

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 December 2006 (28.12.2006)

PCT

(10) International Publication Number
WO 2006/136958 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/IB2006/002373

(22) International Filing Date: 24 January 2006 (24.01.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/646,511 25 January 2005 (25.01.2005) US

(71) Applicant (for all designated States except US): **DSPV, LTD.** [IL/IL]; 20 Wisotzky Street, Apt. 1, 62338 Tel Aviv (IL).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **LEV, Zvi Haim** [IL/IL]; 20 Wisotzky Street, Apt. 1, 62338 Tel Aviv (IL).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 2006/136958 A2

(54) Title: SYSTEM AND METHOD OF IMPROVING THE LEGIBILITY AND APPLICABILITY OF DOCUMENT PICTURES USING FORM BASED IMAGE ENHANCEMENT

(57) Abstract: A system and method for imaging a document, and using a reference document to place pieces of the document in their correct relative position and resize such pieces in order to generate a single unified image, including the electronic capturing a document with one or multiple images using an imaging device, the performing of pre-processing of said images to optimize the results of subsequent image recognition, enhancement, and decoding, the comparing of said images against a database of reference documents to determine the most closely fitting reference document, and the applying of knowledge from said closely fitting reference document to adjust geometrically the orientation, shape, and size of said electronically captured images so that said images correspond as closely as possible to said reference document.

**SYSTEM AND METHOD OF IMPROVING THE LEGIBILITY AND
APPLICABILITY OF DOCUMENT PICTURES USING FORM BASED IMAGE
ENHANCEMENT**

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial Number 60/646,511, filed on January 25, 2005, entitled, "System and method of improving the legibility and applicability of document pictures using form based image enhancement",
5 which is incorporated herein by reference in its entirety.

BACKGROUND OF THE NON-LIMITING EMBODIMENTS OF THE INVENTION

1. Field of the Exemplary Embodiments of the Invention

Exemplary embodiments of the present invention relates generally to the field of imaging, storage and transmission of paper documents, such as predefined forms.
10 Furthermore, these exemplary embodiment s of the invention is for a system that utilizes low quality ubiquitous digital imaging devices for the capture of images/video clips of documents. After the capture of these images/video clips, algorithms identify the form and page in these documents, position of the text in these images/video clips of these documents, and perform special processing to improve the legibility and utility of these documents for the
15 end-user of the system described in these exemplary embodiments of the invention.

2. Definitions

Throughout this document, the following definitions apply. These definitions are provided to merely define the terms used in the related art techniques and to describe non-limiting, exemplary embodiments of the present invention. It will be appreciated that the
20 following definitions are not limitative of any claims in any way.

“Computational facility” means any computer, combination of computers, or other equipment performing computations, that can process the information sent by the imaging device. Prime examples would be the local processor in the imaging device, a remote server, or a combination of the local processor and the remote server.

5 “Displayed” or “printed”, when used in conjunction with an imaged document, is used extensively to mean that the document to be imaged is captured on a physical substance (as by, for example, the impression of ink on a paper or a paper-like substance, or by embossing on plastic or metal), or is captured on a display device (such as LED displays, LCD displays, CRTs, plasma displays, ATM displays, meter reading equipment or cell phone displays).

10 “Form” means any document (displayed or printed) where certain designated areas in this document are to be filled by handwriting or printed data. Some examples of forms are: a typical printed information form where the user fills in personal details, a multiple choice exam form, a shopping web-page where the user has to fill in details, and a bank check.

 “Image” means any image or multiplicity of images of a specific object, including, for
15 example, a digital picture, a video clip, or a series of images. Used alone without a modifier or further explanation, “Image” includes both “still images” and “video clips”, defined further below.

 “Imaging device” means any equipment for digital image capture and sending, including, for example, a PC with a webcam, a digital camera, a cellular phone with a
20 camera, a videophone, or a camera equipped PDA.

 “Still image” is one or a multiplicity of images of a specific object, in which each image is viewed and interpreted in itself, not part of a moving or continuous view.

 “Video clip” is a multiplicity of images in a timed sequence of a specific object viewed together to create the illusion of motion or continuous activity.

3. Description of the Related Art

There are numerous existing methods and systems for the imaging and digitization of scanned documents. These imaging and digitization systems include, among others:

1. Special purpose flatbed scanners where the document is placed on a fixed planar
5 imaging system.

2. Handheld scanners where the document of interest is placed on a flat surface and the handheld scanners are manually moved while in close contact with this document.

3. High-resolution cameras on fixtures. These fixtures provide a fixed imaging geometry of the imaging being fixed. Furthermore, special lighting may be provided to
10 enable high quality uniform contrast and illumination conditions.

4. Facsimile machines and other special purpose scanners where the document of interest is moved mechanically through the scanning element of the scanner.

These existing systems provide a cost effective, reliable solution to the problem of scanning documents, but these systems require special hardware that is costly, and additional
15 hardware that is both costly and not very portable (that is, hardware which must be carried by the user). Furthermore, these existing systems are suited mainly for the imaging of non-glossy planar paper documents. Thus, they cannot serve for the imaging of glossy paper, of plastic documents, or of other displays that are not non-glossy paper. They are also not suited for the imaging of non planar objects.

20 The popularity of mobile imaging devices such as camera phones has led to the development of solutions that attempt to perform similar document scanning using such present-day camera phones as the imaging device. The raw images of documents taken by a camera phone are typically not useful for sending via fax, for archiving, for reading, or for other similar uses, due primarily to the following effects:

25 1. As a result of limited imaging device resolution, physical distance limitations, and imaging angles, the capture of a readable image of a full one page document in a single photo

is very difficult. With some imaging devices, the user may be forced to capture several separate still images of different parts of the full document. With such devices, the parts of the full document must be assembled in order to provide the full coherent image of the document. (It may be noted, however, with other imaging devices, notably some scanners, fax machines, and high resolution cameras for taking fixed images, multiple images are typically not required, but this equipment is expensive, often not easily portable, and generally incapable of dealing with quality issues where the document to be captured is not of high quality, or is not on glossy paper, or suffers other optical defects, as discussed above.) The resolution limitation of mobile devices is a result of both the imaging equipment itself, and of the network and protocol limitations. For example, a 3G mobile phone can have a multi-megapixel camera, yet in a video call the images in the captured video clip are limited to a resolution of 176 by 144 pixels due to the video transmission protocol.

2. Since there is no fixed imaging angle common to all still images of the parts of the full document, the multiple still images suffer from variable skewing, scaling, rotation and other effects of projective geometry. Hence, these still images cannot be simply “put together” or printed conveniently using the technologies commonly available for regular planar document such as faxes.

3. The still images of the full document or parts of it are subject to several optical effects and imaging degradations. The optical effects include: variable lighting conditions, shadowing, defocusing effects due to the optics of the imaging devices, fisheye distortions of the camera lenses. The imaging degradations are caused by image compression and pixel resolution. These optical effects and imaging degradations affect the final quality of the still images of the parts of the full document, making the documents virtually useless for many of the purposes documents typically serve.

4. In addition to all limitations applying to still images, video clips suffer from blocking artifacts, varying compression between frames, varying imaging conditions between

frames, lower resolution, frame registration problems and a higher rate of erroneous image data due to communication errors.

The limited utility of the images/ video clips of parts of the full document is manifest in the following:

5 1. These images of parts of the full document cannot be faxed because of a large dynamic range of imaging conditions within each image, and also between the images. For example, one of the partial images may appear considerably darker or brighter than the other because the first image was taken under different illumination than the second image. Furthermore, without considerable gray level reduction operations the images will not be
10 suitable for faxing.

 2. To read hand-printed writing in these images of parts of the full document even on a high quality computer screen, is very difficult, mainly due to dynamic range of the imaging device, imaging device resolution, compression artifacts, and color contrast of the text versus the background.

15 3. These images of parts of the full document cannot be stored and later retrieved in a uniform manner since several images of the same document may contain duplicities and some parts of the document may be missing from the complete image set.

In order to improve the utility of imaging devices as document capture tools, some existing systems provide extra processing on these images of a full document or parts of it.

20 Some examples of such products are:

 1. The RealEyes3D™ Phone2Fun™ product. This product is composed of software residing on the phone with the camera. This software enables conversion of a single image taken by the phone's camera into a special digitized image. In this digital image, the hand printed text and/or pictures/drawings are highlighted from the background to create more
25 legible image which could potentially be faxed.

2. US Patent Application 20020186425, to Dufaux, Frederic, and Ulichney, Robert Alan, entitled "Camera-based document scanning system using multiple-pass mosaicking", filed June 1, 2001, describes a concept of taking a video file containing the results of a scan of a complete document, and converting it into a digitized and processed image which can be faxed or stored.

3. There are numerous other "panoramic stitching" products for digital cameras which supposedly enable the creation of a single large image from several smaller images with partial overlap. Examples of such products are PanoramaTM from Picture Works Technology, Inc. and QuickStitchTM software from Enroute Imaging.

The image processing products outlined above suffer from certain fundamental limitations that make their widespread adoption problematic and doubtful. Among these limitations are:

1. It is hard to automatically differentiate between the text and the background without prior information. Therefore in some cases the resulting image is not legible and/or the background contains many details resulting from incorrect segmentation between background and text. A good example appears in Figure 2. In Figure 2, an image 201 is the original image, and an image 202 shows the effects of the prior art processing when attempting to convert such an image into a bitonal image suitable for sending via fax.

2. Since it is hard to automatically estimate the imaging angles of the document in a given image, the resulting processed document may contain geometric distortions altering the reading experience of the end-user.

3. The automatic registration of multiple images / frames with partial overlap is technically difficult. Traditional image registration (also known as "stitching" or "panorama generation") methods assume that the images are taken at a large distance from the imaging apparatus, and that there are no significant projective or lighting variations between the different images to be stitched. These conditions are not fulfilled when document imaging is

performed by a portable imaging device. In the typical use of a portable imaging device, the imaging distances are short, and therefore projective geometry and illumination variations between images (in particular due to the effect of the user and the portable device itself on illumination) are very prominent. Furthermore, there is no guarantee that the visual overlap
5 between subsequent images will contain sufficient information to uniquely combine the images in the right way. For example, in Figure 7, discussed further below, an example is provided of two images of parts of a document with no overlap, which could be mistaken to be overlapping images by prior art stitching software.

A different approach to document capture, sending and processing is based on
10 dedicated non-imaging products that directly capture the user's entries into the document. Some examples of such devices are:

1. Personal Digital Assistants with touch-sensitive screens. Notable examples include the Palm family of PDAs, and the "Tablet PC" which is a complete personal computer with a touch-sensitive screen.
- 15 2. "E-pens" – devices where the precise location, speed and sometimes also pressure of the pen used for writing, are continuously monitored/measured using special hardware. Notable examples include the Anoto design implemented in the LogitechTM, HPTM and NokiaTM E-pens, etc.
- 20 3. Pressure based and location based "tablets" that connect to a PC and provide tracking of a stylus, or of a normal pen, on a pre-defined area. A notable example is the pad used in many point-of-sale locations and by some delivery couriers to record the signature of the customer.

These non-imaging solutions require special hardware, require writing with or on special hardware, and introduce a different writing experience for the end-user.

SUMMARY OF THE EXEMPLARY EMBODIMENTS OF THE INVENTION

An aspect of the exemplary embodiments of the present invention is to introduce a new and better way of converting displayed or printed documents into electronic ones that can be the read, printed, faxed, transmitted electronically, stored and further processed for
5 specific purposes such as document verification, document archiving and document manipulation. Unlike prior art, where special purpose equipment is required, another aspect of the exemplary embodiments of the present invention is to utilize the imaging capability of a standard portable wireless device. Such portable devices, such as camera phones, camera enabled PDAs, and wireless webcams, are often already owned by users. By utilizing special
10 recognition capabilities that exist today and some additional available information on the layout and contents of the imaged document, the exemplary embodiments of the present invention may allow documents of full one page (or larger) to be reliably scanned into a usable digital image.

According to an aspect of the exemplary embodiments of the present invention, a
15 method for converting displayed or printed documents into an electronic form, is provided. The first stage of the method includes comparing the images obtained by the user to a database of reference documents. Throughout this document, the "reference electronic version of the document" shall refer to a digital image of a complete single page of the document. This reference digital image can be the original electronic source of the document
20 as used for the document printing (e.g., a TIFF or PhotoshopTM file as created by a graphics design house), or a photographic image of the document obtained using some imaging device (e.g., a JPEG image of the document obtained using a 3G video phone), or a scanned version of the document obtained via a scanning or faxing operation. This electronic version may have been obtained in advance and stored in the database, or it may have been provided by
25 the user as a preparatory stage in the imaging process of this document and inserted into the same database. Thus, the method includes recognizing the document (or a part thereof)

appearing in the image via visual image cues appearing in the image, and using a priori information about the document. This a priori information includes the overall layout of the document and the location and nature of image cues appearing in the document.

The second stage of the method involves performing dedicated image processing on various parts of the image based on knowledge of which document has been imaged and what type of information this document has in its various parts. The document may contain sections where handwritten or printed information is expected to be entered, or places for photos or stamps to be attached, or places for signatures or seals to be applied, etc. For example, areas of the image that are known to include handwritten input may undergo different processing than that of areas containing typed information. Additionally, the knowledge of the original color and reflectivity of the document can serve to correct the apparent illumination level and color of the imaged document. As an example, areas in the document known to be simple white background can serve for white reference correction of the whole document. As another example, areas of the document which have been scanned in separate images or video frames in different resolutions and from different angles can all be combined into one document of unified resolution, orientation and scale. Another example would be selective application of a dust or dirt removal operator to areas in the image known to contain plain background, so as to improve the overall document appearance.

The third stage of the method (which is optional) includes recognition of characters, marks or other symbols entered into the form – e.g. Optical mark recognition (OMR), Intelligent character recognition (ICR) and the decoding of machine readable codes (e.g. barcodes).

The fourth stage of the method includes routing of the information based on the form type, the information entered into the form, the identity of the user sending the image and other similar data.

According to another aspect of the exemplary embodiments of the present invention, a system and a method for converting displayed or printed documents into an electronic form, is provided. The system and the method includes capturing an image of a printed form with printed or handwritten information filled in it, transmitting the image to a remote facility, pre-
5 processing the image in order to optimize the recognition results, searching the image for image cues taken from an electronic version of this form which has been stored previously in the database, utilizing the existence and position of such image cues in the image in order to determine which form it is and the utilization of these recognition results in order to process the image into a higher quality electronic document which can be faxed, and the sending of
10 this fax to a target device such as a fax machine or an email account or a document archiving system.

