



(51) International Patent Classification:

G06Q 30/06 (2012.01) G06K 7/12 (2006.01)

(21) International Application Number:

PCT/CN2017/083143

(22) International Filing Date:

05 May 2017 (05.05.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant: SYMBOL TECHNOLOGIES, LLC  
[US/US]; 1 Zebra Plaza, Holtsville, NY 11742 (US).

(72) Inventors: ZHANG, Yan; 520 Newtown Drive, Buffalo Grove, IL 60089 (US). O'CONNELL, Kevin, J.; 724 E. Mill Valley Road, Palatine, IL 90074 (US). FAN, Zhi-Gang; 4003 Chuanshou Road, Pudong District, Shanghai (CN). ZHAO, Mingxi; 402 Room, 59#, #237 Bohua Road, Pudong New District, Shanghai 201204 (CN).

(74) Agent: SHANGHAI PATENT & TRADEMARK LAW OFFICE, LLC; 435 Guiping Road, Shanghai 200233 (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: METHOD AND APPARATUS FOR DETECTING AND INTERPRETING PRICE LABEL TEXT

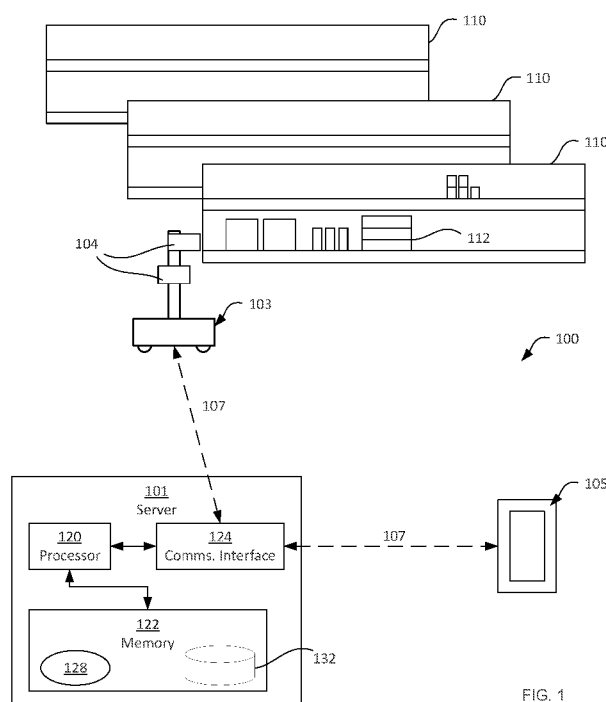


FIG. 1

(57) Abstract: A method of price text detection by an imaging controller comprises obtaining, by the imaging controller, an image of a shelf supporting labels bearing price text, generating, by the imaging controller, a plurality of text regions containing candidate text elements from the image, assigning, by the imaging controller, a classification to each of the text regions, selected from a price text classification and a non-price text classification. The imaging controller, within each of a subset of the text regions having the price text classification: detects a price text sub-region and generates a price text string by applying character recognition to the price text sub-region. The method further includes presenting, by the imaging controller, the locations of the subset of text regions, in association with the corresponding price text strings.

METHOD AND APPARATUS FOR DETECTING  
AND INTERPRETING PRICE LABEL TEXT

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is related to U.S. Patent Application Nos. 15/583,801, entitled “METHOD AND APPARATUS FOR EXTRACTING AND PROCESSING PRICE TEXT FROM AN IMAGE SET” by Zhang et al. and 15/583,786, entitled “METHOD AND APPARATUS FOR LABEL DETECTION” by Lam, as well as U.S. Provisional Patent Application No. 62/492,670, entitled “PRODUCT STATUS DETECTION SYSTEM” by Perrella et al., all having the filing date of May 1, 2017. The contents of the above-reference applications are incorporated herein by reference in their entirety.

BACKGROUND

[0002] Environments in which inventories of objects are managed, such as products for purchase in a retail environment, may be complex and fluid. For example, a given environment may contain a wide variety of objects with different attributes (size, shape, price and the like). Further, the placement and quantity of the objects in the environment may change frequently. Still further, imaging conditions such as lighting may be variable both over time and at different locations in the environment. These factors may reduce the accuracy with which information concerning the objects may be collected within the environment.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0003] The accompanying figures, where like reference numerals refer to identical or functionally similar elements throughout the separate views, together with the detailed description below, are incorporated in and form part of the specification, and serve to further illustrate embodiments of concepts that include the claimed invention, and explain various principles and advantages of those embodiments.

[0004] FIG. 1 is a schematic of a mobile automation system.

[0005] FIG. 2 is a block diagram of certain internal hardware components of the server in the system of FIG. 1.

[0006] FIG. 3 is a flowchart of a method of price text detection and interpretation.

[0007] FIG. 4 is a shelf image employed as input to the method of FIG. 3.

[0008] FIG. 5A illustrates example label formats processed in the method of FIG. 3.

[0009] FIG. 5B depicts a portion of the image of FIG. 4 during the performance of block 310 of the method of FIG. 3.

[0010] FIG. 6 depicts the image portion of FIG. 5B following the performance of block 310 of the method of FIG. 3.

[0011] FIGS. 7A-7B illustrate the generation of feature descriptors during the performance of the method of FIG. 3.

[0012] FIGS. 8-10A depict the processing of a text region during the performance of block 325 of the method of FIG. 3.

[0013] FIG. 10B is a price text sub-region detected during the performance of the method of FIG. 3.

[0014] Skilled artisans will appreciate that elements in the figures are illustrated for simplicity and clarity and have not necessarily been drawn to scale. For example, the dimensions of some of the elements in the figures may be exaggerated relative to other elements to help to improve understanding of embodiments of the present invention.

[0015] The apparatus and method components have been represented where appropriate by conventional symbols in the drawings, showing only those specific details that are pertinent to understanding the embodiments of the present invention so as not to obscure the disclosure with details that will be readily apparent to those of ordinary skill in the art having the benefit of the description herein.

#### DETAILED DESCRIPTION

[0016] Environments such as warehouses, retail locations (e.g. grocery stores) and the like typically contain a wide variety of products supported on shelves, for selection and purchase by customers. The products are generally labelled – for example, via a

label placed on a shelf edge, or directly on the product itself – with information such as the price of the product, an identifier such as a SKU number, and the like. Such environments typically also store reference data relating to the products, for example in a central database, which is consulted by point-of-sale terminals during customer checkout to retrieve price information for the products being purchased. In some cases, the price physically labelled on or near the product on the shelves may not match the price stored in the above-mentioned database, leading to a conflict between the label price and the reference price at the point-of-sale terminal.

**[0017]** Mismatches between label and reference prices may require corrective action at one or both of the label and the central database. However, detecting such mismatches in order to allow corrective action to be taken is conventionally performed by human employees, via visual assessment of the shelves and manual barcode scanning. This form of detection is labor-intensive and therefore costly, as well as error-prone.

**[0018]** Attempts to automate the detection of such mismatches require the detection of the labelled price before an assessment may be conducted as to whether the label price matches the reference price. Various factors impede the accurate autonomous detection and interpretation of labelled prices, however. For example, in a retail environment in which a wide variety of products are arranged on shelves, many of the products themselves bear text (such as brand names, product names, ingredient lists and so on) that does not represent the price of the product. Further, the characters or text that do indicate the price of a product may be displayed in close proximity to other text on a label, rendering the detection and correct interpretation of the price by machine vision techniques difficult.

**[0019]** Examples disclosed herein are directed to a method of price text detection, comprising: obtaining an image of a shelf supporting labels bearing price text; generating a plurality of text regions containing candidate text elements from the image; assigning a classification to each of the text regions, selected from a price text classification and a non-price text classification; within each of a subset of the text regions having the price text classification: detecting a price text sub-region; and generating a price text string by applying character recognition to the price text sub-

region; and presenting the locations of the subset of text regions, in association with the corresponding price text strings.

**[0020]** FIG. 1 depicts a mobile automation system 100 in accordance with the teachings of this disclosure. The system 100 includes a server 101 in communication with at least one mobile automation apparatus 103 (also referred to herein simply as the apparatus 103) and at least one mobile device 105 via communication links 107, illustrated in the present example as including wireless links. The system 100 is deployed, in the illustrated example, in a retail environment including a plurality of shelf modules 110 each supporting a plurality of products 112. The shelf modules 110 are typically arranged in a plurality of aisles, each of which includes a plurality of modules aligned end-to-end. More specifically, the apparatus 103 is deployed within the retail environment, and communicates with the server 101 (via the link 107) to at least partially autonomously navigate the length of at least a portion of the shelves 110. The apparatus 103 is equipped with a plurality of navigation and data capture sensors 104, such as image sensors (e.g. one or more digital cameras) and depth sensors (e.g. one or more lidar sensors), and is further configured to employ the sensors to capture shelf data. In the present example, the apparatus 103 is configured to capture a series of digital images of the shelves 110, as well as a series of depth measurements, each describing the distance and direction between the apparatus 103 and one or more points on a shelf 110.

