



(22) Date de dépôt/Filing Date: 1999/03/12

(41) Mise à la disp. pub./Open to Public Insp.: 1999/09/13

(45) Date de délivrance/Issue Date: 2002/10/15

(30) Priorité/Priority: 1998/03/13 (09/042,016) US

(51) Cl.Int.⁶/Int.Cl.⁶ G06K 9/62

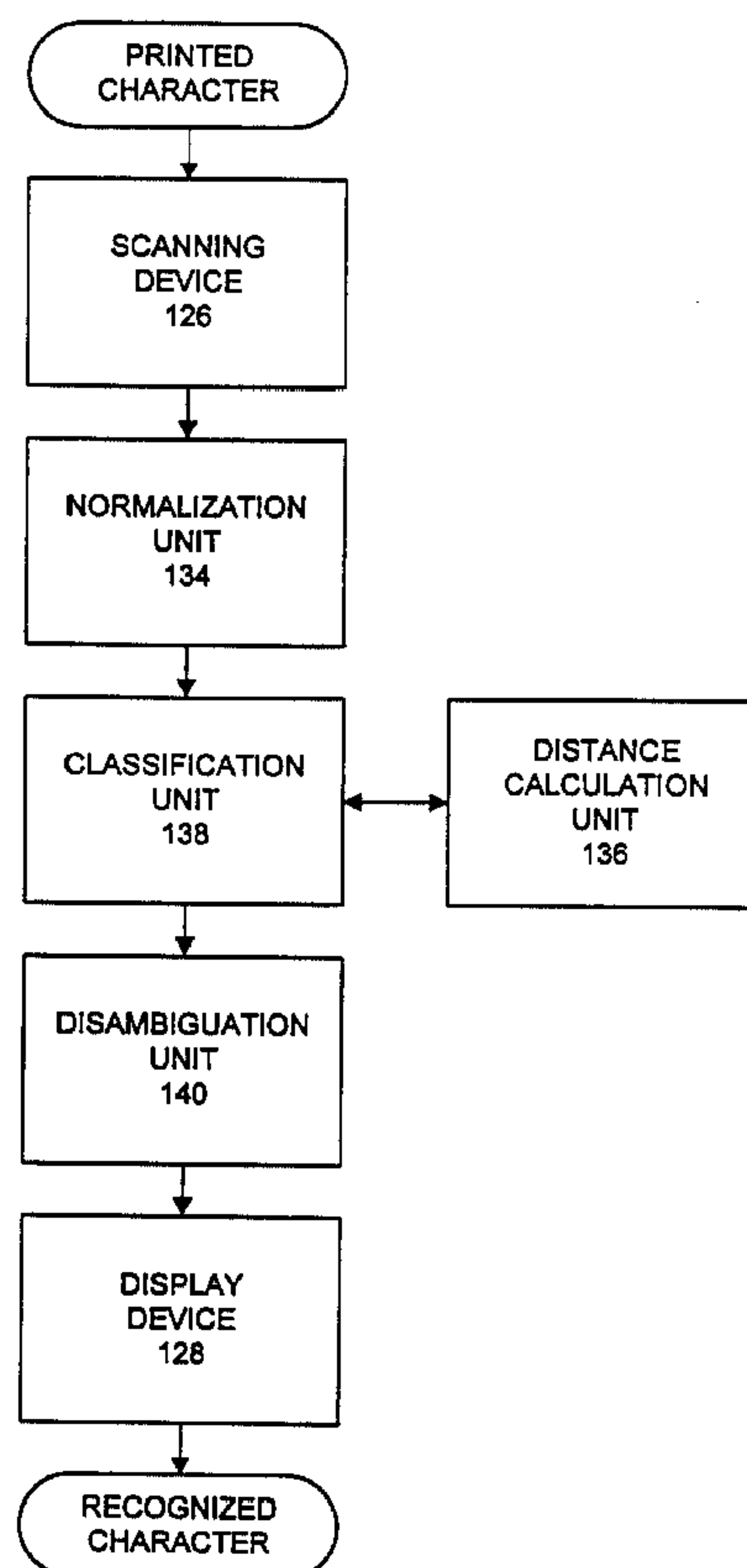
(72) Inventeurs/Inventors:
KRTOLICA, Radovan V., US;
MELEN, Roger D., US

(73) Propriétaire/Owner:
CANON KABUSHIKI KAISHA, JP

(74) Agent: RIDOUT & MAYBEE LLP

(54) Titre : OPERATION DE SEUILLAGE PAR CLASSIFICATION D'UNE IMAGE EN NIVEAUX DE GRIS NORMALISES

(54) Title: CLASSIFICATION-DRIVEN THRESHOLDING OF A NORMALIZED GRayscale IMAGE



(57) Abrégé/Abstract:

A sample image (142) is recognized by normalizing (404) the size of a sample image (142) to the size of a referent images (146); and determining (406) a set of candidate images (147) from a set of referent images (146), wherein each of the candidate images (147) is within an acceptable distance from a different binarization (145) of the sample image (142). A system (120) for

(57) **Abrégé(suite)/Abstract(continued):**

image recognition includes a scanning device (126), a normalization unit (134), a distance calculation unit (136), a classification unit (138), a disambiguation unit (140), and a display device (128).

**CLASSIFICATION-DRIVEN THRESHOLDING OF A NORMALIZED
GRAYSCALE IMAGE**

Abstract of the Disclosure

A sample image (142) is recognized by normalizing (404) the size of a sample image (142) to the size of a referent images (146); and determining (406) a set of candidate images (147) from a set of referent images (146), wherein each of the candidate images (147) is within an acceptable distance from a different binarization (145) of the sample image (142).

A system (120) for image recognition includes a scanning device (126), a normalization unit (134), a distance calculation unit (136), a classification unit (138), a disambiguation unit (140), and a display device (128).

**CLASSIFICATION-DRIVEN THRESHOLDING OF A NORMALIZED
GRAYSCALE IMAGE**

Technical Field

The present invention relates generally to image recognition, and more particularly, to a method and system for optical character recognition by classification-driven thresholding of a normalized grayscale image.

Background Art

In the art of optical character recognition, an image classifier is a functional unit that attempts to match sample images against a set of referent images or templates. Although most character images are sampled in grayscale, which results in multiple data bits per image pixel, image classifiers are generally limited to binary (bi-level) input data. Analyzing grayscale data is substantially more complicated, and requires time-consuming, sophisticated techniques. Thus, although some grayscale classifiers exist, most readily-available image classifiers accept only binary input data. A variety of binary classifiers for optical character recognition are known in the art, such as the system described in U.S. Patent No. 5,539,840 to Krtolica et al. for "Multifont Optical Character Recognition Using a Box Connectivity Approach,".

Because the sample image comprises grayscale data, but the image classifier accepts only binary data, the sample image must be converted initially from grayscale into black and white. This step normally requires a process called thresholding or binarization, which

1 includes selecting a median gray level (usually called a "binarization threshold" or
2 "threshold") and changing the value of each image pixel to either zero or one, depending on
3 whether the original gray level of the pixel had a value greater or less than that of the
4 threshold. In conventional systems, binarization of the sample image is generally performed
5 once, using a single threshold, after which the binary output data is provided to the image
6 classifier.

8 As conventionally implemented, however, thresholding often dramatically reduces
9 recognition accuracy. When an image is thresholded, much useful information about the
10 image is lost. For example, an eight bit grayscale image contains eight times more data than
11 the same thresholded image. Such data assist the human eye in recognizing the image, but
12 are lost to conventional image recognition systems because of thresholding.

14 In addition, thresholding introduces harmful noise into the image. Slight deviations
15 in the image's gray levels are often manifest after thresholding in the form of jagged edges,
16 stray pixels, gaps, and other artifacts that reduce recognition accuracy. Moreover, after
17 thresholding, the sample image is typically normalized to the size of the referent images.
18 However, normalizing binary data generally compounds the noise, reducing recognition
19 accuracy to an even greater degree. What is needed, then, is a method and system for
20 providing binary data to a binary image classifier while retaining as much information as
21 possible about the original grayscale image and reducing the noise associated with the
22 processes of thresholding and normalization.

25 As noted earlier, in conventional systems, thresholding is normally performed as a
26 separate step from image classification. Thus, in such systems, thresholding is merely a
27 simplification or quantizing step. However, as shown in Figure 1, thresholding is central to
28

1 classification and is not so easily separable therefrom. For example, matrix (a) of Figure 1
2 represents a grayscale image sampled at eight bits (256 gray levels) per pixel. If the
3 binarization threshold ("T") is selected to be 128, matrix (b) illustrates the resulting binary
4 image, which would be interpreted by a binary image classifier as the letter "U." If,
5 however, the threshold is selected to be 140, matrix (c) illustrates the resulting binary image,
6 which would be interpreted to be the letter "L." Both interpretations are valid. However, in
7 each case, the selection of the binarization threshold determines which pixels are in the
8 foreground ("1") and which pixels are in the background ("0"). Thus, the thresholding step
9 effectively determines the classification of the image.
10

11
12 The foregoing situation often occurs where there is poor contrast between the
13 foreground and background, and where the foreground or background gray levels are not
14 uniform throughout the sampled image. The human eye can easily compensate for these
15 anomalies. However, a conventional image recognition system that separately thresholds the
16 image before classification will frequently produce inaccurate results. Indeed, as shown
17 above, an arbitrary selection of either threshold will often eliminate valid, and possibly
18 correct, interpretations of the character image.
19

20 Conventionally, a binary image classifier cannot detect such alternative
21 interpretations based on different thresholds since the thresholding step is performed
22 separately from classification. If thresholding could be performed with a foreknowledge of
23 the referent images, then a number of possible interpretations of the sample image, based on
24 different thresholds, could be determined. Moreover, only those interpretations having an
25 acceptable "distance" from the binarized sample image could be selected.
26
27
28

Disclosure of Invention

In accordance with the present invention, a system (120) for image recognition includes a scanning device (126), a normalization unit (134), a distance calculation unit (136), a classification unit (138), a disambiguation unit (140), and a display device (128).

