(14)

"A Method for Compressing Character or Pictorial Image Data".

The present invention relates to a method of compressing character or pictorial data by storing outline specifying information of the character, pictorial image or the like.

It is well known that binary data determined by subjecting a character to dot-decomposition entails a very high degree of redundancy.

In this respect, various data compression methods have heretofore been proposed for decreasing such redundancy. One example of these methods is a so-called outline method wherein the shape of a character is grasped by its outline, and information for specifying the outline is stored thereby compressing the quantity of data stored.

For compressing data according to such outline method, a vector approximating method as illustrated in Figure 1, or an m-order curve approximating method as illustrated in Figure 2 has already been proposed.

The vector approximating method illustrated in Figure 1 corresponds to those disclosed in Japanese Patent Laid-open No. 149522/1979 (U.S.P. 4,254,468) and Japanese Patent Laid-open No. 79154/1980.

In such a vector approximating method an outline 1 of an arbitrary character indicated by a dotted line is approximated with a set of two-dimensional vectors 2, and specified information (position of starting point, length and inclination, or components in horizontal and vertical directions) for the respective vectors is stored as data, whereby data compression becomes possible.

The m-order curve approximating method illustrated in Figure 2 has already been described by the applicant in Japanese Patent Application No. 116,160/1980 and G/B/ 2,115,657.

This m-order curve approximating method compression of a quantity of data is achieved by storing coordinates of a group of points P suitably established on the outline of a character, and at the same time, approximating a desired outline by means of a set of m-order curve elements 3 connecting (m + 1) of arbitrary points with one another.

These outline data compression methods have such that when interpolation processing or scale factor conversion processing of a vector is carried out on compressed data for a character image, the methods can cope with the reproduction of character images of various scale factors.

On the other hand, however, these conventional data compression outline methods involve an essential disadvantage in that the optimum result is not guaranteed in respect of smoothness of the outline (continuity in inclination of the outline). More specifically, for instance, the inclination of segments involving either each terminal outline points P of the vectors in Figure 1, or each central connecting points P_c of the m-order curves 3 in Figure 2 is inevitably discontinuous.

In this respect, an outline of a character is generally such that not only the outline itself is continuous but, also the primary derived function (inclination of outline) as well as the secondary derived function (rate of change in inclination) thereof vary continuously, if peculiar points such as intersecting portions of lines constituting a character, or extreme ends of so-called breaks corresponding to "hane" in Chinese or katakana character are excluded. For this reason, conventional data compressing outline methods cannot provide a faithful character outline, and unnaturalness (discontinuity in inclination) of a character image reproduced on the basis of such data cannot completely be removed.

According to the present invention there is provided, a method of compressing character or pictorial image data which compresses a quantity of data by storing outline specifying information of said character, pictorial image or the like, comprising:

as a first step, determining the outline of said character, pictorial image or the like in terms of orthogonal X- and Y-coordinates;

as a second step, splitting said outline into a set of blocks $[P_1 : P_n]$: where P_1 and P_n are start and end points of any one block;

as a third step, subdividing each of any one of said blocks with $(n - 2)$ break-points P_i $(x_i : y_i)$: where $1 = 2$ to $(n - 1)$, to obtain $(n - 1)$ of segments S_i: where $i = 1$ to $(n - 1)$; and

as a fourth step, computing for each segment S_i a constant σ_i corresponding to 1/6 a value of second derivative of the third order curve passing through the start point $P_i(x_1,y_i)$ of the segment at the start point; and

storing for each segment the constant σ_i and coordinate values $(x_i : y_i)$: where $i = 1$ to n, of the points P_i defining the segments S_i as the outline data corresponding to said blocks.