According to yet another aspect of the exemplary embodiments of the present invention, a system and a method may also present capturing several partial and potentially overlapping images of a printed document, transmitting the image to a remote facility, pre-
15 processing the images in order to optimize the recognition results, searching each of the images for image cues taken from a reference electronic version of this document which has been stored in the database, utilizing the existence and position of such image cues in each image in order to determine which part of the document and which document is imaged in each such image, and the utilization of these recognition results and of the reference version
20 in order to process the images into a single unified higher quality electronic document which can be faxed, and the sending of this fax to a target device.

Thus, part of the utility of the system is the enabling of a capture of several (potentially partial and potentially overlapping) images of the same single document, such that these images, by being of just a part of the whole document, each represent a higher
25 resolution and/or superior image of some key part of this document (e.g. the signature box in a form). The resulting final processed and unified image of the document would thus have a

higher resolution and higher quality in those key parts than could be obtained with the same capture device if an attempt was made to capture the full document in a single image. The prior art presented a dilemma between, on the one hand, limited resolution requiring costly special purpose high resolution imaging capture devices (such as flatbed scanners), or, on the other hand, acceptance of a single low quality image of the whole document as in the RealEyes™ product. A high resolution imaging may be provided without special purpose high resolution imaging capture devices.

Another part of the utility of the system is that if a higher resolution or otherwise superior reference version of a form exists in the database, it is possible to use this reference version to complete parts of the document which were not captured (or were captured at low quality) in the images obtained by the user. For example, it is possible to have the user take image close-ups of the parts of the form with handwritten information in them, and then to complete the rest of the form from the reference version in order to create a single high quality document.

Another part of the utility of the exemplary embodiments of the present invention is that by using information about the layout of a form (e.g., the location of boxes for handwriting/signatures, the location of checkboxes, the location places for attaching a photograph) it is possible to apply different enhancement operators to different locations. This may result in a more legible and useful document.

The exemplary embodiments of the present invention thus enable many new applications, including ones in document communication, document verification, and document processing and archiving.

BRIEF DESCRIPTION OF THE DRAWINGS

Various other objects, features and attendant advantages of the exemplary embodiments of the present invention will become fully appreciated as the same become

better understood when considered in conjunction with the accompanying detailed description, the appended claims, and the accompanying drawings, in which:

FIG. 1 illustrates a typical prior art system for document scanning.

FIG. 2 illustrates a typical result of document enhancement using prior art products
5 that have no a priori information on the location of handwritten and printed text in the document.

FIG. 3 illustrates one exemplary embodiment of the overall method of the present invention.

FIG. 4 illustrates an exemplary embodiment of the processing flow of the present
10 invention.

FIG. 5 illustrates an example of the process of document type recognition according to an exemplary embodiment of the present invention. FIG. 5A is an example of a document retrieved from a database of reference documents. FIG. 5B represents an imaged document which will be compared to the document retrieved from the database of reference documents.

FIG. 6 illustrates how an exemplary embodiment of the present invention may be used
15 to create a single higher resolution document from a set of low resolution images obtained from a low resolution imaging device.

FIG. 7 illustrates the problem of determining the overlap and relative location from
20 two partial images of a document, without any knowledge about the shape and form of the complete document. This problem is paramount in prior art systems that attempt to combine several partial images into a larger unified document.

FIG. 8 shows a sample case of the projective geometry correction applied to the images or parts of the images as part of the document processing according to an exemplary embodiment of the present invention.

FIG. 9 illustrates the different processing stages of an image segment containing printed or handwritten text on a uniform background and with some prior knowledge of the approximate size of the text according to an exemplary embodiment of the present invention.

DETAILED DESCRIPTION OF THE EXEMPLARY EMBODIMENTS

5 An exemplary embodiment of the present invention presents a system and method for document imaging using portable imaging devices. The system is composed of the following main components:

1. A portable imaging device, such as a camera phone, a digital camera, a webcam, or a memory device with a camera. The device is capable of capturing digital images and/or
10 video, and of transmitting or storing them for later transmission.

2. Client software running on the imaging device or on an attached communication module (e.g., a PC). This software enables the imaging and the sending of the multimedia files to a remote server. It can also perform part of or all of the required processing detailed in this application. This software can be embedded software which is part of the device, such
15 as an email client, or an MMS client, or an H.324 or IMS video telephony client. Alternatively, the software can be downloaded software running on the imaging device's CPU.

3. A processing and routing computational facility which receives the images obtained by the portable imaging device and performs the processing and routing of the
20 results to the recipients. This computational facility can be a remote server operated by a service provider, or a local PC connected to the imaging device, or even the local CPU of the imaging device itself.

4. A database of reference documents and meta-data. This database includes the reference images of the documents and further descriptive information about these
25 documents, such as the location of special fields or areas on the document, the routing rules for this document (e.g., incoming sales forms should be faxed to +1-400-500-7000), and the

preferred processing mode for this document (e.g., for ID cards the color should be retained in the processing, paper forms should be converted to grayscale).

Figure 1 illustrates a typical prior art system enabling the scanning of a document from single image and without additional information about the document. The document
5 **101** is digitally imaged by the imaging device **102**. Image processing then takes place in order to improve the legibility of the document. This processing may also include also data reduction in order to reduce the size of the document for storage and transmission – for example reduction of the original color image to a black and white “fax” like image. This processing may also include geometric correction to the document based on estimated angle
10 and orientation extracted from some heuristic rules.

The scanned and potentially processed image is then sent through a wire-line/wireless network **103** to a server or combination of servers **104** that handle the storage and/or processing and /or routing and/or sending of the document. For example, the server may be a digital fax machine that can send the document as a fax over phone lines **105**. The recipient
15 **106** could for example be an email account, a fax machine, a mobile device, a storage facility.

Figure 2 displays typical limitations of prior art in text enhancement. A complex form containing both printed text in several sizes and fonts and handwritten text is processed. Since the algorithms of prior art do not have additional information about which parts of the image contain each type of text, they apply some average processing rule which causes the
20 handwritten text, which is actually the most important part of the document, to become completely unreadable. Element **201** demonstrates that the original writing is legible, while element **202** shows that the processed image is unreadable.

Figure 3 illustrates one exemplary embodiment of the present invention. The input
25 **301** is no longer necessarily a single image of the whole document, but rather can be a plurality of N images that cover various parts of the document. Those images are captured by the portable imaging device **302**, and sent through the wire-line or wireless network **303** to a

computational facility **304** (e.g., a server, or multiple servers) that handles the storage and/or processing and/or routing and/or sending of the document. The image(s) can be first captured and then sent using for example an email client, an MMS client or some other communication software. The images can also be captured during an interactive session of the user with the
5 backend server as part of a video call. The processed document is then sent via a data link **305** to a recipient **306**.

The document database **307** includes a database of possible documents that the system expects the user of **302** to image. These documents can be, for example, enterprise forms for filling (e.g., sales forms) by a mobile sales or operations employee, personal data forms for a
10 private user, bank checks, enrollment forms, signatures, or examination forms. For each such document the database can contain any combination of the following database items:

1. Images of the document – which can be used to complete parts of the document which were not covered in the image set **301**. Such images can be either a synthetic original or scanned or photographed versions of a printed document.
- 15 2. Image cues – special templates that represent some parts of the original document, and are used by the system to identify which document is actually imaged by the user and/or which part of the document is imaged by the user in each single image such as **309**, **310**, and **311**.
- 20 3. Additional information about special fields or areas in the document, e.g. boxes for handwritten input, ticker boxes, places for a photo ID, pre-printed information, barcode location, etc. This information is used in the processing stage to optimize the resulting image quality by applying different processing to the different parts of the document.
- 25 4. Routing information – this information can include commands and rules for the system's business logic determining the routing and handling appropriate for each document type. For example, in an enterprise application it is possible that incoming "new customer" forms will be sent directly to the enrollment department via email, incoming equipment

orders will be faxed to the logistics department fax machine, and incoming inventory list documents may be stored in the system archive. Routing information may also include information about which users may send such a form, and about how certain marks (e.g., check boxes) or printed information on the form (e.g. printed barcodes or alphanumeric information) may affect routing. For example, a printed barcode on the document may be interpreted to determine the storage folder for this document.

The reference document **308** is a single database entry containing the records listed above. The matching of a single specific document type and document reference **308** to the image set **301** is done by the computational facility **304** and is an image recognition operation. An exemplary embodiment of this operation is described with reference to Figure 4.

It is important to note that the reference document **308** may also be an image of the whole document obtained by the same device **302** used for obtaining the image data set **301**. Hence the dotted line connecting **302** and **308**, indicating that **308** may be obtained using **302** as part of the imaging session. For example, a user may start the document imaging operation for a new document by first taking an image of the whole document, potentially also adding manually information about this document, and then taking additional images of parts of the document with the same imaging device. This way, the first image of the whole document serves as the reference image, and the server **304** uses it to extract from it image cues and thus to determine for each image in the image set **301** what part of the full document it represents. A typical use of such a mode would be when imaging a new type of document with a low resolution imaging device. The first image then would serve to give the server **304** the layout of the document at low resolution, and the other images in image set **301** would be images of important parts of the document. This way, even a low resolution imaging device **302** could serve to create a high resolution image of a document by having

the server 304 combine each image in the image set 301 into its respective place. An example of such a placement is depicted in Figure 6.

Thus, the exemplary embodiment of the present invention is different from prior art in the utilization of images of a part of a document in order to improve the actual resolution of the important parts of the document. The exemplary embodiment of the present invention also differs from prior art in that it uses a reference image of the whole document in order to place the images of parts of the document in relation to each other. This is fundamentally different from prior art which relies on the overlap between such partial images in order to combine them. The exemplary embodiment of the present invention has the advantage of not requiring such overlap, and also of enabling the different images to be combined (301) to be radically different in size, illumination conditions etc. Thus the user of the imaging device 302 has much greater freedom in imaging angles and is freed from following any special order in taking the various images of parts of the document. This greater freedom simplifies the imaging process and makes the imaging process more convenient.

Figure 4 illustrates the method of processing according to an exemplary embodiment of the present invention. Each image (of the multiple images as denoted in the previous figure as image set 301) is first pre-processed 401 to optimize the results of subsequent image recognition, enhancement, and decoding operations. The preprocessing can include operations for correcting unwanted effects of the imaging device and of the transmission medium. It can include lens distortions correction, sensor response correction, compression artifact removal and histogram stretching. At this pre-processing stage the server 304 did not determine yet which type of document is in the image, and hence the pre-processing does not utilize such knowledge.

The next stage of processing is to recognize which document or part thereof appears in the image. This is accomplished in the loop construct of elements 402, 403, and 404. Each reference document stored in the database is searched, retrieved, and compared to the

image at hand. This comparison operation is a complex operation in itself, and relies upon the identification of image cues, which exist in the reference image, in the image being processed. The use of image cues, which represent small parts of the document, and their relative location, is especially useful in the present case for several reasons:

- 5 1. The imaged document may be a form in which certain fields are filled in with handwriting or typing. Thus, this imaged document is not really identical to the reference document, since it has additional information printed or handprinted or marked on it. Thus, a comparison operation has to take this into account and only compare areas where the imaged form would still be identical to the reference “empty” form.
- 10 2. Since the image may be of a small part of the full reference document, a full comparison of the reference document to the image would not be meaningful. At the same time, image cues that exist in the reference document may still be located in the image even if the image is only of a segment of the full document. This ambiguity is illustrated in Figures **5A** and **5B**.
- 15 3. Due to the differences in scale, imaging angles, illumination variations and image degradations introduced by the limited resolution of the imaging sensor and image compression, the reliable comparison of a reference image of a document to an image obtained by a portable imaging device is in general a difficult endeavor. The utilization of image cues which are small in relation to the whole reference image is, according to an
20 exemplary embodiment of the invention, a reliable and proven solution to this problem of image comparison.

The method used in the present embodiment to perform the search of the image cues in **403** and for determining the match in **404** is described in great detail in US Non Provisional Patent Application number 11/293,300, to the applicant herein Lev, Tsvi, entitled
25 “SYSTEM AND METHOD OF GENERIC SYMBOL RECOGNITION AND USER AUTHENTICATION USING A CELLULAR/WIRELESS DEVICE WITH IMAGING

CAPABILITIES”, filed on December 5, 2005. The disclosure of such Application is hereby incorporated by reference in its entirety, and is provided below in an Addendum A. This Application describes in great detail a possible method of reliably detecting image cues in digital images in order to recognize whether certain objects (including documents, as
5 discussed herein) do indeed appear in those images.

There are many different variations of “image cues” that can serve for reliable matching of a processed image to a reference document from the database. Some examples are:

1. High contrast, preferably unique image patches from the reference document.

10 2. Special marks which have been inserted into the document on purpose to enable reliable recognition, such as, for example, “cross” signs at or near the boundaries of the document.

3. Areas of the document that are of a distinct color or texture or combination thereof – for example, blue lines on a black and white document.

15 4. Unique alphanumeric codes, graphics or machine readable codes printed on the document in a specific location or plurality of locations.

The determination of the location, size and nature of the image cues is to be performed manually or automatically by the server at the time of insertion of document insertion into the database.

20 A typical criterion for automatic selection of image cues would be a requirement the areas used as image cues are different from most of the rest of the document in shape, grayscale values, texture etc.

Assuming that the processed image has indeed been matched with a reference document or a part thereof, stage 405 then employs the knowledge about the reference
25 document in order to geometrically correct the orientation, shape and size of the image so that they will correspond to a reference orientation, shape and size. This correction is performed

by applying a transformation on the original image, aiming to create an image where the relative positions of the transformed image cue points are identical to their relative positions in the reference document. For example, where the only main distortion of the image is due to projective geometry effects (created by the imaging device's angles and distance from the document) a projective transformation would suffice. Or as another example, in cases where the imaging device's optics create effects such as fisheye distortion, such effects can also be corrected using a different transformation. The estimation of the parameters for these corrective transformations is derived from the relative positions of the image cues. Hence, the more image cues located in the image, the more precise the corrective transformation is. For example, in Figure 5B an image is presented where only three image cues were located, hence it can be corrected using an affine transform but not by a full projective transform. Furthermore, typically the transform would not be applied to the original image but rather to an enlarged (and rescaled) version of the original image, in order to avoid or at least minimize the unwanted smoothing effects of image interpolation.

In stage 406, the image is already in the reference orientation and size, hence the metadata in the database about the location, size and type of different areas in the document can be used to selectively and optimally process the data in each such area. Some examples of such optimized processing are:

1. Replacing an area in the image with a clean reference version of it. In a form, there are typically many printed marks and fields which are part of the form and are not supposed to be influenced by the filling-out process of the form. Since the exact layout and content of the form itself are known in advance and stored in the database, it is possible to thus improve the overall legibility and utility of the resulting document. As a pertinent example, small font text typical of contractual forms and containing the exact terms and conditions of the deal signed may be hard to read from the image obtained by the user, yet the

same exact text is stored in the database and can be used to fill in those hard-to-read parts of the document.