Brief Description of the Drawings

These and other more detailed and specific objects and features of the present invention are more fully disclosed in the following specification, reference being had to the accompanying drawings, in which:

Figure 1 is an example of a classification problem caused by the choice of different binarization thresholds;

Figure 2 is a physical block diagram of a system 120 for recognizing a character image 142 in accordance with the present invention;

Figure 3 is a dataflow diagram of an image recognition system 120 in accordance with the present invention;

Figure 4 is a flow diagram of a preferred method for recognizing a character image 142 in accordance with the present invention;

Figure 5 is a graphical illustration of the process of normalization;

Figure 6 is a flow diagram for an exhaustive method of determining a candidate set 147 in accordance with the present invention;

Figures 6A and 6B are a flow diagram for a preferred method of determining a candidate set 147 in accordance with the present invention;

Figure 6C is a flow diagram for a preferred method of disambiguating a candidate set 147 in accordance with the present invention;

Figure 7A is an example of a gray matrix 144;

Figures 7B and 7C are examples of referent images 146;

1 Figure 7D is a graph of Hamming distances calculated between a gray matrix 144, at
2 a number of thresholds, and two referent images 146a and 146b, respectively;

3 Figure 7E is a graph of Hamming distances between portions of a gray matrix 144, at
4 a number of thresholds, and portions of two referent images 146a and 146b, respectively;
5

6 Figures 8A and 8B are examples of referent images 146;

7 Figure 8C is a partitioning of a gray matrix 144 according to the foreground and
8 background of the referent image 146a in Figure 8A; and

9 Figure 8D is a partitioning of a gray matrix 144 according to the foreground and
10 background of the referent image 146b in Figure 8B.
11

12 **Detailed Description of the Preferred Embodiments**

14 The following description is made in the context of recognizing images of
15 handwritten or machine-printed characters. However, the principles and methods disclosed
16 herein may be used to recognize a variety of images types, in addition to feature vectors that
17 are used to characterize images. Thus, the scope of the present invention should not be
18 limited only to character recognition. In addition, although the following description is in the
19 context of recognizing grayscale images, one skilled in the art will recognize that color
20 images may also be recognized using the method and system of the present invention.
21
22

23 **System Architecture**

24 Referring now to Figure 2, there is shown a physical block diagram of a system 120
25 for recognizing a character image in accordance with the present invention. In one
26 embodiment, the present invention is implemented as software running on a conventional
27
28

1 personal computer such as an IBM® PC or compatible. Thus, the hardware architecture of
2 system 120 as shown in Figure 2 is preferably implemented as a combination of components
3 of such computer, although other implementations are possible.
4

5 A central processing unit (CPU) 122 executes software instructions and interacts with
6 other components to perform the methods of the present invention. A storage device 124
7 provides long term storage of data and software programs, and may be implemented as a hard
8 disk drive or other suitable mass storage device. A scanning device 126 obtains a two-
9 dimensional array of pixel values representing the character to be recognized. In a preferred
10 embodiment, scanning device 126 is an "HP ScanJet IIc" model digital scanner from Hewlett
11 Packard Co., which provides a resolution of 400 dots (pixels) per inch, each pixel being
12 quantized with an eight-bit grayscale resolution. Input device 127, such as a mouse or
13 keyboard, facilitates user control of the operation of system 120. A display device 128 is an
14 output device such as a cathode-ray tube or printer for the display of text and graphics under
15 the control of CPU 122. System 120 may also include communication device 130 such as,
16 for example, a modem or ethernet card for connection to a computer network such as the
17 Internet.
18

19
20 System 120 also includes a memory 132 for storing software instructions to be
21 executed by CPU 122. Memory 132 is implemented using a standard memory device, such
22 as a random access memory (RAM). In a preferred embodiment, memory 132 includes a
23 number of software objects or modules, including a normalization unit 134, a distance
24 calculation unit 136, a classification unit 138, and a disambiguation unit 140. Throughout
25 this discussion, the foregoing modules are assumed to be separate functional units, but those
26
27
28

1 skilled in the art will recognize that the functionality of various units may be combined and
2 even integrated into a single software application.

3 In a preferred embodiment, the memory 132 is also used to store a sample image 142,
4 a gray matrix 144, a binary matrix 145, a plurality of referent images 146, and a candidate set
5 147. The sample image 142 is a bitmapped grayscale image that is captured by the scanning
6 device 126. The gray matrix 144 is preferably a two-dimensional array of integers, and is
7 used to characterize the grayscale information in the sample image 142. The binary matrix
8 145 is a binarized version of the gray matrix 144, and is used in calculating the distance
9 between the gray matrix 144 and the referent images 146. The referent images 146 are a set
10 of previously-obtained, known templates that are compared against the gray matrix 144.
11 Like the gray matrix 144, the referent images 146 are preferably stored as two-dimensional
12 arrays, but contain only bi-level (binary) data. The candidate set 147 is a set of referent
13 images 146 selected as potential candidates for recognition.
14
15

16 Finally, memory 132 includes an operating system 148, for managing, and providing
17 system resources to, the above-mentioned software objects or modules. Preferably, operating
18 system 148 is the Windows 95 operating system manufactured by Microsoft Corporation of
19 Redmond, Washington, although other operating systems, such as UNIX, may be used within
20 the scope of the present invention.
21
22

23 Dataflow of the Image Recognition System

24 Referring now to Figure 3, there is shown a dataflow diagram of an image recognition
25 system 120 in accordance with the present invention. Initially, a sample image 142 is
26 obtained by means of scanning device 126, which obtains a pixel-by-pixel representation of
27
28

1 an image on a scanned object, such as a piece of paper. One skilled in the art will recognize
2 that, typically, an entire page of text is scanned at one time. This generally necessitates the
3 additional step of segmentation, wherein the page is segmented into individual characters as
4 described in U.S. Patent No. 5,539,840 to Krtolica et al. for "Multifont Optical Character
5 Recognition Using a Box Connectivity Approach,". However, for purposes of
6 discussion, it will be assumed that the image 142 represents only a single character, as
7 produced by the segmentation process.
8

9
10 Coupled to the scanning device 126 is the normalization unit 134, which generates the
11 gray matrix 144 by normalizing the size of the sample image 142 to the size of the referent
12 images 146. This process will be described in greater detail below with reference to Figure
13 5. Briefly, however, the normalization unit 134 maps the sample image 142 onto a grid
14 corresponding to gray matrix 144 and calculates a pixel "weight" for the pixels mapped by
15 each region of the grid.
16
17

18 In a preferred embodiment, the gray matrix 144, the binary matrix 145, and the
19 referent images 146 have dimensions of 16 x 16 elements. The precise dimensions are not
20 crucial to the invention, although they are preferably identical for each matrix. In addition,
21 the dimensions should be chosen with several factors in mind. For example, one skilled in
22 the art will recognize that the dimensions should be chosen in order to minimize noise, while
23 capturing sufficient information about the edges of a character. Additionally, pragmatic
24 concerns are relevant, such as selecting dimensions that are a power of two in order to
25 simplify implementation in hardware and software, and selecting small enough dimensions
26 so as not to overburden conventional memories and CPUs.
27
28

1 Coupled to the normalization unit 134 is the classification unit 138, which compares
2 the gray matrix 144 with the referent images 146 in order to determine a candidate set 147.
3 This process will be described in greater detail below with respect to Figures 6A-B. Unlike
4 conventional systems, the classification 138 unit, in conjunction with the distance calculation
5 unit 136, determines a candidate set 147 by classification-driven thresholding of the gray
6 matrix 144. In one embodiment, this is accomplished by calculating a set of distances
7 between the gray matrix 144, binarized at a plurality of thresholds, and the set of referent
8 images 146. Thereafter, a candidate set 147 is selected wherein the distance between a
9 candidate image 147 and a binarization of the sample image 142 is within an acceptable
10 range.
11

12
13 Coupled to the classification unit 138 is the disambiguation unit 140, which is used to
14 select a preferred candidate 147 if more than one is found by the classification unit 138.
15 Briefly, this is accomplished by spatially normalizing each candidate 147 to the size of the
16 original sample image 142, and repeating the classification method as will be described
17 hereafter with reference to Figure 6C.
18

19 Finally, coupled to the disambiguation unit 140 is the display device 128, such as a
20 cathode-ray tube or printer. Display device 128 outputs the recognized candidate image 147
21 to a user of the system 120. If no unique candidate image 147 was identified, then the
22 display device 128 preferably outputs an error message.
23

24 **Preferred Methods for Image Recognition**

25 Referring now to Figure 4, there is shown a flow diagram of a preferred method for
26 recognizing a character image in accordance with the present invention. The method begins
27
28

1 by obtaining 402 the sample image 142 of the character to be recognized. This is
2 accomplished by means of the scanning device 126, which obtains a pixel-by-pixel
3 representation of the character. Because the physical dimensions of a character varies, the
4 pixel dimensions of the image 142 will also vary. Generally, the dimensions of a handwritten
5 character will vary widely, and may be as large as 200 pixels high by 100 pixels wide. In
6 contrast, a machine-printed, fourteen point character will have approximate dimensions of 48
7 x 32 pixels. Referring also to Figure 5, there is shown a graphical illustration of the process
8 of normalization, including a sample image 142 of the letter "M." In this example, the
9 sample image 142 has pixel dimensions of 64 x 64.
10

11
12 The method continues by normalizing 404 the size of the sample image 142 to the
13 size of the referent images 146. In a preferred embodiment, the referent images 146 are
14 stored as two-dimensional matrices, each having dimensions of 16 x 16 elements. It is
15 advantageous, therefore, to convert the sample image 142 into a similarly-sized matrix for
16 purposes of comparison.
17

18 As shown in Figure 5, the normalization step is accomplished by mapping the sample
19 image 142 onto a 16 x 16 grid 500 of boxes 502. The grid 500 is used for assigning values to
20 the gray matrix 144, each box 502 being associated with a corresponding element of the gray
21 matrix 144. As noted above, the choice of dimensions is not crucial to the invention.
22 However, in a preferred embodiment, the sample image 142 should be mapped onto a grid
23 500 with the same dimensions as the referent images 146.
24

25 As further shown in Figure 5, because the sample image 142 has dimensions of 64 x
26 64 pixels, each box 502 is associated with 16 pixels. Additionally, as illustrated in an
27 enlarged grid portion 506, each pixel has a corresponding gray level, which is an integer
28

1 value between zero and 255. The gray level is determined by the scanning device 126 when
2 the image 142 is obtained. Preferably, a gray level of zero corresponds to a completely white
3 pixel, and a gray level of 255 corresponds to a completely black pixel. Values between zero
4 and 255 represent shades of gray. However, one skilled in the art will recognize that some
5 scanning devices 126 invert these color mappings.
6

7 After the sample image 142 is mapped onto the grid 500, a pixel "weight" is
8 calculated for each box 502 by summing the pixel gray levels contained therein. Next, each
9 sum is stored in the corresponding element of the gray matrix 144, as illustrated in the
10 enlarged matrix portion 508 of Figure 5. When the normalization process is complete, each
11 element of the gray matrix 144 comprises an integer value corresponding to the appropriate
12 pixel weight of the region mapped by grid 500.
13

14 One skilled in the art will recognize that the vertical dimensions of the grid 500 and
15 the sample image 142 may be co-prime, as may the horizontal dimensions. For example, a
16 sample image 142 with dimensions of 61 x 47 may be partitioned into a grid 500 with
17 dimensions of 16 x 16. When this occurs, boxes 502 are likely to bisect pixels. Although
18 approximations can be made to compensate for the problem, recognition accuracy may be
19 impaired. Therefore, in a preferred embodiment, if a pixel is bisected, the gray level of the
20 affected pixel is proportionately divided between the two or more boxes 502 containing the
21 pixel. In that way, an exact calculation of the relative weight of each box 502 is enabled.
22

23 Moreover, in a preferred embodiment, the summation process described above
24 employs rational number arithmetic implemented with integer operations, using techniques
25 well known to those skilled in the art. In contrast, if values are calculated using floating
26 point arithmetic, approximations are still necessary, and speed and memory problems may
27
28