A preferred embodiment of the invention will now be described with reference to the accompanying drawings, wherein:

Figures 1 and 2 are diagrams which each illustrate a general sketch of a conventional data compressing method according to the "outline" method;

Figure 3 is a diagram illustrating the relationship between a character outline and blocks B on X- and Y-coordinates;

Figure 4 is a diagram illustrating the relationship between any one block B and a segment S_i on X- and Y-coordinates;

Figure 5 is a diagram showing an example of a preferred data memory format utilised in case of effecting a method according

to the present invention;

Figure 6 is a schematic view illustrating a constructional example for determining memory data by a method according to the present invention; and

Figure 7 is an view illustrating a relationship between a master font and an output character area.

The method of the present invention is applied to row and column matrix-form binary character data supplied from, for example, a scanner or the like. In another example, the method of the invention is applied to coded character data which can specify a row and column matrix-formed dot pattern and which may be in accordance with any method.

Namely, for instance, a character image is subjected to dot-decomposition in row and column matrix-form by means of raster scanning of a scanner to determine bit pattern data, and such bit pattern data are supplied as original character data which come to be objects to be processed in the method of the present invention.

In the method according to the present invention, first, a processing of outline extraction is effected, as the first data processing step, on the aforesaid original data supplied, and positions of the outline on X- and Y-coordinates are determined.

In this case, outline extraction from such original data which can specify original data being dot-decomposed in the aforesaid row and column matrix-form or bit patterns being dot-decomposed in the row and column matrix-form is processed by determining a dot position at which binary data corresponding to the respective decomposed dots change from "0" to "1" or from "1" to "0" in the row or column direction.

If the data relating to the position of the outline thus extracted are the ones which can specify the positions of the outline on X- and Y-coordinates, such data are sufficient for the method according to the present invention and the data having any type of storage may be utilized.

Therefore, the data may be stored in any such suitable manner that X-, Y-coordinate data in respect of the respective outline dots are stored in accordance with table type manner, that bit datum corresponding to the outline position is stored on a frame memory as "1", or the like manner.

Then, a second data processing for making the outline specified in the aforesaid first process to be a set of blocks B is practiced.

Namely, in the second process, point P is established on the outline, and the outline is split in the interval of a single-valued function involving x (row direction ... The same shall apply hereinafter.) as the variable, whereby suitable blocks B are determined, respectively.

In this case, the establishment of the aforesaid blocks B can generally be processed by extracting an optional interval in which X-coordinate values of the respective decomposed dots on a series of outline increase or decrease monotonously as block B.

As a more specific method for establishing blocks B, there is proposed such a method wherein as illustrated in Fig. 3, decomposed dots in which X-coordinate values of the decomposed dots on a series of outlines come to be extreme values are determined, thereafter, the decomposed dots corresponding to the extreme values are utilized as initial break-points P, and further second break-points P' are established suitably on the respective outlines split by the break-points P, whereby intervals split by the respective break-points (P or P') may be employed as the blocks B, respectively.

Physical meaning of each block B thus established is in that each block B corresponds to a set of points making bright points to be scanned in Y-direction (column direction ... The same shall apply hereinafter.) turn-on or turn-off in case of interpreting the data compressed in accordance with the method of the present invention to

reproduce character image on a raster scanning type character display unit such as cathode ray tube or the like.

A method according to the present invention is the one for specifying the outline of a whole character as a set of plural blocks B determined by means of the above described second process.

And in the present invention, a third process for further spliting any one of the blocks B into a plurality of segments S_i as a pretreatment for determining specifically coded data corresponding to the blocks B is practiced.

Fig. 4 is a typical graphical representation indicating a relationship between any one of the blocks B determined in the aforesaid second process and the segment S_i on X- and Y-coordinates wherein points P_1 and P_n are start point and end point, respectively.

Namely, in the third process, as indicated in Fig. 4, any one of the blocks B is delimited with (n - 2) of break-points P_i $(x_i : y_i)$ where i = 2 to (n - 1) to determine (n - 1) of segments $S_i\left[P_i : P_{i+1}\right]$ where i = 1 to (n - 1), and P_i and P_{i+1} are start and end points of the segment S_i, respectively, in respect of each block B.

