US 20040208372A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2004/0208372 A1

Boncyk et al. (43) Pub. Date: Oct. 21, 2004

(54) IMAGE CAPTURE AND IDENTIFICATION SYSTEM AND PROCESS

(76) Inventors: **Wayne C. Boncyk**, Fontana, CA (US); **Ronald H. Cohen**, Pasadena, CA (US)

Correspondence Address:
ROBERT D. FISH
RUTAN & TUCKER LLP
611 ANTON BLVD 14TH FLOOR
COSTA MESA, CA 92626-1931 (US)

(21) Appl. No.: **10/492,243**

(22) PCT Filed: **Nov. 5, 2002**

(86) PCT No.: **PCT/US02/.5407**

(30) **Foreign Application Priority Data**

Nov. 5, 2001 (US)..... 09992942

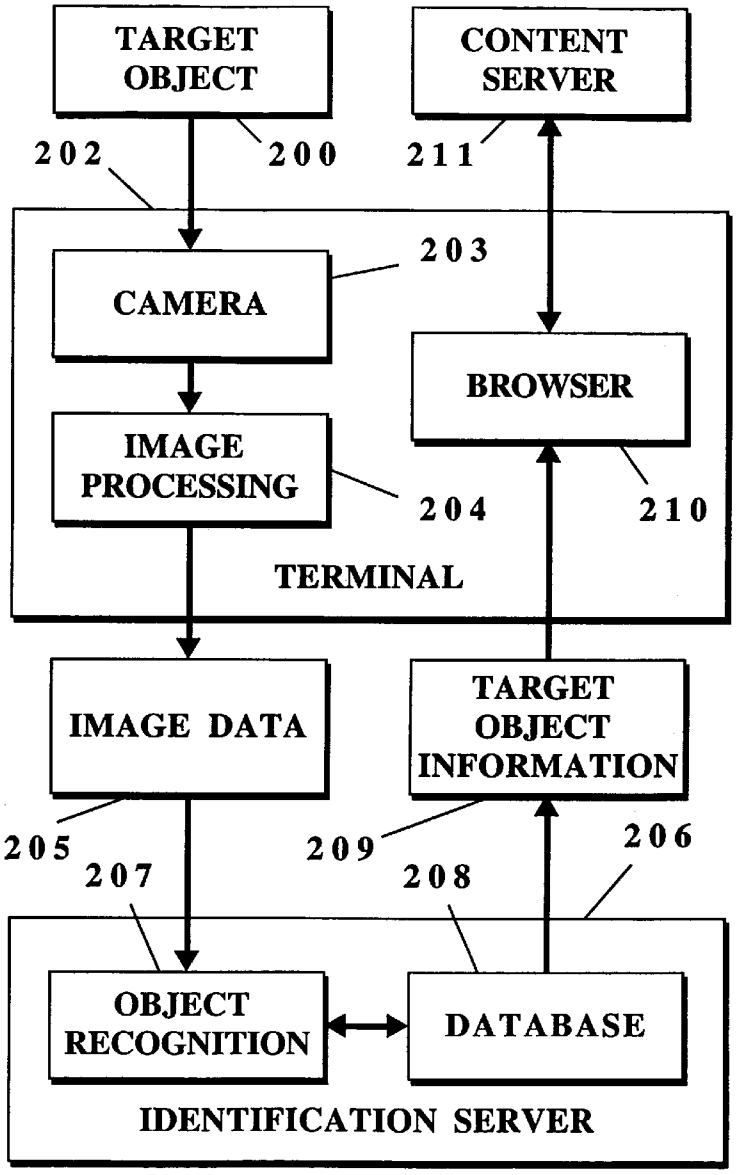
Publication Classification

(51) **Int. Cl.**[7] **G06K 9/00**; G06F 7/00

(52) **U.S. Cl.** **382/181**; 707/104.1

(57) **ABSTRACT**

The invention is a method by which information and communication pertinent to an object is provided based on imagery of the object. A digital image of the object is captured and the object is recognized from a plurality of objects in a database. An information address corresponding to the object is then used to access information and initiate communication pertinent to the object.



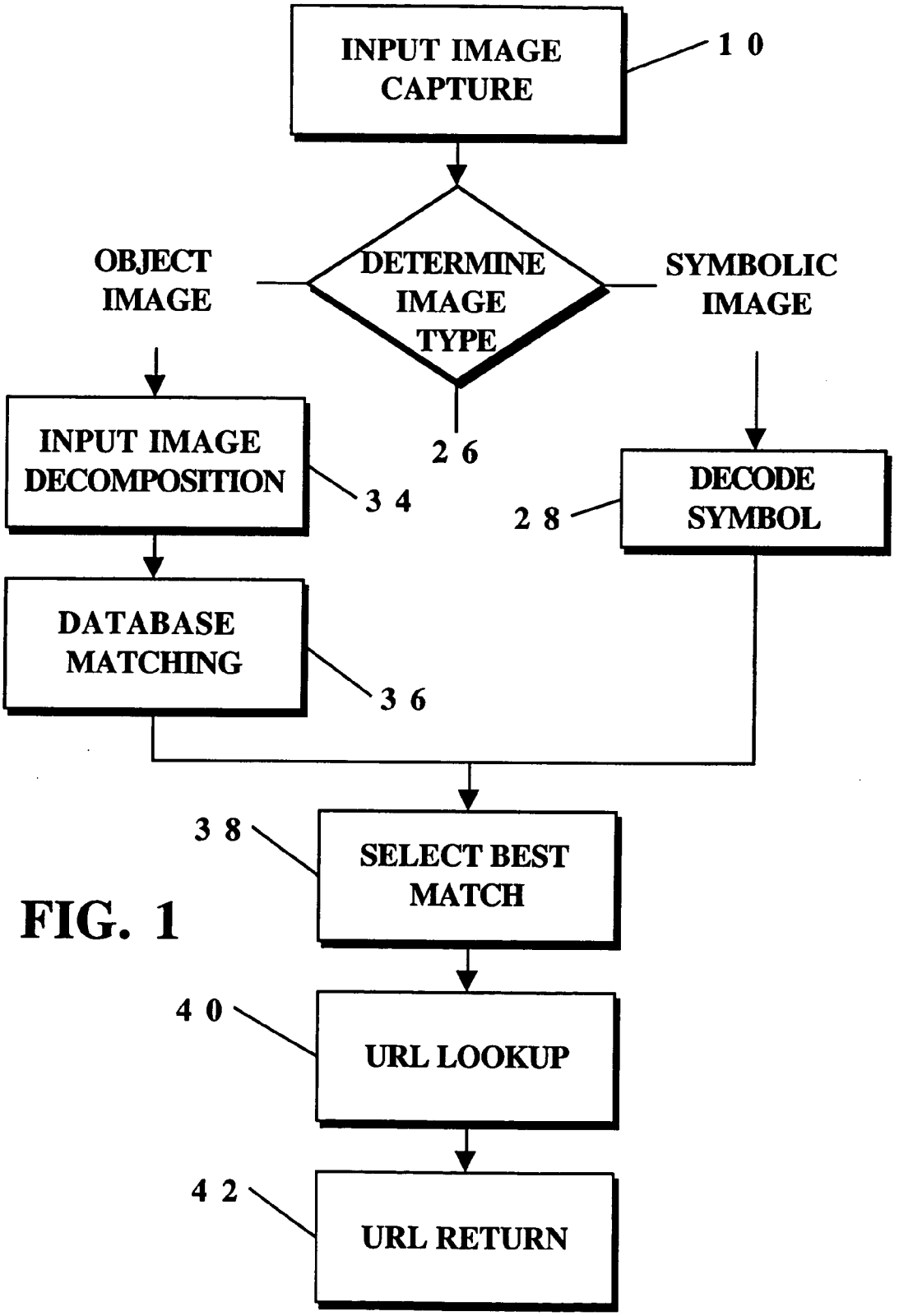

INPUT IMAGE CAPTURE
10
DETERMINE IMAGE TYPE
OBJECT IMAGE
SYMBOLIC IMAGE
26
INPUT IMAGE DECOMPOSITION
34
DECODE SYMBOL
28
DATABASE MATCHING
36
38
SELECT BEST MATCH
40
URL LOOKUP
42
URL RETURN