1 result. Therefore, all values are stored as rational numbers, and all functions are adapted to
2 process rational numbers. This can be done, for example, by custom-developing such
3 functions, or by using any of a number of standard packages for rational number arithmetic
4 implemented with integer operations.
5

6 Referring again to Figure 4, after the normalization step 404 is complete, the gray
7 matrix 144 is a normalized version of the sample image 142 and comprises grayscale data.
8 However, since the referent images 146 comprise bi-level data, the gray matrix 144 must be
9 binarized in order to be compared with the referent images 146. Conventionally, a threshold
10 is selected, and the grayscale data is binarized using standard thresholding techniques.
11 However, as illustrated in Figure 1, the selection of a single threshold may eliminate valid,
12 and possibly correct, interpretations of the image.
13

14 Therefore, in a preferred embodiment, the classification unit 138, in conjunction with
15 the distance calculation unit 136, determines 406 a candidate set 147 using a classification-
16 driven thresholding of the sample image 142. The following description outlines the theory
17 of classification-driven thresholding as well as an exhaustive method for using the technique
18 to determine a candidate set 147. Thereafter, a more efficient, preferred method for
19 classification-driven thresholding is presented.
20

21 Exhaustive Classification-Driven Thresholding

22

23 Conceptually, classification-driven thresholding is an integration of the thresholding
24 and classification steps of image recognition wherein a candidate set is determined by
25 comparing the gray matrix 144, binarized at all possible thresholds, with the set of referent
26 images 146. Figure 6 is a flow diagram of the exhaustive method of classification-driven
27
28

1 thresholding. The method begins by setting 602 the initial threshold ("T") to a minimum
 2 value. In a preferred embodiment, the minimum value is zero. However, in alternative
 3 embodiments, T may start at other desired values, such as the lowest actual value in the gray
 4 matrix 144.
 5

6 After the value of T is reset, the next referent image 146 is selected 604. Thereafter,
 7 the method continues by binarizing 606 the gray matrix 144 using T as the binarization
 8 threshold, and storing the resultant bi-level data in the binary matrix 145. In a preferred
 9 embodiment, the binarization process is performed by changing those values of gray matrix
 10 144 that are less than or equal to T into zeros, and the values that are greater than T into ones.
 11 However, the assignment of ones and zeros could be reversed without departing from the
 12 spirit of the invention.
 13

14 Thereafter, method continues by calculating 608 a distance between the binary matrix
 15 145 and the selected referent image 146. In a preferred embodiment, a Hamming distance is
 16 used as the measure of the distance. Calculating Hamming distances is well known in the art.
 17 Given two matrices, a referent matrix R and a sample matrix S, each matrix being of the
 18 same size, and each having a total number N elements valued at either 1 (representing black)
 19 or 0 (representing white), the number of corresponding elements equal to the value 1 in both
 20 R and S being represented by C[b], the number of corresponding elements equal to the value
 21 0 in both R and S being represented by C[w], a Hamming distance H(R,S) between the
 22 referent matrix R and the sample matrix S is defined as:
 23
 24

$$25 \quad H(R,S) = N - (C[b] + C[w]) \quad \text{Eq. 1}$$

26 One skilled in the art will recognize that a Hamming distance of zero means that the matrices
 27 R and S are identical. As the matrices become more divergent, the Hamming distance
 28

1 increases, until the distance reaches a maximum where one image is entirely black and the
2 other image is entirely white. The maximum Hamming distance, therefore, between the
3 preferred 16 x 16 matrices is 256.
4

5 After the distance is calculated 608, a determination 610 is made whether the distance
6 is within an acceptable range. In one embodiment, a distance of zero is the only acceptable
7 distance, which would require a perfect match. However, in a preferred embodiment, a low,
8 non-zero value is chosen experimentally since noise and other effects can create stray pixels
9 and gaps resulting in non-zero distances. In a preferred embodiment, an acceptable distance
10 is a Hamming distance of 2.
11

12 If the distance is determined 610 to be within the acceptable range, an indication of
13 the referent image 146 is stored 612 in the candidate set 147 in the memory 132. As noted
14 earlier, the referent images 146 are known templates, having characters associated with them.
15 Therefore, preferably, only the associated character and the Hamming distance need to be in
16 the candidate set 147.
17

18 Additionally, in a preferred embodiment, if the Hamming distance associated with the
19 current candidate 147 is acceptable but greater than that of a stored candidate 147, then the
20 current candidate 147 should not be stored since it is less likely to be correct. Likewise, if the
21 current candidate's 147 distance is lower than that of a stored candidate 147, then the stored
22 candidate 147 should be discarded.
23

24 Next, a determination 614 is made whether more referent images 146 remain to be
25 compared with the binary matrix 145. If more referent images 146 remain, then the method
26 returns to step 602; otherwise, the method is complete.
27
28

1 If in step 610 it is determined that the distance is not within an acceptable range, then
2 the method continues by incrementing 616 the value of T. In a preferred embodiment, the
3 threshold increment is equal to the maximal weight of the elements of the gray matrix 144
4 divided by the number of gray levels. In the example illustrated in Figure 5, the threshold
5 increment is 16 ($4080 / 255$). However, the incremental value may be 1 or any other number,
6 although one skilled in the art will recognize the trade-off between speed and accuracy in the
7 choice of the threshold increment.
8

9 After T is incremented 616, a determination 618 is made whether T exceeds a pre-
10 determined maximum. In a preferred embodiment, the maximum is the largest possible value
11 of an element of the gray matrix 144, and will vary according to the number of pixels
12 contained in each box 502. For example, in Figure 5, each box 502 contains 16 pixels.
13 Therefore, the maximum value for an element of the gray matrix 144 is 4080 ($255 * 16$). In
14 an alternative embodiment, the maximum could be set to the largest actual value stored in
15 the gray matrix 144, since thresholding beyond that value will produce identical binary
16 matrices 145. If the maximum was exceeded in step 616, then the method continues at step
17 614; otherwise, the method returns to step 606 in order to binarize the gray matrix 144 at the
18 new threshold.
19
20

21 This "exhaustive" method of classification-driven thresholding is illustrated in Figure
22 7D, which is a graph of Hamming distances calculated between the gray matrix 144 of Figure
23 7A, binarized at a number of thresholds, and two referent images 146a and 146b shown in
24 Figures 7B and 7C, respectively. For ease of illustration, the gray matrix 144 and the two
25 referent images 146a-b each have dimensions of 5 x 4 elements. Moreover, in a preferred
26 embodiment, the matrix 144 would be binarized at each integer value between zero and 255.
27
28

1 However, for simplicity, the distances shown in Figure 7D are calculated only at the points
2 indicated, and the lines, therefore, are provided merely for visualization purposes.

3 As shown in Figure 7D, the Hamming distance between the binarized gray matrix 144
4 and the referent image 146a reaches a minimum, in this case zero, at $T = 153$. Therefore, if
5 the gray matrix 144 is binarized at $T = 153$, the resulting binary matrix 145 is identical to
6 referent image 146a. Thereafter, as T increases, the distance starts to increase beginning at T
7 $= 166$. A similar graph is provided for the referent image 146b, although the minimum
8 distance is 1, and is reached at $T = 199$.

10 Each of the graphs in Figure 7D may be expressed mathematically by the function

$$12 \quad F(T; \mathbf{B}, \mathbf{N}) := \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |\mathbf{B}_{ij} - \mathbf{I}(\mathbf{N}_{ij} - T)| \quad \text{Eq. 2}$$

14 where $\mathbf{B}_{ij}$ and $\mathbf{N}_{ij}$ are elements of the $m \times n$ matrices $\mathbf{B}$ (referent image 146) and $\mathbf{N}$ (gray
15 matrix 144), respectively, T is the threshold, and $\mathbf{I}$ is the step function:

$$17 \quad \mathbf{I}(x) := \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \text{Eq. 3}$$

19 By application of the foregoing method, a set 147 of candidate images with
20 acceptable distances from the gray matrix 144 may be determined. However, this method is
21 "exhaustive" because it must calculate the distances between the gray matrix 144, binarized
22 at each integer value between zero and 255, and each referent image 146. One skilled in the
23 art will recognize a large number of calculations is required. For example, assuming there
24 are 256 gray levels and 256 referent images 146, then 65536 (256×256) distances would need
25 to be calculated in order to determine the candidate set 147. For many computers, this
26 process would be too slow to be practical.
27
28

Efficient Method for Classification-Driven Thresholding

A more efficient method for classification-driven thresholding can be achieved by recognizing that the set of referent images 146 is known. Thus, not all distance calculations need to be made. Indeed, only a few distance calculations may be required in order to find the minimum Hamming distance. In order to achieve this goal, it must be understood that each graph in Figure 7D represents, not single a function, but the sum of two functions—one representing the distance between the “foreground” of the reference image 146 and the corresponding elements of the gray matrix 144, and one representing the distance between the “background” of the referent image 146 and the corresponding elements of the gray matrix 144. In a preferred embodiment, the “foreground” is that part of the referent image 146 that contains ones, whereas the “background” is that part of the image 146 that contains zeros.

Referring now to Figure 7E, there are shown four graphs 702, 703, 704, and 705, corresponding to the distance functions described above. Graph 702 represents the distance between the background of image 146a and the corresponding elements of the gray matrix 144. Graph 703 represents the distance between the foreground of image 146a and the corresponding elements of the gray matrix 144. Graphs 704 and 705 are similar to graphs 702 and 703, except that they are associated with image 146b.

It can be shown mathematically that the functions represented by graphs 702 and 704 are monotonically nonincreasing functions of T . In addition, it can be shown mathematically that the functions represented by graphs 703 and 705 are monotonically nondecreasing functions of T . In the description that follows, the value of T where the nonincreasing

1 graphs 702 and 704 intersect the zero-threshold axis (T) will be called hereafter " T_0 ."

2 Likewise, the value of T where the nondecreasing functions intersect the T-axis will be called
3 hereafter " T_1 ." Mathematically, these values may be calculated as follows:

$$\begin{aligned} T_0 &:= \max\{N_{ij} \mid \text{for all } i, j \text{ such that } B_{ij} = 0\}; \\ T_1 &:= \min\{N_{ij} \mid \text{for all } i, j \text{ such that } B_{ij} = 1\} - 1; \end{aligned} \quad \text{Eq. 4}$$

7 Figure 7E illustrates T_0 and T_1 for the matrices shown.

8 Based on the mathematical properties of these functions, the values for T_0 and T_1 can
9 be used to determine whether the minimum Hamming distance is zero, or, alternatively, to
10 determine the range of thresholds wherein the distance is minimized. For example, if $T_0 \leq$
11 T_1 , then the Hamming distance will be zero for all values of T between T_0 and T_1 . On the
12 other hand, if $T_0 > T_1$, then the Hamming distance will be greater than zero for all values of
13 T, although the minimum distance will be found between T_1 and T_0 . Thus, by knowing the
14 values of T_0 and T_1 , it is possible to avoid many binarization and distance calculations. This
15 process is much more efficient and, therefore, provides a preferred method for classification-
16 driven thresholding.