**[0021]** The server 101 includes a special purpose imaging controller, such as a processor 120, specifically designed to control the mobile automation apparatus 103 to capture data, obtain the captured data via the communications interface 124 and store the captured data in a repository 132 in the memory 122. The server 101 is further configured to perform various post-processing operations on the captured data and to detect the status of the products 112 on the shelves 110. When certain status indicators are detected by the imaging processor 120, the server 101 is also configured to transmit status notifications (e.g. notifications indicating that products are out-of-stock, low stock or misplaced) to the mobile device 105. The processor 120 is interconnected with a non-transitory computer readable storage medium, such as a memory 122, having stored thereon computer readable instructions for executing price text detection and interpretation, as discussed in further detail below. The

memory 122 includes a combination of volatile (e.g. Random Access Memory or RAM) and non-volatile memory (e.g. read only memory or ROM, Electrically Erasable Programmable Read Only Memory or EEPROM, flash memory). The processor 120 and the memory 122 each comprise one or more integrated circuits. In an embodiment, the processor 120, further includes one or more central processing units (CPUs) and/or graphics processing units (GPUs). In an embodiment, a specially designed integrated circuit, such as a Field Programmable Gate Array (FPGA), is designed to perform the price text detection and interpretation discussed herein, either alternatively or in addition to the imaging controller/processor 120 and memory 122. As those of skill in the art will realize, the mobile automation apparatus 103 also includes one or more controllers or processors and/or FPGAs, in communication with the controller 120, specifically configured to control navigational and/or data capture aspects of the apparatus 103.

**[0022]** The server 101 also includes a communications interface 124 interconnected with the processor 120. The communications interface 124 includes suitable hardware (e.g. transmitters, receivers, network interface controllers and the like) allowing the server 101 to communicate with other computing devices – particularly the apparatus 103 and the mobile device 105 – via the links 107. The links 107 may be direct links, or links that traverse one or more networks, including both local and wide-area networks. The specific components of the communications interface 124 are selected based on the type of network or other links that the server 101 is required to communicate over. In the present example, a wireless local-area network is implemented within the retail environment via the deployment of one or more wireless access points. The links 107 therefore include both wireless links between the apparatus 103 and the mobile device 105 and the above-mentioned access points, and a wired link (e.g. an Ethernet-based link) between the server 101 and the access point.

**[0023]** The memory 122 stores a plurality of applications, each including a plurality of computer readable instructions executable by the processor 120. The execution of the above-mentioned instructions by the processor 120 configures the server 101 to perform various actions discussed herein. The applications stored in the memory 122 include a control application 128, which may also be implemented as a suite of

logically distinct applications. In general, via execution of the control application 128 or subcomponents thereof, the processor 120 is configured to implement various functionality. The processor 120, as configured via the execution of the control application 128, is also referred to herein as the controller 120. As will now be apparent, some or all of the functionality implemented by the controller 120 described below may also be performed by preconfigured hardware elements (e.g. one or more ASICs) rather than by execution of the control application 128 by the processor 120.

**[0024]** In the present example, in particular, the server 101 is configured via the execution of the control application 128 by the processor 120, to process image and depth data captured by the apparatus 103 to identify portions of the captured data depicting price labels, and to detect and interpret the text indicative of product prices on such labels.

**[0025]** Turning now to FIG. 2, before describing the operation of the application 128 to identify and interpret price text from captured image data, certain components of the application 128 will be described in greater detail. As will be apparent to those skilled in the art, in other examples the components of the application 128 may be separated into distinct applications, or combined into other sets of components. Some or all of the components illustrated in FIG. 2 may also be implemented as dedicated hardware components, such as one or more Application-Specific Integrated Circuits (ASICs) or FPGAs. For example, in one embodiment, to improve reliability and processing speed, at least some of the components of FIG. 2 are programmed directly into the imaging controller 120, which may be an FPGA or an ASIC having circuit and memory configuration specifically designed to optimize image processing and detection of high volume price label image data being received from the mobile automation apparatus 103. In such an embodiment, some or all of the control application 128, discussed below, is an FPGA or an ASIC chip.

**[0026]** The control application 128, in brief, includes components configured to identify text within a shelf image that is likely to represent a label price, and to interpret (i.e. recognize the characters in, or “read”) that text. The control application 128 includes a text region generator 200 (also referred to herein simply as a region generator 200) and a classifier 204 that are configured to detect text regions in the

image that are likely to correspond to price text labels (since the image may contain a multitude of text regions that are not prices). The control application 128 also includes a sub-region detector 208 configured, for the regions that are considered likely to be price text labels as identified by the region generator 200 and the classifier 204, to detect price text sub-regions within the text regions. The price text sub-regions are specific areas identified as containing price text characters rather than other characters also printed on a price label. The control application 128 also includes an interpreter 212 configured to interpret the text within the above-mentioned sub-regions, in order to generate a machine-readable representation of the price text printed on the physical label.

**[0027]** The functionality of the control application 128 will now be described in greater detail, with reference to the components illustrated in FIG. 2. Turning to FIG. 3, a method 300 of price text detection is shown. The method 300 will be described in conjunction with its performance on the system 100 as described above; however, it will be apparent to those skilled in the art that the method 300 may also be performed on other systems.

**[0028]** The performance of the method 300 begins at block 305, at which the controller 120 is configured to obtain a digital image of the shelf 110, for example captured by the apparatus 103 and stored in the repository 132. An example image 400 is illustrated in FIG. 4, depicting a portion of a shelf 110. In particular, the image 400 depicts a shelf edge 404 and a shelf back 408, as well as a support surface 412 extending the between the shelf edge 404 and the shelf back 408 and supporting products 112-1 and 112-2. As seen in FIG. 4, the products 112 bear text such as the name or brand of a product, a weight, a calorie count, and the like. The image also depicts labels 416-1 and 416-2 (corresponding to the products 112-1 and 112-2, respectively), which in the present example are placed on the shelf edge 404, but may also be placed on the products 112 themselves or in their vicinity in other examples.

**[0029]** The labels 416 include price text, and may also carry various other text. Turning to FIG. 5A, the labels 416 are shown in isolation. As is evident from the example label formats of FIG. 5A, the labels 416 bear a variety of text in addition to a price, and arrange the price and other text in different positions. For example, the

labels 416-1 and 416-2 both include a price text element 500-1, 500-2. However, the price text 500-1 has a different font size for the price itself than for the currency symbol “\$”. Further, the price text 500-2 uses a larger font for a portion of the price (including the currency symbol) than the font size employed by the label 416-1, but a smaller font size for another portion of the price. Still further, both labels 416 include barcodes 504-1 and 504-2, but the locations of the barcodes 504 relative to the price text 500-1, 500-2 is different between the labels 416. In further examples, some or all labels may omit barcodes entirely. In addition, both labels 416 include additional non-price text elements, some of which are in close proximity to the price text 500, such as the string “ACM crch” above the price text 500-2. The examples in FIG. 5A are merely illustrative; as will be apparent to those skilled in the art, a wide variety of label formats exist and may therefore be depicted in any given shelf image. Variations among label formats include variations not only in text positioning and size as shown in FIG. 5A, but also in label shape, colors, and the like.

**[0030]** Returning to FIG. 3, at block 310 the region generator 200 is configured to generate a plurality of text regions containing candidate text elements from the image. In the present example, the region generator 200 is configured to apply a suitable blob detection operation to the image obtained at block 305. For example, the region generator 200 can apply a maximally stable extremal regions (MSER) operation on the image to identify elements in the image likely to be characters of text. As those of skill in the art will realize, a variety of blob detection operations are within the scope of the present disclosure.

**[0031]** Turning to FIG. 5B, a portion of the image 400 is shown with candidate text elements indicated in dashed lines, as identified by the region generator 200. As will be apparent, the exact nature of the candidate text elements identified by the region generator 200 depends on the specific blob detection operation performed and the configuration parameters employed for the blob detection operation. For example, while the strings “48” and “cal” are indicated as having been detected as distinct candidate text elements in FIG. 5B, in other examples the region generator 200 may identify the string “48 cal” as a single contiguous candidate text element. Also note that, in the illustrated example, the barcode of the label 416-2 is detected as a candidate text region.

**[0032]** The region generator 200 is further configured, in the present example, to refine the identification of text regions by grouping the candidate text elements. In particular, the region generator 200 is configured, for each pair of the candidate text elements, to determine whether a distance between the pair is below a predetermined distance threshold. The distance threshold may be preconfigured, for example as a number of pixels, a percentage of the width or height of the candidate text elements under consideration, or the like. When the distance between two candidate text elements is below the threshold, the region generator 200 is configured to generate a text region that encompasses both of the candidate text elements. When the distance between candidate text elements exceeds the threshold, the elements are not encompassed within a single text region. That is, the region generator 200 is configured to generate text regions specific to each of the candidate elements.

**[0033]** In the present example, the region generator 200 is configured to generate the text region as a rectangular region with dimensions selected to encompass the entirety of both candidate text elements. The above process is repeated for each pair of candidate text elements, or where a candidate text element has already been incorporated into a text region, between the text region and another candidate text element (or even between two text regions).

**[0034]** The region generator 200 is also configured, in some examples, to apply a size difference threshold to respective pairs of candidate text elements, text regions, or both. In other words, for each pair of candidate text elements the region generator 200 is configured to determine a size attribute (e.g. a height, width, area or the like) for the elements, and to determine whether a difference between the size attributes exceeds a preconfigured threshold (e.g. a percentage of the above-mentioned size attributes). When the difference between size attributes does not exceed the threshold, the candidate elements are not combined in a single text region. As above, the thresholding process is repeated until no further change is dictated in the extent of the text regions (that is, until the comparisons between text regions no longer result in the combination or extension of any of the text regions to encompass additional candidate text elements).