2. Scale optimized handwriting and printed text enhancement. In areas of a form which are to be filled in, the knowledge of the exact size and background (typically white) in this area, coupled with knowledge of the typical handwriting size or font size to be used in printed information, allow for better enhancement of the text in these areas. A typical subject of document processing research is the reliable differentiation between background and print in documents. In a general document, with no prior knowledge of whether a certain area contains a picture, text or graphics, this is indeed a very difficult problem. On the other hand, by using the information that the pixels in a certain segment of the image are composed of, for example, a white background and some text, this distinction between text and background becomes a much simpler problem that can be resolved with effective algorithms. A, exemplary technique for such enhancement is described below, in the text accompanying Figure 9. It is important to note that most algorithms for enhancing the legibility and appearance of text rely to some extent on the text size and stroke width to be in some pre-determined range. Hence, a priori knowledge of the size of the text box and of the expected handwritten/printed text size is very useful for optimally applying such text enhancement algorithms. The use of such a priori knowledge in the exemplary embodiment of the current invention is an advantage over prior art systems that have no such a priori knowledge regarding the expected size of the text in the image.

3. Optimized adaptation taking into account both a priori knowledge of the image area and of the target device the document is to be routed to. For example, the form could include a photo of a person at some designated area, and the person's signature at another designated area. Thus, the processing of those respective areas can take into account both the expected input there (color photo, handwriting) and the target device – e.g., a bitonal fax, and thus different processing would be applied to the photo area and the signature area. At the

same time, if the target device is an electronic archive system, the two areas could undergo the same processing since no color reduction is required.

In stage 407, optional symbol decoding takes place if this is specified in the document metadata. This symbol decoding relies on the fact that the document is now of a fixed geometry and scale identical to the reference document, hence the location of the symbols to be decoded is known. The symbol decoding could be any combination of existing symbol decoding methods, comprising:

1. Alphanumeric strings recognition and decoding – also known as Optical Character Recognition (OCR).

2. Recognition and decoding of known commercial symbols – also known as Optical Mark Recognition (OMR).

3. Machine code decoding – as in barcode or other machine codes.

4. Graphics Recognition –examples include the recognition of some sticker or stamp used in some part of the document – e.g. to verify the identity of the document.

5. Photo recognition – for example, facial ID could be applied to a photo of a person attached to the document in a specific place (as in passport request forms).

A sample algorithm for the decoding of alphanumeric codes and symbols is described in US Non Provisional Application number 11/266,378, to the applicant herein Lev, Tsvi, entitled “SYSTEM AND METHOD OF ENABLING A CELLULAR/WIRELESS DEVICE WITH IMAGING CAPABILITIES TO DECODE PRINTED ALPHANUMERIC CHARACTERS”, filed November 4, 2005. The disclosure of this Application is hereby incorporated by reference in its entirety, and is provided below in an Addendum B.

In stage 408, the document, having undergone the previous processing steps, is routed to one or several destinations. The business rules of the routing process can take into considerations the following information pieces:

1. The identity of the portable imaging device and the identity of the user operating this imaging device, and additional information provided by the user along with the image.

2. The meta-data for the recognized document which can contain business logic rules specific to this document.

5 3. The results of the symbol decoding stage 407.

4. Indications about image quality such as image noise, focus, angle. Some indications such as imaging angle and imaging distance can be derived from the knowledge of the actual reference document size in comparison to the image being currently processed. For example, if the document is known to be 10 centimeters wide at some point, a measure of
10 the same distance in the recognized image can yield the imaging distance of the camera at the time the image was taken.

Some specific examples of routing are:

1. The user imaging the document attaches to the message containing the image a phone number of a target fax machine. Thus, the processed image is converted to black and
15 white and faxed to this target number.

2. The document in the image is recognized as the “incoming order” document. The meta-data for this document type specifies it should be sent as a high-priority email to a defined address as well as trigger an SMS to the sales department manager.

3. The document includes a printed digital signature in hexadecimal format. This
20 signature is decoded into a digital string and the identity of the person who printed this signature is verified using a standard public-key-infrastructure (PKI) digital signature verification process. The result of the verification is that the document is sent to, and stored in, this person’s personal storage folder.

It should be stressed that the different processing stages described in figure 4 can take
25 place either after the user has sent the image(s) for processing (as in an off-line processing mode) or during the imaging session itself (as in on-line processing). On line processing is

particularly useful when the user is in an interactive session with the server – e.g., in a videotelephony session or a SIP/IMS session. Examples of such interactivity include:

1. Adding the initial picture taken by the user of the whole document to the document database and using it during the session to correctly place further images taken by the user into their respective positions.
2. Informing the user that he or she forgot to take images of some important parts of the document (such as, for example, a signature field).
3. Guiding the user to the proper areas and proper imaging distance in order to optimally capture some parts of the document (for example, “move camera to the right and closer please”), based on the recognition of the part of the document the camera is currently pointing at and the image cue location.
4. Notifying the user if the images obtained so far are of sufficient illumination and sharpness, or if they should be re-captured.
5. Giving further instructions to the user based on the results of the OCR/OMR/symbol recognition. For example, if the form is recognized to contain a serial number that is known to be no longer valid, the user could be warned of this and instructed to use a newer form at the time of document capture.

Figures **5A** and **5B** illustrate a sample process of recognition of a specific image. A certain document **500** is retrieved from the database. It contains several image cues **501**, **502**, **503**, **504** and **505**, which are searched for in the obtained image **506**. A few of them are found and in the proper geometric relation. A sample search and comparison algorithm for the image cues is described in US Non Provisional Application number 11/293,300, cited above and attached hereto as Addendum A.. The occurrence of the image cues in **503**, **504**, and **505** in the image, in areas **507**, **508**, and **509**, thus serve to recognize which part of which document the image **506** contains. It is important to note that the same process could be applied when the image has been itself obtained by the user as e.g. the first image in the

sequence. In such a case, the recognition for image **506** would be relevant for locating the part of original image **500** which appears in it, but there would not be any “metadata” in the database unless the user has specifically provided it. It should be noted that the image cues can be based on color and texture information - for example, a document in specific color
5 may contain segments of a different color that have been added to it or were originally a part of it. Such segments can serve as very effective image cues.

Figure 6 illustrates how the exemplary embodiment of the present invention can be used to create a single high resolution and highly legible image from several lower quality images of parts of the document. Images **601** and **602** were taken by a typical portable
10 imaging device. They can represent photos taken by a camera phone separately, photos taken as part of a multi-snapshot mode in such a camera phone or digital camera, or frames from a video clip or video transmission generated by a camera phone. These images have been recognized by the system as parts of a reference document entitled “US Postal Service Form #1”, and accordingly the images have been corrected and enhanced. Only the parts of these
15 images that contain handwritten input have been used, and the original reference document has been used to fill in the rest of the resulting document **603**. It can be clearly seen that the original images suffered from some fisheye distortion, bad contrast, graininess and non-uniform lighting, but due to the correction and enhancement applied, the resulting final document **603** is free from all of these effects. The system can thus also be applied to
20 signatures in particular, optimally processing the image of a human signature, and potentially comparing it to an existing database of signatures for verification or comparison purposes.

Figure 7 illustrates the deficiencies of prior art. Images **701** and **702** have been sent via the imaging device, and cover different and non-overlapping areas of the document. However, the upper left part of image **701** is virtually identical to the lower right part of
25 image **702**. Hence, any image matching algorithm which works by comparing images and combining them would assume, incorrectly in this case, that these images should be

combined. (An exemplary embodiment of the present invention, conversely, locates images **701** and **702** in the larger framework of the reference image of the whole document, and would therefore not make such a mistake, but would place all images in their correct position, as described further below). Furthermore, the requirement of prior art to maintain substantial
5 overlap between consecutive images in a sequence implies that only specific “scanning” movements are allowed, and that the user’s imaging angles, speed of movement of the mobile device, and distance from the document are severely constrained, resulting in a lengthy and inconvenient process. Furthermore, the user is forced to image the whole document for correct registration, even if the important information contained in the document is
10 concentrated in just a few small areas of the document (e.g. the signature at the bottom of the document).

Figure **8** illustrates how a segment of the image is geometrically corrected once the image **800** has been correlated with the proper reference document. The area **809**, bounded by points **801**, **802**, **803**, and **804**, is identified using the metadata of the reference document
15 as a “text box”, and is geometrically corrected using for example a projective transformation to be of the same size and orientation as the reference text box **810** bounded by points **805**, **806**, **807**, and **808**. The utilization of the image cues provides the correspondence points which are necessary to calculate the parameters of the projective transformation.

Figure **9** illustrates the different processing stages of an image segment containing
20 printed or handwritten text on a uniform background and with some prior knowledge about the approximate size of the text. This algorithm represents one of the processing stages that can be applied in **406**.

In order to correct for lighting non-uniformities in the image, the illumination level in the image is estimated from the image at **901**. This is done by calculating the image
25 grayscale statistics in the local neighborhood of each pixel, and using some estimator on that neighborhood. For example, in the case of dark text on lighter background, this estimator

could be the n th percentile of pixels in the M by M neighborhood of each pixel. Since the printed text does not occupy more than a few percents of the image, estimators such as the 90th percentile of gray scale values would not be affected by it and would represent a reliable estimate of the background grayscale which represents the local illumination level. The neighborhood size M would be a function of the expected size of the text and should be considerably larger than the expected size of a single letter of that text.

Once the local illumination level has been estimated, the image can be normalized to eliminate the lighting non uniformities in **902**. This can be accomplished by dividing the value of each pixel by the estimated illumination level in the pixel's neighborhood as estimated in the previous stage **901**.

In **903**, histogram stretching is applied to the illumination corrected image obtained in **902**. This stretching enhances the contrast between the text and the background, and thereby also enhances the legibility of the text. Such stretching could not be applied before the illumination correction stage since in the original image the grayscale values of the text pixels and background pixels could be overlapping.

In stage **904**, the system again utilizes the knowledge that the handprinted or printed text in the image is known to be in a certain range of size in pixels. Each image block is examined to determine how many pixels it contains whose grayscale value is in the range of values associated text pixels. If this number is below a certain threshold, the image block is declared as pure background and all the pixels in that block are set to some default background pixel value. The purpose of this stage is to eliminate small marks in the document which could be caused by dirt, pixel nonuniformity in the imaging sensor, compression artifacts and similar image degrading effects.

It is important to note that the processing stages described in **901**, **902**, **903**, and **904**, are composed of image processing operations which may be used , in different combinations, in related art techniques of document processing. In an exemplary, non-limiting embodiment

of the present invention, however, these operations utilize the additional knowledge about the document type and layout, and incorporate that knowledge into the parameters that control the different image processing operations. The thresholds, neighborhood size, spectral band used and similar parameters can be all optimized to the expected text size and type, and the
5 expected background.

In stage 905 the image is processed once again in order to optimize it to the routing destination(s). For example, if the image is to be faxed it can be converted to a bitonal image. If the image is to be archived, it can be converted into grayscale and to the desired file format such as JPEG or TIFF. It is also possible that the image format selected will
10 reflect the type of the document as recognized in 404. For example, if the document is known to contain photos, JPEG compression may be better than TIFF. If the document on the other hand is known to contain monochromatic text, then a grayscale or bitonal format such as bitonal TIFF could be used in order to save storage space.

Other variations and modifications are possible, given the above description. All
15 variations and modifications which are obvious to those skilled in the art to which the present invention pertains are considered to be within the scope of the protection granted by this letter patent.

ADDENDUM A:**SYSTEM AND METHOD OF GENERIC SYMBOL RECOGNITION AND USER
AUTHENTICATION USING A COMMUNICATION DEVICE WITH IMAGING
CAPABILITIES****CROSS-REFERENCE TO RELATED APPLICATIONS**

5 This application claims the benefit of U.S. Provisional Application Serial Number
60/632,953, filed on December 6, 2004, entitled, "System and Method of Identifying a User
Viewing Content on a Screen Using a Cellular/Wireless Device with Imaging Capabilities."

BACKGROUND OF THE INVENTION**1. Field of the Invention**

10 The present invention relates generally to the field of digital imaging, digital image
recognition, and utilization of image recognition to applications such as authentication and
access control. The device utilized for the digital imaging is a portable wireless device with
imaging capabilities.

 The invention utilizes an image of a display showing specific information which may
15 be open (that is clear) or encoded. The imaging device captures the image on the display, and
a computational facility will interpret the information (including prior decoding of encoded
information) to recognize the image. The recognized image will then be used for purposes
such as user authentication, access control, expedited processes, security, or location
identification.

20 Throughout this invention, the following definitions apply:

- "Computational facility" means any computer, combination of computers, or other
equipment performing computations, that can process the information sent by the imaging
device. Prime examples would be the local processor in the imaging device, a remote server,
or a combination of the local processor and the remote server.

- “Displayed” or “printed”, when used in conjunction with an object to be recognized, is used expansively to mean that the object to be imaged is captured on a physical substance (as by, for example, the impression of ink on a paper or a paper-like substance, or by engraving upon a slab of stone), or is captured on a display device (such as LED displays, LCD displays, CRTs, plasma displays, or cell phone displays).

- “Image” means any image or multiplicity of images of a specific object, including, for example, a digital picture, a video clip, or a series of images.

- “Imaging device” means any equipment for digital image capture and sending, including, for example, a PC with a webcam, a digital camera, a cellular phone with a camera, a videophone, or a camera equipped PDA.

- “Trusted” means authenticated, in the sense that “A” trusts “B” if “A” believes that the identity of “B” is verified and that this identity holder is eligible for the certain transactions that will follow. Authentication may be determined for the device that images the object, and for the physical location of the device based on information in the imaged object.

2. Description of the Related Art

There exist a host of well documented methods and systems for applications involving mutual transfer of information between a remote facility and a user for purposes such as user authentication, identification, or location identification. Some examples are:

1. Hardware security tokens such as wireless smart cards, USB tokens, Bluetooth tokens/cards, and electronic keys, that can interface to an authentication terminal (such as a PC, cell phone, or smart card reader). In this scheme, the user must carry these tokens around and use them to prove the user’s identity. In the information security business, these tokens are often referred to as “something you have”. The tokens can be used in combination with other security factors, such as passwords (“something you know”) and biometric devices (“something you are”) for what is called “multiple factor authentication”. Some leading

companies in the business of hardware security tokens include RSA Security, Inc., Safenet, Inc., and Aladdin, Inc.

2. The utilization of a mobile phone for authentication and related processes (such as purchase or information retrieval), where the phone itself serves as the hardware token, and the token is verified using well known technology called “digital certificate” or “PKI technology”. In this case, the authentication server communicates with the CPU on the phone to perform challenge-response authentication sequences. The phone can be used both for the identification of the user, and for the user to make choices regarding the service or content he wishes to access. For example, this authentication method is used in the WAP browsers of some current day phones via digital certificates internal to the phone, to authenticate the WAP site and the phone to each other.