FIG. 1

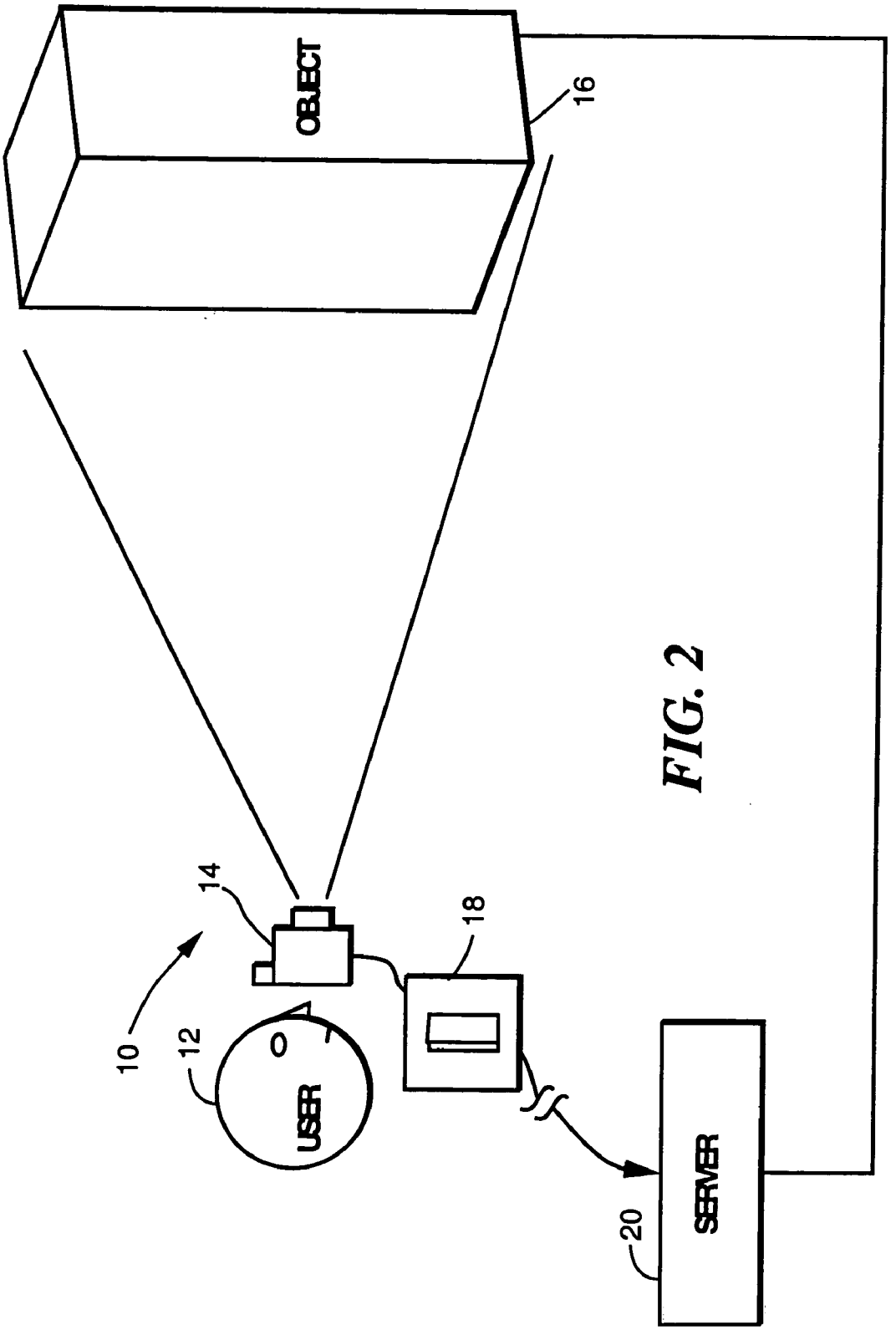


FIG. 2

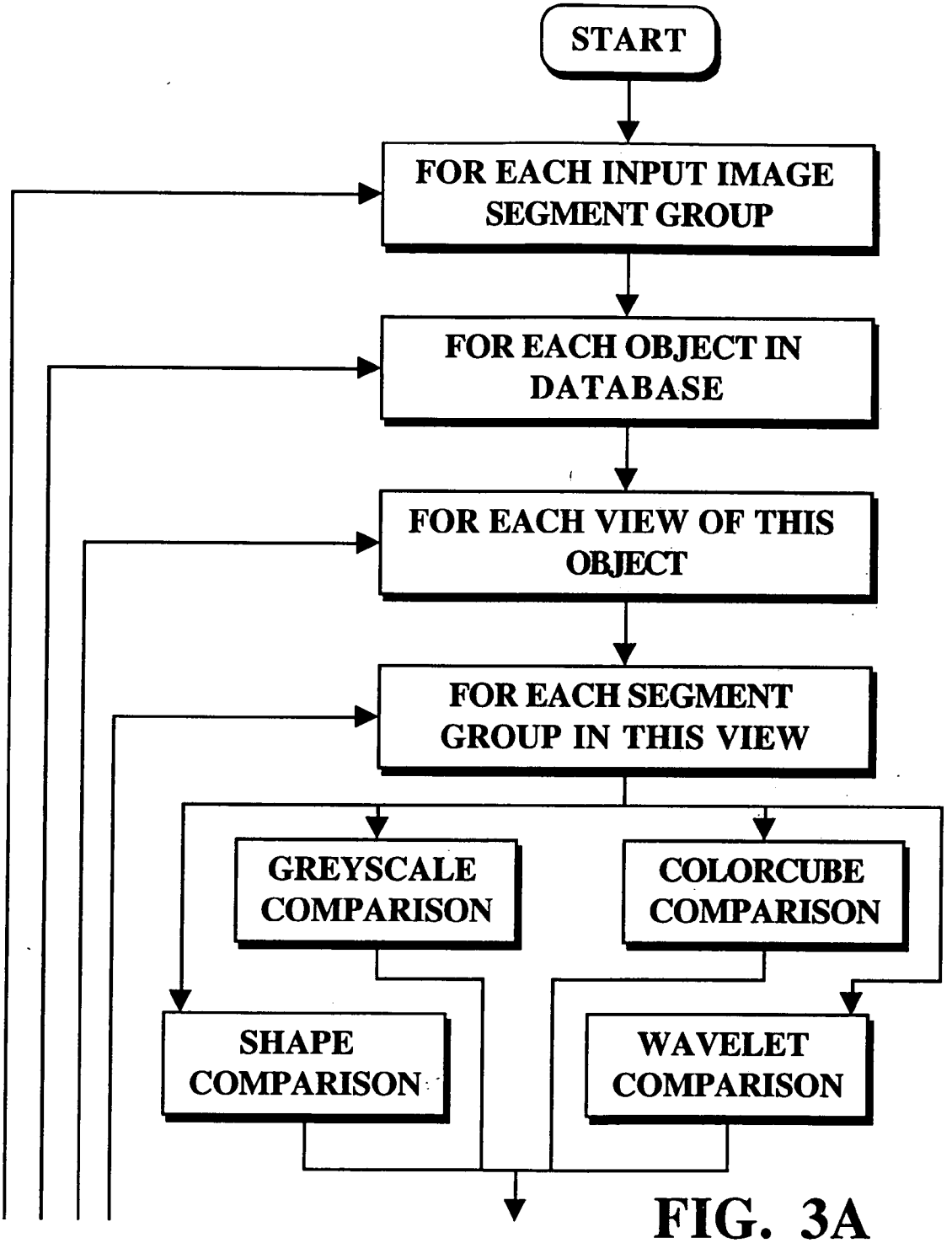

START
FOR EACH INPUT IMAGE SEGMENT GROUP
FOR EACH OBJECT IN DATABASE
FOR EACH VIEW OF THIS OBJECT
FOR EACH SEGMENT GROUP IN THIS VIEW
GREYSCALE COMPARISON
COLORCUBE COMPARISON
SHAPE COMPARISON
WAVELET COMPARISON

FIG. 3A

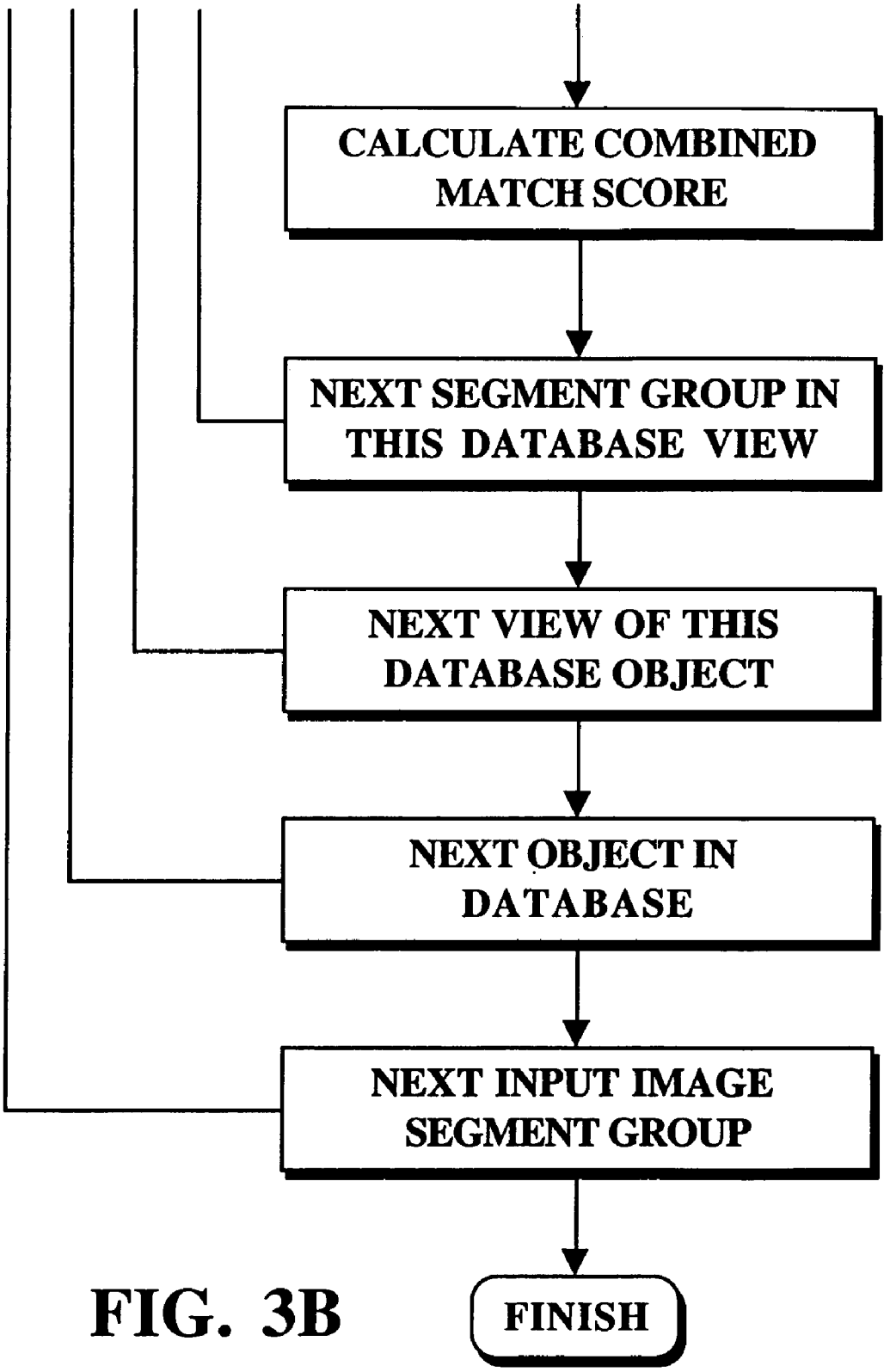

CALCULATE COMBINED MATCH SCORE
NEXT SEGMENT GROUP IN THIS DATABASE VIEW
NEXT VIEW OF THIS DATABASE OBJECT
NEXT OBJECT IN DATABASE
NEXT INPUT IMAGE SEGMENT GROUP
FINISH

FIG. 3B

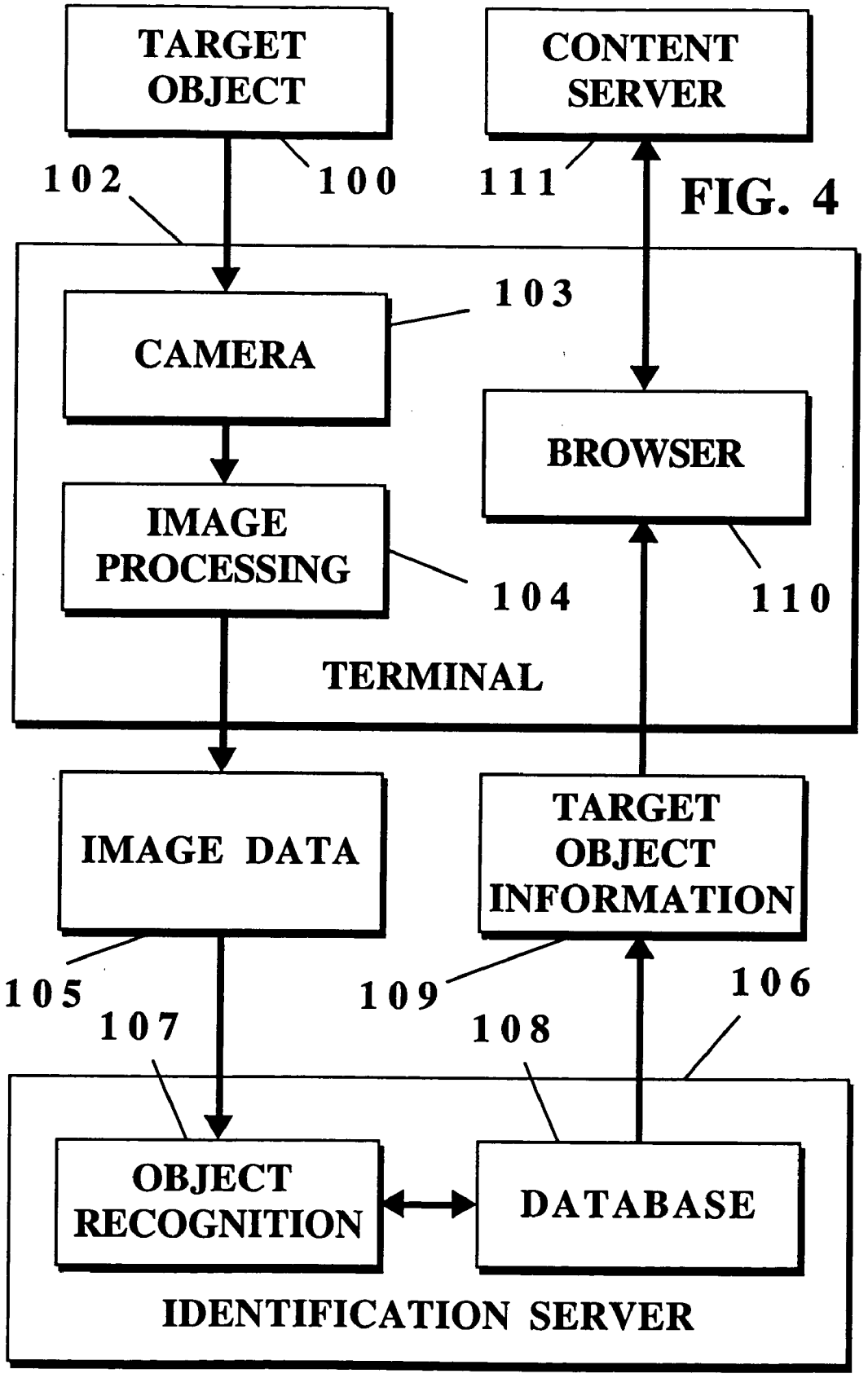

TARGET OBJECT
CONTENT SERVER
102
100
111
103
CAMERA
BROWSER
IMAGE PROCESSING
104
110
TERMINAL
IMAGE DATA
TARGET OBJECT INFORMATION
105
107
109
108
106
OBJECT RECOGNITION
DATABASE
IDENTIFICATION SERVER

