The paper considers the problem of fitting the image of a text with letters of different sizes and different fonts in a given bounded connected region of arbitrary configuration on a plane.

1. INTRODUCTION

Computer systems often have to fit a text in an allocated field, e.g., on an imaging device for temporary display (a monitor screen or an electronic tablet) or on a "hard" carrier for durable recording (paper, photographic material). In what follows, we consider a field on a rectangle in the coordinate space defined by the axes x and y. The origin of the coordinates is in the top left corner of the rectangle. The x-axis points from the top left corner to the top right corner, and the y-axis is directed from the top left corner to the bottom left corner. Lines of text are written inside the rectangle parallel to the x-axis. While constructing the text image, the output device advances over the rectangle with a discrete step e in the direction of the y-axis.

The original description of the text stored in the system actually contains the text proper with codes that determine the location of the characters on the imaging device for temporary display or on the "hard" carrier for permanent recording. For each character, the codes define the language, the point size, the font, and the case: capital or lower-case. Special codes identify places in text where the spacing between letters is increased by leading and indicate the amount of leading. Paragraphs are identified and the paragraph indent is specified. Spacing between adjacent lines of text is specified whenever it is nonstandard.

The system also stores the description of the region allocated to the text. The region may be defined by tracing with a mouse, by moving an optical pen over a field on the monitor screen, or by direct specification of the coordinates of the boundary points.

Fitting the text in the allocated region involves locating each character inside the specified region in space. Given the original text description and the description of the allocated region, it is required to partition the text into lines and to indicate the y-coordinate of each base line. Once the lines are located, the coordinates of the characters within the lines can be determined without difficulty. Then spacing between the base lines should be calculated so that the text uniformly fills the allocated region.

When editing office documents and technical documentation, a relatively short text with infrequently changing point sizes and fonts has to be fitted in a rectangular region. In a larger article set in a newspaper column or on a journal page, point sizes and fonts change more frequently. The region allocated to such text consists of polygons in which the boundary between any pair of adjacent vertices is parallel either to the x-axis or to the y-axis. These are usually rectangles of equal width. More seldom, different-width rectangles are used. Complex polygons are always representable as a collection of rectangles of different widths contiguous in the direction of the y-axis. A region of such complex configuration arises, for instance, when text "flows" around graphics. In all these cases, by first representing the text as a sequence of lines of a certain format, we can reduce the problem of fitting the text in the allocated region to the problem of fitting a given sequence of lines in this region [1].

We consider the case when a short text with frequently changing point sizes and fonts must be fitted in a region of arbitrary configuration. This case often arises in practice when describing a region on a map or an object in an advertising leaflet. In such cases, it is difficult or even impossible to guess in advance the format of each line of text. The format of the

current line can be computed only after the previous line has been located and with it the entire text preceding the given line. This has motivated the development of a new approach to the text location problem, which is proposed in the present paper.

2. DESCRIPTION OF THE TEXT AND THE FIELD

We assume that each character is represented in an electronic system by its description in a rectangular cell. The cell height $\mu$ depends on the point size $k$ of the character and the cell length (or width) $\nu$ depends on the point size $k$ and the font $p$. Rectangular cells displaying capitals and lower case characters are of identical size. The location of the base line in these cells is also the same. The distance $z$ from the lower boundary of the cell to the base line depends only on the point size of the character. In order to describe the main ideas of the proposed solution, it suffices to consider the case when the base lines of all the characters in a line of text coincide with the base line of the text. Figure 1 shows two cells with the images of a capital and a lower case character of the same point size and font.

Consider an imaginary rectangle such that any text may be fitted in it in a single line. We represent the original text description in the form of a strip partitioned into a finite number of cells. Each cell of the strip may contain either a character or a code. The original text description starts with the code BEGIN and ends with the code END. The remaining codes generate the text image on the imaginary rectangle.

The image of each character is defined by the codes SIZE[$k$], FONT[$p$], LANGUAGE[$h$], CAP, LOWERCASE directly preceding the character. The codes CAP and LOWERCASE alternate. The characters that follow the code CAP in the original text are represented by the images of capital characters, and those that follow the code LOWERCASE are represented by images of lower-case letters.

The text is made up of paragraphs. In the original text description, the beginning of a paragraph is identified by the code PARAGRAPH[$a$] and its end is identified by END PARAGRAPH. The value of the parameter $a$ specifies directly the paragraph indent. The indent may be specified indirectly, e.g., by specifying the number of characters BLANK.

With standard location of characters in a line of text in the imaginary rectangle, the rectangular cells with the images of adjacent characters are contiguous. The standard location of the character cells on a line of text may be changed. If a sequence of characters in the original text description if enclosed between the codes SPACING[$g$] and END SPACING, then the distance between the cells with the images of the adjacent characters in this sequence is determined by the value of the parameter $g$. This is usually the absolute value of the spacing. The spacing also can be defined in percent relative to the length of the cell containing the character BLANK. If the code SHIFT RIGHT[$d$] is inserted between two adjacent characters in the original text description, then the cells with the images of these characters are condensed to a spacing determined by the value of the parameter $d$. The value of the parameter $d$ is the absolute magnitude of the right shift. These codes provide a sufficiently broad set of text description instructions for our purposes.

Consider some text $T$ consisting of $I$ paragraphs. Partition the base line of the text $T$ in the imaginary rectangle into $N$ segments. Each segment is occupied by the images of characters from one paragraph, which have the same size and font.