3. Authentication by usage of the cellular networks’ capability to reliably detect the phone number (also called the “MSISDN”) and the phone hardware number (also called the “IMEI”) of a cellular device. For example, suppose an individual’s MSISDN number is known to be +1-412-333-942-1111. That individual can call a designated number and, via an IVR system, type a code on the keypad. In this case, cellular network can guarantee with high reliability that the phone call originated from a phone with this particular MSISDN number – hence from the individual’s phone. Similar methods exist for tracing the MSISDN of SMS messages sent from a phone, or of data transmission (such as, for example, Wireless Session Protocol “WSP” requests).

These methods and systems can be used for a wide variety of applications, including:

1. Access control for sensitive information or for physical entrance to sensitive locations.
2. Remote voting to verify that only authorized users can vote, and to ensure that each user votes only once (or up to a certain amount of times as permitted). Such usage is widespread currently in TV shows, for example, in rating a singer in a contest.

3. Password completion. There exist web sites, web services and local software utilities, that allow a user to bypass or simplify the password authorization mechanism when the user has a hardware token.

4. Charging mechanism. In order to charge a user for content, the user's identity must be reliably identified. For example, some music and streaming video services use premium SMS sent by the user to a special number to pay for the service – the user is charged a premium rate for the SMS, and in return gets the service or content. This mechanism relies on the reliability of the MSISDN number detection by the cellular network.

Although there are a multitude of approaches to providing authentication or authenticated services, these approaches have several key shortcomings, which include:

1. Cost and effort of providing tokens. Special purpose hardware tokens cost money to produce, and additional money to send to the user. Since these tokens serve only the purpose of authentication, they tend to be lost, forgotten or transferred between people. Where the tokens are provided by an employer to an employee (which is frequently but not always the specific use of such tokens), the tokens are single purpose devices provided to the employee with no other practical benefits to the employee (as compared to, for example, cellular phones which are also sometimes provided by the employer but which serve the employee for multiple purposes). It is common for employees to lose tokens, or forget them when they travel. For all of these reasons, hardware tokens, however they are provided and whether or not provided in an employment relationship, need to be re-issued often. Any organization sending out or relying upon such tokens must enforce token revocation mechanisms and token re-issuance procedures. The organization must spend money on the procedures as well as on the procurement and distribution of new tokens.

2. Limited flexibility of tokens. A particular token typically interface only to a certain set of systems and not to others – for example, a USB token cannot work with a TV screen, with a cellular phone or with any Web terminal/PC that lacks external USB access.

3. Complexity. The use of cellular devices with SMS or IVR mechanisms is typically cumbersome for users in many circumstances. The users must know which number to call, and they need to spend time on the phone or typing in a code. Additionally, when users must choose one of several options (e.g., a favorite singer out of a large number of alternatives) the choice itself by a numeric code could be difficult and error prone – especially if there are many choices. An implementation which does not currently exist but which would be superior, would allow the user to direct some pointing device at the desired option and press a button, similar to what is done in the normal course of web browsing.

4. Cost of service. Sending a premium SMS or making an IVR call is often more expensive than sending data packets (generally more expensive even than sending data packets of a data-rich object such as a picture).

5. Cost of service enablement. Additionally, the service provider must acquire from the cellular or landline telecom operator, at considerable expense, an IVR system to handle many calls, or a premium SMS number.

6. Difficulty in verification of user physical presence. When a user uses a physical hardware token in conjunction with a designated reader, or when the user types a password at a specific terminal, the user's physical presence at that point in time at that particular access point is verified merely by the physical act. The current scheme does not require the physical location of the sending device, and is therefore subject to user counterfeiting. For example, the user could be in a different location altogether, and type an SMS or make a call with the information provided to the user by someone who is at the physical location. (Presumably the person at the physical location would be watching the screen and reporting to the user what to type or where to call.) Thus, for example, in SMS based voting, users can "vote" to their favorite star in a show without actually watching the show. That is not the declared intention of most such shows, and defeats the purpose of user voting.

SUMMARY OF THE INVENTION

The present invention presents a method and system of enabling a user with an imaging device to conveniently send digital information appearing on a screen or in print to a remote server for various purposes related to authentication or service request.

5 The invention presents, in an exemplary embodiment, capturing an image of a printed object, transmitting the image to a remote facility, pre-processing the image in order to optimize the recognition results, searching the image for alphanumeric characters or other graphic designs, and decoding said alphanumeric characters and identification of the graphic designs from an existing database.

10 The invention also presents, in an exemplary embodiment, the utilization of the image recognition results of the image (that is, the alphanumeric characters and/or the graphic designs of the image) in order to facilitate dynamic data transmission from a display device to an imaging device. Thus, information can be displayed on the screen, imaged via the imaging device, and decoded into digital data. Such data transmission can serve any purpose
15 for which digital data communications exist. In particular, such data transmission can serve to establish a critical data link between a screen and the user's trusted communication device, hence facilitating one channel of the two channels required for one-way or mutual authentication of identity or transmission of encrypted data transmission.

 The invention also presents, in an exemplary embodiment, the utilization of the image
20 recognition results of the image in order to establish that the user is in a certain place (that is, the place where the specific object appearing in the image exists) or is in possession of a certain object.

 The invention also presents, in an exemplary embodiment a new and novel algorithm, which enables the reliable recognition of virtually any graphic symbol or design, regardless
25 of size or complexity, from an image of that symbol taken by a digital imaging device. Such

algorithm is executed on any computational facility capable of processing the information captured and sent by the imaging device.

BRIEF DESCRIPTION OF THE DRAWINGS

Various other objects, features and attendant advantages of the present invention will
5 become fully appreciated as the same become better understood when considered in
conjunction with the accompanying detailed description, the appended claims, and the
accompanying drawings, in which:

FIG. 1 is a block diagram of a prior art communication system for establishing the
identity of a user and facilitating transactions.

10 FIG. 2 is a flowchart diagram of a typical method of image recognition for a generic
two-dimensional object.

FIG. 3 is a block diagram of the different components of an exemplary embodiment
of the present invention.

15 FIG. 4 is a flowchart diagram of a user authentication sequence according to one
embodiment of the present invention.

FIG. 5 is a flow chart diagram of the processing flow used by the processing and
authentication server in the system in order to determine whether a certain two-dimensional
object appears in the image.

20 FIG. 6 is a flow chart diagram showing the determination of the template permutation
with the maximum score value, according to one embodiment of the present invention.

FIG. 7 is a diagram of the final result of a determination of the template permutation
with the maximum score value, according to one embodiment of the present invention.

FIG. 8 is an illustration of the method of multiple template matching which is one
algorithm used in an exemplary embodiment of the invention.

25 FIG. 9 is an example of an object to be recognized, and of templates of parts of that
object which are used in the recognition process.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

This invention presents an improved system and method for user interaction and data exchange between a user equipped with an imaging device and some server/service.

The system includes the following main components:

5 - A communication imaging device (wireless or wireline), such as a camera phone, a webcam with a WiFi interface, or a PDA (which may have a WiFi or cellular card). The device is capable of taking images, live video clips, or off-line video clips.

 - Client software on the device enabling the imaging and the sending of the multimedia files to a remote server. This software can be embedded software which is part of
10 the device, such as an email client, or an MMS client, or an H.324 video telephony client. Alternatively, the software can be downloaded software, either generic software such as blogging software (e.g., the Picoblogger™ product by Picostation™, or the Cognima Snap™ product by Cognima™, Inc.), or special software designed specifically and optimized for the imaging and sending operations.

15 - A remote server with considerable computational resources or considerable memory. “Considerable computational resources” in this context means that this remote server can perform calculations faster than the local CPU of the imaging device by at least one order of magnitude. Thus the user’s wait time for completion of the computation is much smaller when such a remote server is employed. “Considerable memory” in this context means that
20 the server has a much larger internal memory (the processor’s main memory or RAM) than the limited internal memory of the local CPU of the imaging device. The remote server’s considerable memory allows it to perform calculations that the local CPU of the imaging device cannot perform due to memory limitations of the local CPU. The remote server in this context will have considerable computational resources, or considerable memory, or both.

25 - A display device, such as a computer screen, cellular phone screen, TV screen, DVD player screen, advertisement board, or LED display. Alternatively, the display device can be

just printed material, which may be printed on an advertisement board, a receipt, a newspaper, a book, a card, or other physical medium.

The method of operation of the system may be summarized as follows:

5 - The display device shows an image or video clip (such as a login screen, a voting menu, or an authenticated purchase screen) that identifies the service, while also showing potentially other content (such as an ongoing TV show, or preview of a video clip to be loaded.).

10 - The user images the display with his portable imaging device, and the image is processed to identify and decode the relevant information into a digital string. Thus, a de-facto one way communication link is established between the display device and the user's communication device, through which digital information is sent.

- The information decoded in the previous stage is used for various purposes and applications, such as for example two way authentication between the user and the remote service.

15 Figure 1 illustrates a typical prior art authentication system for remote transactions. A server **100** which controls access to information or services, controls the display of a web browser **101** running in the vicinity of the user **102**. The user has some trusted security token **103**. In some embodiments, the token **103** is a wireless device that can communicate through a communication network **104** (which may be wireless, wireline, optical, or any other
20 network that connects two or more non-contiguous points). The link **105** between the server and the web browser is typically a TCP/IP link. The link **106** between the web browser and the user is the audio/visual human connectivity between the user and the browser's display. The link **107** between the user and the token denotes the user-token interface, which might be a keypad, a biometric sensor, or a voice link. The link **108** between the token and the web
25 browser denotes the token's interaction channel based on infra red, wireless, physical electric connection, acoustic, or other methods to perform a data exchange between the token **103** and

the web browsing device **101**. The link **109** between the token and the wireless network can be a cellular interface, a WiFi interface, a USB connector, or some other communication interface. The link **110** between the communication network and the server **100** is typically a TCP/IP link.

5 The user **102** reads the instructions appearing on the related Web page on browser **101**, and utilizes some authentication token **103** in order to validate the user's identity and/or the identity and validity of the remote server **100**. The token can be, for example, one of the devices mentioned in the Description of the Related Art, such as a USB token, or a cellular phone. The interaction channel **107** of the user with the token can involve the user typing a
10 password at the token, reading a numeric code from the token's screen, or performing a biometric verification through the token. The interaction between the token **103** and the browser **101** is further transferred to the remote server **100** for authentication (which may be performed by comparison of the biometric reading to an existing database, password verification, or cryptographic verification of a digital signature). The transfer is typically
15 done through the TCP/IP connection **105** and through the communication network **104**.

 The key factor enabling the trust creation process in the system is the token **103**. The user does not trust any information coming from the web terminal **101** or from the remote server **100**, since such information may have been compromised or corrupted. The token **103**, carried with the user and supposedly tamper proof, is the only device that can signal to
20 the user that the other components of the system may be trusted. At the same time, the remote server **100** only trusts information coming from the token **103**, since such information conforms to a predefined and approved security protocol. The token's existence and participation in the session is considered a proof of the user's identity and eligibility for the service or information (in which "eligible" means that the user is a registered and paying user
25 for service, has the security clearance, and meets all other criteria required to qualify as a person entitled to receive the service).

In the embodiments where the token **103** is a mobile device with wireless data communication capabilities, the communication network **104** is a wireless network, and may be used to establish a faster or more secure channel of communication between the token **103** and the server **100**, in addition to or instead of the TCP/IP channel **105**. For example, the server **100** may receive a call or SMS from the token **103**, where wireless communication network **104** reliably identifies for the server the cellular number of the token/phone. Alternatively, the token **103** may send an inquiry to the wireless communication network **104** as to the identity and eligibility of the server **100**.

A key element of the prior art are thus the communication links **106**, **107**, and **108**, between the web browser **101**, the user **102**, and the token **103**. These communication links require the user to manually read and type information, or alternatively require some form of communication hardware in the web browser device **101** and compatible communication hardware in the token **103**.

Figure **2** illustrates a typical prior art method of locating an object in a two-dimensional image and comparing it to a reference in order to determine if the objects are indeed identical. A reference template **200** (depicted in an enlarged view for clarity) is used to search an image **201** using the well known and established technology of “normalized cross correlation method” (also known as “NCC”). Alternatively, other similarity measures such as the “sum of absolute differences” (“SAD”) and its variants may be used. The common denominator of all of these methods (NCC, SAD, and their variants) is that the methods get a fixed size template, compare that template to parts of the image **201** which are of identical size, and return a single number on some given scale where the magnitude of the number indicates whether or not there is a match between the template and the image. For example, a 1.0 would denote a perfect match and a 0.0 would indicate no match. Thus, if a “sliding window” of a size identical to the size of the template **200** is moved horizontally and vertically over the image **201**, and the results of the comparison method – the “match values”

(e.g. NCC, SAD) are registered for each position of the sliding window, a new “comparison results” image is created in which for each pixel the value is the result of the comparison of the area centered around this pixel in the image **201** with the template **200**. Typically, most pixel locations in the image **201** would yield low match values. The resulting matches, 5 determined by the matching operation **202** are displayed in elements **203**, **204**, and **205**. In the example shown in Figure 2, pixel location denoted in **203** (the center of the black square) has yielded a low match value (since the template and the image compared are totally dissimilar), pixel location denoted in **204** has yielded an intermediate match value (because both images include the faces and figures of people, although there is not a perfect match), 10 and the pixel location denoted in **205** has yielded a high match value. Therefore, application of a threshold criterion to the resulting “match values” image generates image **206**, where only in specific locations (here **207**, **208**, **209**) is there a non-zero value. Thus, image **206** is not an image of a real object, but rather a two dimensional array of pixel values, where each pixel’s value is the match. Finally, it should be noted that in the given example we would 15 expect the value at pixel **209** to be the highest since the object at this point is identical to the template.

The prior art methods are useful when the image scale corresponds to the template size, and when the object depicted in the template indeed appears in the image with very little change from the template. However, if there is any variation between the template and the 20 image, then prior art methods are of limited usefulness. For example, if the image scale or orientation are changed, and/or if the original object in the image is different from the template due to effects such as geometry or different lighting conditions, or if there are imaging optical effects such as defocusing and smearing, then in any of these cases the value at the pixel of the “best match” **209** could be smaller than the threshold or smaller than the 25 value at the pixel of the original “fair match” **208**. In such a case, there could be an incorrect

detection, in which the algorithm has erroneously identified the area around location **208** as containing the object depicted in the template **200**.

A further limitation of the prior art methods is that as the template **200** becomes larger (that is to say, if the object to be searched is large), the sensitivity of the match results to the effects described in the previous paragraph is increased. Thus, application of prior art methods is impractical for large objects. Similarly, since prior art methods lack sensitivity, they are less suitable for identification of graphically complicated images such as a complex graphical logo.

In typical imaging conditions of a user with an imaging device performing imaging of a screen or of printed material, the prior art methods fail for one or more of the deficiencies mentioned above. Thus, a new method and system are required to solve these practical issues, a method and system which are presented here as exemplary embodiments of the present invention.