In this case, after all, if the aforesaid break point P_i has been established on a character outline corresponding to the block B, it is satisfactory as is also apparent from Fig. 4 so that the number and space for the establishment are generally arbitrary so far as they are within a range where whose character image reproducibility is allowable.

However, situations for establishing the block B and the segment S_i processed in the above described second and third processes are important factors exerting directly influence upon compressibility of the character data compressed by the method according to the present invention.

In this connection, the present inventors have made also the invention in respect of preferred methods for establishing these blocks B and the segments S_i as mentioned hereinbefore.

Meanwhile, when attention is paid to Fig. 4, besides an m-order curve passing through each start point $P_i(x_i : y_i)$ of (n - 1) of the aforesaid segments S_i is imagined, the curve is specified by the following equation:

$$f_i(x) = y_i + \sum_{j=1}^{m} A_{i,j} (x - x_i)^j \quad \ldots\ldots \quad (1)$$

Accordingly, if all the values of effective coefficient $A_{i,j}$ where i = 1 to (n - 1), and j = 1 to m, in the segment S_i of the interval (x_i to X_{i+1}) have previously been determined in accordance with any method, a y-coordinate value in a range (x_1 - x_n) corresponding to any one block B can be successively computed from the above equation (1). Furthermore, it becomes possible to approximate the y-coordinate value thus computed as y-coordinate value of a desired character outline in the aforesaid range (x_1 to x_n) of block B.

In the following description, an embodiment wherein each segment S_i is made to be approximated by means of the following third-order curve where m = 3:

$$f_i(x) = y_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad \ldots\ldots (2)$$

will be described.

Now, in the case where m = 3, three unknown coefficients b_i, c_i and d_i exist in respect of each segment S_i. As a result, 3 (n - 1) of unknown coefficients exist with respect to any one of the blocks B consisting of (n - 1) of the aforesaid segments S_i.

Accordingly, in the case where the aforesaid unknown coefficients required for approximating the character outline in any one block B have been determined beforehand, 3 (n - 1) of relational expressions are needed, even if any method is adopted.

Meantime, an outline of character, pictorial image or the like possesses, as mentioned already, the following general properties such that (1) the outline itself is naturally continuous, besides (2) primary derived function of the outline (inclination of the outline) varies continuously, and (3) secondary derived function of the outline (rate of change of the inclination) also varies continuously.

Under the circumstances, in the method according to the present invention, attention is paid to the several properties (1) - (3) involved in the outline of a character, pictorial image or the like, besides such attention is

also paid to such a fact that these several properties can afford some of the relational expressions required as described above.

Namely, the following three relational expressions:

$$\left.\begin{aligned} f_{i-1}(x_i) &= f_i(x_i) \\ f_{i-1}'(x_i) &= f_i'(x_i) \\ f_{i-1}''(x_i) &= f_i''(x_i) \end{aligned}\right\} \quad \ldots\ldots \quad (3)$$

are given as a condition by which the adjacent two segments S_{i-1} and S_i satisfy the aforesaid properties (1) - (3), respectively, in the aforesaid optional one break-point $P_i(x_i : y_i)$.

In this case, the number of break-points P_i is, as mentioned above, $(n - 2)$ in total in case of $i = 2 - (n - 1)$ so that after all, $3(n - 2)$ of relational expressions can be obtained from the aforesaid properties of character outline.

Furthermore, for example, the following three relational expressions;

$$\left.\begin{aligned} f_{n-1}(x_n) &= y_n \\ f_1'''(x_1) &= C_1'''(x_1) \\ f_{n-1}'''(x_n) &= C_n'''(x_n) \end{aligned}\right\} \quad \ldots\ldots \quad (4)$$

where C_1 (x) and C_n (x) are third-order equations passing through points P_1, P_2, P_3 and P_4, and points P_{n-3}, P_{n-2}, P_{n-1} and P_n, respectively, are given as a boundary condition at the start point P_1 and the end point P_n in the block B.