19 Referring now to Figure 6A, there is shown a flow diagram for a preferred method of
20 determining a candidate set 147 by classification-driven thresholding in accordance with the
21 present invention. As explained above, the method must first determine values for T_0 and T_1
22 with respect to a referent image 146. Therefore, in a preferred embodiment, the method
23 begins by selecting 650 the next referent image 146. Thereafter, as illustrated in Figures 8C
24 and 8D, the gray matrix 144 is partitioned 652 into a foreground portion 710 (shown in
25 boldface), and a background portion 712, corresponding to the foreground and background of
26 the selected referent image 146. In the examples shown, partitioned matrix 144a corresponds
27
28

1 to a partitioning according to selected referent image 146a of Figure 8A, and partitioned
2 matrix 144b corresponds to a partitioning according to selected referent image 146b of Figure
3 8B.

4
5 After the gray matrix 144 is partitioned 652, the method continues by determining
6 654 values for T_0 and T_1 . In a preferred embodiment, T_0 is determined by finding the
7 maximum value in the background portion 712 of the gray matrix 144. In the case of
8 partitioned matrix 144a, T_0 is 153; with respect to partitioned matrix 144b, T_0 is 199.
9 Thereafter, T_1 is determined by finding the minimum value in the foreground portion 710
10 minus one. In the case of partitioned matrix 144a, T_1 is 165 (i.e., 166 - 1); with respect to
11 partitioned matrix 144b, T_1 is 198 (i.e., 199 - 1).
12

13 Next, the method continues by determining 656 whether $T_0 \leq T_1$. As explained
14 above, if $T_0 \leq T_1$, then the Hamming distance between the binary matrix 145 and the referent
15 image 146 is zero for all values of T between and including T_0 and T_1 . This is true, for
16 example, in the case of partitioned matrix 144a of Figure 8C, where T_0 (153) is less than T_1
17 (165). Thus, if partitioned matrix 144a is binarized with $T = 153$, for example, the resultant
18 binary matrix 145 will be identical to referent image 146a, and a perfect match is found. This
19 may be verified graphically in Figure 7D, wherein the Hamming distance at $T = 153$ is shown
20 to be zero.
21

22
23 If it is determined 656 that $T_0 \leq T_1$, the method continues at step 658 by storing 658
24 an indication of the current referent image 146 in the candidate set 147 and removing any
25 candidates 147 with distances greater than zero. In a preferred embodiment, if an image 146
26 perfectly matches the binary matrix 145, any candidates 147 having non-zero distances from
27 the binary matrix 145 should be discarded since they are less likely to be the correct
28

1 interpretation. As will be explained below, candidates 147 with non-zero distances (i.e. $T_0 >$
2 T_1) are provisionally stored with their associated T_0 and T_1 values. Therefore, they may be
3 easily identified and discarded.

4 If, however, it is determined in step 656 that $T_0 > T_1$, the method continues by
5 determining whether a candidate 147 was already found with a minimum Hamming distance
6 of zero. Preferably, if a perfect match was already found, there is no need to store a less-
7 than-perfect match. Therefore, if a candidate 147 has already been found with a zero
8 distance, the method continues at step 662; otherwise, the method continues by provisionally
9 storing 661 an indication of the current referent image with its associated values for T_0 and
10 T_1 .

11 In a preferred embodiment, the actual Hamming distance is preferably not calculated
12 at this point in order to reduce unnecessary calculations. It is enough to recognize that the
13 minimum Hamming distance for the candidate 147 will be non-zero for all thresholds. Later,
14 if a candidate 147 with a zero minimum distance is identified, all non-zero candidates 147 are
15 discarded, regardless of the distance.

16 An example of a case where $T_0 > T_1$ is seen in partitioned matrix 144b, wherein the
17 value for T_0 (199) exceeds the value of T_1 (198). As explained above, this means that the
18 minimum Hamming distance will be greater than zero for all values of T , and the minimum
19 will be found between T_0 and T_1 . This can be verified graphically with reference to Figure
20 7D. As noted above, the minimum distance does not need to be calculated at this point.
21 Instead, an indication of the image 146b is provisionally stored 661 along with its associated
22 values for T_0 and T_1 , i.e., 199 and 198, respectively.
23
24
25
26
27
28

1 After either steps 658 or 661, the method continues by determining 662 whether more
2 referents 146 remain to be compared with the binary matrix 145. If more referents 146
3 remain, the method returns to step 650; otherwise, the method continues at step 664 by
4 determining whether at least one perfect match ($T_0 \leq T_1$) was found by step 656. If at least
5 one perfect match was found, the method 406 terminates because a highly-probable candidate
6 set 147 has been determined.
7

8 Referring now to Figure 6B, if no perfect match was found, candidate set 147
9 comprises a number of provisionally stored candidates with T_0 and T_1 values. Therefore, it is
10 necessary to find the minimum Hamming distances for each provisionally stored candidate
11 147. Thereafter, in a preferred embodiment, only those candidates 147 with acceptable
12 distances will be retained; all others will be discarded.
13

14 The method continues after step 664 by selecting 666 the next stored candidate 147.
15 Preferably, only those candidates 147 for which a distance has not been calculated are
16 selected. Next, the minimum distance is calculated 668 between the candidate 147 and the
17 gray matrix 144. Unlike the exhaustive method, however, not all possible binarizations need
18 to be tested. As noted earlier, the minimum Hamming distance may be found by
19 thresholding the gray matrix 144 at values between T_1 and T_0 . Accordingly, the gray matrix
20 144 is binarized at each integer value between T_1 and T_0 , and the distance from the candidate
21 147 is calculated. The lowest distance found is selected as the minimum distance.
22
23

24 Although this step requires some distance calculations, one skilled in the art will
25 recognize that far fewer calculations are required than in the exhaustive method. Indeed, if
26 image 146b of Figure 8B is the selected candidate, only two distance calculations are
27 required (at $T = 198$ and 199), as opposed to the 256 calculations required in the exhaustive
28

1 method. Moreover, it is possible to reduce the number of distance calculations even more by
2 using a vector quantization technique based on triangle inequalities as described by Allen
3 Gersho and Robert M. Gray in "Vector Quantization and Signal Compression", Kluwer
4 Academic Publications, Boston, 1992 (fifth printing 1996).

5
6 After the minimum distance is calculated 668 for the selected candidate 147, the
7 method continues by determining 670 whether the distance is within an acceptable range.
8 As explained above, the acceptable range is preferably determined experimentally in order to
9 select a small, highly-probable candidate set 147. If the distance is not acceptable, the
10 candidate 147 is discarded 674; otherwise, a determination 672 is made whether the
11 calculated minimum distance is less than or equal to those of the candidates 147 whose
12 distances were previously determined. In a preferred embodiment, if the calculated minimum
13 distance is greater than a previously determined distance, the selected candidate 147 is less
14 likely to be correct and is therefore discarded 674. However, if the calculated minimum
15 distance is less than or equal to that of a previously determined distance, the method
16 continues by storing 676 the distance in the candidate set 147.

17
18
19 After either steps 674 or 676, the method continues by determining whether any
20 candidates remain for which the minimum distance has not been determined. If more
21 candidates remain, the method returns to step 666; otherwise the method is complete.

22 Referring again to Figure 4, after the candidate set 147 is determined 406 using
23 classification-driven thresholding, the method continues by determining 408 whether the size
24 of the candidate set 147 is greater than one. Preferably, no more than one character is
25 selected as the recognized character. Therefore, if more than one candidate 147 was found in
26 step 406, the method continues by disambiguating 408 the candidate set 147.
27
28

1 An example of an ambiguous candidate set 147 is illustrated in Figure 1, wherein two
2 interpretations of the grayscale image 142 are possible depending on the threshold selected.
3 For example, if $T = 128$, then the image 142 is interpreted as the letter "U." However, if $T =$
4 140, then the image 142 is interpreted as the letter "L." Both interpretations are valid, and
5 each interpretation provides a Hamming distance of zero. A conventional image classifier
6 would miss one of the interpretations because, conventionally, the image 142 is binarized
7 with a single threshold prior to classification. However, in the present invention, a candidate
8 set 147 would be determined 406 comprising both interpretations. Nevertheless, only one
9 interpretation is correct if the human eye can recognize the character as either "U" or "L."
10

11 As noted earlier, the original source image 142 contains more data about the scanned
12 object than the normalized grayscale matrix 144. Indeed, as illustrated in Figure 5, 16 pixels
13 of the source image 142 are used to create one element of the gray matrix 144. Therefore, the
14 additional information in image 142 may assist in disambiguating the candidate set 147 and
15 selecting the preferred character.
16

17 The process of disambiguation is similar to method described in Figures 6A-B.
18 However, as explained in greater detail below, each of the referent images 146 of the
19 candidates 147 are first spatially normalized (enlarged) to the size of the source image 142,
20 rather than reducing the source image 142 to the size of the referent images 146 as was done
21 previously. Thereafter, the classification-driven thresholding method of Figures 6A-B is
22 repeated to calculate a new candidate set 147.
23

24 Because of the additional image data available in the sample image 142, it is expected
25 that repeating the classification-driven thresholding technique will eliminate some of the
26 original candidates 147. One skilled in the art, however, will recognize that this process can
27
28

1 result in a large number of calculations, given the relatively larger size of the source image
2 142. Nevertheless, in a preferred embodiment, the candidate set 147 will likely be small,
3 such that only a few interpretations need to be reevaluated.
4

5 Referring now to Figure 6C, there is shown a preferred method for disambiguating
6 410 the candidate set 147 and selecting a preferred candidate. The method begins by
7 spatially normalizing 682 the referent images 146 of the selected candidates 147 to the size of
8 the source image 142. Techniques for spatial normalization of binary images are well known
9 in the art. For example, if the referent image 146 with dimensions of 16 x 16 is spatially
10 normalized (enlarged) to the size of the sample image 142 with dimensions of 64 x 64, one
11 skilled in the art will recognize that each pixel of the referent image 146 must be reproduced
12 16 times in the enlarged image. When complete, a set of normalized referent images 146 is
13 created.
14

15 In a preferred embodiment, the method continues by determining 406 a new candidate
16 set 147 using substantially the same method described in Figures 6A-B. However, this time,
17 the gray matrix 144 is, in fact, the sample image 142, and the referent images 146 are the set
18 of normalized referent images 146 produced in step 682. Preferably, one modification that
19 should be made to the method described is to increase the acceptable distance in step 670 to
20 take into account the relatively larger size of the new gray matrix 144 (sample image 142).
21

22 When step 406 is complete, a determination 684 is made whether only one candidate
23 147 remains. If so, the remaining candidate 147 is recognized as the preferred candidate
24 147. However, if more than one candidate 147 is found, in one embodiment, all of the
25 candidates 147 are rejected. However, in an preferred embodiment, the ambiguity is
26 resolved 688 using either contextual or geometric analysis.
27
28

1 Briefly, contextual analysis is the process of resolving ambiguities by referring to the
2 letters or words surrounding the character in question, possibly with reference to a dictionary
3 or the like. For example, in deciding between the letters "U" and "L," the fact that the first
4 seven letters in the word were with "GRAYSCA" would suggest that the correct
5 interpretation is "L" (in order to spell the word "GRAYSCALE")
6

7 Similarly, geometric analysis is the process of eliminating impossible candidate
8 choices by examining the geometric properties of the scanned character. For example, in
9 deciding between the letters "W" and "I," by referring back to the size of the original sample
10 image 142, the letter "I" could be eliminated if the image 142 was relatively wide. Both
11 techniques are well known to one skilled in the art of optical character recognition. For
12 example, U.S. Patent No. 5,539,840 to Krtolica et al. for Multifont Optical Character
13 Recognition Using a Box Connectivity Approach, discloses a process of eliminating
14 characters based on geometric properties of the minimal bounding frame. After either step
15 686 or step 688, the method terminates.
16
17

18 Referring again to Figure 4, after either steps 408 or 410, the preferred candidate 147
19 is output, for example, by means of the display device 128 or the communications device
20 130. Alternatively, the preferred candidate 147 may be stored in the storage device 124. If
21 no candidate 147 was identified, or if the ambiguity was not resolved, the system 120
22 preferably outputs an error message.
23

24 The above description is included to illustrate the operation of the preferred
25 embodiments and is not meant to limit the scope of the invention. The scope of the invention
26 is to be limited only by the following claims. From the above discussion, many variations
27
28

1 will be apparent to one skilled in the art that would yet be encompassed by the spirit and
2 scope of the present invention.