In Figure 3, the main components of an exemplary embodiment of the present invention are described. As in the prior art described in Figure 1, a remote server **300** is used. (Throughout this application, the term "remote server" **300** means any combination of servers or computers.) The remote server **300** is connected directly to a local node **301**. (Throughout this application, the term "local node" **301** means any device capable of receiving information from the remote server and displaying it on a display **302**.) Examples of local nodes include a television set, a personal computer running a web browser, an LED display, or an electronic bulletin board.

The local node is connected to a display **302**, which may be any kind of physical or electronic medium that shows graphics or texts. In some embodiments, the local node **301** and display device **302** are a static printed object, in which case their only relation to the server **300** is off-line in the sense that the information displayed on **302** has been determined

by or is known by the server **300** prior to the printing and distribution process. Examples of such a local node include printed coupons, scratch cards, or newspaper advertisements.

The display is viewed by an imaging device **303** which captures and transmits the information on the display. There is a communication module **304** which may be part of the
5 imaging device **303** or which may be a separate transmitter, which sends the information (which may or may not have been processed by a local CPU in the imaging device **303** or in the communication module **304**) through a communication network **305**. In one embodiment, the communication network **305** is a wireless network, but the communication network may be also a wireline network, an optical network, a cable network, or any other network that
10 creates a communication link between two or more nodes that are not contiguous.

The communication network **305** transmits the information to a processing and authentication server **306**. The processing and authentication server **306** receives the transmission from the communication network **305** in whatever degree of information has been processed, and then completes the processing to identify the location of the display, the
15 time the display was captured, and the identity of the imaging device (hence, also the service being rendered to the user, the identity of the user, and the location of the user at the time the image or video clip was captured by the imaging device). The processing and authentication server **306** may initiate additional services to be performed for the user, in which case there will be a communication link between that server **306** and server **300** or the local node **301**,
20 or between **306** and the communication module **304**.

The exact level of processing that takes place at **304**, **305**, and **306** can be adapted to the desired performance and the utilized equipment. The processing activities may be allocated in any combination among **304**, **305**, and **306**, depending on factors such as the processing requirements for the specific information, the processing capabilities of these
25 three elements of the system, and the communication speeds between the various elements of the system. As an example, components **303** and **304** could be parts of a 3G phone making a

video call through the a cellular network **305** to the server **306**. In this example, video frames reach **306** and must be completely analyzed and decoded there, at server **306**, to decode the symbols, alphanumerics and/or machine codes in the video frames. An alternative example would be a “smartphone” (which is a phone that can execute local software) running some decoding software, such that the communication module **304** (which is a smartphone in this example) performs symbol decoding and sends to server **306** a completely parsed digital string or even the results of some cryptographic decoding operation on that string.

In Figure 3, a communication message has been transmitted from server **300** to the processing and authentication server **306** through the chain of components **301**, **302**, **303**, **304**, and **305**. Thus, one key aspect of the current invention, as compared to the prior art depicted in Figure 1, is the establishment of a new communication channel between the server **300** and the user’s device, composed of elements **303** and **304**. This new channel replaces or augments (depending on the application) the prior art communication channels **106**, **107**, and **108**, depicted in Figure 1.

In Figure 4, a method of operative flow of a user authentication sequence is shown. In stage **400**, the remote server **300** prepares a unique message to be displayed to a user who wishes to be authenticated, and sends that message to local node **301**. The message is unique in that at a given time only one such exact message is sent from the server to a single local node. This message may be a function of time, presumed user’s identity, the local node’s IP address, the local node’s location, or other factors that make this particular message singular, that is, unique. Stage **400** could also be accomplished in some instances by the processing and authentication server **306** without affecting the process as described here.

In stage **401**, the message is presented on the display **302**. Then, in stage **402**, the user uses imaging device **303** to acquire an image of the display **302**. Subsequently, in stage **403**, this image is processed to recover the unique message displayed. The result of this recovery is some digital data string. Various examples of a digital data string could be an

alphanumeric code which is displayed on the display **302**, a URL, a text string containing the name of the symbol appearing on the display (for example "Widgets Inc. Logo"), or some combination thereof. This processing can take place within elements **304**, **305**, **306**, or in some combination thereof.

5 In stage **404**, information specific to the user is added to the unique message recovered in stage **403**, so that the processing and authentication server **306** will know who is the user that wishes to be authenticated. This information can be specific to the user (for example, the user's phone number or MSISDN as stored on the user's SIM card), or specific to the device the user has used in the imaging and communication process (such as, for
10 example, the IMEI of a mobile phone), or any combination thereof. This user-specific information may also include additional information about the user's device or location supplied by the communication network **305**.

In stage **405**, the combined information generated in stages **403** and **404** is used for authentication. In the authentication stage, the processing and authentication server **306**
15 compares the recovered unique message to the internal repository of unique messages, and thus determines whether the user has imaged a display with a valid message (for example, a message that is not older than two days, or a message which is not known to be fictitious), and thus also knows which display and local node the user is currently facing (since each local node receives a different message). In stage **405**, the processing and authentication
20 server **306** also determines from the user's details whether the user should be granted access from this specific display and local node combination. For example, a certain customer of a bank may be listed for remote Internet access on U.S. soil, but not outside the U.S. Hence, if the user is in front of an access login display in Britain, access will not be granted. Upon completion of the authentication process in **405**, access is either granted or denied in stage
25 **406**. Typically a message will be sent from server **306** to the user's display **302**, informing the user that access has been granted or denied.

In order to clarify further the nature and application of the invention, it would be valuable to consider several examples of the manner in which this invention may be used. The following examples rely upon the structure and method as depicted in Figures 3 and 4:

Example 1 of using the invention is user authentication. There is displayed **401** on the display **302** a unique, time dependent numeric code. The digits displayed are captured **403**, decoded (**403**, **404**, **405**, and **406**), and sent back to remote server **300** along with the user's phone number or IP address (where the IP address may be denoted by "X"). The server **300** compares the decoded digital string (which may be denoted as "M") to the original digits sent to local node **301**. If there is a match, the server **300** then knows for sure that the user holding the device with the phone number or IP address X is right now in front of display device **302** (or more specifically, that the imaging device owned or controlled by the user is right now in front of display device **302**). Such a procedure can be implemented in prior art by having the user read the digits displayed by the web browser **101** and manually type them on the token **103**. Alternatively in prior art, this information could be sent on the communication channel **108**. Some of the advantages of the invention over prior art, is that the invention avoids the need for additional hardware and avoids also the need for the user to type the information. In the embodiment of the invention described herein, therefore, the transaction is faster, more convenient, and more reliable than the manner in which transaction is performed according to prior art. Without limitation, the same purpose accomplished here with alphanumeric information could be accomplished by showing on the display **302** some form of machine readable code or any other two-dimensional and/or time changing figure which can be compared to a reference figure. Using graphic information instead of alphanumerics has another important security advantage, in that another person (not the user) watching the same display from the side will not be able to write down, type, or memorize the information for subsequent malicious use. A similar advantage could be achieved by using a very long alphanumeric string.

Example 2 of using the invention is server authentication. The remote server 300 displays 401 on the display 302 a unique, time dependent numeric code. The digits displayed appear in the image captured 403 by imaging device 303 and are decoded by server 306 into a message M (in which "M" continues to be a decoded digital string). The server 306 also
5 knows the user's phone number or IP address (which continues to be denoted by "X"). The server 306, has a trusted connection 307 with the server 300, and makes an inquiry to 300, "Did you just display message M on a display device to authenticated user X?" The server 300 sends transmits the answer through the communication network 305 to the processing and authentication server 306. If the answer is yes, the server 306 returns, via
10 communication network 305, to the user on the trusted communication module 304 an acknowledgement that the remote server 300 is indeed the right one. A typical use of the procedure described here would be to prevent ip-address spoofing, or prevent pharming/phishing. "Spoofing" works by confusing the local node about the IP address to which the local node is sending information. "Pharming" and "Phishing" attacks work by
15 using a valid domain name which is not the domain name of the original service, for example, by using www.widgetstrick.com instead of the legitimate service www.widgetsinc.com. All of these different attack schemes strive in the end to cause the user who is in front of local node 301 to send information and make operations while believing that the user is communicating with legitimate server 300 while in fact all the information is sent to a
20 different, malicious server. Without limitation, the server identification accomplished here with alphanumeric information, could be accomplished by showing on the display 302 some form of machine readable code or any other two-dimensional and/or time changing figure which can be compared to a reference figure.

Example 3 of using the invention is coupon loading or scratch card activation. The
25 application and mode of usage would be identical to Example 1 above, with the difference that the code printed on the card or coupon is fixed at the time of printing (and is therefore

not, as in Example 1, a decoded digital string). Again, advantages of the present invention over prior art would be speed, convenience, avoidance of the potential user errors if the user had to type the code printed on the coupon/card, and the potential use of figures or graphics that are not easily copied.

5 Example 4 of using the invention is a generic accelerated access method, in which the code or graphics displayed are not unique to a particular user, but rather are shared among multiple displays or printed matter. The server 300 still receives a trusted message from 306 with the user identifier X and the decoded message M (as is described above in Examples 1 and 3), and can use the message as an indication that the user is front of a display of M.

10 However, since M is shared by many displays or printed matters, the server 300 cannot know the exact location of the user. In this example, the exact location of the user is not of critical importance, but quick system access is of importance. Various sample applications would be content or service access for a user from a TV advertisement, or from printed advertisements, or from a web page, or from a product's packaging. One advantage of the invention is in

15 making the process simple and convenient for the user, avoiding a need for the user to type long numeric codes, or read complex instructions, or wait for an acknowledgment from some interactive voice response system. Instead, in the present invention the user just takes a picture of the object 403, and sends the picture somewhere else unknown to the user, where the picture will be processed in a manner also unknown to the user, but with quick and

20 effective system access.

As can be understood from the discussion of Figures 3 and 4, one aspect of the present invention is the ability of the processing software in 304 and/or 306 to accurately and reliably decode the information displayed 401 on the display device 302. As has been mentioned in the discussion of Figure 2, prior art methods for object detection and

25 recognition are not necessarily suitable for this task, in particular in cases where the objects to

be detected are extended in size and/or when the imaging conditions and resolutions are those typically found in portable or mobile imaging devices.

Figure 5 illustrates some of the operating principles of one embodiment of the invention. A given template, which represents a small part of the complete object to be searched in the image, is used for scanning the complete target image acquired by the imaging device 303. The search is performed on several resized versions of the original image, where the resizing may be different for the X,Y scale. Each combination of X,Y scales is given a score value based on the best match found for the template in the resized image. The algorithm used for determining this match value is described in the description of Figure 6 below.

The scaled images 500, 501, and 502, depict three potential scale combinations for which the score function is, respectively, above the minimum threshold, maximal over the whole search range, and below the minimum threshold. Element 500 is a graphic representation in which the image has been magnified by 20% on the y-scale. Hence, in element 500 the x-scale is 1.0 and y-scale is 1.2. The same notation applies for element 501 (in which the y-scale is 0.9) and element 502 (in which each axis is 0.8). These are just sample scale combinations used to illustrate some of the operating principles of the embodiment of the invention. In any particular transaction, any number and range of scale combinations could be used, balancing total run time on the one hand (since more scale combinations require more time to search) and detection likelihood on the other hand (since more scale combinations and a wider range of scales increase the detection probability).

Accordingly, in stage 503 the optimal image scale (which represents the image scale at which the image's scale is closest to the template's scale) is determined by first searching among all scales where the score is above the threshold (hence element 502 is discarded from the search, while elements 500 and 501 are included), and then choosing 501 as the optimal image scale. Alternatively, the optimal image scale may be determined by other score

functions, by a weighting of the image scales of several scale sets yielding the highest scores, and/or by some parametric fit to the whole range of scale sets based on their relative scores. In addition to searching over a range of image scales for the X and Y axes, the search itself could be extended to include image rotation, skewing, projective transformations, and other
5 transformations of the template.

In stage **504**, the same procedure performed for a specific template in stage **503** is repeated for other templates, which represent other parts of the full object. The scale range can be identical to that used in **503** or can be smaller, as the optimal image scale found in stage **503** already gives an initial estimate to the optimal image scale. For example, if at
10 stage **503** the initial search was for X and Y scale values between 0.5 to 1.5, and the optimal scale was at X=1.0, Y=0.9, then the search in stage **504** for other templates may be performed at a tighter scale range of between 0.9 and 1.1 for both the X and Y scales.

It is important to note that even at an “optimal scale” for a given template search, there may be more than one candidate location for that template in the image. A simple
15 example can be **Figure 2**. Although the best match is in element **205**, there is an alternative match in element **204**. Thus, in the general case, for every template there will be several potential locations in the image even in the selected “optimal scale”. This is because several parts of the image may be sufficiently similar to the template to yield a sufficiently high match value.

In stage **505**, the different permutations of the various candidates are considered to determine whether the complete object is indeed in the image. (This point is further explained in **Figure 6** and **Figure 7**.) Hence, if the object is indeed in the image, all of these templates should appear in the image with similar relative positions between them. Some
20 score function, further explained in the discussion of **Figures 6** and **7**, is used to rate the relative likelihood of each permutation, and a best match (highest score) is chosen in stage
25 **506**. Various score functions can be used, such as, for example, allowing for some template

candidates to be missing completely (e.g., no candidate for template number 3 has been located in the image).

In stage 507 the existence of the object in the image is determined by whether best match found in stage 506 has met exceeded some threshold match. If the threshold match has been met or exceeded, the a match is found and the logo (or other information) is identified 509. If the threshold is not met, then the match has not been found 508, and the process must be repeated until a match is found.

There are some important benefits gained by searching for various sub-parts of the complete object instead of directly searching for the complete object as is done in prior art.

10 For example:

- Parts of the object may be occluded, shadowed, or otherwise obscured, but nevertheless, as long as enough of the sub-templates are located in the image, the object's existence can be determined and identified.

- By searching for small parts of the object rather than for the whole object, the sensitivity of the system to small scale variations, lighting non-uniformity, and other geometrical and optical effects, is greatly reduced. For example, consider an object with a size of 200 by 200 pixels. In such an image, even a 1% scale error/difference between the original object and the object as it appears in the image could cause a great reduction in the match score, as it reflects a change in size of 2 pixels. At the same time, sub-templates of the full object, at a size of 20 by 20 pixels each, would be far less sensitive to a 1% scale change.

- A graphic object may include many areas of low contrast, or of complex textures or repetitive patterns. Such areas may yield large match values between themselves and shifted, rotated or rescaled versions of themselves. This will confuse most image search algorithms. At the same time, such an object may contain areas with distinct, high contrast patterns (such as, for example, an edge, or a symbol). These high contrast, distinct patterns would serve as good templates for the search algorithm, unlike the fuzzy, repetitive or low

contrast areas. Hence, the present invention allows the selection of specific areas of the object to be searched, which greatly increases the precision of the search.