As described above, it is arranged that

3 (n - 1) in total

of relational expressions are given on the basis of the properties of character outline and a boundary condition at the both ends of block B.

Consequently, as mentioned above, when simultaneous equations being composed of these expressions are solved, the aforesaid (n - 1) sets of unknown coefficients b_i, c_i and d_i can be determined, and the solution thereof is as follows:

$$\left.\begin{aligned} b_i &= (y_{i+1} - y_i)/h_i - h_i\ (\sigma_{i+1} + 2\sigma_i) \\ c_i &= 3\sigma_i \\ d_i &= (\sigma_{i+1} - \sigma_i)/h_i \end{aligned}\right\} \quad \ldots\ldots \quad (5)$$

where

$$\sigma_i = (\beta_i - h_i \cdot \sigma_{i+1})/\alpha_i \quad \ldots\ldots \quad (6),$$

and value of the aforesaid σ_i is as follows:

$$\sigma_n = \beta_n/\alpha_n$$

$$h_i = x_{i+1} - x_i$$

$$\beta_1 = h_1^2 \cdot \Delta^{(3)}{}_1$$

$$\beta_i = (\Delta_i - \Delta_{i-1}) - h_{i-1} \cdot \beta_{i-1}/\alpha_{i-1}$$

$$\beta_n = - h_{n-1}^2 \cdot \Delta^{(3)}_{n-3} - h_{n-1} \cdot \beta_{n-1}/\alpha_{n-1}$$

$$\alpha_1 = - h_1$$

$$\alpha_i = 2 (h_{i-1} + h_i) - h_{i-1}^2/\alpha_{i-1}$$

$$\alpha_n = - h_{n-1} - h_{n-1}^2/\alpha_{n-1}$$

$$\Delta_i = (y_{i+1} - y_i)/h_i$$

$$\Delta^{(2)}_i = (\Delta_{i+1} - \Delta_i)/(x_{i+2} - x_i)$$

$$\Delta^{(3)}_i = (\Delta^{(2)}_{i+1} - \Delta^{(2)}_i)/(x_{i+3} - x_i).$$

Accordingly, values of unknown coefficients b_i, c_i and d_i have previously been computed from the aforesaid equation (5) as a forth data processing step, and the coefficients thus computed as well as coordinates $(x_i : y_i)$ of points P_i defining respective segments S_i are stored as compressed, coded data corresponding to the aforesaid blocks B.

Of course, interpretation of these compressed data, that is, reproduction of the character image is practiced by operating values of f_i (x) in the equation (2) in respect of the segments S_i corresponding to arbitrary x-coordinate values.

In this case, such outline determined by means of values in accordance with the aforesaid operation is the one in which the outline itself is not only continuous, but also primary derived function (inclination of the outline) as well as secondary derived function (rate of change of

the inclination) of the aforesaid outline change continuously so that such an output being more approximate to a natural character image can be obtained.

Fig. 5 is a view for illustrating one example of a preferred data memory format which is utilized in case of practicing the method according to the present invention as described above wherein N is the number of break-points P_i existing in any one of blocks, XST and YST indicate X- and Y-coordinate values of start point P_1 in any one block, $\Delta_{i.x}$ and $\Delta_{i.y}$ are differences along X- and Y-directions between two adjacent points P_i and P_{i-1}, respectively, and coordinates of an arbitrary break-point P_i are successively specified by means of these differences.

The data compression method as mentioned above is the one which relates to a method for storing coefficients b_i, c_i and d_i as a part of compression data.

In view of the above, another method in which data other than the above described coefficients are stored will be described hereinbelow.