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

Claims

1. In an image recognition system, a method for recognizing a sample image comprising a plurality of pixels from a set of binary referent images, each pixel of the sample image comprising a multi-bit value, the method comprising:
normalizing the size of the sample image to the size of the referent images;
and
determining a set of candidate images from the set of referent images,
wherein each of the candidate images is within a predesignated distance from plural different binarizations of the normalized sample image.
2. The method of claim 1, further comprising: responsive to the set of candidate images comprising more than one image, disambiguating the candidate set by selecting a preferred candidate image for recognition.
3. The method of claim 2, wherein the disambiguation step comprises:
spatially normalizing the candidate images to the size of an unnormalized sample image; and
determining a second set of candidate images from the normalized candidate images, wherein each of the second set of candidate images is within a predesignated distance from plural different binarizations of the unnormalized sample image.
4. The method of claim 3, further comprising:
responsive to the second set of candidate images comprising more than one image, selecting a preferred candidate image by a method selected from the group of methods comprising geometric analysis and contextual analysis.
5. The method of claim 1, wherein the determining step comprises:
binarizing the sample image at a plurality of different thresholds to obtain plural binarized sample images;
calculating a set of distances between the plural binarized sample images

and the set of referent images;

selecting as a candidate a first referent image that is within a predesignated distance from a first one of the plural binarized sample images; and

selecting as a candidate a second referent image that is within a predesignated distance from a second one of the plural binarized sample images, the first binarization comprising a different threshold than the second binarization.

6. The method of claim 5, wherein the distance is a Hamming distance.

7. The method of claim 5, wherein the image to be recognized is a character image.

8. The method of claim 1, wherein, for each referent image, the determining step comprises:

partitioning the pixels of the sample image into a foreground portion and a background portion corresponding to the foreground and background of the referent image;

determining a first value corresponding to a maximum pixel value in the background portion;

determining a second value corresponding to a minimum pixel value in the foreground portion minus one;

responsive to the first value being no greater than the second value, selecting the referent image as a candidate; and

responsive to the first value being greater than the second value:

calculating a minimum distance between the sample image and the referent image; and

responsive to the minimum distance being within a predesignated range, selecting the referent image as a candidate.

9. The method of claim 8, wherein the referent image is not selected as a candidate when a previously selected candidate has an associated minimum distance less than the minimum distance of the referent image.

10. The method of claim 8, further comprising:
responsive to the minimum distance of the referent image being zero,
discarding any previously selected candidates having associated minimum distances
greater than zero.
11. The method of claim 1, wherein the normalization step comprises:
mapping the sample image onto a grid of boxes;
summing the pixel values contained within each box; and
storing the sum of the pixel values within a corresponding location of a
normalized sample image.
12. The method of claim 11, wherein the value of a pixel bisected by the grid is
proportionately divided between the boxes containing the pixel, and the proportional
values are calculated using rational number arithmetic implemented with integer
operations.
13. A system for recognizing a sample image comprising a plurality of pixels from
a set of binary referent images, each pixel of the sample image comprising a multi-
bit value, the system comprising:
a normalization unit for normalizing the size of the sample image to the size
of the referent images;
a distance calculation unit for calculating a set of distances between the
normalized sample image, binarized at a plurality of thresholds, and the set of
referent images; and
a classification unit for determining a set of candidate images from the set of
referent images, wherein each candidate image is within a predesignated distance
from a different binarization of the normalized sample image.
14. The system of claim 13, further comprising:
a disambiguation unit for disambiguating the candidate set and selecting a
preferred candidate image for recognition by comparing the candidate images with

an unnormalized sample image.

15. The system of claim 13, wherein the image to be recognized is a character image.

16. The system of claim 13, wherein the distance is a Hamming distance.

17. A computer-readable medium having computer-readable code executable to recognize a sample image comprising a plurality of pixels from a set of binary referent images, each pixel of the sample image comprising a multi-bit value, the computer-readable code comprising:

computer-readable normalization code to normalize the size of the sample image to the size of the referent images; and

computer-readable determination code to determine a set of candidate images from the set of referent images, wherein each of the candidate images is within a designated distance from plural different binarizations of the normalized sample image.

18. The computer-readable medium of claim 17, wherein the computer-readable determination code comprises:

computer-readable code to partition the pixels of the sample image into a foreground portion and a background portion corresponding to the foreground and background of the referent image;

computer-readable code to determine a first value corresponding to the maximum pixel value in the background portion;

computer-readable code to determine a second value corresponding to the minimum pixel value in the foreground portion minus one;

computer-readable code responsive to the first value being no greater than the second value, to select the referent image as a candidate; and

computer-readable code responsive to the first value being greater than the second value:

(a) to calculate a minimum distance between the sample image and

the referent image; and

(b) responsive to the minimum distance being within a predesignated range, select the referent image as a candidate.

19. The computer-readable medium of claim 17, further comprising:
computer-readable disambiguation code to disambiguate the candidate set by selecting a preferred candidate image for recognition.

20. The computer-readable medium of claim 19, wherein the computer-readable disambiguation code comprises:

computer-readable code to spatially normalize the candidate images to the size of an unnormalized sample image; and

computer-readable code to determine a second set of candidate images from the normalized candidate images, wherein each of the second set of candidate images is within a predesignated distance from a different binarization of the unnormalized sample image.

GRAYSCALE IMAGE

145	75	36	128
168	23	2	135
195	48	14	138
154	173	163	182

(a)

BINARY IMAGE
(T=128)

1	0	0	1
1	0	0	1
1	0	0	1
1	1	1	1

(b)

RECOGNIZED
AS "U"

BINARY IMAGE
(T=140)

1	0	0	0
1	0	0	0
1	0	0	0
1	1	1	1

(c)

RECOGNIZED
AS "L"

