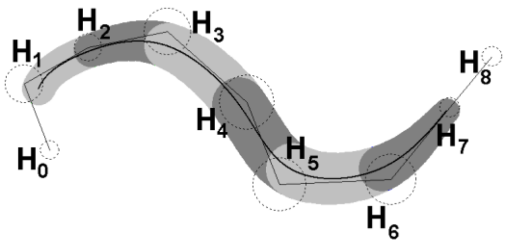


Figure 6. The compound B-spline fat curve.

3. METHOD DESCRIPTION

The presented method to create computer cursive (obtaining of fat-linear representation) consists of the following steps:

- an author of handwriting writes by hand on a paper separate characters, for example, the alphabet;
- an image is entered into computer using scanner;
- a designer selected characters in the scanned image in form of binary bitmaps (black characters on white background);
- for each character the continuous skeleton (as a set of middle axes) [Lee82,Mes98] is constructed;
- each branch of the skeletal graph is approximated by compound B-spline fat curve;
- fat curves unite in strokes representing a character.

Let's call an association of finite number of fat curves as a circular shape. A representation (an approximation) of raster images of characters by circular figures lays on basis of a method. The problem of approximation of the raster sample of handwriting by a circular shape can be solved in several ways of different parts of manual work. Completely manual way assumes direct input and editing of fat curves in the form of sequence of control circles by the designer. Automatic construction of a circular shape is provided using a method described in [Mes00]. The idea of this approach consists in construction of a continuous skeleton and approximation of its branches by fat curves. A problem of primary adjustment is the most labor-consuming. But after it is solved it is necessary to receive qualitative representation of a font where the manual technique is fundamental. However, ample opportunities of editing of fat curves allow developing a special graphic editor for construction of letter descriptions, supporting operations of alignment, editing of axes and width, "interchange" of difficult crossings and overlapping of strokes.

There are stages of transformation of the character given on a raster to the set of fat curves below.

Figure 7. The transformation of the raster character to the set of fat curves.

 ISBN 978-86943-15-2

Some parts of characters (letters) can contain noise which will be visually expressed in not so smooth border and axis of approximating fat curves, as is obvious from figure 7. This noise can be an effect of errors arising during scanning and binarization. Deliverance of a fat curve from such artifacts we will name cleanup. Editing of B-spline fat curve is made by changing positions of centers and radiuses of its control circles. A designer can easily correct the form of a fat curve by moving of control circles and changing their radiuses in a manual mode or he can try to do this operation automatically. For each control circle H_i, $i = 1,..., n-2$, despite of the first one and the last one, we will recount value of its radius using the following formula: $H^{'}_{ir} = k_1 H_{i-1r} + k_2 H_{ir} + k_3 H_{i+1r}$, where $H^{'}_{ir}$ - the new value of radius of circle H_i, k_i - weighting coefficients, satisfying the condition $k_1 + k_2 + k_3 = 1$. The given procedure represents an operation of discrete convolution for a one-dimensional signal with a kernel (k_1, k_2, k_3). A choice of convolution kernel (k_1, k_2, k_3) will characterize a degree of smoothing of control-circle radiuses in a fat curve. An operation of smoothing of control-circle radiuses with the convolution kernel $(1/3, 1/3, 1/3)$ is demonstrated in fig. 8. There is the initial fat curve on the left and the smoothed fat curve on the right. Corresponding circles before and after smoothing operation are presented by polylines.

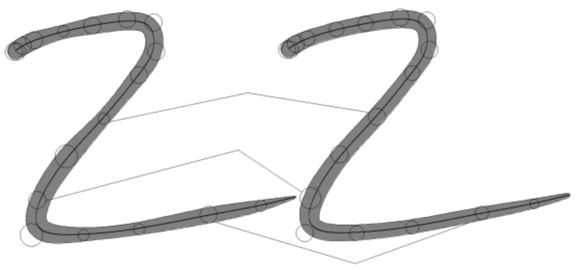


Figure 8. A smoothing of control-circle radiuses.

Similarly, such step can be done for an axis of fat curve. In this case a formula of smoothing operation is the following:

$$H^{'}_{ix} = k_1 H_{i-1x} + k_2 H_{ix} + k_3 H_{i+1x},$$
$$H^{'}_{iy} = k_1 H_{i-1y} + k_2 H_{iy} + k_3 H_{i+1y},$$
$$i = 1,..., n-2.$$

An example of smoothing of axis function of fat curve with the convolution kernel $(1/3, 1/3, 1/3)$ is presented in fig. 9. There is the initial fat curve on the left and the smoothed one on the right.