First of all, as a premise of the explanation, when secondary derived function of the aforesaid equation (2) concerning each segment S_i is determined, it is expressed by the following equation:

$$f_i''(x) = 2c_i + 6d_i(x - x_i) \quad \ldots\ldots (7)$$

As a consequence, a value of the secondary derived function in the start point P_i $(x_i : y_i)$ for segment S_i is as follows:

$f_i''(x_i) = 2c_i$ = constant ..... (8).

In this connection, σ_i in the aforesaid equation (6) is such a constant possessing the following property being derived from the aforesaid equations (5) and (8):

$\sigma_i = 2c_i/6$ = (value of secondary derived function in the start point P_i of the segment S_i)/6.

Accordingly, in the other method according to the present invention, a value of σ_i possessing such property as mentioned above has previously been computed as the fourth data processing step, and the constant σ_i thus computed as well as coordinates $(x_i : y_i)$ of points P_i defining respective segments S_i are stored as compressed, coded data corresponding to the aforesaid blocks B.

Interpretation of these compressed data is practiced by operating values of $f_i(x)$ from the aforesaid equations (5) and (2) in respect of the segments S_i corresponding to arbitrary x-coordinate values.

Of course, in also this case, primary and secondary derived functions of outline vary continuously so that a natural character image is reproduced and outputted similarly as in the case where the aforesaid coefficients are utilized as the compression data.

On one hand, in case of the latter method where the constant σ_i is stored, an additional operation of the equation (5) must be executed in case of interpreting compression data as mentioned above.

Therefore, it is inevitable that a period for reproducing character image becomes longer than that of the former method in which coefficients are directly stored. On the other hand, however, it is sufficient to store only one constant for the latter method unlike the former method which requires to store three coefficients. Hence, the latter method has been further improved as compared with the former method in view of data compressibility.

Fig. 6 is a schematic view illustrating a constructional example for determining the memory data as described above wherein reference numeral 61 designates a scanner for determining row and column matrix-formed binary data corresponding to two-dimensional character image. The scanner 61 is arranged in such that transmitted light or reflected light from a character image pattern 63 being irradiated by means of scanning light from, for example, a flying-spot tube 62 is detected by a detector 64, and the signal detected is subjected to restoration sampling, whereby desired binary data are determined.

Reference numeral 65 designates a memory for once storing original binary data determined in the scanner 61. Reference numeral 66 designates an arithmetic control part wherein the aforesaid binary data supplied in accordance with online or batch processing are subjected to data processing to output desired memory data to a compressed data memory 67, and it is preferable that such arithmetic control part is generally constructed by means of electronic computer.

Principal processing order in the aforesaid arithmetic control part 66 is as follows.

(1) Input of original binary data

(2) Extraction of outline

(3) Establishment of block B

(4) Establishment of segment S_i

(5) Computation of desired memory data

(6) Output of memory data

As a result of an examination wherein the character or pictorial image data compression method according to the present invention as described hereinbefore is applied to one of Japanese characters "あ" being composed of 800 x 800 dots, if allowable error with respect to a desired character image is assumed to be 1 dot, data compressibilities 1.385 % and 0.75 % could be obtained in the former method for storing coefficients and the latter method for storing constant σ_i, respectively.

Next, such a case in which character images of various scale factors are reproduced on the basis of memory data obtained by the character or pictorial image compressing method as mentioned hereinbefore will be described hereunder.

Fig. 7 is an explanatory view indicating a relationship between a master font 71 composed of FXF dots and an output character area 72 composed of GXG dots wherein output character image is outputted to the master font 71 with scale factor of a = G/F times. Accordingly, x-coordinate value X_c on the output character area 72 corresponds to X_c/a on the master font 71.

For this reason, first, a segment S_i which comes to be $x_i \leq (X_c/a) \leq x_{i+1}$ on the master font 71 is specified, and a value of $y_c = a \cdot f_i\ (X_c/a)$ is determined from the aforesaid equation (2) on the basis of memory data relating to the segment S_i thus specified, whereby y-coordinate value y_c of the character outline corresponding to X-coordinate X_c on the output character area 72 can be determined.