FIG. 4

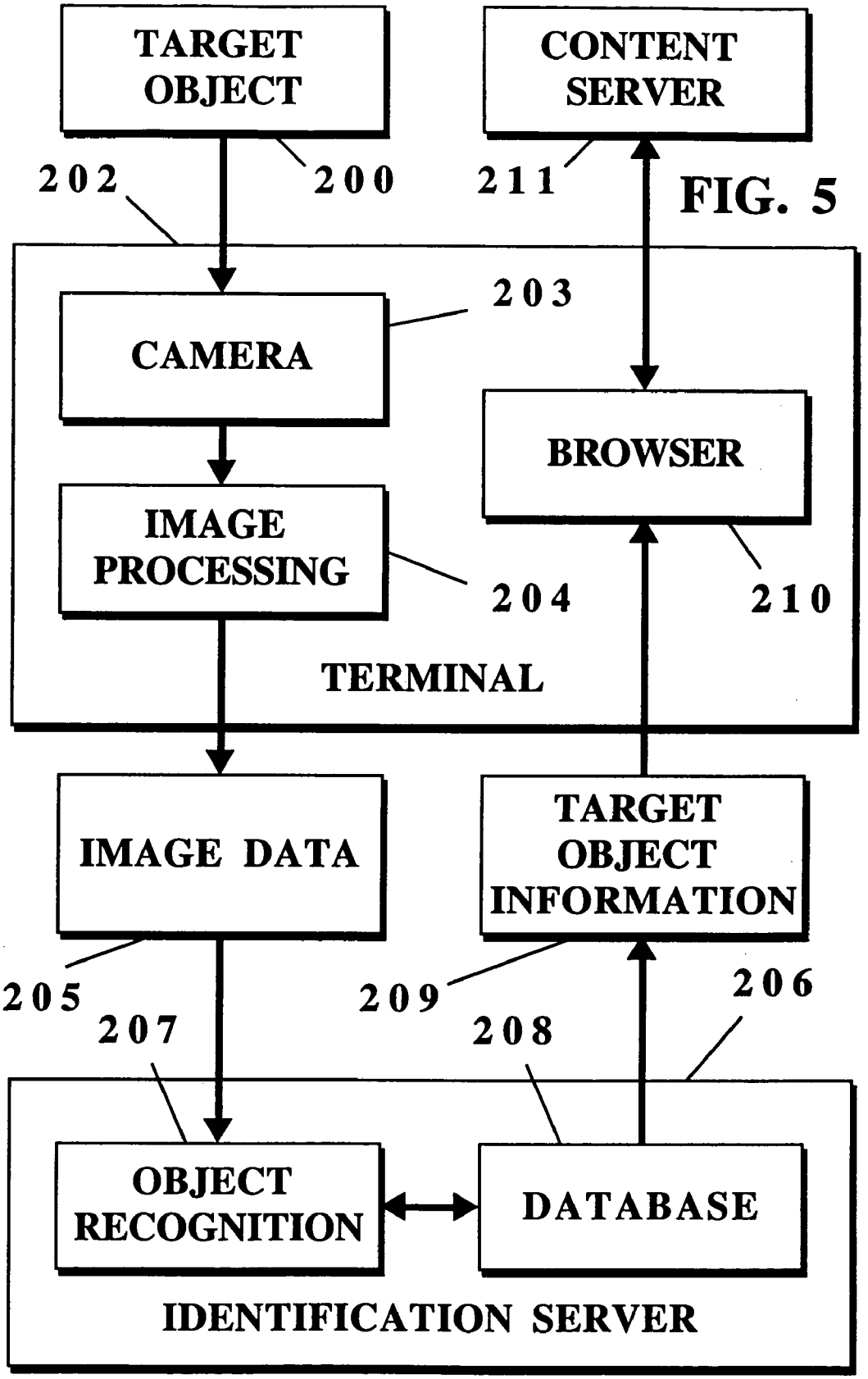

TARGET OBJECT
CONTENT SERVER
202
200
211
203
CAMERA
BROWSER
IMAGE PROCESSING
204
210
TERMINAL
IMAGE DATA
TARGET OBJECT INFORMATION
205
207
209
208
206
OBJECT RECOGNITION
DATABASE
IDENTIFICATION SERVER

FIG. 5

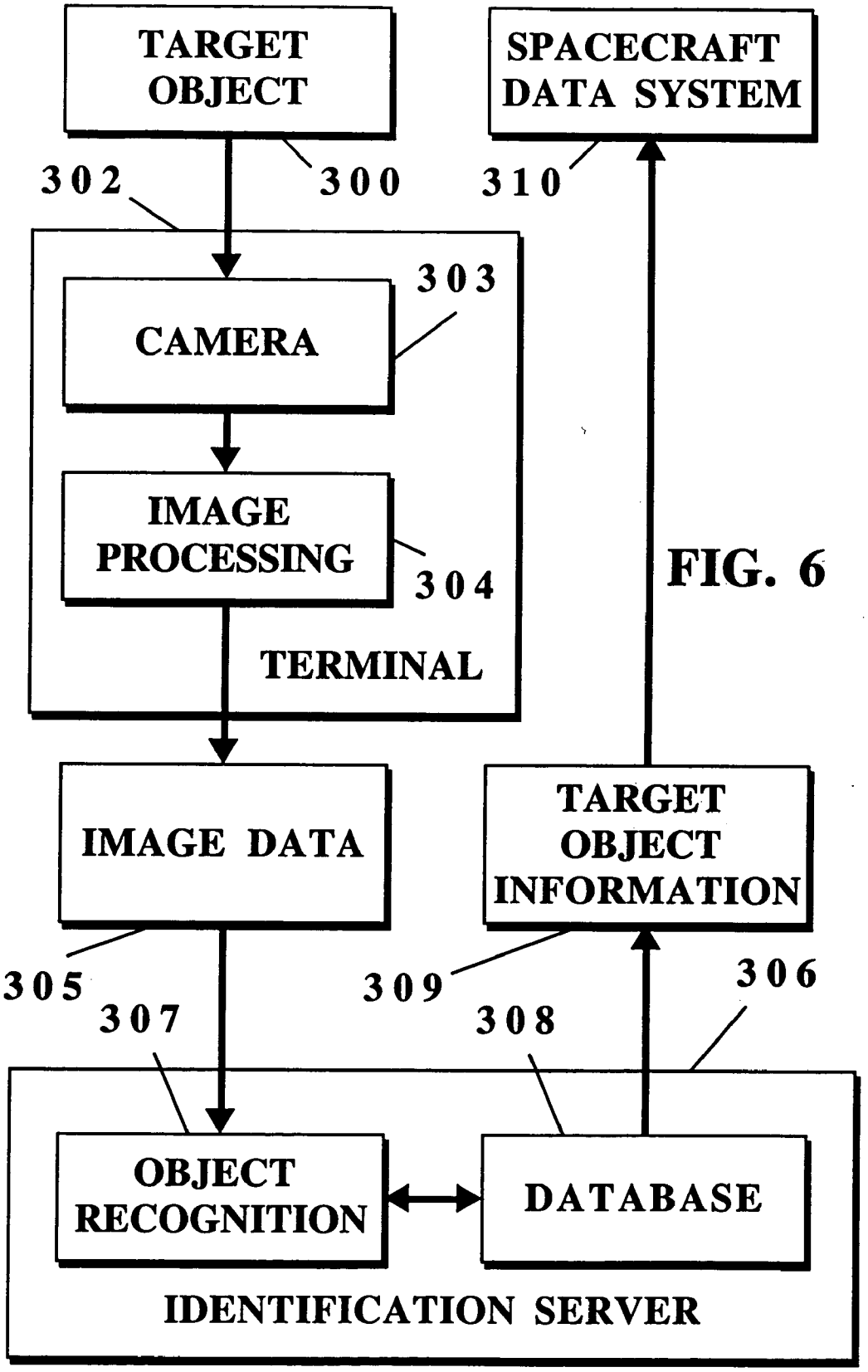

TARGET OBJECT
SPACECRAFT DATA SYSTEM
302
300
310
303
CAMERA
IMAGE PROCESSING
304
TERMINAL
IMAGE DATA
TARGET OBJECT INFORMATION
305
307
309
308
306
OBJECT RECOGNITION
DATABASE
IDENTIFICATION SERVER

FIG. 6

IMAGE CAPTURE AND IDENTIFICATION SYSTEM AND PROCESS

TECHNICAL FIELD

[0001] The invention relates an identification method and process for objects from digitally captured images thereof that uses image characteristics to identify an object from a plurality of objects in a database.

BACKGROUND ART

[0002] There is a need to provide hyperlink functionality in known objects without modification to the objects, through reliably detecting and identifying the objects based only on the appearance of the object, and then locating and supplying information pertinent to the object or initiating communications pertinent to the object by supplying an information address, such as a Uniform Resource Locator (URL), pertinent to the object.

[0003] There is a need to determine the position and orientation of known objects based only on imagery of the objects.

[0004] The detection, identification, determination of position and orientation, and subsequent information provision and communication must occur without modification or disfigurement of the object, without the need for any marks, symbols, codes, barcodes, or characters on the object, without the need to touch or disturb the object, without the need for special lighting other than that required for normal human vision, without the need for any communication device (radio frequency, infrared, etc.) to be attached to or nearby the object, and without human assistance in the identification process. The objects to be detected and identified may be 3-dimensional objects, 2-dimensional images (e.g., on paper), or 2-dimensional images of 3-dimensional objects, or human beings.

[0005] There is a need to provide such identification and hyperlink services to persons using mobile computing devices, such as Personal Digital Assistants (PDAs) and cellular telephones.

[0006] There is a need to provide such identification and hyperlink services to machines, such as factory robots and spacecraft.

[0007] Examples Include:

[0008] Identifying pictures or other art in a museum, where it is desired to provide additional information about such art objects to museum visitors via mobile wireless devices;

[0009] provision of content (information, text, graphics, music, video, etc.), communications, and transaction mechanisms between companies and individuals, via networks (wireless or otherwise) initiated by the individuals "pointing and clicking" with camera-equipped mobile devices on magazine advertisements, posters, billboards, consumer products, music or video disks or tapes, buildings, vehicles, etc.;

[0010] establishment of a communications link with a machine, such a vending machine or information kiosk, by "pointing and clicking" on the machine with a camera-equipped mobile wireless device and then execution of communications or transactions between the mobile wireless device and the machine;

[0011] identification of objects or parts in a factory, such as on an assembly line, by capturing an image of the objects or parts, and then providing information pertinent to the identified objects or parts;

[0012] identification of a part of a machine, such as an aircraft part, by a technician "pointing and clicking" on the part with a camera-equipped mobile wireless device, and then supplying pertinent content to the technician, such maintenance instructions or history for the identified part;

[0013] identification or screening of individual(s) by a security officer "pointing and clicking" a camera-equipped mobile wireless device at the individual(s) and then receiving identification information pertinent to the individuals after the individuals have been identified by face recognition software;

[0014] identification, screening, or validation of documents, such as passports, by a security officer "pointing and clicking" a camera-equipped device at the document and receiving a response from a remote computer;

[0015] determination of the position and orientation of an object in space by a spacecraft nearby the object, based on imagery of the object, so that the spacecraft can maneuver relative to the object or execute a rendezvous with the object;

[0016] identification of objects from aircraft or spacecraft by capturing imagery of the objects and then identifying the objects via image recognition performed on a local or remote computer;

[0017] watching movie previews streamed to a camera-equipped wireless device by "pointing and clicking" with such a device on a movie theatre sign or poster, or on a digital video disc box or videotape box;

[0018] listening to audio recording samples streamed to a camera-equipped wireless device by "pointing and clicking" with such a device on a compact disk (CD) box, videotape box, or print media advertisement;

[0019] purchasing movie, concert, or sporting event tickets by "pointing and clicking" on a theater, advertisement, or other object with a camera-equipped wireless device;

[0020] purchasing an item by "pointing and clicking" on the object with a camera-equipped wireless device and thus initiating a transaction;

[0021] interacting with television programming by "pointing and clicking" at the television screen with a camera-equipped device, thus capturing an image of the screen content and having that image sent to a remote computer and identified, thus initiating interaction based on the screen content received (an example is purchasing an item on the television screen by "pointing and clicking" at the screen when the item is on the screen);

Inter IKEA Systems B.V.
Exhibit 1005 Page 9 of 24

[0022] interacting with a computer-system based game and with other players of the game by "pointing and clicking" on objects in the physical environment that are considered to be part of the game;

[0023] paying a bus fare by "pointing and clicking" with a mobile wireless camera-equipped device, on a fare machine in a bus, and thus establishing a communications link between the device and the fare machine and enabling the fare payment transaction;

[0024] establishment of a communication between a mobile wireless camera-equipped device and a computer with an Internet connection by "pointing and clicking" with the device on the computer and thus providing to the mobile device an Internet address at which it can communicate with the computer, thus establishing communications with the computer despite the absence of a local network or any direct communication between the device and the computer;

[0025] use of a mobile wireless camera-equipped device as a point-of-sale terminal by, for example, "pointing and clicking" on an item to be purchased, thus identifying the item and initiating a transaction;

DISCLOSURE OF INVENTION

[0026] The present invention solves the above stated needs. Once an image is captured digitally, a search of the image determines whether symbolic content is included in the image. If so the symbol is decoded and communication is opened with the proper database, usually using the Internet, wherein the best match for the symbol is returned. In some instances, a symbol may be detected, but non-ambiguous identification is not possible. In that case and when a symbolic image can not be detected, the image is decomposed through identification algorithms where unique characteristics of the image are determined. These characteristics are then used to provide the best match or matches in the data base, the "best" determination being assisted by the partial symbolic information, if that is available.