Figure 9. A smoothing of axis function of fat curve.

It is very convenient to use smoothing operation for radiuses of control circles and an axis of fat curve simultaneously.

The letter breaks into separate fat curves, and their number coincides with number of branches of the corresponding skeletal graph, as is evident from figure 7. It occurs because an automatic adjustment of a circular shape to a raster image gives as a set of fat curves only noncrossing fragments of trajectories. These fragments meet in endpoints, in threes or more branches in a point. Therefore there is a necessity of association of two pairs of branches in one common trajectory for more adequate modeling of pen movement. We will call an operation of joining of one fat line to another as sewing. Sewing procedure is the operation that demands work of a designer. This sewing procedure will allow reducing a number of fat curves, of which a character consists, and making strokes smoother. Solution of a problem of reducing of fat curve from the given set of circles lays on basis of the sewing method. Let $G_0,..., G_{n-1}$ be a consequence of circles, each of them is described as $G_j = (G_{jx}, G_{jy}, G_{jr})$, where (G_{jx}, G_{jy}) - coordinates of the center of the circle and G_{jr} - its radius. Let's call this set of circles as basic. It is required to construct a compound B-spline fat curve $C(t) = (u(t), v(t), r(t))$, that exactly passes all circles G_j (an interpolation problem) or sufficiently near from circles G_j (an approximation problem). Constructing of the required fat curve consists in finding its control circles. Formally, an interpolation problem consists in choice of such description $C(t)$ that the following condition is satisfied:

$$(\exists\ t = t_i : C(t_i) = G_i,\ i = 0,..., n-1).$$

 ISBN 978-86943-15-2

A in a an approximation problem – the condition

$$(\exists\, t = t_i : \mu(C(t_i), G_i) < \varepsilon,\ i = 0,\ldots,n-1)\ ,$$

where ε is a preassigned accuracy.

As stated above, both these problems can be solved using maximal empty circles of skeletal graph as a set of basic circles. From the algorithmic point of view sewing procedure consists in the following. Sets of basic circles of skeletal branches unite in one sequence for which an approximation of the derived set of circles is made using the method described in [Mes00]. In fig. 10 the letter after "sewing" procedure is shown. After each step of "sewing" it is possible to make smoothing and editing procedure of radiuses and centers of control circles of a new fat curve. This procedure allows adequate simulating of strokes and also reducing a dimension of shape description and gives the convenient tool to change character forms.

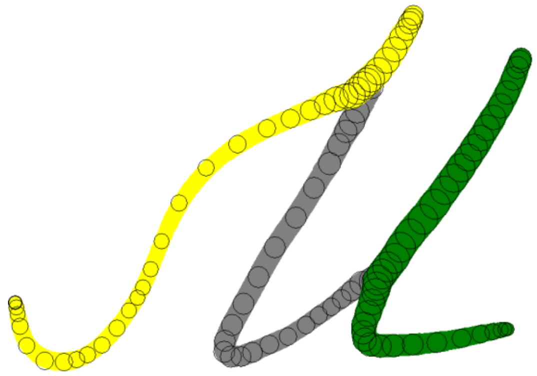


Figure 10. A step of "sewing" fat curves.

It is required to consider a case when two different strokes are partially or completely overlapped by each other during a process of writing for correct modeling of letter characters by fat curves. Visible form of a stroke in the scanned image is an association of two or more trajectories of pen movement. Therefore it is necessary to select them from this association for the adequate description of these trajectories. It is difficult to automatically consider this feature of human writing during approximation of such characters by fat curves. The case when two pen trajectories look like one is shown in fig. 11

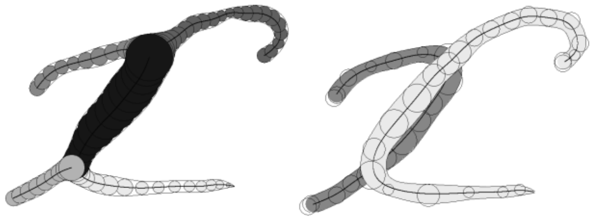


Figure 11. A step of cloning and "sewing" fat curves.

Participation of a designer is required to solve such situations. The presented decision consists in copying (cloning) of that fat curve which has united in itself two strokes. Each of two clones is integrated (“sewed”) in a corresponding trajectory. The further step consists in sewing, smoothing and editing derived fat curves. The result of this operation is demonstrated in fig. 11. Such character “structure” consisting of fat curves is closer to a real writing of letter by person.