Fig. 1

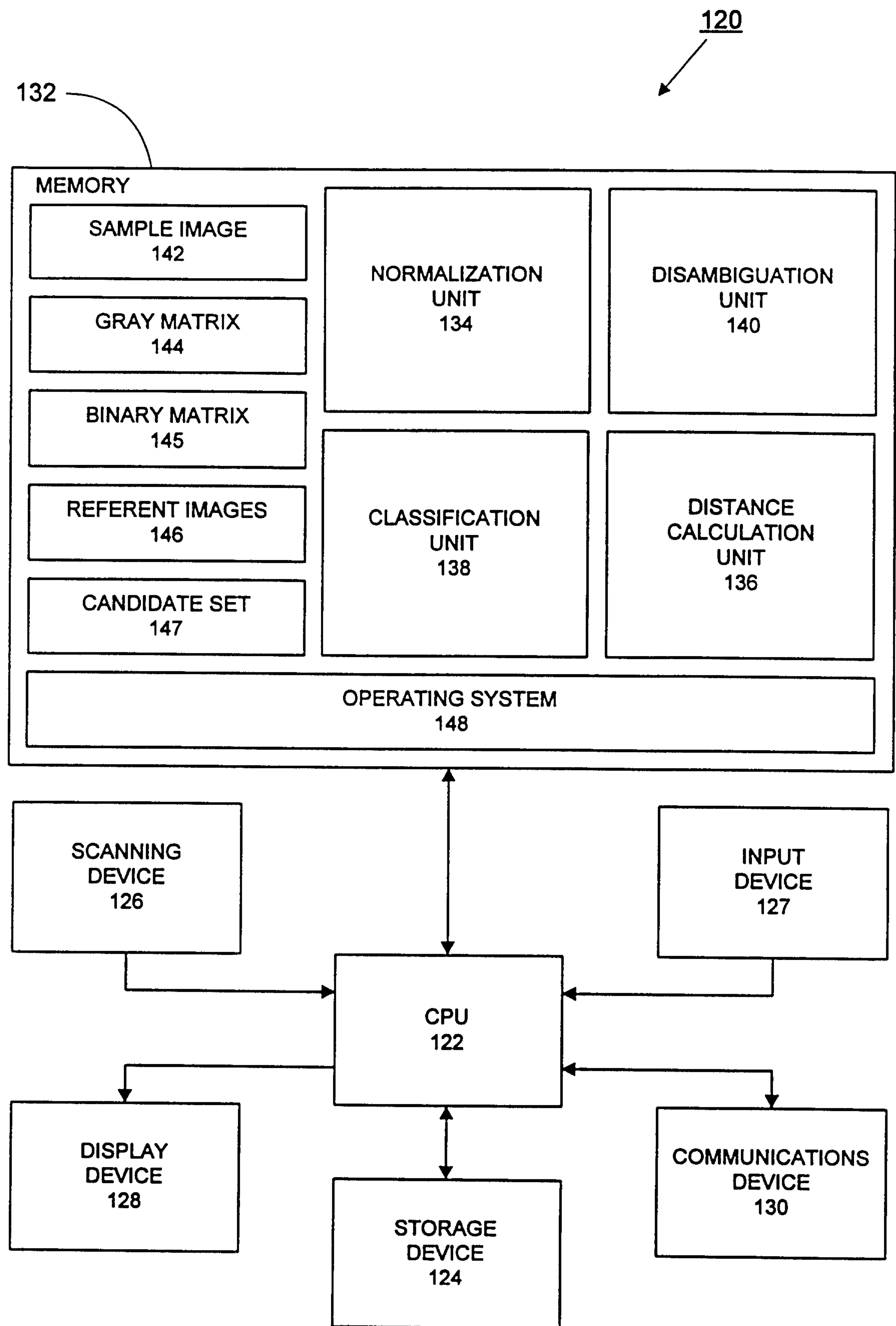


FIG. 2

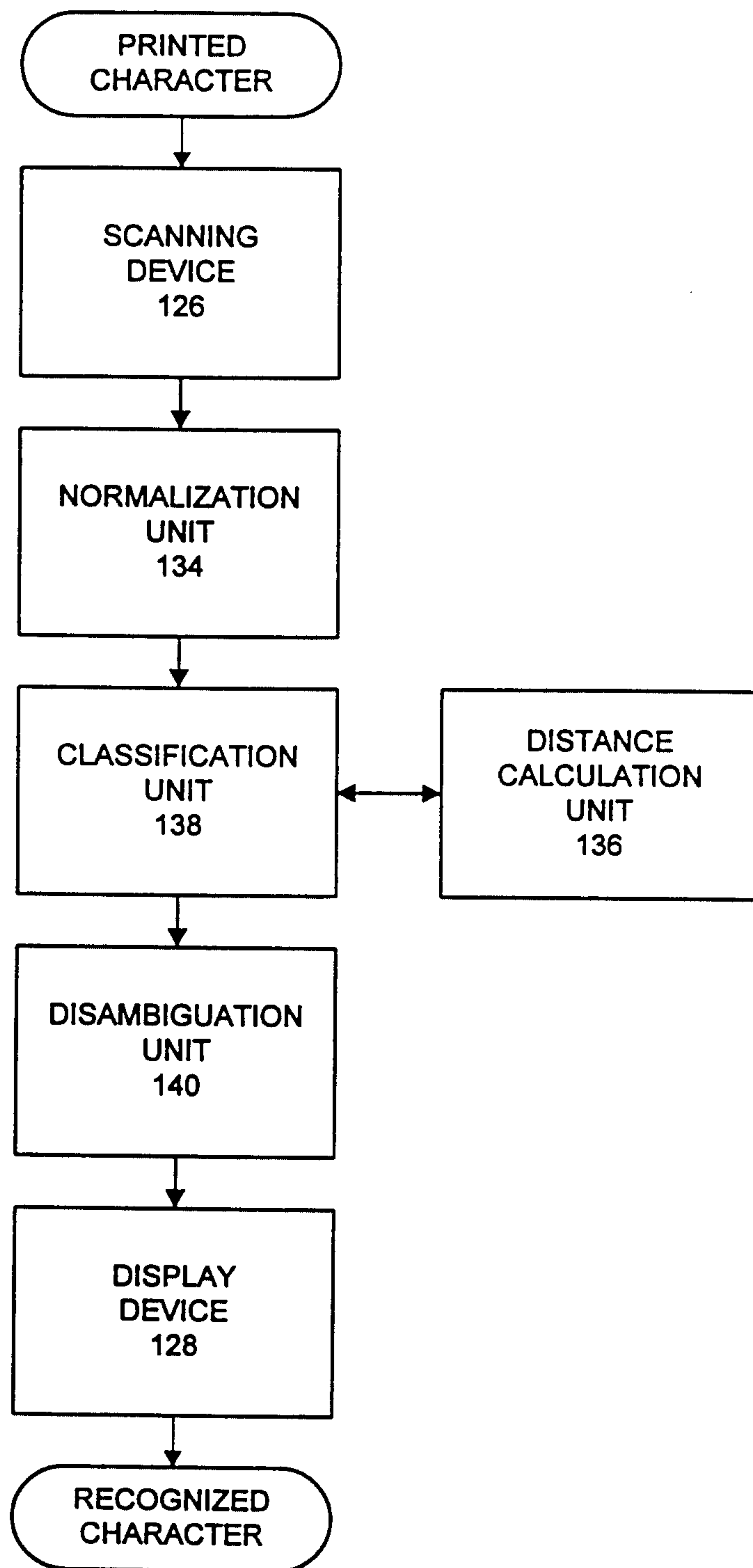


Fig. 3

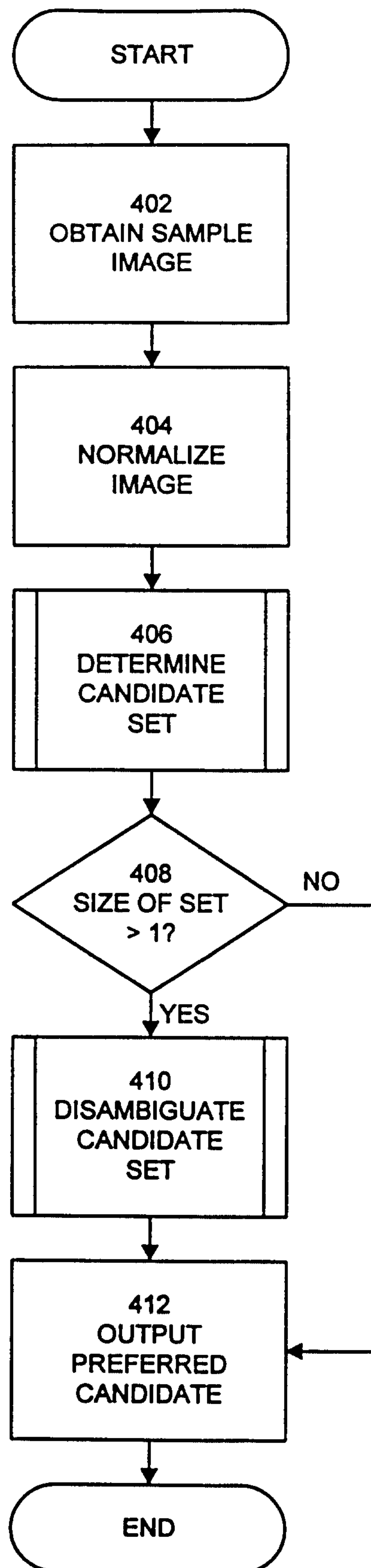


Fig. 4

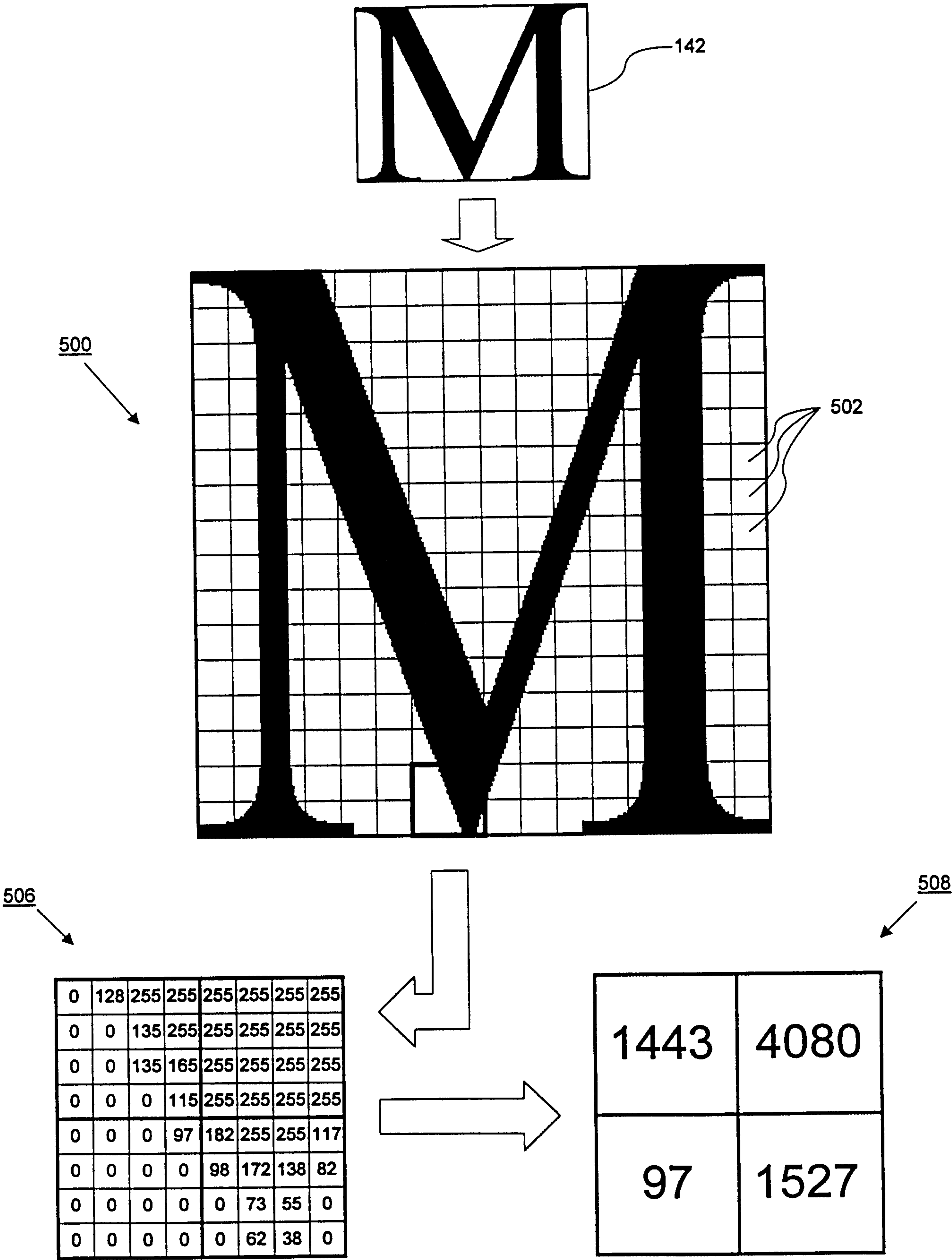


Fig. 5

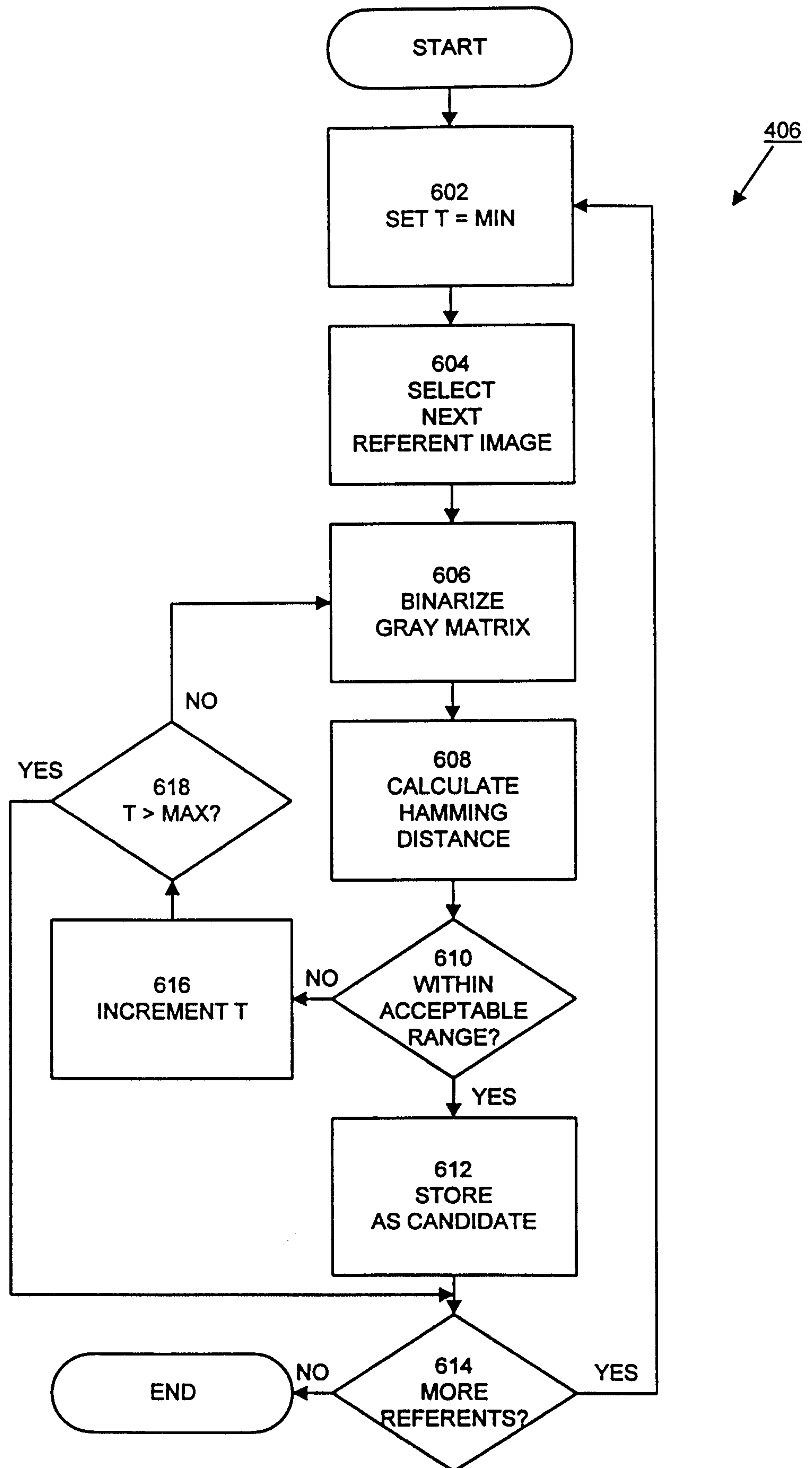


Fig. 6

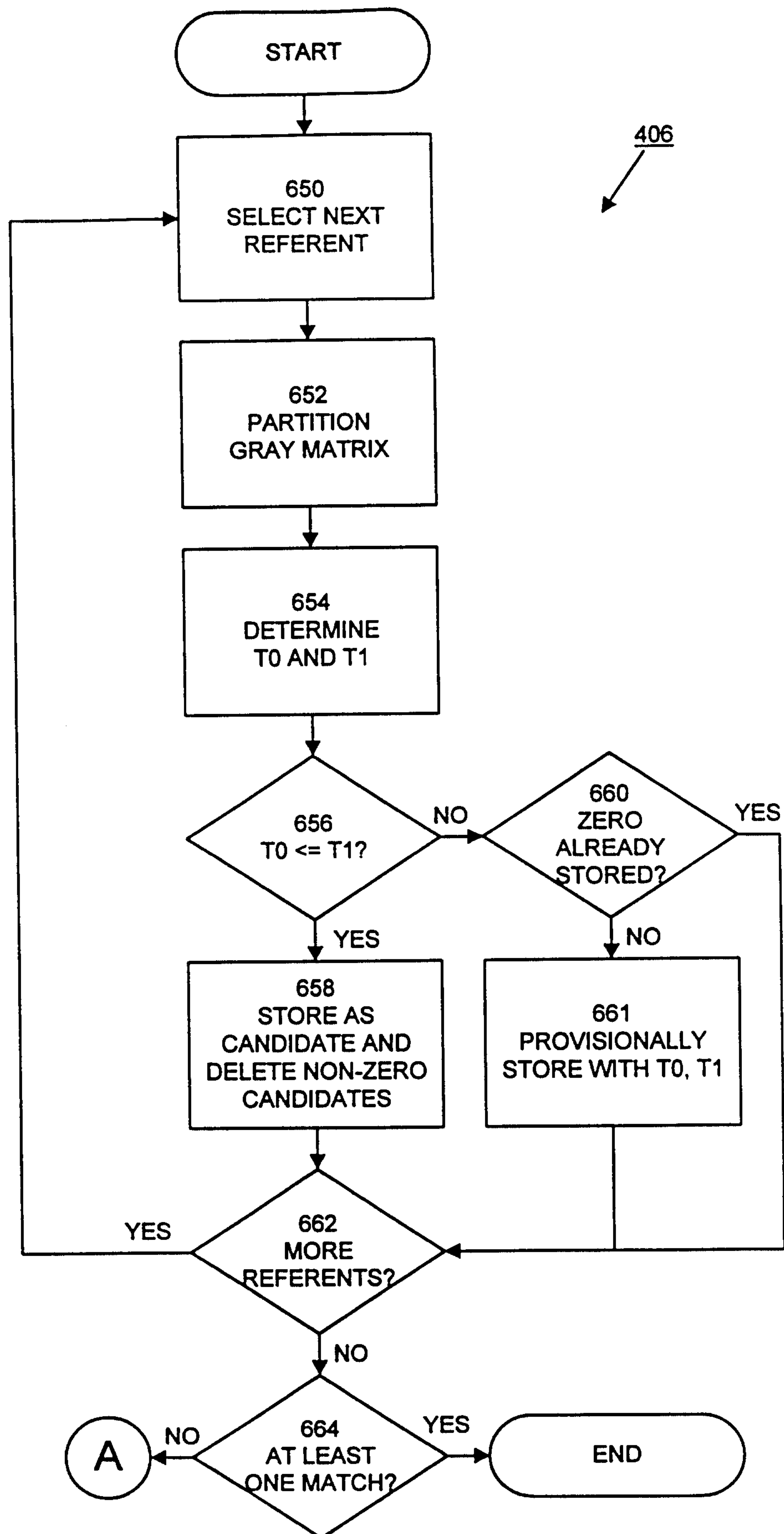


Fig. 6A

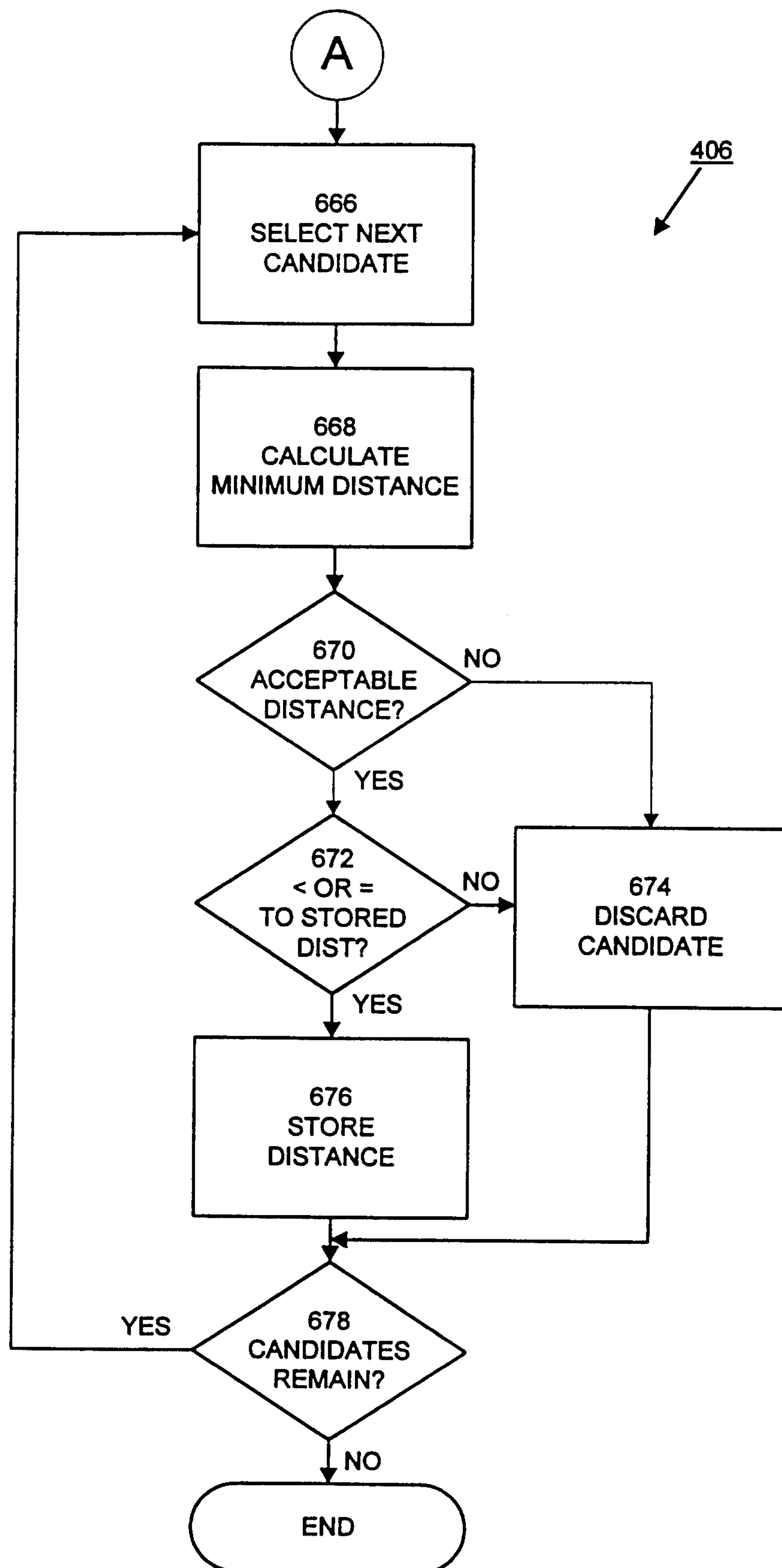


Fig. 6B