[0027] Therefore the present invention provides technology and processes that can accommodate linking objects and images to information via a network such as the Internet, which requires no modification to the linked object. Traditional methods for linking objects to digital information, including applying a barcode, radio or optical transceiver or transmitter, or some other means of identification to the object, or modifying the image or object so as to encode detectable information in it, are not required because the image or object can be identified solely by its visual appearance. The users or devices may even interact with objects by "linking" to them. For example, a user may link to a vending machine by "pointing and clicking" on it. His device would be connected over the Internet to the company that owns the vending machine. The company would in turn establish a connection to the vending machine, and thus the user would have a communication channel established with the vending machine and could interact with it.

[0028] The decomposition algorithms of the present invention allow fast and reliable detection and recognition of images and/or objects based on their visual appearance in an image, no matter whether shadows, reflections, partial obscuration, and variations in viewing geometry are present. As stated above, the present invention also can detect, decode, and identify images and objects based on traditional symbols which may appear on the object, such as alphanumeric characters, barcodes, or 2-dimensional matrix codes.

[0029] When a particular object is identified, the position and orientation of an object with respect to the user at the time the image was captured can be determined based on the appearance of the object in an image. This can be the location and/or identity of people scanned by multiple cameras in a security system, a passive locator system more accurate than GPS or usable in areas where GPS signals cannot be received, the location of specific vehicles without requiring a transmission from the vehicle, and many other uses.

[0030] When the present invention is incorporated into a mobile device, such as a portable telephone, the user of the device can link to images and objects in his or her environment by pointing the device at the object of interest, then "pointing and clicking" to capture an image. Thereafter, the device transmits the image to another computer ("Server"), wherein the image is analyzed and the object or image of interest is detected and recognized. Then the network address of information corresponding to that object is transmitted from the ("Server") back to the mobile device, allowing the mobile device to access information using the network address so that only a portion of the information concerning the object need be stored in the systems database.

[0031] Some or all of the image processing, including image/object detection and/or decoding of symbols detected in the image may be distributed arbitrarily between the mobile (Client) device and the Server. In other words, some processing may be performed in the Client device and some in the Server, without specification of which particular processing is performed in each, or all processing may be performed on one platform or the other, or the platforms may be combined so that there is only one platform. The image processing can be implemented in a parallel computing manner, thus facilitating scaling of the system with respect to database size and input traffic loading.

[0032] Therefore, it is an object of the present invention to provide a system and process for identifying digitally captured images without requiring modification to the object.

[0033] Another object is to use digital capture devices in ways never contemplated by their manufacturer.

[0034] Another object is to allow identification of objects from partial views of the object.

[0035] Another object is to provide communication means with operative devices without requiring a public connection therewith.

[0036] These and other objects and advantages of the present invention will become apparent to those skilled in the art after considering the following detailed specification, together with the accompanying drawings wherein:

BRIEF DESCRIPTION OF THE DRAWINGS

[0037] **FIG. 1** is a schematic block diagram top-level algorithm flowchart;

[0038] **FIG. 2** is an idealized view of image capture;

[0039] **FIGS. 3A and 3B** are a schematic block diagram of process details of the present invention;

[0040] **FIG. 4** is a schematic block diagram of a different explanation of invention;

[0041] **FIG. 5** is a schematic block diagram similar to **FIG. 4** for cellular telephone and personal data assistant (PDA**0** applications; and

[0042] **FIG. 6** is a schematic block diagram for spacecraft applications.

BEST MODES FOR CARRYING OUT THE INVENTION

[0043] The present invention includes a novel process whereby information such as Internet content is presented to a user, based solely on a remotely acquired image of a physical object. Although coded information can be included in the remotely acquired image, it is not required since no additional information about a physical object, other than its image, needs to be encoded in the linked object. There is no need for any additional code or device, radio, optical or otherwise, to be embedded in or affixed to the object. Image-linked objects can be located and identified within user-acquired imagery solely by means of digital image processing, with the address of pertinent information being returned to the device used to acquire the image and perform the link. This process is robust against digital image noise and corruption (as can result from lossy image compression/decompression), perspective error, rotation, translation, scale differences, illumination variations caused by different lighting sources, and partial obscuration of the target that results from shadowing, reflection or blockage.

[0044] Many different variations on machine vision "target location and identification" exist in the current art. However, they all tend to provide optimal solutions for an arbitrarily restricted search space. At the heart of the present invention is a high-speed image matching engine that returns unambiguous matches to target objects contained in a wide variety of potential input images. This unique approach to image matching takes advantage of the fact that at least some portion of the target object will be found in the user-acquired image. The parallel image comparison processes embodied in the present search technique are, when taken together, unique to the process. Further, additional refinement of the process, with the inclusion of more and/or different decomposition-parameterization functions, utilized within the overall structure of the search loops is not restricted. The detailed process is described in the following.igure **1** shows the overall processing flow and steps. These steps are described in further detail in the following sections.

[0045] For image capture **10**, the User **12** (**FIG. 2**) utilizes a computer, mobile telephone, personal digital assistant, or other similar device **14** equipped with an image sensor (such as a CCD or CMOS digital camera). The User **12** aligns the sensor of the image capture device **14** with the object **16** of interest. The linking process is then initiated by suitable means including: the User **12** pressing a button on the device **14** or sensor; by the software in the device **14** automatically recognizing that an image is to be acquired; by User voice command; or by any other appropriate means. The device **14** captures a digital image **18** of the scene at which it is pointed. This image **18** is represented as three separate 2-D matrices of pixels, corresponding to the raw RGB (Red, Green, Blue) representation of the input image. For the purposes of standardizing the analytical processes in this embodiment, if the device **14** supplies an image in other than RGB format, a transformation to RGB is accomplished. These analyses could be carried out in any standard color format, should the need arise.

[0046] If the server **20** is physically separate from the device **14**, then user acquired images are transmitted from the device **14** to the Image Processor/Server **20** using a conventional digital network or wireless network means. If the image **18** has been compressed (e.g. via lossy JPEG DCT) in a manner that introduces compression artifacts into the reconstructed image **18**, these artifacts may be partially removed by, for example, applying a conventional despeckle filter to the reconstructed image prior to additional processing.

[0047] The Image Type Determination **26** is accomplished with a discriminator algorithm which operates on the input image **18** and determines whether the input image contains recognizable symbols, such as barcodes, matrix codes, or alphanumeric characters. If such symbols are found, the image **18** is sent to the Decode Symbol **28** process. Depending on the confidence level with which the discriminator algorithm finds the symbols, the image **18** also may or alternatively contain an object of interest and may therefore also or alternatively be sent to the Object Image branch of the process flow. For example, if an input image **18** contains both a barcode and an object, depending on the clarity with which the barcode is detected, the image may be analyzed by both the Object Image and Symbolic Image branches, and that branch which has the highest success in identification will be used to identify and link from the object.

[0048] The image is analyzed to determine the location, size, and nature of the symbols in the Decode Symbol **28**. The symbols are analyzed according to their type, and their content information is extracted. For example, barcodes and alphanumeric characters will result in numerical and/or text information.

[0049] For object images, the present invention performs a "decomposition", in the Input Image Decomposition **34**, of a high-resolution input image into several different types of quantifiable salient parameters. This allows for multiple independent convergent search processes of the database to occur in parallel, which greatly improves image match speed and match robustness in the Database Matching **36**. The Best Match **38** from either the Decode Symbol **28**, or the inage Database Matching **36**, or both, is then determined. If a specific URL (or other online address) is associated with the image, then an URL Lookup **40** is performed and the Internet address is returned by the URL Return **42**.

[0050] The overall flow of the Input Image Decomposition process is as follows:

Radiometric Correction
Segmentation
Segment Group Generation
FOR each segment group
  Bounding Box Generation

-continued

```
Geometric Normalization
Wavelet Decomposition
Color Cube Decomposition
Shape Decomposition
Low-Resolution Grayscale Image Generation
FOR END
```

[0051] Each of the above steps is explained in further detail below. For Radiometric Correction, the input image typically is transformed to an 8-bit per color plane, RGB representation. The RGB image is radiometrically normalized in all three channels. This normalization is accomplished by linear gain and offset transformations that result in the pixel values within each color channel spanning a full 8-bit dynamic range (256 possible discrete values). An 8-bit dynamic range is adequate but, of course, as optical capture devices produce higher resolution images and computers get faster and memory gets cheaper, higher bit dynamic ranges, such as 16-bit, 32-bit or more may be used.

[0052] For Segmentation, the radiometrically normalized RGB image is analyzed for "segments," or regions of similar color, i.e. near equal pixel values for red, green, and blue. These segments are defined by their boundaries, which consist of sets of (x, y) point pairs. A map of segment boundaries is produced, which is maintained separately from the RGB input image and is formatted as an x, y binary image map of the same aspect ratio as the RGB image.

[0053] For Segment Group Generation, the segments are grouped into all possible combinations. These groups are known as "segment groups" and represent all possible potential images or objects of interest in the input image. The segment groups are sorted based on the order in which they will be evaluated. Various evaluation order schemes are possible. The particular embodiment explained herein utilizes the following "center-out" scheme: The first segment group comprises only the segment that includes the center of the image. The next segment group comprises the previous segment plus the segment which is the largest (in number of pixels) and which is adjacent to (touching) the previous segment group. Additional segments are added using the segment criteria above until no segments remain. Each step, in which a new segment is added, creates a new and unique segment group.

[0054] For Bounding Box Generation, the elliptical major axis of the segment group under consideration (the major axis of an ellipse just large enough to contain the entire segment group) is computed. Then a rectangle is constructed within the image coordinate system, with long sides parallel to the elliptical major axis, of a size just large enough to completely contain every pixel in the segment group.

[0055] For Geometric Normalization, a copy of the input image is modified such that all pixels not included in the segment group under consideration are set to mid-level gray. The result is then resampled and mapped into a "standard aspect" output test image space such that the corners of the bounding box are mapped into the corners of the output test image. The standard aspect is the same size and aspect ratio as the Reference images used to create the database.

[0056] For Wavelet Decomposition, a grayscale representation of the full-color image is produced from the geometrically normalized image that resulted from the Geometric Normalization step. The following procedure is used to derive the grayscale representation. Reduce the three color planes into one grayscale image by proportionately adding each R, G, and B pixel of the standard corrected color image using the following formula:

$$L_{x,y}=0.34*R_{x,y}+0.55*G_{x,y}+0.11*B_{x,y}$$

[0057] then round to nearest integer value. Truncate at 0 and 255, if necessary. The resulting matrix L is a standard grayscale image. This grayscale representation is at the same spatial resolution as the full color image, with an 8-bit dynamic range. A multi-resolution Wavelet Decomposition of the grayscale image is performed, yielding wavelet coefficients for several scale factors. The Wavelet coefficients at various scales are ranked according to their weight within the image.