**[0035]** The above-mentioned size and distance-based thresholds may also be combined by the region generator 200, in order to generate text regions encompassing candidate text elements that are sufficiently close to each other in both size and distance. In other examples, the region generator 200 applies only one or the other of the above-mentioned thresholds. In further examples, additional properties of the candidate text elements may also be considered, such as color. Referring to FIG. 6, the portion of the image 400 illustrated in FIG. 5B is shown again, following the application of size and distance-based thresholds. In particular, the candidate text elements shown in FIG. 5B are encompassed within two text regions 600-1 and 600-2.

**[0036]** Referring again to FIG. 3, at block 315 the classifier 204 is configured to assign a classification to each of the text regions 600, selected from a price text classification and a non-price text classification. In general, the classifier 204 is configured to distinguish between text regions likely to contain price text, and text regions that, although containing text, are not likely to contain price text. In the present example, the classifier 204 is configured to assign one of the above-mentioned classifications to the text regions 600 by generating a feature descriptor corresponding to each text region 600, providing the feature descriptor as an input to a trained classification engine, and receiving a classification as an output from the trained classification engine.

**[0037]** In the present example, the classifier 204 is configured to generate a feature descriptor for each text region 600 in the form of a combined feature vector generated from a histogram of oriented gradients (HOG) and a local binary pattern (LBP). In other examples, the feature descriptor is based on either the HOG or the local binary pattern. In further examples, as will be apparent to those skilled in the art, other suitable feature descriptors or combinations of feature descriptors can be employed.

**[0038]** To generate the above-mentioned HOG descriptor for a text region 600, the classifier 204 extracts each text region 600 from the image 400, and divides the extracted text region 600 into cells 700-1, 700-2, 700-3, 700-4 and so on, shown in FIG. 7A. The cells 700 can have predetermined dimensions (e.g. 4 x 4, 8 x 8 pixels or other suitable dimensions), or dimensions that vary based on the dimensions of the text regions 600 (for example, each text region 600 can be divided into four cells 700

of equal size). For each pixel of each cell 700, the classifier 204 then generates a gradient vector 704 indicating the angle of the greatest change in intensity between the pixel and its neighbors, as well as the magnitude of the change in intensity. Having obtained the above-mentioned vectors 704, the classifier 204 is configured to build a histogram, with bins corresponding to ranges of angles (e.g. 8 bins each accounting for an unsigned range of 20 degrees). The magnitude of each vector is added to the bin encompassing the vector's angle; in some example implementations, vectors with angles near the boundary between two adjacent bins may have their magnitudes divided between those bins. The resulting histogram for each cell is employed to construct a  $1 \times N$  vector 708 (vectors 708-1, 708-2, 708-3, 708-4 are illustrated in FIG. 7A), where  $N$  is the number of histogram bins (8 in the present example, though other numbers of bins may also be employed), containing the magnitudes assigned to each of the bins. The classifier 204 is then configured to concatenate the feature vectors 708 of the cells 700 for each text region 600 into a single vector.

[0039] To generate the LBP descriptor for a text region 600, as illustrated in FIG. 7B, the classifier 204, as above, extracts each text region 600 from the image 400, and divides the extracted text region 600 into cells 700. For each pixel 712 (two pixels 712-1 and 712-2 are labelled for illustrative purposes in FIG. 7B) within each cell 700, the classifier 204 is then configured to traverse the eight neighbors of that pixel in a predetermined direction (typically clockwise or counterclockwise). For each neighboring pixel, the classifier 204 is configured to determine a difference 716 between the intensities of the central pixel and the neighboring pixel (the difference between the intensities of the pixels 712-1 and 712-2 is labelled in FIG. 7B). The classifier 204 is further configured to determine whether the above-mentioned difference 716 is greater or smaller than zero (i.e. whether the neighboring pixel 712 has a smaller or greater intensity than the central pixel 712). Binary values 720 are selected based on the above determination, and assembled into a feature vector 724 for the central pixel. Specifically, when the neighboring pixel 712 has a greater intensity than the central pixel, a "1" is appended to a feature vector 724 for the central pixel (otherwise, a "0" is appended to the feature vector 724). The result is, for each pixel, an eight-digit binary number, as seen in FIG. 7B. The classifier 204 is then

configured to generate a histogram for the cell 700 based on the set of above-mentioned eight-digit numbers, in which the bins correspond to the positions of neighboring pixels in the above-mentioned clockwise or counterclockwise traverses. The histogram indicates the frequency, within the cell, with which each of the eight neighboring positions has a greater intensity than the central pixel (e.g. the frequency of “1” values in the present example). The histogram is employed to construct a vector 728, and the vectors 728 for all cells are then concatenated to produce a feature descriptor for the text region 600.

**[0040]** The classifier 204 is further configured, in the present example, to combine the HOG and LBP descriptors, for example by concatenation. Following the generation of feature descriptors as discussed above, the classifier 204 is configured to assign a classification to each text region 600 based on the feature descriptors. The classification, as noted above, is one of price text classification and a non-price text classification, and may be assigned in a variety of ways.

**[0041]** In the present example, the classifier 204 – specifically a preconfigured (i.e. trained) classification engine of the classifier 204 – is configured to accept the above-mentioned feature descriptor as an input and to generate, as an output, a score for each text region 600. The score (for example, a percentage or a value between zero and one) indicates a level of confidence that the text region 600 represents a label containing price text. The classification engine can be a suitable classification engine, such as a neural network, support vector machine (SVM), or the like. The classification engine, as will be apparent to those skilled in the art, is trained prior to performance of the method 300. Training the classification engine is typically conducted by providing a plurality of ground truth examples (i.e. “correct” text regions that are known to contain price text) and a plurality of negative examples (i.e. “incorrect” text regions that are known not to contain price text). The classification engine is configured to construct model parameters allowing it to correctly identify price text-containing samples.

**[0042]** In some environments, as noted earlier, various label formats may be employed. The classification engine is trained, in such environments, to identify each label type separately. Therefore, in addition to a classification score, the output of the

classifier 204 can include a label type identifier indicating the label format that best matches the text region 600. In other embodiments, classification may be performed by template matching or another suitable mechanism, rather than by a trained classification engine as discussed above.

**[0043]** The outcome of the performance of block 315 is a subset of the text regions 600 to which a price text classification was assigned at block 315. The subset of price text-classified text regions (region 600-2, in the example illustrated in FIG. 6) may be provided by the classifier 204 as, for example, rectangular bounding boxes associated with the above-mentioned scores.

**[0044]** At block 320, the control application is configured to determine whether all the price text-classified text regions have been processed. When the determination is negative, the performance of the method 300 proceeds to block 325. At block 325, the sub-region detector 208 is configured, for each of the subset of text regions 600 classified as containing price text, to detect a price text sub-region within the text region.

**[0045]** Turning to FIG. 8, the sub-region detector 208 is configured, in some examples, to begin the performance of block 320 by binarizing the extracted text region. FIG. 8 illustrates an example text region 800 extracted from a shelf image, in which the contrast ratio between the various text elements of the text region 800 is low. The sub-region detector 208 is therefore configured to generate a binarized version 804 of the text region 800 by applying any suitable binarization operation (e.g. an adaptive binarization operation). In other examples, binarization is omitted. In further examples, the sub-region detector 208 is configured to determine a contrast ratio for the text region under consideration (that is, the ratio of the intensity of the brightest pixel in the text region to the intensity of the least bright pixel in the text region). The sub-region detector 208 is then configured to apply binarization only when the contrast ratio fails to exceed a predetermined threshold.

**[0046]** Within the binarized text region 804 generated, the sub-region detector 208 is configured to group candidate text elements within the binarized text region 804 by size. In FIG. 8, a plurality of candidate text elements 808-1, 808-2, 808-3, 808-4, 808-5, 808-6 are illustrated (for example, as identified earlier at block 310) in dashed lines.

In other embodiments, a distinct candidate text element operation may be performed at block 320, with parameters selected to increase the likelihood of individual characters of the price text being selected as distinct text elements.

**[0047]** As seen in FIG. 8, the candidate text elements 808 have different dimensions. The sub-region detector 208 is configured to group the text elements 808 based on their dimensions; in particular, in the present example the sub-region generator is configured to group the text elements 808 by height (that is, the substantially vertical dimension as illustrated in FIG. 8). That is, the sub-region generator is configured to compare the height of each pair of text elements 808, and when the difference between the heights is below a predefined threshold, the sub-region detector 208 is configured to group the text elements 808 together. FIG. 9 illustrates the completion of the grouping process for the binarized text region 804. In particular, the three text elements 808-3, 808-4 and 808-5 have been grouped into a text group 900. The group 900 is defined, in the present example, as a rectangular bounding box dimensioned to encompass each of the candidate text elements 808-3, 808-4 and 808-5. Of particular note, the candidate text element 808-2 is greater in height than the elements of the group 900, because the character “2” and the characters “png” in the element 808-2 have been identified as a single candidate text element. Therefore, the element 808-2 has not been grouped with the group 900. The remaining elements 808 have also not been grouped, and are therefore considered independent groups.