The next step to create of a personal cursive is a reduction of all letters (characters) to a base line (a line of writing). A designer should do this step for each character himself. Five characters reduced to a base line of a font are shown in fig. 12. A base line is easily entered into description of a circular shape. It is possible to enter it both manually and automatically, for example, using a requirement that the sample of handwriting is presented in a form with a base line

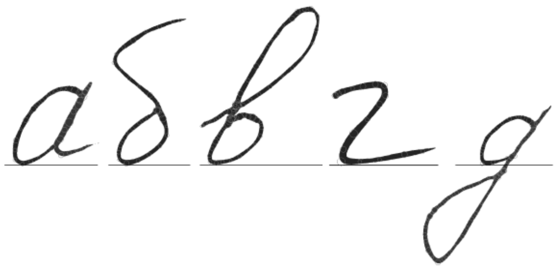


Figure 12. A reduction of all letters to a base line.

4. METHOD OF AUTOMATIC KERNING

A modern computer font consists of a set of character images, and also of a collection of instructions and rules for correct display of letters during rasterization. One of examples of such rules is a visual alignment of letter spacing (kerning) in process of text on raster devices [Kar01]. There can be such combinations of signs and characters which will form visual clearances and clottings, that introduce a certain discomfort in a rhythm of text reading, practically in any text. This problem becomes especially serious when a font size is large. A feature of a kerning problem is in squared relationship of number of formed kerning pairs from a general number of characters in a font. Though actually kerning pairs which practically can’t be found in texts exist. Nevertheless, processing of such volume of information in manual mode is obviously a labor-consuming problem. Therefore development of an automatic method of visual alignment of letter spacing is a rather actual problem. It is necessary to notice, that for a cursive the kerning problem has more importance than for typographical type-setting

 ISBN 978-86943-15-2

fonts. The example of hand-written words without the kerning mechanism and with it is presented below (fig. 13). For clearness the minimum rectangles of each character are also shown.

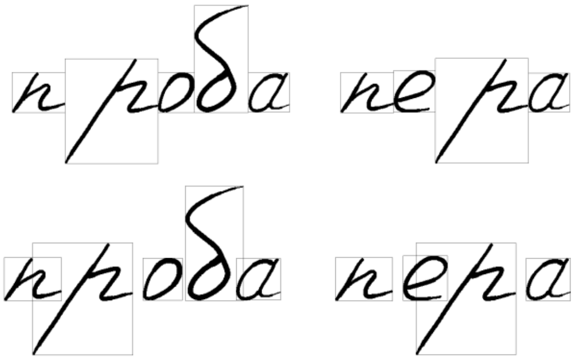


Figure 13. A hand-written words without/with the kerning.

The kerning operation is applied to achieve optical balance of the text. Because of features of the form of some character pairs there is an illusion that the distance between them is more or less than real. Kerning can be manual and automatic. In manual kerning subjective feeling of a designer is the criterion of optical balance of a text. The presented method of automatic kerning of hand-written characters includes the following steps:

1. rasterization of kerning pair of characters on a common base line of raster;
2. scanning by a horizontal line "from top to bottom" with step of 1 pixel;
3. the most east and the most west cross points are defined for left and right character of kerning pair respectively at each position of a scanning line;
4. position of a scanning line, at which distance between the points found on the previous step is minimum, is defined;
5. the right character of kerning pair is displaced horizontally (to the right/to the left) in such a way that the minimum distance between the characters was equaled to preassigned constant (fig. 14).

This value will characterize the minimum permissible horizontal distance between two characters of kerning pair in rasterization. The more is this value, the more will be visual “clearance” between characters in the text. Set of all character pairs form a so-called kerning table. Value of an element of kerning table, located on crossing of line i and column j, specifies on how many pixels to the right/to the left it is necessary to move the right character of kerning pair in process of printed output.