- By searching for smaller templates instead of the complete object as a single template, the number of computations is significantly reduced. For example, a normalized cross correlation search for a 200 by 200 pixel object would be more than 100 times more computationally intensive than a similar normalized cross correlations search for a 20 by 20 sub template of that object.

Figures 6 and 7 illustrate in further detail the internal process of element 505. In stage 600, all candidates for all templates are located and organized into a properly labeled list. As an example, in a certain image, there may be 3 candidates for template #1, which are depicted graphically in Figure 7, within 700. The candidates are, respectively, 701 (candidate a for template #1, hence called 1a), 702 (candidate b for template #1, hence called 1b), and 703 (candidate c for template #1, hence called 1c). These candidates are labeled as 1a, 1b, and 1c, since they are candidates of template #1 only. Similarly 704 and 705 denote candidate locations for template #2 in the same image which are hence properly labeled as 2a and 2b. Similarly for template #3, in this example only one candidate location 706 has been located and labeled as 3a. The relative location of the candidates in the figure correspond to their relative locations in the original 2D image.

In stage 601, an iterative process takes place in which each permutation containing exactly one candidate for each template is used. The underlying logic here is the following: if the object being searched indeed appears in the image, then not only should the image include templates 1, 2, and 3, but in addition it should also include them with a well defined, substantially rigid geometrical relation among them. Hence, in the specific example, the potentially valid permutations used in the iteration of stage 601 are {1a,2a,3a}, {1a,2b,3a}, {1b,2a,3a}, {1b,2b,3a}, {1c,2a,3a}, {1bc,2a,3a}.

In stage **602**, the exact location of each candidate on the original image is calculated using the precise image scale at which it was located. Thus, although the different template candidates may be located at different image scales, for the purpose of the candidates' relative geometrical position assessment, they must be brought into the same geometric scale.

5 In stage **603**, the angles and distance among the candidates in the current permutation are calculated for the purpose of later comparing them to the angles and distances among those templates in the searched object.

As a specific example, Figure 7 illustrates the relative geometry of {1a,2b,3a}. Between each of the two template candidates there exists a line segment with specific
10 location, angle and length. In the example in Figure 7, these are, respectively, element **707** for 1a and 2b, element **708** for 2b and 3a, and element **709** for 1a and 2a.

In stage **604**, this comparison is performed by calculating a "score value" for each specific permutation in the example. Continuing with the specific example, the lengths, positions and angles of line segments **707**, **708**, and **709**, are evaluated by some mathematical
15 score function which returns a score value of how similar those segments are to the same segments in the searched object. A simple example of such a score function would be a threshold function. Thus, if the values of the distance and angles of **707**, **708**, and **709**, deviate from the nominal values by a certain amount, the score function will return a 0. If they do not so deviate, then the score function will return a 1. It is clear to those experienced
20 in the art of score function and optimization searches that many different score functions can be implemented, all serving the ultimate goal of identifying cases where the object indeed appears in the image and separating those cases from cases those where the object does not appear in the image.

In stage **605**, the score values obtained in all the potential permutations are compared
25 and the maximum score is used to determine if the object does indeed appear in the image. It is also possible, in some embodiments, to use other results and parameters in order to make

this determination. For example, an occurrence of too many template candidates (and hence many permutations) might serve as a warning to the algorithm that the object does not indeed appear in the image, or that multiple copies of the object are in the same image.

It should be understood that the reliance on specific templates implies that if those
5 templates are not reliably located in the image, or if the parts of the object belonging to those templates are occluded or distorted in some way (as for example by a light reflection), then in the absence of any workaround, some embodiments invention may not work optimally. A potential workaround for this kind of problem is to use many more templates, thereby improving robustness while increasing the run time of the algorithm.

10 It should also be understood that some embodiments of the invention are not completely immune to warping of the object. If, for example, the object has been printed on a piece of paper, and that piece of paper is imaged by the user in a significantly warped form, the relative locations and angles of the different template candidates will be also warped and the score function thus may not enable the detection of the object. This is a kind of problem
15 that is likely to appear in physical/printed, as opposed to electronic, media.

It should also be understood that some embodiments of the invention can be combined with other posterior criteria used to ascertain the existence of the object in the image. For example, once in stage 605 the maximum score value exceeds a certain threshold, it is possible to calculate other parameters of the image to further verify the object's
20 existence. One example would be criteria based on the color distribution or texture of the image at the points where presumably the object has been located.

Figure 8 illustrates graphically some aspects of the multi-template matching algorithm, which is one important algorithm used in an exemplary embodiment of the present invention (in processing stages 503 and 504). The multi-template matching algorithm is
25 based on the well known template matching method for grayscale images called "Normalized Cross Correlation" (NCC), described in Figure 2 and in the related prior art discussion. A

main deficiency of NCC is that for images with non-uniform lighting, compression artifacts, and/or defocusing issues, the NCC method yields many “false alarms” (that is, incorrect conclusions that a certain status or object appears) and at the same time fails to detect valid objects. The multi-template algorithm described as part of this invention in Figure 5, extends
5 the traditional NCC by replacing a single template for the NCC operation with a set of N templates, which represent different parts of an object to be located in the image. The templates **805** and **806** represent two potential such templates, representing parts of the digit “1” in a specific font and of a specific size. For each template, the NCC operation is performed over the whole image **801**, yielding the normalized cross correlation images **802**
10 and **803**. The pixels in these images have values between -1 and 1 , where a value of 1 for pixel (x,y) indicates a perfect match between a given template and the area in image **801** centered around (x,y) . At the right of **802** and **803**, respectively, sample one-dimensional cross sections of those images are shown, showing how a peak of 1 is reached exactly at a certain position for each template. One important point is that even if the image indeed has
15 the object to be searched for centered at some point (x,y) , the response peaks for the NCC images for various templates will not necessarily occur at the same point. For example, in the case displayed in Figure 8, there is a certain difference **804** of several pixels in the horizontal direction between the peak for template **805** and the peak for template **806**. These differences can be different for different templates, and the differences are taken into account by the
20 multi-template matching algorithm. Thus, after the correction of these deltas, all the NCC images (such as **802** and **803**) will display a single NCC “peak” at the same (x,y) coordinates which are also the coordinates of the center of the object in the image. For a real life image, the values of those peaks will not reach the theoretical “1.0” value, since the object in the image will not be identical to the template. However, proper score functions and thresholds
25 allow for efficient and reliable detection of the object by judicious lowering of the detection thresholds for the different NCC images. It should be stressed that the actual templates can

be overlapping, partially overlapping or with no overlap. Their size, relative position, and shape can be changed, as long as the templates continue to correspond to the same object that one wishes to locate in the image. Furthermore, masked NCC, which are well known extension of NCC, can be used for these templates to allow for non-rectangular templates.

5 As can be understood from the previous discussion, the results of the NCC operation for each sub-template out of N such sub-templates generates a single number per each pixel in the image (x,y). Thus, for each pixel (x,y) there are N numbers which must be combined in some form to yield a score function indicating the match quality. Let us denote by $T_i^A(x,y)$ the value of the normalized cross correlation value of sub-template i of the object "A" at pixel
 10 x,y in the image I. A valid score function then could be $f(x,y)=\text{Prod}_{i=1..N}T_i^A(x,y)$ – namely, the scalar product of these N values. Hence for example, if there is a perfect match between the object "A" and the pixels centered at (x_0,y_0) in the image I, then $T_i^A(x_0,y_0)=1.0$ for any i and our score function $f(x,y)=1$ at $\{x=x_0,y=y_0\}$. It is clear to someone familiar with the art of score function design and classification that numerous other score functions could be used,
 15 e.g. a weighted average of the N values, or a neural network where the N values are the input, or many others which could be imagined.

Thus, after the application of the chosen score function, the result of the multi-template algorithm is an image identical in size to the input image I, where the value of each pixel (x,y) is the score function indicating the quality of the match between the area centered
 20 around this pixel and the searched template.

It is also possible to define a score function for a complete image, indicating the likelihood that the image as a whole contains at least one occurrence of the searched template. Such a score function is used in stages 503 and 504 to determine the optimal image scale. A simple yet effective example of such a score function is $\max_{(x,y)}\{\text{Prod}_{i=1..N}T_i^A(x,y)\}$
 25 where (x,y) represents the set of all pixels in I. This function would be 1.0 if there is a perfect match between some part of the image I and the searched template. It is clear to

someone familiar with the art of score function design, that numerous other score functions could be used, such as, for example, a weighted sum of the values of the local score function for all pixels.

Figure 9 illustrates a sample graphic object 900, and some selected templates on it
5 901, 902, 903, 904, and 905. In one possible application of the present invention, to search for this object in a picture, the three templates 901, 902, and 903, are searched in the image, where each template in itself is searched using the multi-template algorithm described in Figure 8. After determination of the candidate locations for templates 901, 902, and 903 in Figure 7 (template 901 candidates are 701, 702, and 703, template 902 candidates are 704
10 and 705, and template 903 candidate is 706), the relative distances and angles for each potential combination of candidates (one for each template, e.g. {701, 705, 706}) are compared to the reference distances and angles denote by line segments 906, 907, and 908. Some score function is used to calculate the similarity between line segments 707, 708, and 709 on the one hand, and line segments 906, 907, and 908 on the other hand. Upon testing all
15 potential combinations (or a subset thereof), the best match with the highest score is used in stage 507 to determine whether indeed the object in the image is our reference object 900.

It is clear to someone familiar with the art of object recognition that the reliability, run time, and hit/miss ratios of the algorithm described in this invention can be modified based on the number of different templates used, their sizes, the actual choice of the templates, and the
20 score functions. For example, by employing all five templates 901, 902, 903, 904, and 905, instead of just three templates, the reliability of detection would increase, yet the run time would also increase. Similarly, template 904 would not be an ideal template to use for image scale determination or for object search in general, since it can yield a good match with many other parts of the searched object as well as with many curved lines which can appear in any
25 image. Thus, the choice of optimal templates can be critical to reliable recognition using a

minimum number of templates (although adding a non-optimal template such as 904 to a list of templates does not inherently reduce the detection reliability).

It is also clear from the description of the object search algorithm, that with suitably designed score functions for stages 505 and 506, it is possible to detect an object even if one
5 or more of the searched templates are not located in the image. This possibility enables the recognition of objects even in images where the objects are partially occluded, weakly illuminated, or covered by some other non-relevant objects. Some specific practical examples of such detection include the following:

Example 1: When imaging a CRT display, the exposure time of the digital imaging
10 device coupled to the refresh times of the screen can cause vertical banding to appear. Such banding cannot be predicted in advance, and thus can cause part of the object to be absent or to be much darker than the rest of the object. Hence, some of the templates belonging to such an object may not be located in the image. Additionally, the banding effect can be reduced significantly by proper choices of the colors used in the object and in its background.

15 Example 2: During the encoding and communication transmission stages between components 304 and 305, errors in the transmission or sub-optimal encoding and compression can cause parts of the image of the object to be degraded or even completely non-decodable. Therefore, some of the templates belonging to such an object may not be located in the image.

20 Example 3: when imaging printed material in glossy magazines, product wrappings or other objects with shiny surfaces, some parts of the image may be saturated due to reflections from the surrounding light sources. Thus in those areas of the image it may be impossible or very hard to detect object features and templates. Therefore, some of the templates belonging to such an object may not be located in the image.

25 Hence, the recognition method and system outlined in the present invention, along with other advantages, enable increased robustness to such image degradation effects.

Another important note is that embodiments of the present invention as described here allows for any graphical object – be it alphanumeric, a drawing, a symbol, a picture, or other, to be recognized. In particular, even machine readable codes can be used as objects for the purpose of recognition. For example, a specific 2D barcode symbol defining any specific URL, as for example the URL <http://www.dspv.net>, could be entered as an object to be searched.

Since different potential objects can be recognized using the present invention, it is also possible to use animations or movies where specific frames or stills from the animation or movie are used as the reference objects for the search. For example, the opening shot of a commercial could be used as a reference object, where the capturing of the opening shot of the image indicates the user's request to receive information about the products in this commercial.

The ability to recognize different objects also implies that a single logo with multiple graphical manifestations can be entered in the authentication and processing server's database as different objects all leading to a unified service or content. Thus, for example, all the various graphical designs of the logo of a major corporation could be entered to point to that corporation's web site.

By establishing a communication link based on visual information between a display or printed matter and a portable imaging device (which is one embodiment of imaging device), embodiments of the present invention enable a host of different applications in addition to those previously mentioned in the prior discussion. Some examples of such applications are:

- Product Identification for price comparison/information gathering: The user sees a product (such as a book) in a store, with specific graphics on it (e.g., book cover). The user takes a picture/video of the identifying graphics on the product. Based on

code/name/graphics of the product, the user receives information on the price of this product, its features, its availability, information to order it, etc.

- URL launching. The user snaps a photo of some graphic symbol (e.g., a company's logo) and later receives a WAP PUSH message for the relevant URL.

5 - Prepaid card loading or purchased content loading. The user takes a photo of the recently purchased pre-paid card, and the credit is charged to his/her account automatically. The operation is equivalent to currently inputting the prepaid digit sequence through an IVR session or via SMS, but the user is spared from actually reading the digits and typing them one by one.

10 - Status inquiry based on printed ticket: The user takes a photo of a lottery ticket, a travel ticket, etc., and receives back the relevant information, such as winning status, flight delayed/on time, etc. The graphical and/or alphanumeric information on the ticket is decoded by the system, and hence triggers this operation.

15 - User authentication for Internet shopping: When the user makes a purchase, a unique code is displayed on the screen and the user snaps a photo, thus verifying his identity via the phone. Since this code is only displayed at this time on this specific screen, the photo taken by the user represents a proof of the user's location, which, coupled to the user's phone number, create reliable location-identity authentication.

20 - Location Based Coupons: The user is in a real brick and mortar store. Next to each counter, there is a small sign/label with a number/text on it. The user snaps a photo of the label and gets back information, coupons, or discounts relevant to the specific clothes items (jeans, shoes, etc.) in which he is interested. The label in the store contains an ID of the store and an ID of the specific display the user is next to. This data is decoded by the server and sent to the store along with the user's phone ID.

25 - Digital signatures for payments, documents, or identities. A printed document (such as a ticket, contract, or receipt) is printed together with a digital signature (such as a number

with 20-40 digits) on it. The user snaps a photo of the document and the document is verified by a secure digital signature printed in it. A secure digital signature can be printed in any number of formats, such as, for example, a 40-digit number, or a 20-letter word. This number can be printed by any printer. This signature, once converted again to numerical
5 form, can securely and precisely serve as a standard, legally binding digital signature for any document.