**[0048]** Having grouped the text elements, the sub-region detector 208 is configured to select a primary one of the groups. In general, the selection of a primary group aims to select the group most likely to contain solely price text characters. In the present example, the primary group is selected based on a comparison of the two-dimensional area of the sub-regions. Thus, the area of the group 900 is compared with the respective areas of the remaining text elements 808 shown in FIG. 9 (which, for the present analysis, are considered to be groups). The sub-region detector 208 is configured to select the group with the greatest area as the primary sub-region. In the example of FIG. 9, the group 900 has a greater area than the other sub-regions, and is therefore selected as the primary group.

**[0049]** Following the selection of a primary group, the sub-region detector 208 is configured to fit upper and lower boundary lines to the primary group. Turning to FIG. 10A, upper and lower boundary lines 1000 and 1004, respectively, are illustrated. The sub-region detector 208 is configured to fit the boundary lines 1000 and 1004 by determining a position and slope of the upper and lower edges of the rectangular bounding box of the primary group 900, and generating lines having the same positions and slopes, and extending the entire width of the text region 804.

**[0050]** Having fitted the boundary lines 1000 and 1004 to the primary group 900, the sub-region detector 208 is then configured to extend the bounding box of the primary group 900 along the upper and lower boundary lines to define the price text sub-region. As a result, the sub-region detector 208 generates a price text sub-region 1008 as shown in isolation in FIG. 10B (in practice, the sub-region 1008 need not be extracted from the text region 804). As seen in FIG. 10B, the sub-region 1008 includes the string “5.77” previously encompassed by the group 900, but also includes the characters “\$” and “2” which were previously excluded from the group 900. Still further, the sub-region 1008 excludes the characters “png” which were previously included with the character “2” as a single candidate text element due to their proximity with the “2”.

**[0051]** Returning to FIG. 3, following the detection of a price text sub-region, at block 330 the interpreter 212 is configured to generate a price text string from the price text sub-region. To generate the price text string, the interpreter 212 is configured to apply a suitable optical character recognition (OCR) technique to the price text sub-region. In the present example, the interpreter 212 is configured to apply a set of operations comprising an OCR pipeline, including a feature extraction step to generate a feature descriptor (e.g. an HOG descriptor), followed by a linear discriminant analysis (LDA), a distance classification operation such as a modified quadrant discriminant function (MQDF) and a comparison of the outputs of the above-mentioned operations with a character database stored in the memory 122. The database may be restricted to only numerical digits and currency symbols in some examples. A variety of other suitable OCR techniques may also be applied, as will now be apparent to those skilled in the art.

**[0052]** It has been found that the generation of text regions and detection of a price text sub-region 1008, as shown in FIG. 10B, may result in increased price text interpretation accuracy, even when conventional OCR techniques are applied by the interpreter 212 at block 330. In contrast, the application of a conventional OCR technique, such as that provided by a first common portable document format (PDF) reader application, to the text region 804 in FIG. 8 fails to identify the price text as a text string, and therefore fails to correctly interpret the price text. The OCR function provided by a second common PDF reader application identifies a portion of the price text, although only at a certain scale. The output of this second application is the string “25 77” which lacks the decimal point and a currency unit and is therefore not a correct actionable price text string. At other scales, the second application, like the first, fails to recognize the price digits as text at all. However, applying both of the above-mentioned conventional OCR techniques to the price text sub-region 1008 in FIG. 10B, at a variety of scales, yields the correct price text string “\$25.77”.

**[0053]** Testing of the processing techniques described above has revealed that by performing the method 300, the controller 120 correctly identifies text region locations at a rate of between 93% and 96% at various imaging distances (e.g. distances between the apparatus 103 and the shelves 110), varying between 50 cm and 85 cm. Further, the detection and interpretation of price text from within those text regions was performed correctly for a minimum of 88% of samples (at an imaging distance of 50 cm) and a maximum of 92% (at an imaging distance of 85 cm).

**[0054]** Returning to FIG. 3, following the generation of a price text string at block 330, the control application 128 returns to block 320 to determine whether any price text-classified text regions remain to be processed. When the determination is negative, the performance of the method 300 proceeds to block 335, at which the control application 128 is configured to present the locations of the detected text regions (e.g. regions 600), as well as the price text strings interpreted therefrom. In some examples, the control application 128 is configured to present the text regions and price text strings as an overlay on the image obtained at block 305, for example on a display of the mobile device 105 or another client computing device; the overlaid image can also be stored in the repository 132 instead of, or in addition to, display on the mobile device 105 or other client computing device. In other examples, the control

application 128 presents – for display at the mobile device 105, another client device, or both, alone or in combination with presentation to the repository 132 for storage – the text regions and price text strings as a list of bounding box coordinates (e.g. relative to the image obtained at block 305, or relative to a common frame of reference corresponding to the retail environment in which the shelves 110 are located). The list includes, with each bounding box, the corresponding price text string, as well as a label type indicator (where multiple label types were detected by the classifier 204) and a confidence value generated by the interpreter 212 at block 330. An example of such a list is shown below in Table 1.

Table 1: Price Text String List

| <b>Bounding Box</b> | <b>Price Text String</b> | <b>Label Type</b> | <b>Confidence</b> |
|---------------------|--------------------------|-------------------|-------------------|
| [X1, Y1] [X2, Y2]   | \$2.45                   | Type-A            | 83%               |
| [X1, Y1] [X2, Y2]   | \$12.99                  | Type-B            | 92%               |

**[0055]** Variations to the techniques set out above are contemplated. For example, in the detection of a price text sub-region at block 325, geometric constraints may be applied by the sub-region detector 208 in addition to the size constraint mentioned above. For example, the sub-region detector 208 is configured in some examples to group the candidate text elements by size and only in certain directions, such that candidate text elements are only grouped in a horizontal direction, whether or not similarly-sized elements are present above or below.

**[0056]** In a further example, the sub-region detector 208 is configured to determine whether the difference in size (e.g. area) of two or more candidate primary groups is below a predetermined threshold. When the determination is affirmative, the sub-region detector 208 is configured to apply one or more additional criteria to select the primary group. For example, based on the label type output from the classifier 204, the sub-region detector 208 can be configured to select the primary group candidate closest to an expected price text position (e.g. the center of the label, in the case of the format of the label 416-2).

**[0057]** In the foregoing specification, specific embodiments have been described. However, one of ordinary skill in the art appreciates that various modifications and changes can be made without departing from the scope of the invention as set forth in

the claims below. Accordingly, the specification and figures are to be regarded in an illustrative rather than a restrictive sense, and all such modifications are intended to be included within the scope of present teachings.

**[0058]** The benefits, advantages, solutions to problems, and any element(s) that may cause any benefit, advantage, or solution to occur or become more pronounced are not to be construed as a critical, required, or essential features or elements of any or all the claims. The invention is defined solely by the appended claims including any amendments made during the pendency of this application and all equivalents of those claims as issued.

**[0059]** Moreover in this document, relational terms such as first and second, top and bottom, and the like may be used solely to distinguish one entity or action from another entity or action without necessarily requiring or implying any actual such relationship or order between such entities or actions. The terms "comprises," "comprising," "has", "having," "includes", "including," "contains", "containing" or any other variation thereof, are intended to cover a non-exclusive inclusion, such that a process, method, article, or apparatus that comprises, has, includes, contains a list of elements does not include only those elements but may include other elements not expressly listed or inherent to such process, method, article, or apparatus. An element preceded by "comprises ...a", "has ...a", "includes ...a", "contains ...a" does not, without more constraints, preclude the existence of additional identical elements in the process, method, article, or apparatus that comprises, has, includes, contains the element. The terms "a" and "an" are defined as one or more unless explicitly stated otherwise herein. The terms "substantially", "essentially", "approximately", "about" or any other version thereof, are defined as being close to as understood by one of ordinary skill in the art, and in one non-limiting embodiment the term is defined to be within 10%, in another embodiment within 5%, in another embodiment within 1% and in another embodiment within 0.5%. The term "coupled" as used herein is defined as connected, although not necessarily directly and not necessarily mechanically. A device or structure that is "configured" in a certain way is configured in at least that way, but may also be configured in ways that are not listed.

**[0060]** It will be appreciated that some embodiments may be comprised of one or more generic or specialized processors (or “processing devices”) such as microprocessors, digital signal processors, customized processors and field programmable gate arrays (FPGAs) and unique stored program instructions (including both software and firmware) that control the one or more processors to implement, in conjunction with certain non-processor circuits, some, most, or all of the functions of the method and/or apparatus described herein. Alternatively, some or all functions could be implemented by a state machine that has no stored program instructions, or in one or more application specific integrated circuits (ASICs), in which each function or some combinations of certain of the functions are implemented as custom logic. Of course, a combination of the two approaches could be used.

**[0061]** Moreover, an embodiment can be implemented as a computer-readable storage medium having computer readable code stored thereon for programming a computer (e.g., comprising a processor) to perform a method as described and claimed herein. Examples of such computer-readable storage mediums include, but are not limited to, a hard disk, a CD-ROM, an optical storage device, a magnetic storage device, a ROM (Read Only Memory), a PROM (Programmable Read Only Memory), an EPROM (Erasable Programmable Read Only Memory), an EEPROM (Electrically Erasable Programmable Read Only Memory) and a Flash memory. Further, it is expected that one of ordinary skill, notwithstanding possibly significant effort and many design choices motivated by, for example, available time, current technology, and economic considerations, when guided by the concepts and principles disclosed herein will be readily capable of generating such software instructions and programs and ICs with minimal experimentation.