[0058] For Color Cube Decomposition, an image segmentation is performed (see "Segmentation" above), on the RGB image that results from Geometric Normalization. Then the RGB image is transformed to a normalized Intensity, In-phase and Quadrature-phase color image (YIQ). The segment map is used to identify the principal color regions of the image, since each segment boundary encloses pixels of similar color. The average Y, I, and Q values of each segment, and their individual component standard deviations, are computed. The following set of parameters result, representing the colors, color variation, and size for each segment:

[0059] Y_{avg}=Average Intensity

[0060] I_{avg}=Average In-phase

[0061] Q_{avg}=Average Quadrature

[0062] Y_{sigma}=Intensity standard deviation

[0063] I_{sigma}=In-phase standard deviation

[0064] Q_{sigma}=Quadrature standard deviation

[0065] N_{pixels}=number of pixels in the segment

[0066] The parameters comprise a representation of the color intensity and variation in each segment. When taken together for all segments in a segment group, these parameters comprise points (or more accurately, regions, if the standard deviations are taken into account) in a three-dimensional color space and describe the intensity and variation of color in the segment group.

[0067] For Shape Decomposition, the map resulting from the segmentation performed in the Color Cube Generation step is used and the segment group is evaluated to extract the group outer edge boundary, the total area enclosed by the boundary, and its area centroid. Additionally, the net ellipticity (semi-major axis divided by semi-minor axis of the closest fit ellipse to the group) is determined.

[0068] For Low-Resolution Grayscale Image Generation, the full-resolution grayscale representation of the image that was derived in the Wavelet Generation step is now subsampled by a factor in both x and y directions. For the example of this embodiment, a 3:1 subsampling is assumed. The subsampled image is produced by weighted averaging of pixels within each 3×3 cell. The result is contrast binned, by reducing the number of discrete values assignable to each

Inter IKEA Systems B.V.
Exhibit 1005 Page 12 of 24

pixel based upon substituting a "binned average" value for all pixels that fall within a discrete (TBD) number of brightness bins.

[0069] The above discussion of the particular decomposition methods incorporated into this embodiment are not intended to indicate that more, or alternate, decomposition methods may not also be employed within the context of this invention.

[0070] In other words:

```
FOR each input image segment group
  FOR each database object
    FOR each view of this object
      FOR each segment group in this view of this
      database object
        Shape Comparison
        Grayscale Comparison
        Wavelet Comparison
        Color Cube Comparison
        Calculate Combined Match Score
      END FOR
    END FOR
  END FOR
END FOR
```

[0071] Each of the above steps is explained in further detail below.

[0072] For Each Input Image Segment Group

[0073] This loop considers each combination of segment groups in the input image, in the order in which they were sorted in the "Segment Group Generation" step. Each segment group, as it is considered, is a candidate for the object of interest in the image, and it is compared against database objects using various tests.

[0074] One favored implementation, of many possible, for the order in which the segment groups are considered within this loop is the "center-out" approach mentioned previously in the "Segment Group Generation" section. This scheme considers segment groups in a sequence that represents the addition of adjacent segments to the group, starting at the center of the image. In this scheme, each new group that is considered comprises the previous group plus one additional adjacent image segment. The new group is compared against the database. If the new group results in a higher database matching score than the previous group, then new group is retained. If the new group has a lower matching score then the previous group, then it is discarded and the loop starts again. If a particular segment group results in a match score which is extremely high, then this is considered to be an exact match and no further searching is warranted; in this case the current group and matching database group are selected as the match and this loop is exited.

[0075] FOR Each Database Object

[0076] This loop considers each object in the database for comparison against the current input segment group.

[0077] FOR Each View of This Object

[0078] This loop considers each view of the current database object, for comparison against the current input segment group. The database contains, for each object, multiple views from different viewing angles.

[0079] FOR Each Segment Group In This View of this Database Object

[0080] This loop considers each combination of segment groups in the current view of the database object. These segment groups were created in the same manner as the input image segment groups.

[0081] Shape Comparison

[0082] Inputs:

[0083] For the input image and all database images:

[0084] I. Segment group outline*

[0085] II. Segment group area

[0086] III. Segment group centroid location

[0087] IV. Segment group bounding ellipse ellipticity

[0088] Algorithm:

[0089] V. Identify those database segment groups with an area approximately equal to that of the input segment group, within TBD limits, and calculate an area matching score for each of these "matches."

[0090] VI. Within the set of matches identified in the previous step, identify those database segment groups with an ellipticity approximately equal to that of the input segment group, within TBD limits, and calculate an ellipticity position matching score for each of these "matches."

[0091] VII. Within the set of matches identified in the previous step, identify those database segment groups with a centroid position approximately equal to that of the input segment group, within TBD limits, and calculate a centroid position matching score for each of these "matches."

[0092] VIII. Within the set of matches identified in the previous step, identify those database segment groups with an outline shape approximately equal to that of the input segment group, within TBD limits, and calculate an outline matching score for each of these "matches." This is done by comparing the two outlines and analytically determining the extent to which they match.

[0093] Note: this algorithm need not necessarily be performed in the order of Steps 1 to 4. It could alternatively proceed as follows:

```
FOR each database segment group
  IF the group passes Step 1
    IF the group passes Step 2
      IF the group passes Step 3
        IF the group passes Step 4
          Successful comparison, save
          result
        END IF
      END IF
    END IF
  END IF
END FOR
```

[0094] Grayscale Comparison

[0095] Inputs:

[0096] For the input image and all database images:

[0097] IX. Low-resolution, normalized, contrast-binned, grayscale image of pixels within segment group bounding box, with pixels outside of the segment group set to a standard background color.

[0098] Algorithm:

[0099] Given a series of concentric rectangular "tiers" of pixels within the low-resolution images, compare the input image pixel values to those of all database images. Calculate a matching score for each comparison and identify those database images with matching scores within TBD limits, as follows:

```
FOR each database image
  FOR each tier, starting with the innermost and
  progressing to the outermost
    Compare the pixel values between the input and
    database image
    Calculate an aggregate matching score
    IF matching score is greater than some TBD limit
    (i.e., close match)
      Successful comparison, save result
    END IF
  END FOR
END FOR
```

[0100] Wavelet Comparison

[0101] Inputs:

[0102] For the input image and all database images:

[0103] X. Wavelet coefficients from high-resolution grayscale image within segment group bounding box.

[0104] Algorithm:

[0105] Successively compare the wavelet coefficients of the input segment group image and each database segment group image, starting with the lowest-order coefficients and progressing to the highest order coefficients. For each comparison, compute a matching score. For each new coefficient, only consider those database groups that had matching scores, at the previous (next lower order) coefficient within TBD limits.

```
FOR each database image
  IF input image C0 equals database image C0 within TBD
  limit
    IF input image C1 equals database image C1 within
    TBD limit
      ...
      IF input image CN equals database image CN
      within TBD limit
        Close match, save result and match
        score
      END IF
      ...
    END IF
  END IF
END FOR
```

[0106] Notes:

[0107] I. "C_i" are the wavelet coefficients, with C_0 being the lowest order coefficient and C_N being the highest.

[0108] II. When the coefficients are compared, they are actually compared on a statistical (e.g. Gaussian) basis, rather than an arithmetic difference.

[0109] III. Data indexing techniques are used to allow direct fast access to database images according to their C_i values. This allows the algorithm to successively narrow the portions of the database of interest as it proceeds from the lowest order terms to the highest.

[0110] Color Cube Comparison

[0111] Inputs:

[0112] [Y_{avg}, I_{avg}, Q_{avg}, Ysigma, I_{sigma}, Q_{sigma}, Npixels] data sets ("Color Cube Points") for each segment in:

[0113] I. The input segment group image

[0114] II. Each database segment group image

[0115] Algorithm:

```
FOR each database image
  FOR each segment group in the database image
    FOR each Color Cube Point in database segment group,
    in order of descending Npixels value
      IF Gaussian match between input (Y, I, Q) and
      database (Y, I, Q)
        I. Calculate match score for this segment
        II. Accumulate segment match score into
          aggregate match score for segment group
        III. IF aggregate matching score is greater than
          some TBD limit (i.e., close match)
            Successful comparison, save result
      END IF
    END FOR
  END FOR
END FOR
```

Notes:
I. The size of the Gaussian envelope about any Y, I, Q point is determined by RSS of standard deviations of Y, I, and Q for that point.

[0116] Calculate Combined Match Score

[0117] The four Object Image comparisons (Shape Comparison, Grayscale Comparison, Wavelet Comparison, Color Cube Comparison) each return a normalized matching score. These are independent assessments of the match of salient features of the input image to database images. To minimize the effect of uncertainties in any single comparison process, and to thus minimize the likelihood of returning a false match, the following root sum of squares relationship is used to combine the results of the individual comparisons into a combined match score for an image: CurrentMatch=SQRT-($W_{oc}M_{oc}^2+W_{ccc}M_{ccc}^2+W_{wc}M_{wc}^2+W_{SGC}M_{SGC}^2$), where Ws are TBD parameter weighting coefficients and Ms are the individual match scores of the four different comparisons.

[0118] The unique database search methodology and subsequent object match scoring criteria are novel aspects of the present invention that deserve special attention. Each decomposition of the Reference image and Input image regions represent an independent characterization of salient characteristics of the image. The Wavelet Decomposition, Color Cube Decomposition, Shape Decomposition, and

Inter IKEA Systems B.V.
Exhibit 1005 Page 14 of 24

evaluation of a sub-sampled low-resolution Grayscale representation of an input image all produce sets of parameters that describe the image in independent ways. Once all four of these processes are completed on the image to be tested, the parameters provided by each characterization are compared to the results of identical characterizations of the Reference images, which have been previously calculated and stored in the database. These comparisons, or searches, are carried out in parallel. The result of each search is a numerical score that is a weighted measure of the number of salient characteristics that "match" (i.e. that are statistically equivalent). Near equivalencies are also noted, and are counted in the cumulative score, but at a significantly reduced weighting.

[0119] One novel aspect of the database search methodology in the present invention is that not only are these independent searches carried out in parallel, but also, all but the low-resolution grayscale compares are "convergent." By convergent, it is meant that input image parameters are searched sequentially over increasingly smaller subsets of the entire database. The parameter carrying greatest weight from the input image is compared first to find statistical matches and near-matches in all database records. A normalized interim score (e.g., scaled value from zero to one, where one is perfect match and zero is no match) is computed, based on the results of this comparison. The next heaviest weighted parameter from the input image characterization is then searched on only those database records having initial interim scores above a minimum acceptable threshold value. This results in an incremental score that is incorporated into the interim score in a cumulative fashion. Then, subsequent compares of increasingly lesser-weighted parameters are assessed only on those database records that have cumulative interim scores above the same minimum acceptable threshold value in the previous accumulated set of tests.

[0120] This search technique results in quick completion of robust matches, and establishes limits on the domain of database elements that will be compared in a subsequent combined match calculation and therefore speeds up the process. The convergent nature of the search in these comparisons yields a ranked subset of the entire database.

[0121] The result of each of these database comparisons is a ranking of the match quality of each image, as a function of decomposition search technique. Only those images with final cumulative scores above the acceptable match threshold will be assessed in the next step, a Combined Match Score evaluation.

[0122] Four database comparison processes, Shape Comparison, Grayscale Comparison, Wavelet Comparison, and Color Cube Comparison, are performed. These processes may occur sequentially, but generally are preferably performed in parallel on a parallel computing platform. Each comparison technique searches the entire image database and returns those images that provide the best matches, for the particular algorithm, along with the matching scores for these images. These comparison algorithms are performed on segment groups, with each input image segment group being compared to each segment group for each database image.

[0123] **FIGS. 3A and 3B** show the process flow within the Database Matching operation. The algorithm is presented here as containing four nested loops with four parallel processes inside the innermost loop. This structure is for presentation and explanation only. The actual implementation, although performing the same operations at the innermost layer, can have a different structure in order to achieve the maximum benefit from processing speed enhancement techniques such as parallel computing and data indexing techniques. It is also important to note that the loop structures can be implemented independently for each inner comparison, rather than the shared approach shown in the **FIGS. 3A and 3B**.