- Catalog ordering/purchasing: The user is leafing through a catalogue. He snaps a photo of the relevant product with the product code printed next to it, and this action is equivalent to an “add to cart operation”. The server decodes the product code and the
10 catalogue ID from the photo, and then sends the information to the catalogue company’s server, along with the user’s phone number.

- Business Card exchange: The user snaps a photo of a business card. The details of the business card, possibly in VCF format, are sent back to the user’s phone. The server identifies the phone numbers on the card, and using the carrier database of phone numbers,
15 identifies the contact details of the relevant cellular user. These details are wrapped in the proper “business card” format and sent to the user.

- Coupon Verification: A user receives to his phone, via SMS, MMS, or WAP PUSH, a coupon. At the POS terminal (or at the entrance to the business using a POS terminal) he shows the coupon to an authorized clerk with a camera phone, who takes a picture of the
20 user’s phone screen to verify the coupon. The server decodes the number/string displayed on the phone screen and uses the decoded information to verify the coupon.

ADDENDUM B:**SYSTEM AND METHOD OF ENABLING A CELLULAR/WIRELESS DEVICE
WITH IMAGING CAPABILITIES TO DECODE PRINTED ALPHANUMERIC
CHARACTERS****CROSS-REFERENCE TO RELATED APPLICATIONS**

5 This application claims the benefit of U.S. Provisional Application Serial Number 60/625,632, filed on November 8, 2004, entitled, "System and Method of Enabling a Cellular/Wireless Device with Imaging Capabilities to Decode Printed Alphanumeric Characters", which is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION**10 1. Field of the Invention**

 The present invention relates generally to digital imaging technology, and more specifically it relates to optical character recognition performed by an imaging device which has wireless data transmission capabilities. This optical character recognition operation is done by a remote computational facility, or by dedicated software or hardware resident on the
15 imaging device, or by a combination thereof. The character recognition is based on an image, a set of images, or a video sequence taken of the characters to be recognized. Throughout this patent, "character" is a printed marking or drawing, "characters" refers to "alphanumeric characters", and "alphanumeric" refers to representations which are alphabetic, or numeric, or graphic (typically with an associated meaning, including, for example, traffic signs in which
20 shape and color convey meaning, or the smiley picture, or the copyright sign, or religious markings such as the Cross, the Crescent, the Star of David, and the like) or symbolic (for example, signs such as +, -, =, \$, or the like, which represent some meaning but which are not

in themselves alphabetic or numeric, or graphic marks or designs with an associated meaning), or some combination of the alphabetic, numeric, graphic, and symbolic.

2. Description of the Related Art

Technology for automatically recognizing alphanumeric characters from fixed fonts using scanners and high-resolution digital cameras has been in use for years. Such systems, generally called OCR (Optical Character Recognition) systems, are typically comprised of:

1. A high-resolution digital imaging device, such as a flatbed scanner or a digital camera, capable of imaging printed material with sufficient quality.

2. OCR software for converting an image into text.

3. A hardware system on which the OCR software runs, typically a general purpose computer, a microprocessor embedded in a device or on a remote server connected to the device, or a special purpose computer system such as those used in the machine vision industry.

4. Proper illumination equipment or setting, including, for example, the setup of a line scanner, or illumination by special lamps in machine vision settings.

Such OCR systems appear in different settings and are used for different purposes. Several examples may be cited. One example of such a purpose is conversion of page-sized printed documents into text. These systems are typically comprised of a scanner and software running on a desktop computer, and are used to convert single or multi-page documents into text which can then be digitally stored, edited, printed, searched, or processed in other ways.

Another example of such a purpose is the recognition of short printed numeric codes in industrial settings. These systems are typically comprised of a high end industrial digital camera, an illumination system, and software running on a general purpose or proprietary computer system. Such systems may be used to recognize various machine parts, printed circuit boards, or containers. The systems may also be used to extract relevant information

about these objects (such as the serial number or type) in order to facilitate processing or inventory keeping. The VisionPro™ optical character verification system made by Cognex™ is one example of such a product.

5 A third example of such a purpose is recognition of short printed numeric codes in various settings. These systems are typically comprised of a digital camera, a partial illumination system (in which “partial” means that for some parts of the scene illumination is not controlled by this system, such as, for example, in the presence of outdoor lighting may exist in the scene), and software for performing the OCR. A typical application of such systems is License Plate Recognition, which is used in contexts such as parking lots or tolled
10 highways to facilitate vehicle identification. Another typical application is the use of dedicated handheld scanning devices for performing scanning, OCR, and processing (e.g., translation to a different language) - such as the Quicktionary™ OCR Reading pen manufactured by Seiko which is used for the primary purpose of translating from one language to another language.

15 A fourth example of such a purpose is the translation of various sign images taken by a wireless PDA, where the processing is done by a remote server (such as, for example, the Infoscope™ project by IBM™). In this application, the image is taken with a relatively high quality camera utilizing well-known technology such as a Charge Couple Device (CCD) with variable focus. With proper focusing of the camera, the image may be taken at long range
20 (for a street sign, for example, since the sign is physically much larger than a printed page, allowing greater distance between the object and the imaging device), or at short range (such as for a product label). The OCR processing operation is typically performed by a remote server, and is typically reliant upon standard OCR algorithms. Standard algorithms are sufficient where the obtained imaging resolution for each character is high, similar to the
25 quality of resolution achieved by an optical scanner.

Although OCR is used in a variety of different settings, all of the systems currently in use rely upon some common features. These features would include the following:

First, these systems rely on a priori known geometry and setting of the imaged text. This known geometry affects the design of the imaging system, the illumination system, and the software used. These systems are designed with implicit or explicit assumptions about the physical size of the text, its location in the image, its orientation, and/or the illumination geometry. For example, OCR software using input from a flatbed scanner assumes that the page is oriented parallel to the scanning direction, and that letters are uniformly illuminated across the page as the scanner provides the illumination. The imaging scale is fixed since the camera/sensor is scanning the page at a very precise fixed distance from the page, and the focus is fixed throughout the image. As another example, in industrial imaging applications, the object to be imaged typically is placed at a fixed position in the imaging field (for example, where a microchip to be inspected is always placed in the middle of the imaging field, resulting in fixed focus and illumination conditions). A third example is that license plate recognition systems capture the license plate at a given distance and horizontal position (due to car structure), and license plates themselves are at a fixed size with small variation. A fourth example is the street sign reading application, which assumes imaging at distances of a couple of feet or more (due to the physical size and location of a street sign), and hence assumes implicitly that images are well focused on a standard fixed-focus camera.

Second, the imaging device is a “dedicated one” (which means that it was chosen, designed, and placed for this particular task), and its primary or only function is to provide the required information for this particular type of OCR.

Third, the resulting resolution of the image of the alphanumeric characters is sufficient for traditional OCR methods of binarization, morphology, and/or template matching, to work. Traditional OCR methods may use any combination of these three types of operations and criteria. These technical terms mean the following:

- "Binarization" is the conversion of a gray scale or color image into a binary one. Grey becomes pixels, which are exclusively (0) or (1). Under the current art, grayscale images captured by mobile cameras from short distances are too fuzzy to be processed by binarization. Algorithms and hardware systems that would allow binarization processing for such images or an alternative method would be improvement in the art, and these are one object of the present invention.

- "Morphology" is a kind of operation that uses morphological data known about the image to decode that image. Most of the OCR methods in the current art perform part or all of the recognition phase using morphological criteria. For example, consecutive letters are identified as separate entities using the fact that they are not connected by contiguous blocks of black pixels. Another example is that letters can be recognized based on morphological criteria such as the existence of one or more closed loops as part of a letter, and location of loops in relation to the rest of the pixels comprising the letter. For example, the numeral "0" (or the letter O) could be defined by the existence of a closed loop and the absence of any protruding lines from this loop. When the images of characters are small and fuzzy, which happens frequently in current imaging technology, morphological operations cannot be reliably performed. Algorithms and hardware systems that would allow morphology processing or an alternative method for such images, would be improvement in the art, and these are one object of the present invention

- "Template Matching" is a process of mathematically comparing a given image piece to a scaled version of an alphanumeric character (such as, for example, the letter "A") and giving the match a score between 0 and 1, where 1 would mean a perfect fit. These methods are used in some License Plate Recognition (LPR) systems, where the binarization and morphology operations are not useful due to the small number of pixels for the character. However, if the image is blurred, which may be the case is the image has alternate light and shading, or where number of pixels for a character is very small, template matching will also

fail, given current algorithms and hardware systems. Conversely, algorithms and hardware systems that would allow template matching in cases of blurred images or few pixels per character, would be an improvement in the art, and these are one object of the present invention.

5 Fourth, typically the resolution required by current systems is of on the order of 16 or more pixels on the vertical side of the characters. For example, the technical specifications of a modern current product such as the "Camreader"TM by MediaseekTM indicate a requirement for the imaging resolution to provide at least 16 pixels at the letter height for correct recognition. It should be stressed that the minimum number of pixels require for recognition
10 is not a hard limit. Some OCR systems, in some cases, may recognize characters with pixels below this limit, while other OCR systems, in other cases, will fail to recognize characters even above this limit. Although the point of degradation to failure is not clear in all cases, current art may be characterized such that almost all OCR systems will fail in almost always cases when where the character height of the image is on the order of 10 pixels or less, and
15 almost all OCR systems in almost cases will succeed in recognition where the character height of the image is on the order of 25 pixels or more. Where text is relatively condensed, character heights are relatively short, and OCR systems in general will have great difficulty decoding the images. Alternatively, when the image suffers from fuzziness due to de-
20 focusing (which can occur in, for example, imaging from a small distance using a fixed focus camera) and/or imager movement during imaging, the effective pixel resolution would also decrease below the threshold for successful OCR. Thus, when the smear of a point object is larger than one pixel in the image, the point smear function (PSF) should replace the term pixel in the previous threshold definitions.

Fifth, current OCR technology typically does not, and cannot, take into consideration
25 the typical severe image de-focusing and JPEG compression artifacts which are frequently encountered in a wireless environment. For example, the MediaSeekTM product runs on a cell

phone's local CPU (and not on a remote server). Hence, such a product can access the image in its non-transmitted, pre-encoded, and pristine form. Wireless transmission to a remote server (whether or not the image will be re-transmitted ultimately to a remote location) creates the vulnerabilities of de-focusing, compression artifacts, and transmission
5 degradation, which are very common in a wireless environment.

Sixth, current OCR technology works badly, or not at all, on what might be called "active displays" showing characters, that is, for example, LED displays, LCD displays, CRTs, plasma displays, and cell phone displays, which are not fixed but which have changing information due to type and nature of the display technology used.

10 Seventh, even apart from the difficulties already noted above, particularly the difficulties of wireless de-focusing and inability to deal with active display, OCR systems typically cannot deal with the original images generated by the digital cameras attached to wireless devices. Among other problems, digital cameras in most cases suffer from the following difficulties. First, their camera optics are fixed focus, and cannot image properly at
15 distances of less than approximately 20 centimeters. Second, the optical components are often minimal or of low quality, which causes inconsistency of image sharpness, which makes OCR according to current technology very difficult. For example, the resolution of the imaging sensor is typically very low, with resolutions ranging from 1.3 Megapixel at best down to VGA image size (that is, 640 by 480 or roughly 300,000 pixels) in most models.
20 Some models even have CIF resolution sensors (352 by 288, or roughly 100,000 pixels). Even worse, the current existing standard for 3G (Third Generation cellular) video-phones dictates a transmitted imaging resolution of QCIF (176 by 144 pixels). Third, due to the low sensitivity of the sensor and the lack of a flash (or insufficient light emitted by the existing flash), the exposure times required in order to yield a meaningful image in indoor lighting
25 conditions are relatively large. Hence, when an image is taken indoors, the hand

movement/shake of the person taking the image typically generates motion smear in the image, further reducing the image's quality and sharpness.

SUMMARY OF THE INVENTION

The present invention presents a method for decoding printed alphanumeric characters
5 from images or video sequences captured by a wireless device, the method comprising, in an
exemplary embodiment, pre-processing the image or video sequence to optimize processing
in all subsequent steps, searching one or more grayscale images for key alphanumeric
characters on a range of scales, comparing the key alphanumeric values to a plurality of
template in order to determine the characteristics of the alphanumeric characters, performing
10 additional comparisons to a plurality of templates to determine character lines, line edges,
and line orientation, processing information from prior steps to determine the corrected scale
and orientation of each line, recognizing the identity of each alphanumeric character in string
of such characters, and decoding the entire character string in digitized alphanumeric format.
Throughout this patent, "printed" is used expansively to mean that the character to be imaged
15 is captured on a physical substance (as by, for example, the impression of ink on a paper or a
paper-like substance, or by engraving upon a slab of stone), or is captured on a display device
(such as LED displays, LCD displays, CRTs, plasma displays, or cell phone displays).
"Printed" also includes typed, or generated automatically by some tool (whether the tool be
electrical or mechanical or chemical or other), or drawn whether by such a tool or by hand.

20 The present invention also presents a system for decoding printed alphanumeric
characters from images or video sequences captured by a wireless device, the system
comprising, in a exemplary embodiment, an object to be imaged or to be captured by video
sequence, that contains within it alphanumeric characters, a wireless portable device for
capturing the image video sequence, and transmitting the captured image or video sequence
25 to a data network, a data network for receiving the image or video sequence transmitted by
the wireless portable device, and for retransmitting it to a storage server, a storage receiver

for receiving the retransmitted image or video sequence, for storing the complete image or video sequence before processing, and for retransmitting the stored image or video sequence to a processing server, and a processing server for decoding the printed alphanumeric characters from the image or video sequence, and for transmitting the decoded characters to
5 an additional server.

The present invention also presents a processing server within a telecommunication system for decoding printed alphanumeric characters from images or video sequences captured by a wireless device, the processing server comprising, in an exemplary embodiment, a server for interacting with a plurality of storage servers, a plurality of
10 content/information servers, and a plurality of wireless messaging servers, within the telecommunication system for decoding printed alphanumeric characters from images, the server accessing image or video sequence data sent from a data network via a storage server, the server converting the image or video sequence data into a digital sequence of decoded alphanumeric characters, and the server communicating such digital sequence to an additional
15 server.

The present invention also presents a computer program product, comprising a computer data signal in a carrier wave having computer readable code embodied therein for causing a computer to perform a method comprising, in an exemplary embodiment, pre-processing an alphanumeric image or video sequence, searching on a range of scales for key
20 alphanumeric characters in the image or sequence, determining appropriate image scales, searching for character lines, line edges, and line orientations, correcting for the scale and orientation, recognizing the strings of alphanumeric characters, and decoding the character strings.

BRIEF DESCRIPTION OF THE DRAWINGS

25 Various other objects, features and attendant advantages of the present invention will become fully appreciated as the same become better understood when considered in

conjunction with the accompanying detailed description, the appended claims, and the accompanying drawings, in which:

FIG. 1 is a block diagram of a prior art OCR system which may be implemented on a mobile device.