**[0062]** The Abstract of the Disclosure is provided to allow the reader to quickly ascertain the nature of the technical disclosure. It is submitted with the understanding that it will not be used to interpret or limit the scope or meaning of the claims. In addition, in the foregoing Detailed Description, it can be seen that various features are grouped together in various embodiments for the purpose of streamlining the disclosure. This method of disclosure is not to be interpreted as reflecting an intention that the claimed embodiments require more features than are expressly recited in each claim. Rather, as the following claims reflect, inventive subject matter

lies in less than all features of a single disclosed embodiment. Thus the following claims are hereby incorporated into the Detailed Description, with each claim standing on its own as a separately claimed subject matter.

## Claims

1. A method of price text detection by an imaging controller, comprising:
  - obtaining, by the imaging controller, an image of a shelf supporting labels bearing price text;
  - generating, by the imaging controller, a plurality of text regions containing candidate text elements from the image;
  - assigning, by the imaging controller, a classification to each of the text regions, selected from a price text classification and a non-price text classification;
  - wherein the imaging controller, within each of a subset of the text regions having the price text classification:
    - detects a price text sub-region; and
    - generates a price text string by applying character recognition to the price text sub-region; and
    - presenting, by the imaging controller, the locations of the subset of text regions, in association with the corresponding price text strings.
2. The method of claim 1, further comprising: prior to generating the text regions, identifying the candidate text elements comprises applying a blob detection operation to the image.
3. The method of claim 2, wherein generating the text regions comprises:
  - for each pair of the candidate text elements, determining whether a distance between the pair is below a distance threshold.
4. The method of claim 2, wherein generating the text regions comprises:
  - for each pair of the candidate text elements, determining whether a difference between the size of each of the pair is below a size threshold.
5. The method of claim 1, wherein assigning a classification to each of the text regions comprises:
  - generating a feature descriptor for the text region;

providing the feature descriptor to a classifier; and  
receiving the classification from the classifier.

6. The method of claim 1, wherein detecting a price text sub-region comprises:
  - assigning the candidate text elements within the text region to groups based on the size of the candidate text elements;
  - selecting a primary one of the groups;
  - fitting upper and lower bounding lines to the primary group; and
  - extending the primary group along the bounding lines to define the price text sub-region.
7. The method of claim 6, wherein selecting the primary one of the groups comprises selecting the group having the largest area.
8. The method of claim 6, further comprising: prior to assigning the candidate text elements within the text region to groups, binarizing the text region.
9. The method of claim 6, wherein at least one of the upper and lower bounding lines intersects a candidate text element in a group other than the primary group.
10. The method of claim 1, the presenting further comprising presenting a confidence level corresponding to each price text string.
11. A server for detecting price text, comprising:
  - a memory storing an image of a shelf supporting labels bearing price text;
  - an imaging controller coupled to the memory, the imaging controller comprising:
    - a text region generator configured to generate a plurality of text regions containing candidate text elements from the image;
    - a classifier configured to assign a classification to each of the text regions, selected from a price text classification and a non-price text classification;

a sub-region generator configured to detect a price text sub-region within each of a subset of the text regions having the price text classification; and

an interpreter configured to generate a price text string by applying character recognition to the price text sub-region; and to present the locations of the subset of text regions, in association with the corresponding price text strings.

12. The server of claim 11, the text region generator further configured, prior to generating the text regions, to identify the candidate text elements by applying a blob detection operation to the image.

13. The server of claim 12, the text region generator configured to generate the text regions by:

for each pair of the candidate text elements, determining whether a distance between the pair is below a distance threshold.

14. The server of claim 12, the text region generator configured to generate the text regions by:

for each pair of the candidate text elements, determining whether a difference between the size of each of the pair is below a size threshold.

15. The server of claim 11, the classifier configured to assign a classification to each of the text regions by:

generating a feature descriptor for the text region;  
providing the feature descriptor to a classifier; and  
receiving the classification from the classifier.

16. The server of claim 11, the sub-region generator configured to detect a price text sub-region by:

assigning the candidate text elements within the text region to groups based on the size of the candidate text elements;

selecting a primary one of the groups;  
fitting upper and lower bounding lines to the primary group; and  
extending the primary group along the bounding lines to define the price text sub-region.

17. The server of claim 16, wherein selecting the primary one of the groups comprises selecting the group having the largest area.

18. The server of claim 16, the sub-region generator further configured, prior to assigning the candidate text elements within the text region to groups, to binarize the text region.

19. The server of claim 16, wherein at least one of the upper and lower bounding lines intersects a candidate text element in a group other than the primary group.

20. The server of claim 11, the interpreter further configured to present a confidence level corresponding to each price text string.