[0124] Preferably, parallel processing is used to divide tasks between multiple CPUs (Central Processing Units) and/or computers. The overall algorithm may be divided in several ways, such as:

Sharing the Outer Loop:	In this technique, all CPUs run the entire algorithm, including the outer loop, but one CPU runs the loop for the first N cycles, another CPU for the second N cycles, all simultaneously.
Sharing the Comparisons:	In this technique, one CPU performs the loop functions. When the comparisons are performed, they are each passed to a separate CPU to be performed in parallel.
Sharing the Database:	This technique entails splitting database searches between CPUs, so that each CPU is responsible for searching one section of the database, and the sections are searched in parallel by multiple CPUs. This is, in essence, a form of the "Sharing the Outer Loop" technique described above.

[0125] Actual implementations can be some combination of the above techniques that optimizes the process on the available hardware.

[0126] Another technique employed to maximize speed is data indexing. This technique involves using a priori knowledge of where data resides to only search in those parts of the database that contain potential matches. Various forms of indexing may be used, such as hash tables, data compartmentalization (i.e., data within certain value ranges are stored in certain locations), data sorting, and database table indexing. An example of such techniques is, in the Shape Comparison algorithm (see below), if a database is to be searched for an entry with an Area with a value of A, the algorithm would know which database entries or data areas have this approximate value and would not need to search the entire database.

[0127] Another technique employed is as follows. **FIG. 4** shows a simplified configuration of the invention. Boxes with solid lines represent processes, software, physical objects, or devices. Boxes with dashed lines represent information. The process begins with an object of interest: the target object **100**. In the case of consumer applications, the target object **100** could be, for example, beverage can, a music CD box, a DVD video box, a magazine advertisement, a poster, a theatre, a store, a building, a car, or any other object that user is interested in or wishes to interact with. In security applications the target object **100** could be, for example, a person, passport, or driver's license, etc. In industrial applications the target object **100** could be, for example, a part in a machine, a part on an assembly line, a box in a warehouse, or a spacecraft in orbit, etc.

Inter IKEA Systems B.V.
Exhibit 1005 Page 15 of 24

[0128] The terminal **102** is a computing device that has an "image" capture device such as digital camera **103**, a video camera, or any other device that an convert a physical object into a digital representation of the object. The imagery can be a single image, a series of images, or a continuous video stream. For simplicity of explanation this document describes the digital imagery generally in terms of a single image, however the invention and this system can use all of the imagery types described above.

[0129] After the camera **103** captures the digital imagery of the target object **100**, image preprocessing **104** software converts the digital imagery into image data **105** for transmission to and analysis by an identification server **106**. Typically a network connection is provided capable of providing communications with the identification server **106**. Image data **105** is data extracted or converted from the original imagery of the target object **100** and has information content appropriate for identification of the target object **100** by the object recognition **107**, which may be software or hardware. Image data **105** can take many forms, depending on the particular embodiment of the invention.

[0130] Examples of image data **105** are:

[0131] Compressed (e.g., JPEG2000) form of the raw imagery from camera **103**;

[0132] Key image information, such as spectral and/or spatial frequency components (e.g. wavelet components) of the raw imagery from camera **103**; and

[0133] MPEG video stream created from the raw imagery from camera **103**;

[0134] The particular form of the image data **105** and the particular operations performed in image preprocessing **104** depend on:

[0135] Algorithm and software used in object recognition **107** Processing power of terminal **102**;

[0136] Network connection speed between terminal **102** and identification server **106**;

[0137] Application of the System; and

[0138] Required system response time;

[0139] In general, there is a tradeoff between the network connection speed (between terminal **102** and identification server **106**) and the processing power of terminal **102**. The results all of the above tradeoffs will define the nature of image preprocessing **104** and image data **105** for a specific embodiment. For example, image preprocessing **104** could be image compression and image data **105** compressed imagery, or image preprocessing **104** could be wavelet analysis and image data **105** could be wavelet coefficients.

[0140] The image data **105** is sent from the terminal **102** to the identification server **106**. The identification server **106** receives the image data **105** and passes it to the object recognition **107**.

[0141] The identification server **106** is a set of functions that usually will exist on computing platform separate from the terminal **102**, but could exist on the same computing platform. If the identification server **106** exists on a separate computing device, such as a computer in a data center, then the transmission of the image components **105** to the identification server **106** is accomplished via a network or combination of networks, such a cellular telephone network, wireless Internet, Internet, and wire line network. If the identification server **106** exists on the same computing device as the terminal **102** then the transmission consists simply of a transfer of data from one software component or process to another.

[0142] Placing the identification server **106** on a computing platform separate from the terminal **102** enables the use of powerful computing resources for the object recognition **107** and database **108** functions, thus providing the power of these computing resources to the terminal **102** via network connection. For example, an embodiment that identifies objects out of a database of millions of known objects would be facilitated by the large storage, memory capacity, and processing power available in a data center; it may not be feasible to have such computing power and storage in a mobile device. Whether the terminal **102** and the identification server **106** are on the same computing platform or separate ones is an architectural decision that depends on system response time, number of database records, image recognition algorithm computing power and storage available in terminal **102**, etc., and this decision must be made for each embodiment of the invention. Based on current technology, in most embodiments these functions will be on separate computing platforms.

[0143] The overall function of the identification server **106** is to determine and provide the target object information **109** corresponding to the target object **100**, based on the image data **105**.

[0144] The object recognition **107** and the database **108** function together to:

[0145] 1. Detect, recognize, and decode symbols, such as barcodes or text, in the image.

[0146] 2. Recognize the object (the target object **100**) in the image.

[0147] 3. Provide the target object information **109** that corresponds to the target object **100**. The target object information **109** usually (depending on the embodiment) includes an information address corresponding to the target object **100**.

[0148] The object recognition **107** detects and decodes symbols, such as barcodes or text, in the input image. This is accomplished via algorithms, software, and/or hardware components suited for this task. Such components are commercially available (The HALCON software package from MVTec is an example). The object recognition **107** also detects and recognizes images of the target object **100** or portions thereof. This is accomplished by analyzing the image data **105** and comparing the results to other data, representing images of a plurality of known objects, stored in the database **108**, and recognizing the target object **100** if a representation of target object **100** is stored in the database **108**.

[0149] In some embodiments the terminal **102** includes software, such as a web browser (the browser **110**), that receives an information address, connects to that information address via a network or networks, such as the Internet, and exchanges information with another computing device

at that information address. In consumer applications the terminal **102** may be a portable cellular telephone or Personal Digital Assistant equipped with a camera **103** and wireless Internet connection. In security and industrial applications the terminal **102** may be a similar portable hand-held device or may be fixed in location and/or orientation, and may have either a wireless or wire line network connection.

[**0150**] Other object recognition techniques also exist and include methods that store 3-dimensional models (rather than 2-dimensional images) of objects in a database and correlate input images with these models of the target object is performed by an object recognition technique of which many are available commercially and in the prior art. Such object recognition techniques usually consist of comparing a new input image to a plurality of known images and detecting correspondences between the new input image and one of more of the known images. The known images are views of known objects from a plurality of viewing angles and thus allow recognition of 2-dimensional and 3-dimensional objects in arbitrary orientations relative to the camera **103**.

[**0151**] **FIG. 4** shows the object recognition **107** and the database **108** as separate functions for simplicity. However, in many embodiments the object recognition **107** and the database **108** are so closely interdependent that they may be considered a single process.

[**0152**] There are various options for the object recognition technique and the particular processes performed within the object recognition **107** and the database **108** depend on this choice. The choice depends on the nature, requirements, and architecture of the particular embodiment of the invention. However, most embodiments will usually share most of the following desired attributes of the image recognition technique:

[**0153**] Capable of recognizing both 2-dimensional (i.e., flat) and 3-dimensional objects;

[**0154**] Capable of discriminating the target object **100** from any foreground or background objects or image information, i.e., be robust with respect to changes in background;

[**0155**] Fast;

[**0156**] Autonomous (no human assistance required in the recognition process);

[**0157**] Scalable; able to identify objects from a large database of known objects with short response time; and

[**0158**] Robust with respect to:

[**0159**] Affine transformations (rotation, translation, scaling);

[**0160**] Non-affine transformations (stretching, bending, breaking);

[**0161**] Occlusions (of the target object **100**);

[**0162**] Shadows (on the target object **100**);

[**0163**] Reflections (on the target object **100**);

[**0164**] Variations in light color temperature;

[**0165**] Image noise;

[**0166**] Capable of determining position and orientation of the target object **100** in the original imagery; and

[**0167**] Capable of recognizing individual human faces from a database containing data representing a large plurality of human faces.

[**0168**] All of these attributes do not apply to all embodiments. For example, consumer linking embodiments generally do not require determination of position and orientation of the target object **100**, while a spacecraft target position and orientation determination system generally would not be required to identify human faces or a large number of different objects.

[**0169**] It is usually desirable that the database **108** be scalable to enable identification of the target object **100** from a very large plurality (for example, millions) of known objects in the database **108**. The algorithms, software, and computing hardware must be designed to function together to quickly perform such a search. An example software technique for performing such searching quickly is to use a metric distance comparison technique for comparing the image data **105** to data stored in the database **108**, along with database clustering and multiresolution distance comparisons. This technique is described in "Fast Exhaustive Multi-Resolution Search Algorithm Based on Clustering for Efficient Image Retrieval," by Song, Kim, and Ra, 2000.

[**0170**] In addition to such software techniques, a parallel processing computing architecture may be employed to achieve fast searching of large databases. Parallel processing is particularly important in cases where a non-metric distance is used in object recognition **107**, because techniques such database clustering and multiresolution search may not be possible and thus the complete database must be searched by partitioning the database across multiple CPUs.

[**0171**] As described above, the object recognition **107** can also detect identifying marks on the target object **100**. For example, the target object **100** may include an identifying number or a barcode. This information can be decoded and used to identify or help identify the target object **100** in the database **108**. This information also can be passed on as part of the target object information **109**. If the information is included as part of the target object information **109** then it can be used by the terminal **102** or content server **111** to identify the specific target object **100**, out of many such objects that have similar appearance and differ only in the identifying marks. This technique is useful, for example, in cases where the target object **100** is an active device with a network connection (such as a vending machine) and the content server establishes communication with the target object **100**. A combination with a Global Positioning System can aloso be used to identify like objects by their location.

[**0172**] The object recognition **107** may be implemented in hardware, software, or a combination of both. Examples of each category are presented below.

[**0173**] Hardware object recognition implementations include optical correlators, optimized computing platforms, and custom hardware.

[**0174**] Optical correlators detect objects in images very rapidly by, in effect, performing image correlation calculations with light. Examples of optical correlators are:

Inter IKEA Systems B.V.
Exhibit 1005 Page 17 of 24

[0175] Litton Miniaturized Ruggedized Optical Correlator, from Northrop Grumman Corp;

[0176] Hybrid Digital/Optical Correlator, from the School of Engineering and Information Technology, University of Sussex, UK; and

[0177] OC-VGA3000 and OC-VGA6000 Optical Correlators from INO, Quebec, Canada.

[0178] Optimized computing platforms are hardware computing systems, usually on a single board, that are optimized to perform image processing and recognition algorithms very quickly. These platforms must be programmed with the object recognition algorithm of choice. Examples of optimized computing platforms are

[0179] VIP/Balboa™ Image Processing Board, from Irvine Sensors Corp.; and

[0180] 3DANN™-R Processing System, from Irvine Sensors Corp.

[0181] Image recognition calculations can also be implemented directly in custom hardware in forms such as Application Specific Integrated Circuits (ASICs), Field Programmable Gate Arrays (FPGAs), and Digital Signal Processors (DSPs).

[0182] There are many object and image recognition software applications available commercially and many algorithms published in the literature. Examples of commercially available image/object recognition software packages include:

[0183] Object recognition system, from Sandia National Laboratories;

[0184] Object recognition perception modules, from Evolution Robotics;

[0185] ImageFinder, from Attrasoft;

[0186] ImageWare, from Roz Software Systems; and

[0187] ID-2000, from Imagis Technologies.

[0188] Some of the above recognition systems include 3-dimensional object recognition capability while others perform 2-dimensional image recognition. The latter type are used to perform 3-dimensional object recognition by comparing input images to a plurality of 2-dimensional views of objects from a plurality of viewing angles.