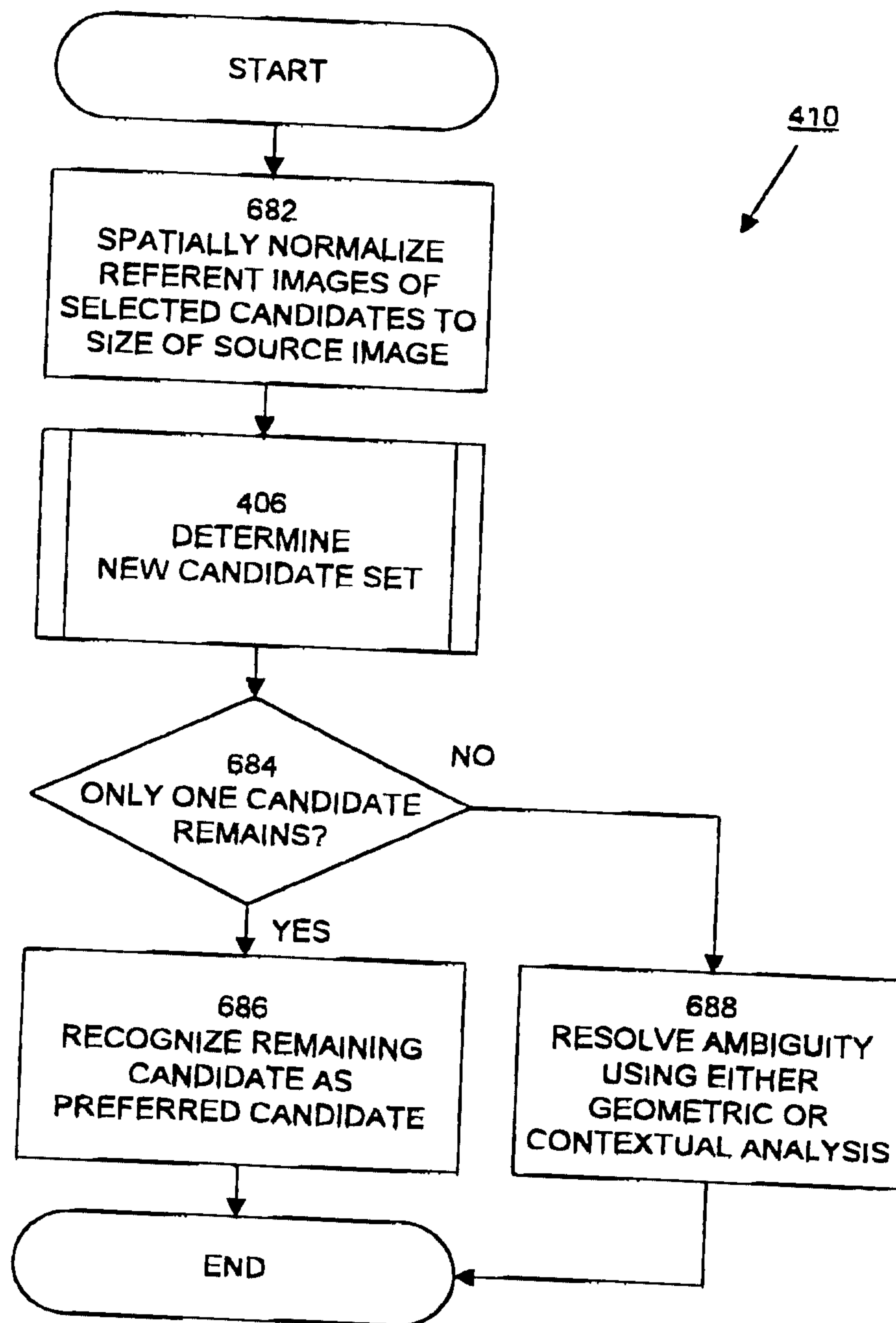


Fig. 6C

144

235	120	127	186
247	130	153	166
229	110	150	184
207	112	143	199
212	201	204	199

Fig. 7A

146a

1	0	0	1
1	0	0	1
1	0	0	1
1	0	0	1
1	1	1	1

Fig. 7B

146b

1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	1	1	1

Fig. 7C

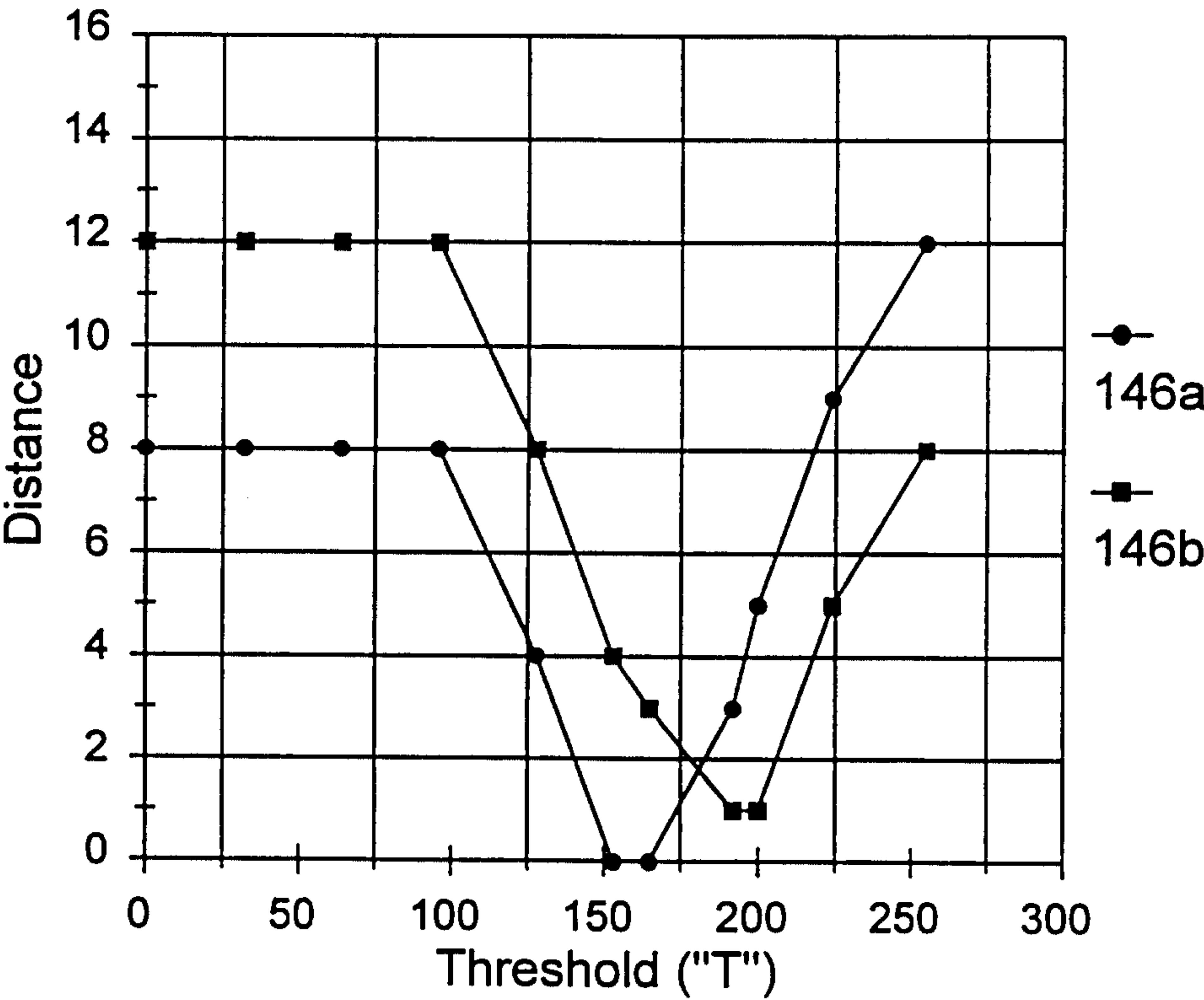


Fig. 7D

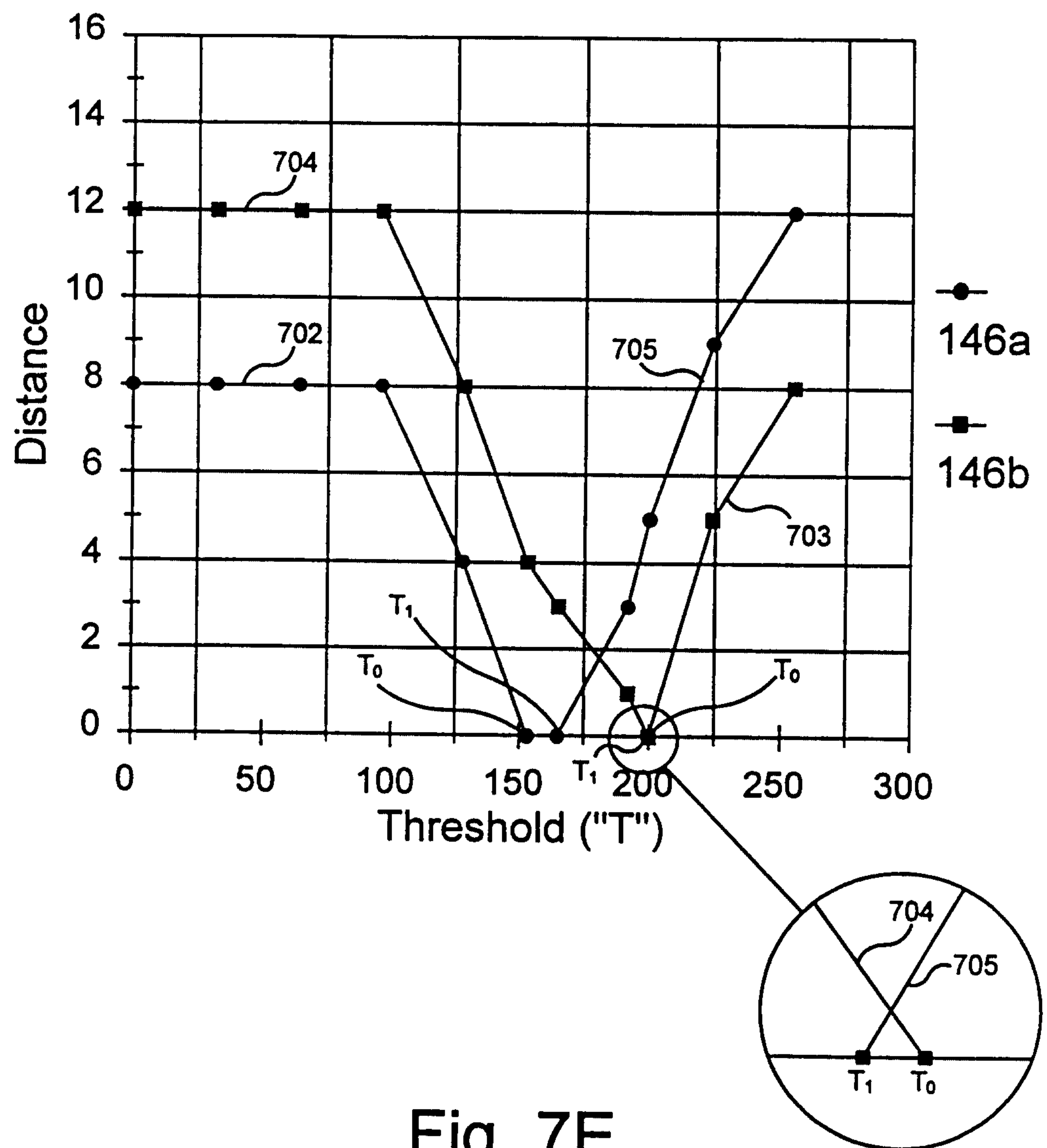


Fig. 7E

146a

1	0	0	1
1	0	0	1
1	0	0	1
1	0	0	1
1	1	1	1

Fig. 8A

146b

1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	1	1	1

Fig. 8B

710 712 144a

235	120	127	186
247	130	153	166
229	110	150	184
207	112	143	199
212	201	204	199

Fig. 8C

710 712 144b

235	120	127	186
247	130	153	166
229	110	150	184
207	112	143	199
212	201	204	199

Fig. 8D