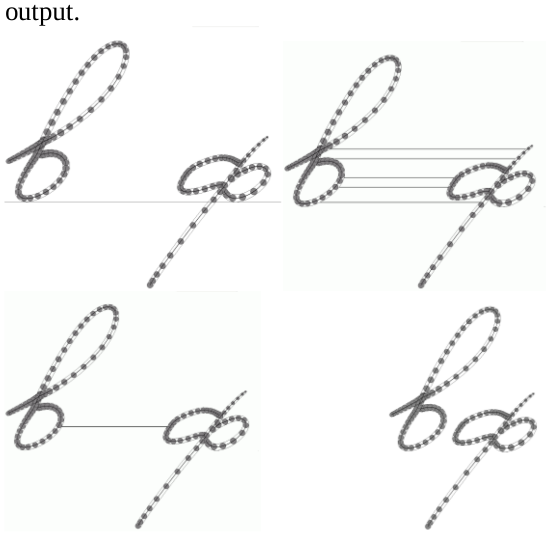
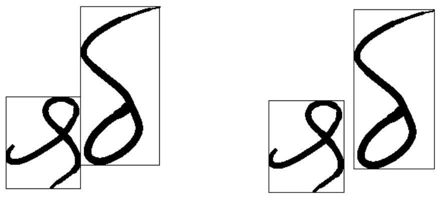


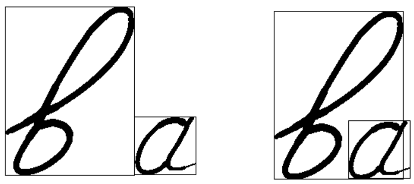
Figure 14. A automatic kerning of hand-written characters.

Positive value of an element of kerning table means that it is necessary to “move aside” (to displace to the right) the right character. Negative value signed that it is necessary to “move up” (to displace to the left) the right character of kerning pair to the left character (fig. 15).

	а	**б**	**в**	**..**	**ю**	**я**
а	19	15	8	..	11	19
б	-13	-17	-5	..	-29	-21
в	**-72**	1	-62	..	-88	-80
..	..	..	..	..	..	..
ю	16	12	6	..	2	12
я	19	**15**	11	..	19	19



KERNING[*я*][*б*] = 15



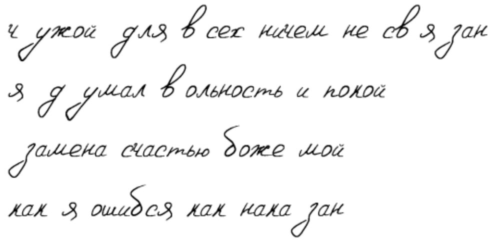
KERNING[*в*][a] = -72

Figure 15. Examples of kerning pair.

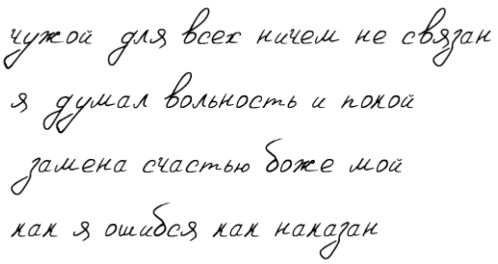
 ISBN 978-86943-15-2

It is necessary to notice, that minimum permissible distance between kerning pairs is a value which can be operated. The same hand-written text with different values of this parameter is shown below.

(a)



(b)



(c)

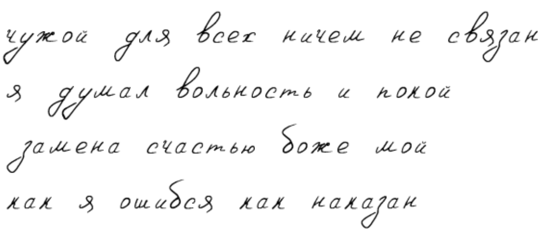


Figure 16. Examples of hand-written text with various kerning
(16a – no kerning, 16b – kerning = 20,
16c – kerning = 40).

5. APPLICATIONS

The mathematical apparatus of letter description using fat curves allows realizing modeling of "real" handwriting by entering of small dynamic changes into the letter forms. The idea of such revival of handwriting consists in the following: having printed the same text using such font several times, we will receive distinct results. There is a software product ParaNoise belonging to Russian firm ParaType which allows changing the form of characters by the given rule. The basic idea of this program is a deformation of a contour (shift, rotation, scaling) using certain algorithm. In terms of fat curves the effect of "randomness" for character forms can be implemented much more variously and easier like this. In process of rasterization (an output on a raster device) of each particular character of all radiuses of control circles, we will add a certain random variate, namely $H_{ir} = H_{ir} + \xi_i$, where H_{ir} – a radius of control circle H_i, ξ_i – a random variate, evenly distributed on a given segment $[-a, a]$. The more is the value of a, the more random and "dissimilar" the corresponding characters will be. There is a letter with modeling of random form (segment $[-a, a] = [-5, 5]$) in fig. 17. Here we use a display pixel as a unit of measurement.