[0189] Examples of object recognition algorithms in the literature and intended for implementation in software are:

[0190] Distortion Invariant Object Recognition in the Dynamic Link Architecture, Lades et al, 1993;

[0191] SEEMORE: Combining Color, Shape, and Texture Histogramming in a Neurally Inspired Approach to Visual Object Recognition, Mel, 1996;

[0192] Probabilistic Affine Invariants for Recognition, Leung et al, 1998;

[0193] Software Library for Appearance Matching (SLAM), Nene at al, 1994;

[0194] Probabilistic Models of Appearance for 3-D Object Recognition, Pope & Lowe, 2000;

[0195] Matching 3D Models with Shape Distributions, Osada et al, 2001;

[0196] Finding Pictures of Objects in Large Collections of Images, Forsyth et al, 1996;

[0197] The Earth Mover's Distance under Transformation Sets, Cohen & Guibas, 1999;

[0198] Object Recognition from Local Scale-Invariant Features, Lowe, 1999; and

[0199] Fast Object Recognition in Noisy Images Using Simulated Annealing, Betke & Makris, 1994.

[0200] Part of the current invention is the following object recognition algorithm specifically designed to be used as the object recognition **107** and, to some extent, the database **108**. This algorithm is robust with respect to occlusions, reflections, shadows, background/foreground clutter, object deformation and breaking, and is scalable to large databases. The task of the algorithm is to find an object or portion thereof in an input image, given a database of multiple objects with multiple views (from different angles) of each object.

[0201] This algorithm uses the concept of a Local Image Descriptor (LID) to summarize the information in a local region of an image. A LID is a circular subset, or "cutout," of a portion of an image. There are various formulations for LIDs; two examples are:

[0202] LID Formulation 1

[0203] The area within the LID is divided into range and angle bins. The average color in each [range, angle] bin is calculated from the pixel values therein.

[0204] LID Formulation 2

[0205] The area within the LID is divided into range bins. The color histogram values within each range bin are calculated from the pixel values therein. For each range bin, a measure of the variation of color with angle is calculated as, for example, the sum of the changes in average color between adjacent small angular slices of a range bin.

[0206] A LID in the input image is compared to a LID in a database image by a comparison technique such the L1 Distance, L2 Distance, Unfolded Distance, Earth Mover Distance, or cross-correlation. Small distances indicate a good match between the portions of the images underlying the LIDs. By iteratively changing the position and size of the LIDs in the input and database images the algorithm converges on the best match between circular regions in the 2 images.

[0207] Limiting the comparisons to subsets (circular LIDs) of the images enables the algorithm to discriminate an object from the background. Only LIDs that fall on the object, as opposed to the background, yield good matches with database images. This technique also enable matching of partially occluded objects; a LID that falls on the visible part of an occluded object will match to a LID in the corresponding location in the database image of the object.

[0208] The iteration technique used to find the best match is simulated annealing, although genetic search, steepest descent, or other similar techniques appropriate for multivariable optimization can also be used individually or in combination with simulated annealing. Simulated annealing is modeled after the concept of a molten substance cooling

Inter IKEA Systems B.V.
Exhibit 1005 Page 18 of 24

and solidifying into a solid. The algorithm starts at a given temperature and then the temperature is gradually reduced with time. At each time step, the values of the search variables are perturbed from the their previous values to a create a new "child" generation of LIDs. The perturbations are calculated statistically and their magnitudes are functions of the temperature. As the temperature decreases the perturbations decrease in size. The child LIDs, in the input and database images, are then compared. If the match is better than that obtained with the previous "parent" generation, then a statistical decision is made regarding to whether to accept or reject the child LIDs as the current best match. This is a statistical decision that is a function of both the match distance and the temperature. The probability of child acceptance increases with temperature and decreases with match distance. Thus, good matches (small match distance) are more likely to be accepted but poor matches can also be accepted occasionally. The latter case is more likely to occur early in the process when the temperature is high. Statistical acceptance of poor matches is included to allow the algorithm to "jump" out of local minima.

[0209] When LID Formulation 1 is used, the rotation angle of the LID need not necessarily be a simulated annealing search parameter. Faster convergence can be obtained by performing a simple step-wise search on rotation to find the best orientation (within the tolerance of the step size) within each simulated annealing time step.

[0210] The search variables, in both the input and database images, are:

[0211] LID x-position;

[0212] LID y-position;

[0213] LID radius;

[0214] LID x-stretch;

[0215] LID y-stretch; and

[0216] LID orientation angle (only for LID Formulation 1).

[0217] LID x-stretch and LID y-stretch are measures of "stretch" distortion applied to the LID circle, and measure the distortion of the circle into an oval. This is included to provide robustness to differences in orientation and curvature between the input and database images.

[0218] The use of multiple simultaneous LIDs provides additional robustness to occlusions, shadows, reflections, rotations, deformations, and object breaking. The best matches for multiple input image LIDs are sought throughout the database images. The input image LIDs are restricted to remain at certain minimum separation distances from each other. The minimum distance between any 2 LIDs centers is a function of the LID radii. The input image LIDs converge and settle on the regions of the input image having the best correspondence to any regions of any database images. Thus the LIDs behave in the manner of marbles rolling towards the lowest spot on a surface, e.g., the bottom of a bowl, but being held apart by their radius (although LIDs generally have minimum separation distances that are less than their radii).

[0219] In cases where the object in the input image appears deformed or curved relative to the known configuration in which it appears in the database, multiple input image LIDs will match to different database images. Each input image LID will match to that database image which shows the underlying portion of the object as it most closely resembles the input image. If the input image object is bent, e.g., a curved poster, then one part will match to one database orientation and another part will match to a different orientation.

[0220] In the case where the input image object appears to be broken into multiple pieces, either due to occlusion or to physical breakage, use of multiple LIDs again provides robust matching: individual LIDs "settle" on portions of the input image object as they match to corresponding portions of the object in various views in the database.

[0221] Robustness with respect to shadows and reflections is provided by LIDs simply not detecting good matches on these input image regions. They are in effect accommodated in the same manner as occlusions.

[0222] Robustness with respect to curvature and bending is accommodated by multiple techniques. First, use of multiple LIDs provides such robustness as described above. Secondly, curvature and bending robustness is inherently provided to some extent within each LID by use of LID range bin sizes that increase with distance from the LID center (e.g., logarithmic spacing). Given matching points in an input image and database image, deformation of the input image object away from the plane tangent at the matching point increases with distance from the matching point. The larger bin sizes of the outer bins (in both range and angle) reduce this sensitivity because they are less sensitive to image shifts.

[0223] Robustness with respect to lighting color temperature variations is provided by normalization of each color channel within each LID.

[0224] Fast performance, particular with large databases, can be obtained through several techniques, as follows:

[0225] 1. Use of LID Formulation 2 can reduce the amount of search by virtue of being rotationally invariant, although this comes at the cost of some robustness due to loss of image information.

[0226] 2. If a metric distance (e.g., L1, L2, or Unfolded) is used for LID comparison, then database clustering, based on the triangle inequality, can be used to rule out large portions of the database from searching. Since database LIDs are created during the execution of the algorithm, the run-time database LIDs are not clustered. Rather, during preparation of the database, sample LIDs are created from the database images by sampling the search parameters throughout their valid ranges. From this data, bounding clusters can be created for each image and for portions of images. With this information the search algorithm can rule out portions of the search parameter space.

[0227] 3. If a metric distance is used, then progressive multiresolution search can be used. This technique saves time by comparing data first at low resolution and only proceeds with successive higher-resolution comparison on candidates with correlations better than the current best match. A discussion of this technique, along with database clustering, can be found in "Fast Exhaustive Multi-Resolution Search Algorithm Based on Clustering for Efficient Image Retrieval,"[1] by Song et al, **2000**.

Inter IKEA Systems B.V.
Exhibit 1005 Page 19 of 24

[0228] 4. The parameter search space and number of LIDs can be limited. Bounds can be placed, for example, on the sizes of LIDs depending on the expected sizes of input image objects relative to those in the database. A small number of LIDs, even **1**, can be used, at the expense of some robustness.

[0229] 5. LIDs can be fixed in the database images. This eliminates iterative searching on database LID parameters, at the expense of some robustness.

[0230] 6. The "x-stretch" and "y-stretch" search parameters can be eliminated, although there is a trade-off between these search parameters and the number of database images. These parameters increase the ability to match between images of the same object in different orientations. Elimination of these parameters may require more database images with closer angular spacing, depending on the particular embodiment.

[0231] 7. Parallel processing can be utilized to increase computing power.

[0232] This technique is similar to that described by Betke & Makris in "Fast Object Recognition in Noisy Images Using Simulated Annealing", **1994**, with the following important distinctions:

[0233] The current algorithm is robust with respect to occlusion. This is made possible by varying size and position of LIDs in database images, during the search process, in order to match non-occluded portions of database images.

[0234] The current algorithm can identify 3-dimensional objects by containing views of objects from many orientations in the database.

[0235] The current algorithm uses database clustering to enable rapid searching of large databases.

[0236] The current algorithm uses circular LIDs.

[0237] In addition to containing image information, the database **108** also contains address information. After the target object **100** has been identified, the database **108** is searched to find information corresponding to the target object **100**. This information can be an information address, such as an Internet URL. The identification server **106** then sends this information, in the form of the target object information **109**, to the terminal **102**. Depending on the particular embodiment of the invention, the target object information **109** may include, but not be limited to, one or more of the following items of information pertaining to the target object **100**:

[0238] Information address (e.g., Internet URL)

[0239] Identity (e.g., object name, number, classification, etc.);

[0240] Position;

[0241] Orientation;

[0242] Size;

[0243] Color;

[0244] Status;

[0245] Information decoded from and/or referenced by symbols (e.g. information coded in a barcode or a URL referenced by such a barcode); and

[0246] Other data (e.g. alphanumerical text)

[0247] Thus, the identification server determines the identity and/or various attributes of the target object **100** from the image data **105**.

[0248] The target object information **109** is sent to the terminal **102**. This information usually flows via the same communication path used to send the image data **105** from the terminal **102** to the identification server **106**, but this is not necessarily the case. This method of this flow information depends on the particular embodiment of the invention.

[0249] The terminal **102** receives the target object information **109**. The terminal **102** then performs some action or actions based on the target object information **109**. This action or actions may include, but not be limited to:

[0250] Accessing a web site.

[0251] Accessing or initiating a software process on the terminal **102**.

[0252] Accessing or initiating a software process on another computer via a network or networks such as the Internet.

[0253] Accessing a web service (a software service accessed via the Internet).

[0254] Initiating a telephone call (if the terminal **102** includes such capability) to a telephone number that may be included in or determined by the target object Information, may be stored in the terminal **102**, or may be entered by the user.

[0255] Initiating a radio communication (if the terminal **102** includes such capability) using a radio frequency that may be included in or determined by the target object Information, may be stored in the terminal **102**, or may be entered by the user.

[0256] Sending information that is included in the target object information **109** to a web site, a software process (on another computer or on the terminal **102**), or a hardware component.

[0257] Displaying information, via the screen or other visual indication, such as text, graphics, animations, video, or indicator lights.

[0258] Producing an audio signal or sound, including playing music.

[0259] In many embodiments, the terminal **102** sends the target object information **109** to the browser **110**. The browser **110** may or may not exist in the terminal **102**, depending on the particular embodiment of the invention. The browser **110** is a software component, hardware component, or both, that is capable of communicating with and accessing information from a computer at an information address contained in target object information **109**.

[0260] In most embodiments the browser **110** will be a web browser, embedded in the terminal **102**, capable of accessing and communicating with web sites via a network or networks such as the Internet. In some embodiments, however, such as those that only involve displaying the identity, position, orientation, or status of the target object **100**, the browser **110** may be a software component or

application that displays or provides the target object information **109** to a human user or to another software component or application.

[**0261**] In embodiments wherein the browser **110** is a web browser, the browser **110** connects to the content server **111** located at the information address (typically an Internet URL) included in the target object information **109**. This connection is effected by the terminal **102** and the browser **110** acting in concert. The content server **111** is an information server and computing system. The connection and information exchanged between the terminal **102** and the content server **111** generally is accomplished via standard Internet and wireless network software, protocols (e.g. HTTP, WAP, etc.), and networks, although any information exchange technique can be used. The physical network connection depends on the system architecture of the particular embodiment but in most embodiments will involve a wireless network and the Internet. This physical network will most likely be the same network used to connect the terminal **102** and the identification server **106**.