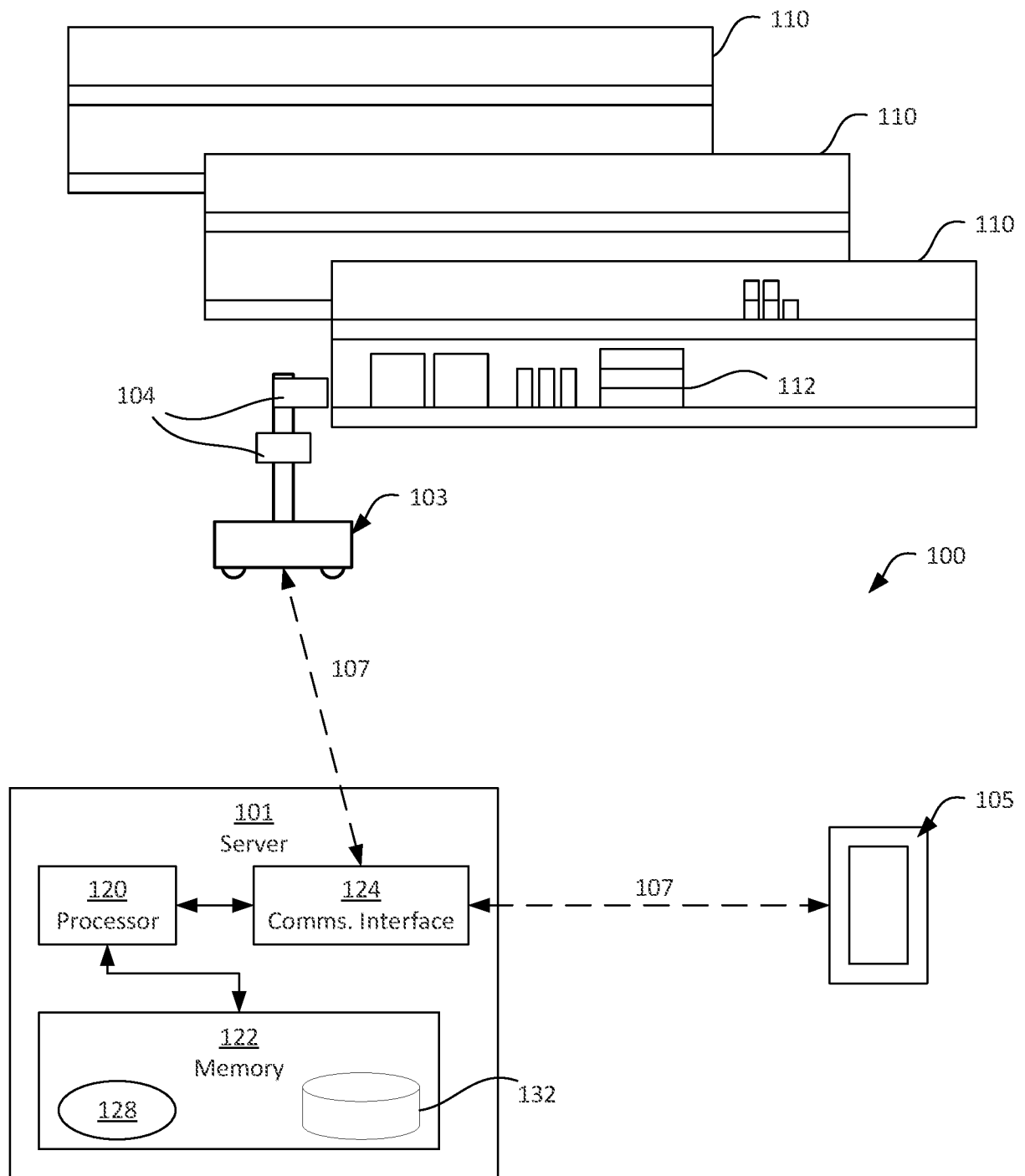


FIG. 1

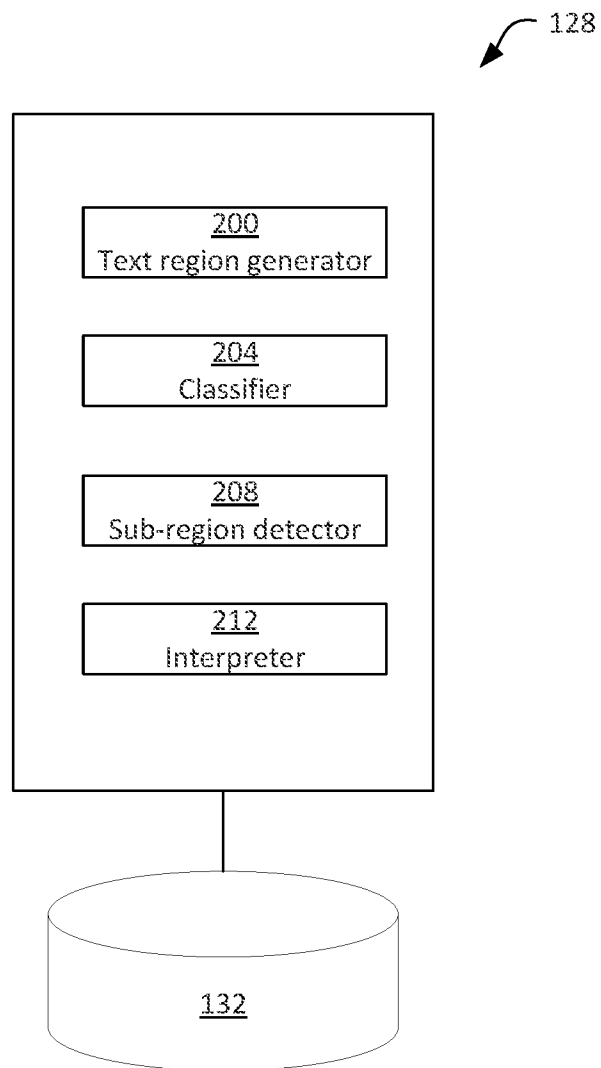


FIG. 2

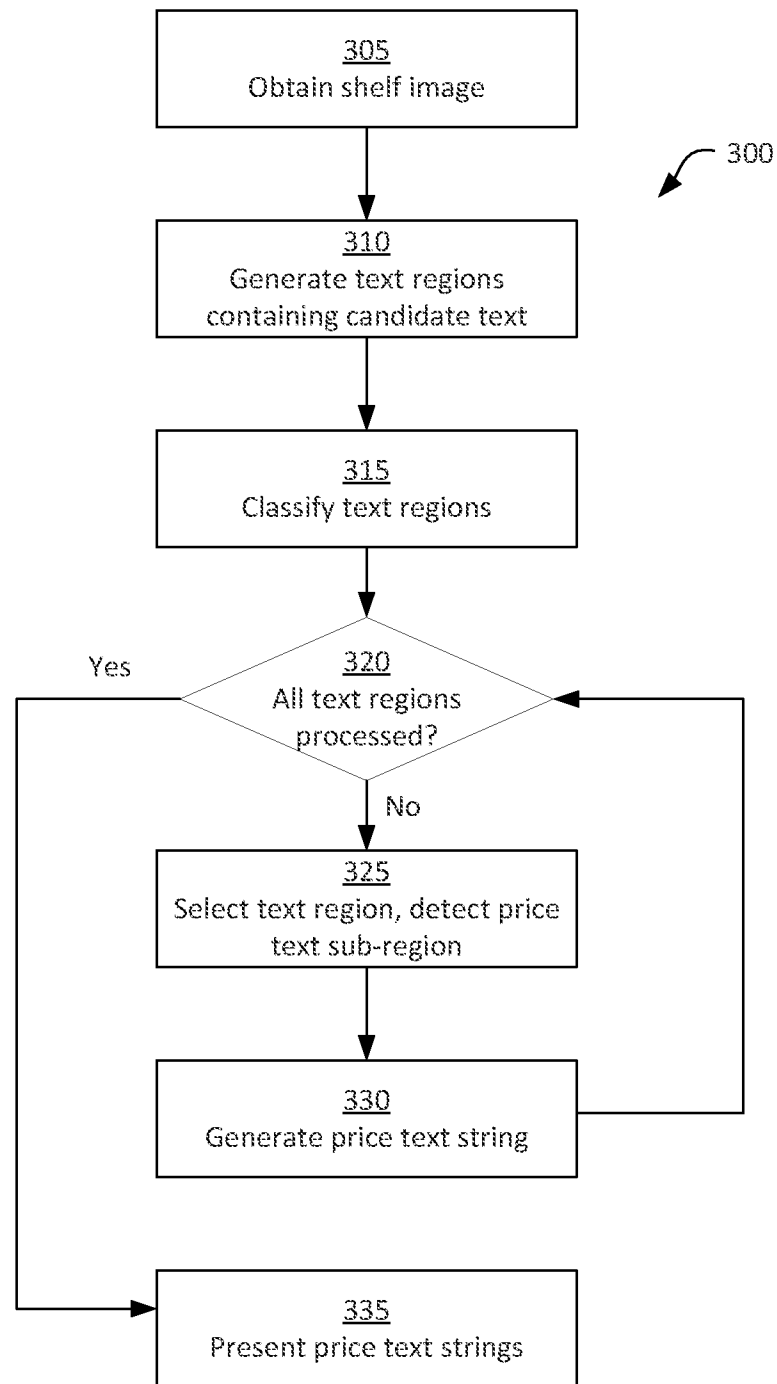


FIG. 3

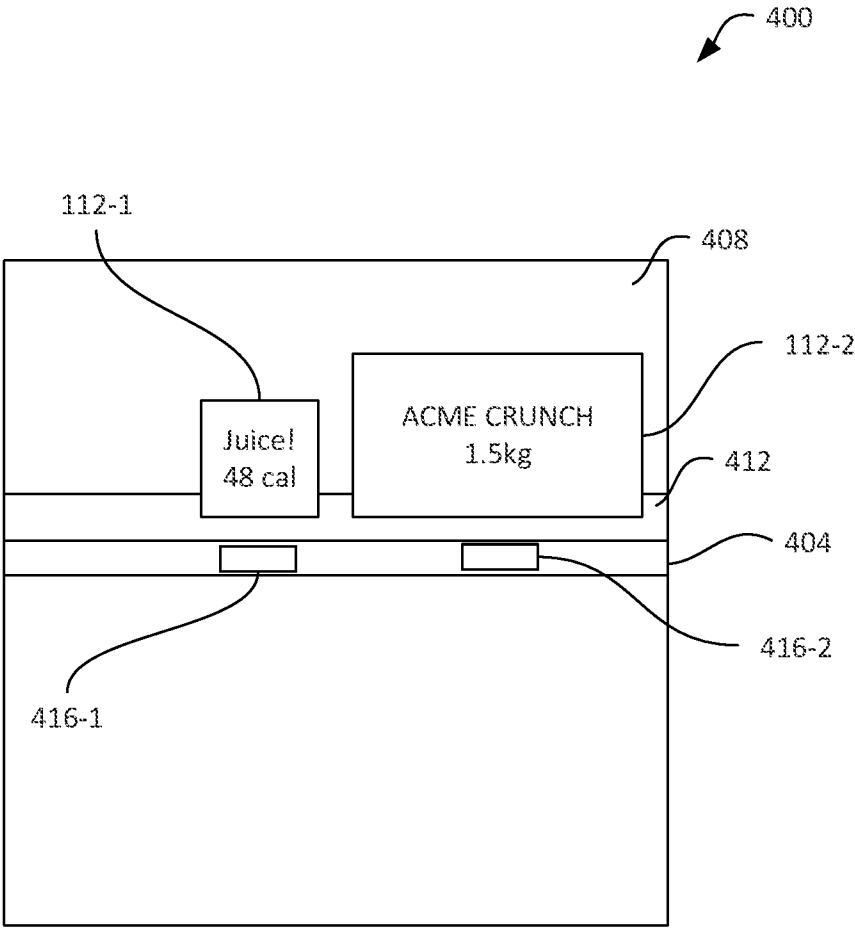


FIG. 4



FIG. 5A

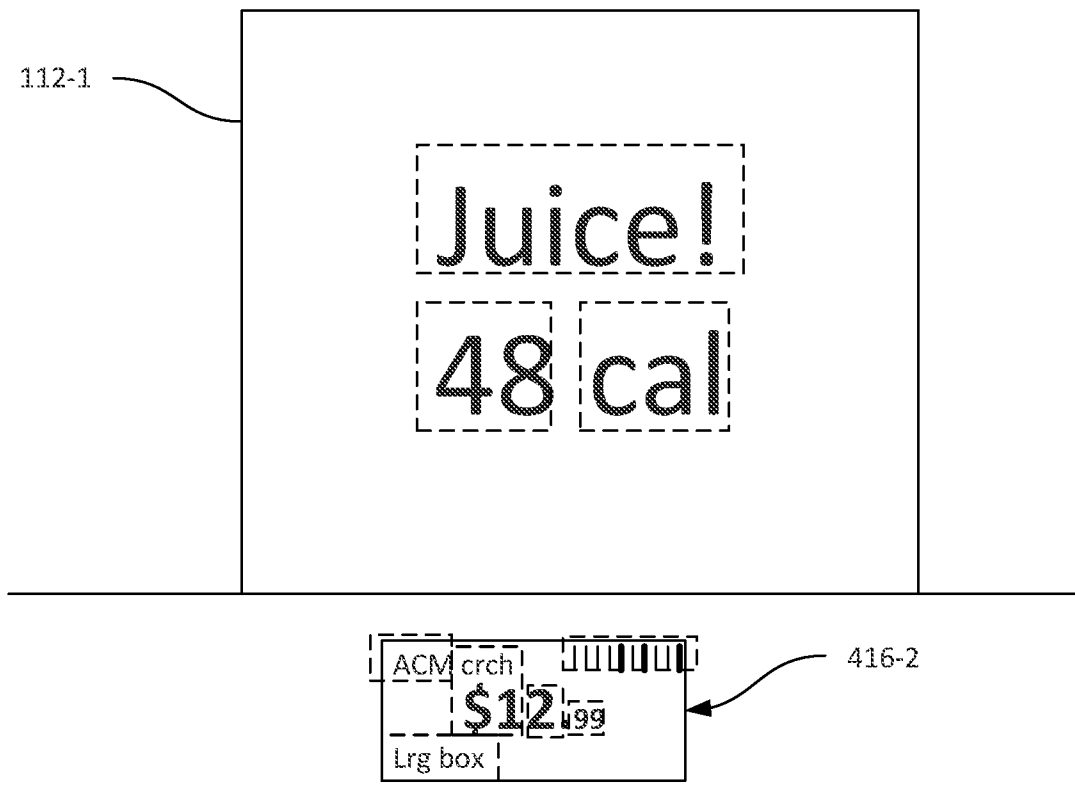


FIG. 5B

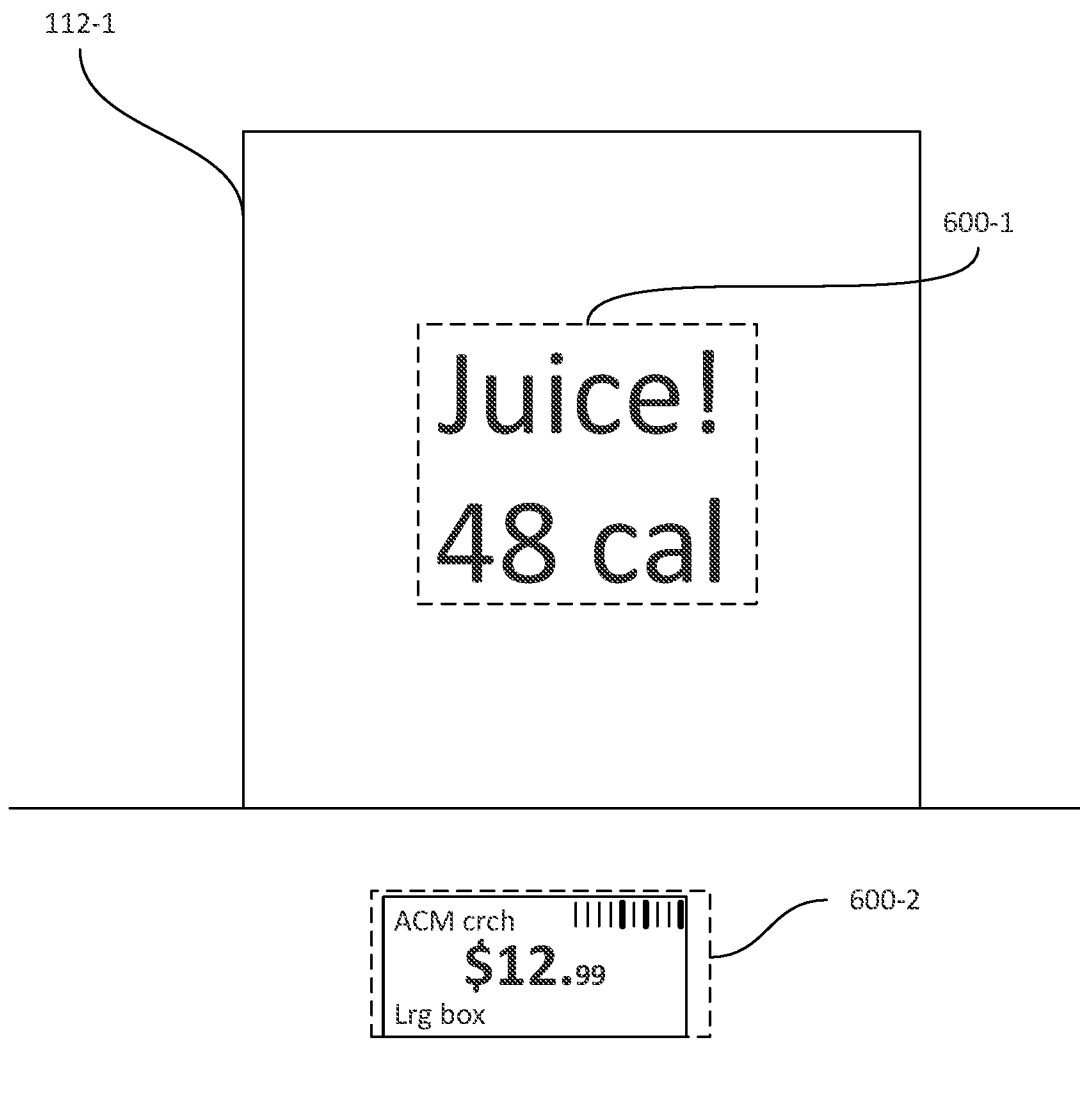


FIG. 6

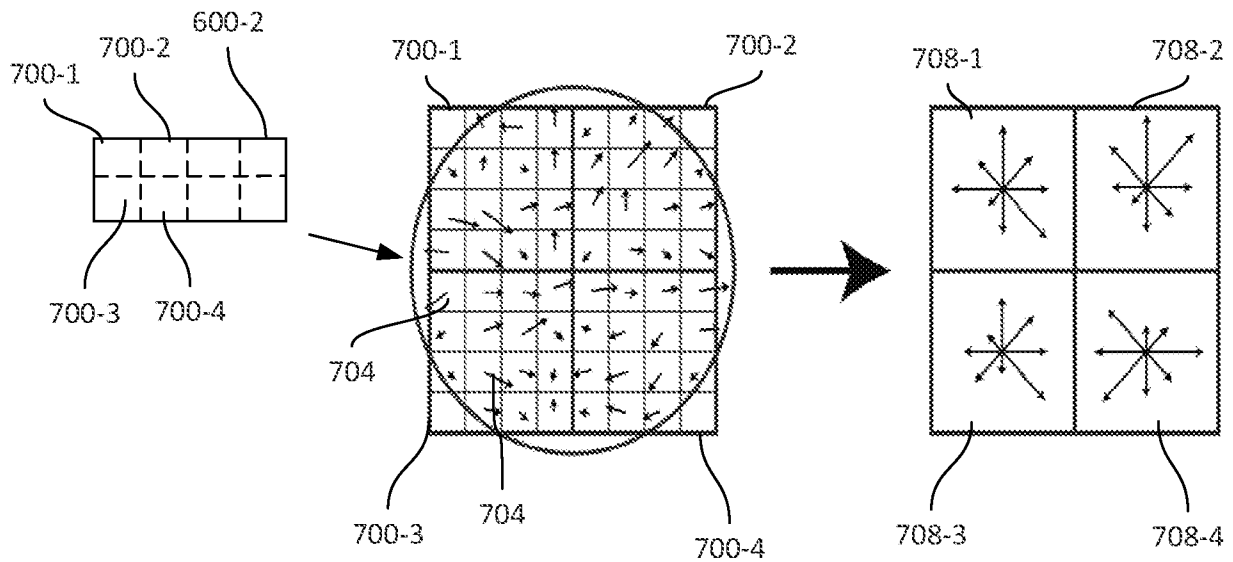


FIG. 7A

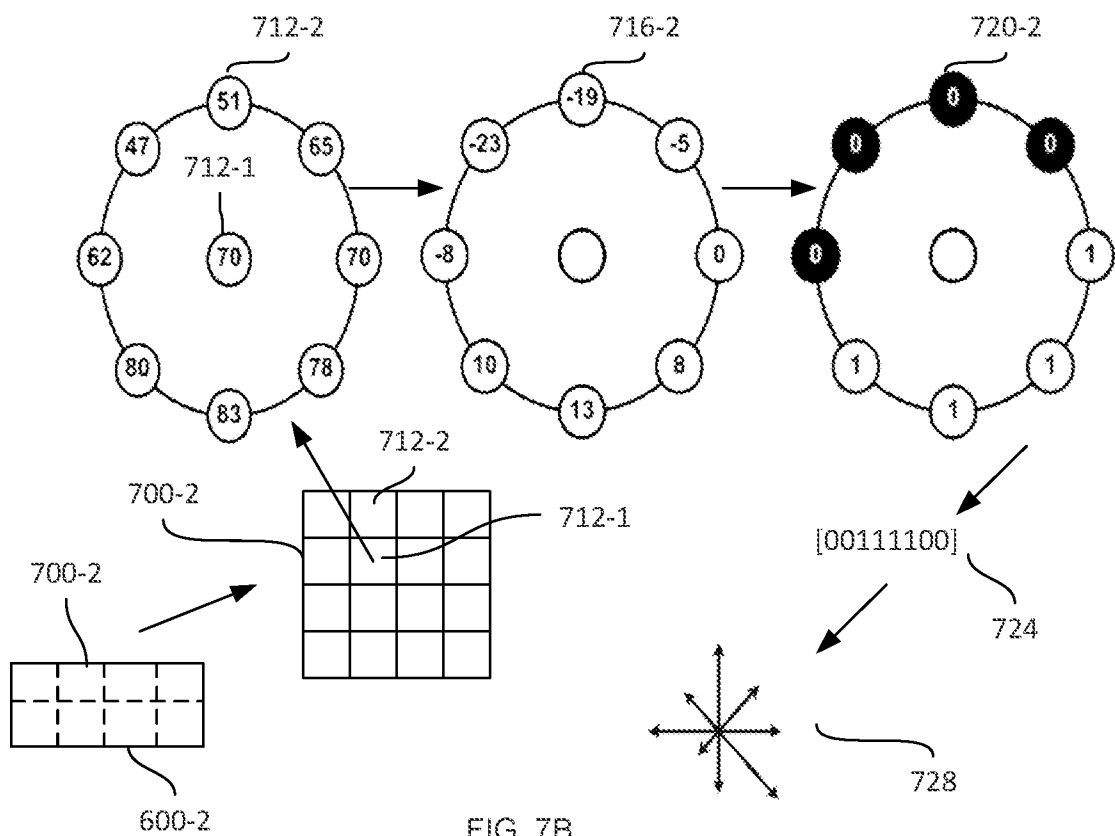


FIG. 7B



FIG. 8

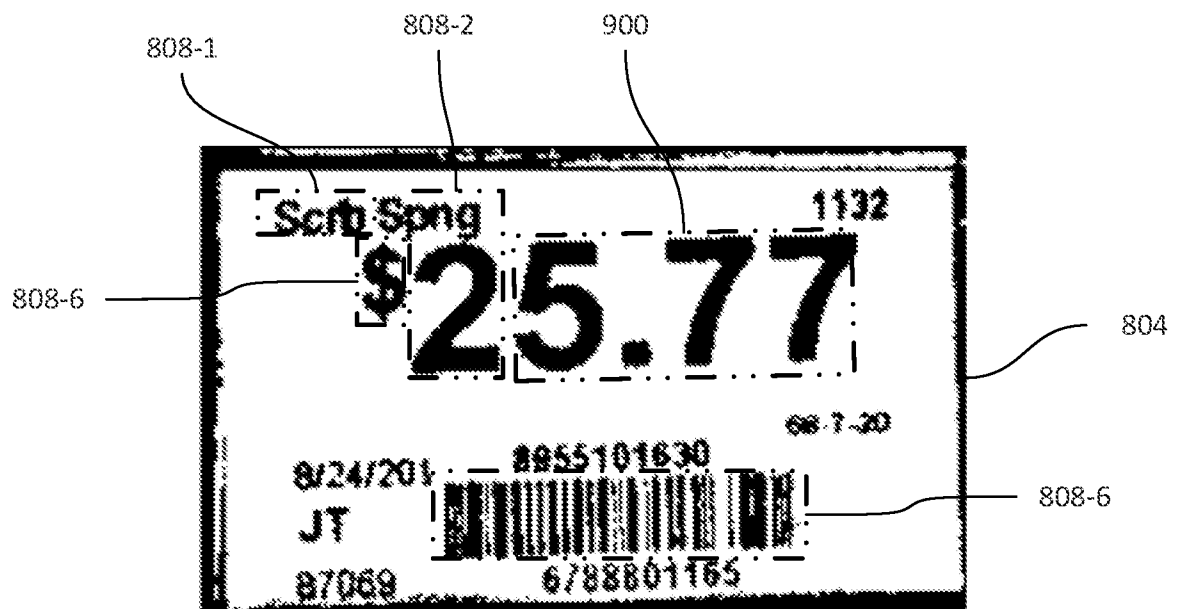


FIG. 9

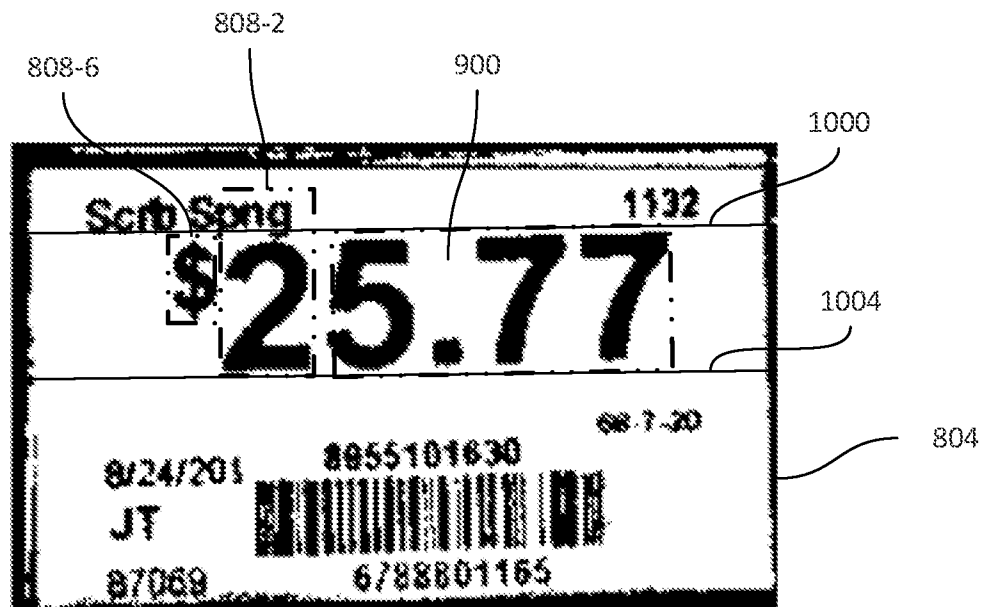


FIG. 10A

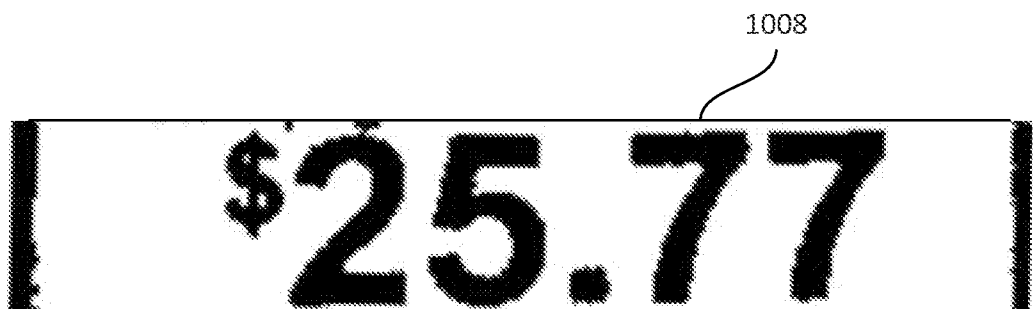


FIG. 10B

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/083143

**A. CLASSIFICATION OF SUBJECT MATTER**

G06Q 30/06(2012.01)i; G06K 7/12(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06K,G06Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT,CNKI,IEEE,WPI,EPODOC,GOOGLE: price,label,tag,detect,identify,OCR,text, region,area,classification,image

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                      | Relevant to claim No. |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | JP 2016194834 A (FUJITSU LTD.) 17 November 2016 (2016-11-17)<br>description, paragraphs 0002-0004, 0008, 0021-0027, figures 1-5         | 1-20                  |
| A         | JP 2017016539 A (NEC CORP.) 19 January 2017 (2017-01-19)<br>the whole document                                                          | 1-20                  |