Figure 17. Modeling of "randomness" forms.

Besides, it is possible to use different types (normal, thin, wide) of pen in process of text output on raster devices. The construction of fat curve allows changing pen width easily using proportional change of radiuses of all control circles. There is the initial symbol in fig. 18a, in fig. 18b radiuses are reduced by 30 %, in fig. 18c - radiuses are increased by 50 %.

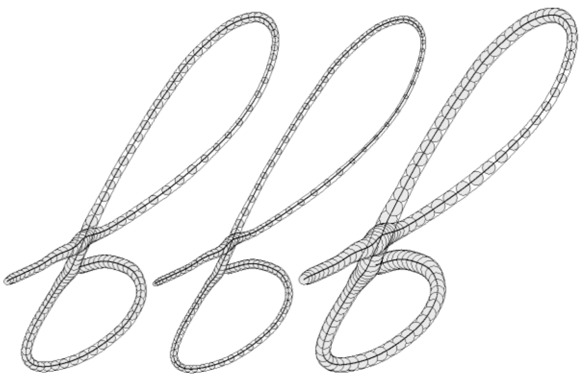


Figure 18. Modeling of "thickness".

The way of character description using the model of a pen trace allows making converting of the received font into standard formats (TrueType and PostScript) in the form of the outline description (fig. 19). It is possible to make by means of construction of border of a fat curve as envelope curve for smooth sets of circles [Mes00] and approximation of these curves by Bézier curves of the second and the third order [Mes06]. The possibility of approximation of characters with controllable accuracy in the form of outlined description by Bézier curves of the third order is implemented by authors.

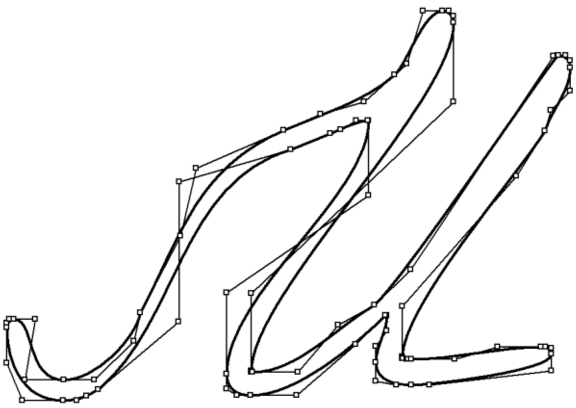


Figure 19. The outline description.

6. CONCLUSIONS

The method presented in this work to describe hand-written characters using curves of variable width was software implemented in the form of prototype of the font editor. All illustrative examples demonstrated in the paper, are received using the developed program. The conducted research has shown that the method of fat curves is potentially convenient and efficient tool in process of modeling of form of hand-written characters. The main advantage of the presented approach consists in more adequate description of a cursive using «a pen with width». The presented approach can be used as embedded module (plug-in) in font editors of hand-written characters.

7. ACKNOWLEDGMENTS

Authors are grateful to the Russian Foundation of Basic Researches for the support of this work (grant № 05-01-00542).

8. REFERENCES

[Boo72] de Boor C. On calculation with B-splines. J. Approx. Theory, Vol. 6, 1972.

[Lee82] Lee D.T. Medial axes transform of planar shape. IEEE Trans. Patt. Anal. Mach. Intell. PAMI-4 (1982), 363-369.

[Kar01] Karov P. Font technologies. The description and toolkit, 2001 (In Russian).

[Kla01] Klassen V. Variable width splines: a possible font representation? Electronic publishing, vol.6 (3), 1993.

[Knu03] Knuth D. All about METAFONT. Publishing house VILJAMS, 2003 (In Russian).

[Knu79] METAFONT, A system for alphabet design, apart 3 of TeX and METAFONT: New directions in typesetting, American Mathematical Society and Digital Press, Bedford, Mass, 1979).

[Mes00] Mestetskiy L. Fat curves and representation of planar figures. *Computers & Graphics*, vol.24, No. 1,2000, pp. 9-21.

[Mes06] Mestetskiy L. Skeleton of a multiply connected polygonal figure. International conference of computer graphics “Graphicon”, Novosibirsk, 2006, pp. 299-314. (In Russian).

[Mes98] Mestetskiy L. Continuous skeleton of binary raster bitmap. In Graphicon’98, International Conference on computer graphics, Moscow, 1998 (in Russian).

[Yao91] Yao C., Rokne J. Fat curves. Computer graphics forum, 10, 1991.

.