The character image thus reproduced on the output character area 72 is formed by multiplying the value determined from the aforesaid equation (2) by a constant of proportionality (scale factor a) so that such reproduced image being always proportional to the master font is obtained irrespective of extent of the scale factor. Accordingly, even if it is in such a case where a character having a larger size than that of the character image specified on the master font 71 is outputted, smoothness of the outline can be maintained.

Furthermore, different points of the method according to the present invention from those of conventional m-order curve approximating method reside in that the conventional method utilizes only coordinate values of the points established on a character outline as whose memory data, and contemplates to approximate the outline by means of the m-order curve specified by points of optionally continuous (m + 1), whilst the method of the present invention utilizes also additional information other than

point coordinate values, and contemplates to approximate an interval between optionally adjacent two points by means of m-order curve. In addition, the aforesaid additional information is the one which is determined to make a connecting condition for two segments at both the ends of an arbitrary break-point consistent with properties of whose character image so that the method according to the present invention makes possible to reproduce the character image which is far closer to whose desired character image as mentioned above.

Attention is drawn to our copending application No. 8508056 which is directed to a method of compressing character or pictorial image data employing features described in this specification.

* C L M S *

CLAIMS:

1. A method of compressing character or pictorial image data which compresses a quantity of data by storing outline specifying information of said character, pictorial image or the like, comprising:

as a first step, determining the outline of said character, pictorial image or the like in terms of orthogonal X- and Y-coordinates;

as a second step, splitting said outline into a set of blocks $[P_1 : P_n]$: where P_1 and P_n are start and end points of any one block;

as a third step, subdividing each of any one of said blocks with $(n - 2)$ break-points P_i $(x_i : y_i)$: where $1 = 2$ to $(n - 1)$, to obtain $(n - 1)$ of segments S_i: where $i = 1$ to $(n - 1)$; and

as a fourth step, computing for each segment S_i a constant σ_i corresponding to 1/6 a value of second derivative of the third order curve passing through the start point $P_i(x_1,y_i)$ of the segment at the start point; and

storing for each segment the constant σ_i and coordinate values $(x_i : y_i)$: where $i = 1$ to n, of the points P_i defining the segments S_i as the outline data corresponding to said blocks.

2. A method for compressing character or pictorial image data as claimed in claim 1, wherein value of said constant σ_i has the following relationship:

$$\sigma_i = (\beta_i - h_i \cdot \sigma_{i+1})/\alpha_i$$

where $\sigma_n = \beta_n/\alpha_n$

$$h_i = x_{i+1} - x_i$$

$$\beta_1 = h_1^2 \cdot \Delta^{(3)}_1$$

$$\beta_i = (\Delta_i - \Delta_{i-1}) - h_{i-1} \cdot \beta_{i-1}/\alpha_{i-1}$$

$$\beta_n = - h_{n-1}^2 \cdot \Delta^{(3)}_{n-3} - h_{n-1} \cdot \beta_{n-1}/\alpha_{n-1}$$

$$\alpha_1 = - h_1$$

$$\alpha_i = 2 (h_{i-1} + h_i) - h_{i-1}^2/\alpha_{i-1}$$

$$\alpha_n = - h_{n-1} - h_{n-1}^2/\alpha_{n-1}$$

$$\Delta_i = (y_{i+1} - y_i)/h_i$$

$$\Delta^{(2)}_i = (\Delta_{i+1} - \Delta_i)/(x_{i+2} - x_i)$$

$$\Delta^{(3)}_i = (\Delta^{(2)}_{i+1} - \Delta^{(2)}_i)/(x_{i+3} - x_i).$$

3. A method of compressing character or pictorial data according to claim 1, and substantially as herein described, with reference to the accompanying drawings.