[**0262**] The content server **111** sends content information to the terminal **102** and browser **110**. This content information usually is pertinent to the target object **100** and can be text, audio, video, graphics, or information in any form that is usable by the browser **110** and terminal **102**. The terminal **102** and browser **110** send, in some embodiments, additional information to the content server **111**. This additional information can be information such as the identity of the user of the terminal **102** or the location of the user of the terminal **102** (as determined from a GPS system or a radio-frequency ranging system). In some embodiments such information is provided to the content server by the wireless network carrier.

[**0263**] The user can perform ongoing interactions with the content server **111**. For example, depending on the embodiment of the invention and the applications, the user can:

[**0264**] Listen to streaming audio samples if the target object **100** is an audio recording (e.g., compact audio disc).

[**0265**] Purchase the target object **100** via on-line transaction, with the purchase amount billed to an account linked to the terminal **102**, to the individual user, to a bank account, or to a credit card.

[**0266**] In some embodiments the content server **111** may reside within the terminal **102**. In such embodiments, the communication between the terminal **102** and the content server **111** does not occur via a network but rather occurs within the terminal **102**.

[**0267**] In embodiments wherein the target object **100** includes or is a device capable of communicating with other devices or computers via a network or networks such as the Internet, and wherein the target object information **109** includes adequate identification (such as a sign, number, or barcode) of the specific target object **100**, the content server **111** connects to and exchanges information with the target object **100** via a network or networks such as the Internet. In this type of embodiment, the terminal **102** is connected to the content server **111** and the content server **111** is connected to the target object **100**. Thus, the terminal **102** and target object **100** can communicate via the content server **111**. This enables the user to interact with the target object **100** despite the lack of a direct connection between the target object **100** and the terminal **102**.

[**0268**] The following are examples of embodiments of the invention.

[**0269**] **FIG. 5** shows a preferred embodiment of the invention that uses a cellular telephone, PDA, or such mobile device equipped with computational capability, a digital camera, and a wireless network connection, as the terminal **202** corresponding to the terminal **102** in **FIG. 4**. In this embodiment, the terminal **202** communicates with the identification server **206** and the content server **211** via networks such as a cellular telephone network and the Internet.

[**0270**] This embodiment can be used for applications such as the following ("User" refers to the person operating the terminal **202**, and the terminal **202** is a cellular telephone, PDA, or similar device, and "point and click" refers to the operation of the User capturing imagery of the target object **200** and initiating the transfer of the image data **205** to the identification server **206**).

[**0271**] The User "points and clicks" the terminal **202** at a compact disc (CD) containing recorded music or a digital video disc (DVD) containing recorded video. The terminal **202** browser connects to the URL corresponding to the CD or DVD and displays a menu of options from which the user can select. From this menu, the user can listen to streaming audio samples of the CD or streaming video samples of the DVD, or can purchase the CD or DVD.

[**0272**] The User "points and clicks" the terminal **202** at a print media advertisement, poster, or billboard advertising a movie, music recording, video, or other entertainment. The browser **210** connects to the URL corresponding to the advertised item and the user can listen to streaming audio samples, purchase streaming video samples, obtain show times, or purchase the item or tickets.

[**0273**] The User "points and clicks" the terminal **202** at a television screen to interact with television programming in real-time. For example, the programming could consist of a product promotion involving a reduced price during a limited time. Users that "point and click" on this television programming during the promotion are linked to a web site at which they can purchase the product at the promotional price. Another example is a interactive television programming in which users "point and click" on the television screen at specific times, based on the on-screen content, to register votes, indicate actions, or connect to a web site through which they perform real time interactions with the on-screen program.

[**0274**] The User "points and clicks" on an object such as a consumer product, an advertisement for a product, a poster, etc., the terminal **202** makes a telephone call to the company selling the product, and the consumer has a direct discussion with a company representative regarding the company's product or service. In this case the company telephone number is included in the target object information **209**. If the target object information **209** also includes the company URL then the User can interact with the company via both voice and Internet (via browser **210**) simultaneously.

[**0275**] The User "points and clicks" on a vending machine (target object **200**) that is equipped with a connection to a

Inter IKEA Systems B.V.
Exhibit 1005 Page 21 of 24

network such as the Internet and that has a unique identifying mark, such as a number. The terminal **202** connects to the content server **211** of the company that operates the vending machine. The identification server identifies the particular vending machine by identifying and decoding the unique identifying mark. The identity of the particular machine is included in the target object information **209** and is sent from the terminal **202** to the content server **211**. The content server **211**, having the identification of the particular vending machine (target object **200**), initiates communication with the vending machine. The User performs a transaction with the vending machine, such as purchasing a product, using his terminal **202** that communicates with the vending machine via the content server **211**.

[**0276**] The User "points and clicks" on part of a machine, such as an aircraft part. The terminal **202** then displays information pertinent to the part, such as maintenance instructions or repair history.

[**0277**] The User "points and clicks" on a magazine or newspaper article and link to streaming audio or video content, further information, etc.

[**0278**] The User "points and clicks" on an automobile. The location of the terminal **206** is determined by a Global Position System receiver in the terminal **206**, by cellular network radio ranging, or by another technique. The position of the terminal **202** is sent to the content server **211**. The content server provides the User with information regarding the automobile, such as price and features, and furthermore, based on the position information, provides the User with the location of a nearby automobile dealer that sells the car. This same technique can be used to direct Users to nearby retail stores selling items appearing in magazine advertisements that Users "point and click" on.

[**0279**] For visually impaired people:

[**0280**] Click on any item in a store and the device speaks the name of the item and price to you (the items must be in the database).

[**0281**] Click on a newspaper or magazine article and the device reads the article to you.

[**0282**] Click on a sign (building, streetsign, etc.) and the device reads the sign to you and provides any addition pertinent information (the signs must be in the database).

[**0283**] **FIG. 6** shows an embodiment of the invention for spacecraft applications. In this embodiment, all components of the system (except the target object **300**) are onboard a Spacecraft. The target object **300** is another spacecraft or object. This embodiment is used to determine the position and orientation of the target object **300** relative to the Spacecraft so that this information can be used in navigating, guiding, and maneuvering the spacecraft relative to the target object **300**. An example use of this embodiment would be in autonomous spacecraft rendezvous and docking.

[**0284**] This embodiment determines the position and orientation of the target object **300**, relative to the Spacecraft, as determined by the position, orientation, and size of the target object **300** in the imagery captured by the camera **303**, by comparing the imagery with views of the target object **300** from different orientations that are stored in the database **308**. The relative position and orientation of the target object **300** are output in the target object information, so that the spacecraft data system **310** can use this information in planning trajectories and maneuvers.

INDUSTRIAL APPLICABILITY

[**0285**] The industrial applicability is anywhere that objects are to be identified by a digital optical representation of the object.

1. A method for identifying an object from a digitally captured image thereof whose image characteristics are present in at least one database including:

decomposing the digitally captured image into a group of characterization parameter sets; and

comparing each of said characterization parameter sets with identical parameters of known objects stored in the at least one database; and

producing the best matches from the at least one database to identify the object.

2. The method as defined in claim 1 further including before decomposing the digitally captured image:

determining if the digitally captured image includes at least one symbolic image, and if such is included;

decoding the at least one symbolic image;

comparing the decoded at least one symbolic image to a database of decoded symbolic images; and

determining a level of confidence that the object has been identified.

3. The method as defined in claim 2 further including:

reducing the database choices in accordance with the determined level of confidence with which said comparing each of said group of characterization parameter sets with identical parameters of known objects stored in the at least one database is performed.

4. The method as defined in claim 1 wherein the at least one database includes:

an information address associated with the objects, and wherein said method includes:

establishing a connection between an information address associated with the identified object and the user.

5. The method as defined in claim 1 wherein the at least one database includes:

operative devices for interaction with the user when identified as the object.

6. The method defined in claim 1 wherein the decomposing the digitally captured image into a group of characterization parameter sets includes:

performing a radiometric correction of the digitally captured image by:

transforming the digitally captured image to one of a set of digital numbers per color plane, RGB representation; and

normalizing in all color planes with linear gain and offset transformations so that pixel values within each color channel span a full dynamic range of the set.

7. The method defined in claim 1 wherein the decomposing the digitally captured image into a group of characterization parameter sets includes:

grouping of image parameters to produce a plurality of related parameter sets.

8. The method defined in claim 1 further including:

further narrowing the classification of the at least one object by decoding at least one symbol in the at least one image.

9. The method defined in claim 1 further including:

obtaining information pertinent to at least one object by capturing at least one image of the at least one object;

recognizing the appearance of the at least one object in the at least one image;

determining at least one information address corresponding to the at least one object; and

communicating with at least one information resource, such as at least one computer, corresponding to the at least one information address.

10. The method defined in claim 1 further including:

establishing communication pertinent to to at least one object by capturing at least one image of the at least one object;

recognizing the appearance of the at least one object in the at least one image;

determining at least one information address corresponding to the at least one object; and

accessing information from at least one information resource, such as at least one computer, corresponding to the at least one information address.

11. The method defined in claim 1 wherein for each determined output:

comparing each against the at least one database; and

producing best matches from the at least one database.

12. A method for identifying an object in an at least one given image including:

iteratively comparing at least one region in the at least one given image with regions in images in an at least one database until the best correspondence is found between the at least one region in the at least one given image and the at least one region in images in the at least one database; wherein the images in the least one database represent views of at least one object in at least one viewing orientation, wherein the comparing is performed by:

extracting and comparing parameters from the regions of the at least one given image and the images in the at least one database.

13. The method for identifying an object in an at least one given image as defined in claim 12 including:

a method for reducing the number of required comparisons, which includes:

determining a priori that there is poor correspondence between the at least one region in the at least one given image and certain images in the at least one database.

14. The method for identifying an object in an at least one given image as defined in claim 12 including:

capturing at least one image; and

providing communications or information that are determined by both the identity of the object and the location of the device used to capture the at least one image, based both on the identity of the object and on refining or modifying the set of candidate identities of the at least one object based on the location of the device used to capture the at least one image.

15. The method for identifying an object in an at least one given image as defined in claim 12 further including:

establishing a connection between a portable device and an information site pertinent to an object by:

determining the address of the information site corresponding to the object using a remote computer,

sending the address to the portable device, and

connecting to the information site via at least one network.

16. The method for identifying an object in an at least one given image as defined in claim 12 further including:

providing audio information pertinent to at least one object based on capturing at least one image of the at least one object with a portable device,

recognizing the at least one object using the portable device or a remote computer; and

providing the pertinent audio information to the portable device.

17. A method for identifying to a user a multi-dimensional object from at least one digitally captured image thereof whose image characteristics are present in at least one database including:

decomposing the at least one digitally captured image into a group of parameters of different types; and

comparing the parameters of each type of the group of different parameter types with parameters of matching types of the different parameter types of known objects stored in the at least one database; and

producing a parameter match to a user from at least one parameter type, with respect to the at least one database for a high probability identification of the object.

18. The method defined in claim 17 wherein the producing a parameter match to a user from at least one parameter type, with respect to the at least one database for a high probability identification of the object includes:

producing the best parameter match from a plurality of parameters.

19. The method defined in claim 17 wherein the producing a parameter match to a user from at least one parameter type, with respect to the at least one database for a high probability identification of the object includes:

producing the best match from at least one group of a plurality of parameter types.

20. The method defined in claim 17 wherein the decomposing the digitally captured image into a group of parameters of different types includes:

Inter IKEA Systems B.V.
Exhibit 1005 Page 23 of 24

performing a radiometric correction of the at least one digitally captured image by:

transforming the at least one digitally captured image to at least one of a set of digital numbers per color plane, RGB representation; and

normalizing in all color planes with linear gain and offset transformations so that pixel values within each color channel span a full dynamic range of the set.

21. The method defined in claim 17 further including:

initiating a telephone call pertinent to the identified object, from a portable telephone;

capturing the at least one digitally captured image object using a camera in the portable telephone;

determining at least one telephone number corresponding to the object, and

initiating a telephone call to the at least one telephone number.

* * * * *