| A         | CN 105989512 A (HUANGJINFU INTELLECTUAL PROPERTY CONSULTATION<br>SHENZHEN CO., LTD.) 05 October 2016 (2016-10-05)<br>the whole document | 1-20                  |
| A         | CN 102214343 A (ACCENTURE GLOBAL SERVICES LTD.) 12 October 2011<br>(2011-10-12)<br>the whole document                                   | 1-20                  |

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&amp;” document member of the same patent family

Date of the actual completion of the international search

15 January 2018

Date of mailing of the international search report

11 February 2018

Name and mailing address of the ISA/CN

STATE INTELLECTUAL PROPERTY OFFICE OF THE  
P.R.CHINA  
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing  
100088  
China

Authorized officer

WANG,Ping

Facsimile No. (86-10)62019451

Telephone No. (86-10)53961372

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2017/083143**

| Patent document<br>cited in search report |            |   | Publication date<br>(day/month/year) | Patent family member(s) |            |    | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|---|--------------------------------------|-------------------------|------------|----|--------------------------------------|
| JP                                        | 2016194834 | A | 17 November 2016                     | None                    |            |    |                                      |
| JP                                        | 2017016539 | A | 19 January 2017                      | None                    |            |    |                                      |
| CN                                        | 105989512  | A | 05 October 2016                      | None                    |            |    |                                      |
| CN                                        | 102214343  | A | 12 October 2011                      | AU                      | 2010246468 | A1 | 27 October 2011                      |
|                                           |            |   |                                      | RU                      | 2452021    | C1 | 27 May 2012                          |
|                                           |            |   |                                      | EP                      | 2375365    | A1 | 12 October 2011                      |
|                                           |            |   |                                      | CA                      | 2720217    | A1 | 08 October 2011                      |
|                                           |            |   |                                      | US                      | 2011011936 | A1 | 20 January 2011                      |
|                                           |            |   |                                      | US                      | 2015317620 | A1 | 05 November 2015                     |