5 FIG. 2 is a flowchart diagram of the processing steps in a prior art OCR system.

FIG. 3 is a block diagram of the different components of an exemplary embodiment of the present invention.

FIG. 4 is flow chart diagram of the processing flow used by the processing server in the system in order to decode alphanumeric characters in the input.

10 FIG. 5 is an illustration of the method of multiple template matching which is one algorithm in an exemplary embodiment of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

This invention presents an improved system and method for performing OCR for images and/or video clips taken by cameras in phones or other wireless devices.

15 The system includes the following main components:

1. A wireless imaging device, which may be a camera phone, a webcam with a WiFi interface, a PDA with a WiFi or cellular card, or some such similar device. The device is capable of taking images or video clips (live or off-line).

2. Client software on the device enabling the imaging and sending of the multimedia files to a remote server. This client software may be embedded software which is part of the device, such as, for example, an email client, or an MMS client, or an H.324 Video telephony client. Alternatively, this client software may be downloaded software, either generic software such as blogging software (for example, the PicobloggerTM product by PicostationTM), or special software designed specifically and optimized for the OCR operation.

20

25

3. A remote server with considerable computational resources. In this context, “considerable” means that the remote server meets either of two criteria. First, the server may perform calculations faster than the local CPU of the imaging device by at least one order in magnitude, that is, 10 times or more faster than the ability of the local CPU. Second, the remote server may be able to perform calculations that the local CPU of the imaging device is totally incapable of due to other limitations, such as limitation of memory or limitation of battery power.

The method of operation of the system may be summarized as follows:

1. The user uses the client software running on the imaging device to acquire an image/video clip of printed alphanumeric information. (In this context, and throughout the application, “alphanumeric information” means information which is wholly numeric, or wholly alphabetic, or a combination of numeric and alphabetic.) This alphanumeric information can be printed on paper (such as, for example, a URL on an advertisement in a newspaper), or printed on a product (such as, for example, the numerals on a barcode printed on a product’s packaging), or displayed on a display (such as a CRT, an LCD display, a computer screen, a TV screen, or the screen of another PDA or cellular device).

2. This image/clip is sent to the server via wireless networks or a combination of wireline and wireless networks. For example, a GSM phone may use the GPRS/GSM network to upload an image, or a WiFi camera may use the local WiFi WLAN to send the data to a local base station from which the data will be further sent via a fixed line connection.

3. The server, once the information arrives, performs a series of image processing and/or video processing operations to find whether alphanumeric characters are indeed contained in the image/video clip. If they are, server extracts the relevant data and converts it into an array of characters. In addition, the server retains the relative positions of

those characters as they appear in the image/video clip, and the imaging angle/distance as measured by the detection algorithm.

4. Based on the characters obtained in the prior step, and based potentially on other information that is provided by the imaging device, and/or resident on external
5 databases, and/or stored in the server itself, the server may initiate one of several applications located on the server or on remote separate entities. Extra relevant information used for this stage may include, for example, the physical location of the user (extracted by the phone's
GPS receiver or by the carrier's Location Based Services-LBS), the MSISDN (Mobile
International Subscriber Directory Number) of the user, the IMEI (International Mobile
10 Equipment Identity) number of the imaging device, the IP address of the originating client application, or additional certificates/PKI (Public Key Infrastructure) information relevant to the user.

Various combinations of the steps above, and/or repetitions of various steps, are possible in the various embodiments of the invention. Thus, there is a combinatorially large
15 number of different complete specific implementations. Nevertheless, for purposes of clarity these implementations may be grouped into two broad categories, which shall be called "multiple session implementations", and "single session implementations", and which are set forth in detail in the Detailed Description of the Exemplary Embodiments.

Figure 1 illustrates a typical prior art OCR system. There is an object which must be
20 imaged **100**. The system utilizes special lighting produced by the illumination apparatus **101**, which illuminates the image to be captured. Imaging optics **102** (such as the optical elements used to focus light on the digital image sensor) and high resolution imaging sensors **103** (typically an IC chip that converts incoming light to digital information) generate digital images of the printed alphanumeric text **104** which have high resolution (in which "high
25 resolution" means many pixels in the resulting image per each character), and where there is a clear distinction between background pixels (denoting the background paper of the text)

and the foreground pixels belonging to the alphanumeric characters to be recognized. The processing software **105** is executed on a local processor **106**, and the alphanumeric output can be further processed to yield additional information, URL links, phone numbers, or other useful information. Such a system can be implemented on a mobile device with imaging capabilities, given that the device has the suitable components denoted here, and that the device has a processor that can be programmed (during manufacture or later) to run the software **105**.

Figure **2** illustrates the key processing steps of a typical prior art OCR system. The digitized image **201** undergoes binarization **202**. Morphological operations **203** are then applied to the image in order to remove artifacts resulting from dirt or sensor defects. Then morphological operations **203** then identify the location of rows of characters and the characters themselves **204**. In step **205**, characters are recognized by the system based on morphological criteria and/or other information derived from the binarized image of each assumed character. The result is a decoded character string **206** which can then be passed to other software in order to generate various actions.

In Figure **3**, the main components of an exemplary embodiment of the present invention are described. The object to be imaged **300**, which presumably has alphanumeric characters in it, may be printed material or a display device, and may be binary (like old calculator LCD screens), monochromatic or in color. There is wireless portable device **301** (that may be handheld or mounted in any vehicle) with a digital imaging sensor **302** which includes optics. Lighting element **101** from Figure **1** is not required or assumed here, and the sensor according to the preferred embodiment of the invention need not be high resolution, nor must the optics be optimized to the OCR task. Rather, the wireless portable device **301** and its constituent components may be any ordinary mobile device with imaging capabilities. The digital imaging sensor **302** outputs a digitized image which is transferred to the communication and image/video compression module **303** inside the portable device **301**.

This module encapsulates and fragments the image or video sequence in the proper format for the wireless network, while potentially also performing compression. Examples of formats for communication of the image include email over TCP/IP, and H.324M over RTP/IP. Examples of compression methods are JPEG compression for images, and MPEG 4 for video sequences.

The wireless network **304** may be a cellular network, such as a UMTS, GSM, iDEN or CDMA network. It may also be a wireless local area network such as WiFi. This network may also be composed of some wireline parts, yet it connects to the wireless portable device **301** itself wirelessly, thereby providing the user of the device with a great degree of freedom in performing the imaging operation.

The digital information sent by the device **301** through the wireless network **304** reaches a storage server **305**, which is typically located at considerable physical distance from the wireless portable device **301**, and is not owned or operated by the user of the device. Some examples of the storage server are an MMS server at a communication carrier, an email server, a web server, or a component inside the processing server **306**. The importance of the storage server is that it stores the complete image/video sequence before processing of the image/video begins. This system is unlike some prior art OCR systems that utilize a linear scan, where the processing of the top of the scanned page may begin before the full page has been scanned. The storage server may also perform some integrity checks and even data correction on the received image/video.

The processing server **306** is one novel component of the system, as it comprises the algorithms and software enabling OCR from mobile imaging devices. This processing server **306** accesses the image or video sequence originally sent from the wireless portable device **301**, and converts the image or video sequence into a digital sequence of decoded alphanumeric characters. By doing this conversion, processing server **306** creates the same kind of end results as provided by prior art OCR systems such as the one in depicted in

Figure 1, yet it accomplishes this result with fewer components and without any mandatory changes or additions to the wireless portable device 301. A good analogy would be comparison between an embedded data entry software on a mobile device on the one hand, and an Interactive Voice Response (IVR) system on the other. Both the embedded software and the IVR system accomplish the decoding of digital data typed by the user on mobile device, yet in the former case the device must be programmable and the embedded software must be added to the device, whereas the IVR system makes no requirements of the device except that the device should be able to handle a standard phone call and send standard DTMF signals. Similarly, the current system makes minimal requirements of the wireless portable device 301.

After or during the OCR decoding process, the processing server 306 may retrieve content or information from the external content/information server 308. The content/information server 308 may include pre-existing encoded content such as audio files, video files, images, and web pages, and also may include information retrieved from the server or calculated as a direct result of the user's request for it (such as, for example, a price comparison chart for a specific product, or the expected weather at a specific site, or a specific purchase deals or coupons offered to the user at this point in time). It will be appreciated that the contents/information server 308 may be configured in multiple ways, including, solely by way of example, one physical server with databases for both content and information, or one physical server but with entirely different physical locations for content versus information, or multiple physical servers, each with its own combination of external content and results. All of these configurations are contemplated by the current invention.

Based on the content and information received from the content/information server 308, the processing server 306 may make decisions affecting further actions. One example would be that, based on the user information stored on some content/information server 308, the processing server 306 may select, for example, specific data to send to the user's wireless

portable device 301 via the wireless messaging server 307. Another example would be that the processing server 306 merges the information from several different content/information servers 308 and creates new information from it, such as, for example, a comparing price information from several sources and sending the lowest offer to the user.

5 The feedback to the user is performed by having the processing server 306 submit the content to a wireless messaging server 307. The wireless messaging server 307 is connected to the wireless and wireline data network 304 and has the required permissions to send back information to the wireless portable device 301 in the desired manner. Examples of wireless messaging servers 307 include a mobile carrier's SMS server, an MMS server, a video
10 streaming server, and a video gateway used for mobile video calls. The wireless messaging server 307 may be part of the mobile carrier's infrastructure, or may be another external component (for example, it may be a server of an SMS aggregator, rather than the server of the mobile carrier, but the physical location of the server and its ownership are not relevant to the invention). The wireless messaging server 307 may also be part of the processing server
15 306. For example, the wireless messaging server 307 might be a wireless data card or modem that is part of the processing server 306 and that can send or stream content directly through the wireless network.

Another option is for the content/information server 308 itself to take charge and manage the sending of the content to the wireless device 301 through the network 304. This
20 could be preferred because of business reasons (e.g., the content distribution has to be controlled via the content/information server 308 for DRM or billing reasons) and/or technical reasons (that is, in this mode the content/information server 308 is a video streaming server which resides within the wireless carrier infrastructure and hence has a superior connection to the wireless network over that of the processing server).

25 Figure 3 demonstrates that exemplary embodiments of the invention includes both "Single Session" and "Multiple Session" operation.

In "Single Session" operation, the different steps of capturing the image/video of the object, the sending and the receiving of data are encapsulated within a single mode of wireless device and network operation. Graphically, the object to be imaged 300 is imaged by the wireless portable device 301, including image capture by the digital imaging sensor 302 and processing by the communication and image/video compression module 303. Data communicated to the wireless and wireline data network 304, hence to the storage server 305, then to the processing server 306, where there may or may not be interaction with the content/information server 308 and/or the wireless messaging server 307. If data is indeed sent back to the user device 301 through the messaging server 307, then by definition of "single session" this is done while the device 301 is still in the same data sending/receiving session started by the user sending the original image and/or video. At the same time, additional data may be sent through the messaging server 307 to other devices/addresses.

The main advantages of the Single Session mode of operation are ease of use, speed (since no context switching is needed by the user or the device), clarity as to the whole operation and the relation between the different parts, simple billing, and in some cases lower costs due to the cost structure of wireless network charging. The Single Session mode may also yield greater reliability since it relies on fewer wireless services to be operative at the same time.

Some modes which enable single session operation are:

20 A 3G H.324M/IMS SIP video Telephony session where the user points the device at the object, and then receives instructions and resulting data/service as part of this single video-telephony session.

A special software client on the phone which provides for image/video capture, sending of data, and data retrieval in a single web browsing, an Instant Messaging Service (IMS) session (also known as a Session Initiation Protocol or SIP session) or other data packet session.

Typically, the total time since the user starts the image/video capture until the user receives back the desired data could be a few seconds up to a minute or so. The 3G 324M scenario is suitable for UMTS networks, while the IMS/SIP and special client scenarios could be deployed on WiFi, CDMA 1x, GPRS, iDEN networks.

5 “Multiple Session” operation is a mode of usage operation the user initiates a session of image/video capture, the user then sends the image/video, the sent data then reaches a server and is processed, and the resulting processed data/services are then sent back to the user via another session. The key difference between Multiple Session and Single Session is that in Multiple Session the processed data/services are sent back to the user in a different
10 session or multiple sessions. Graphically, Multiple Session is the same as Single Session described above, except that communication occurs multiple times in the Multiple Session and/or through different communication protocols and sessions.

The different sessions in Multiple Session may involve different modes of the wireless and wireline and wireline data network 304 and the sending/receiving wireless
15 portable device 301. A Multiple Session operation scenario is more complex typically than a Single Session, but may be the only mode currently supported by the device/network or the only suitable mode due to the format of the data or due to cost considerations. For example, when a 3G user is roaming in a different country, the single session video call scenario may be unavailable or too expensive, while GPRS roaming enabling MMS and SMS data
20 retrieval, with is an example of Multiple Session, would still be an existent and viable option.

Examples of image/video capture as part of a multiple session operation would be:

The user may take one or more photos/video clips using an in-built client of the wireless device.

The user may take one or more photos/video clips using a special software client
25 resident on the device (e.g., a Java MIDLet or a native code application).

The user may make a video call to a server where during the video call the user points the phone camera at the desired object.

Examples of possible sending modes as part of a multiple session operation would be:

The user uses the device's in-built MMS client to send the captured images/video clips to a phone number, a shortcode or an email address.

The user uses the device's in-built Email client to send the captured images/video clips to an email address.

The user uses special software client resident on the device to send the data using a protocol such as HTTP.POST, UDP or some other TCP protocol, etc.

Examples of possible data/service retrieval modes as part of a multiple session operation are :

The data is sent back to the user as a Short Message Service (SMS).

The data is sent back to the user as a Multimedia Message (MMS).

The data is sent back to the user as an email message.

A link to the data (a phone number, an email address, a URL etc.) is sent to the user encapsulated in an SMS/MMS/email message.

A voice call/video call to the user is initiated from an automated/human response center.

An email is sent back to the user's pre-registered email account (unrelated to his wireless portable device 301).

A combination of several of the above listed methods – e.g., a vCARD could be sent in an MMS, at the same time a URL could be sent in an SMS, and a voice call could be initiated to let the user know he/she has won some prize.

Naturally, any combination of the capture methods {a,b,c}, the sending methods {d,e,f} and the data retrieval methods {g,h,i,j,k,l,m} is possible and valid.

Typically, the total time since the user starts the image/video capture until the user received back the desired data could be 1-5 minutes. The multiple session scenario is particularly suitable for CDMA 1x, GPRS, iDEN networks, as well as for Roaming UMTS scenarios. Typically, a multiple session scenario would involve several separate billing events in the user's bill.

Figure 4 depicts the steps by which the processing server 306 converts input into a string of decoded alphanumeric characters. In the preferred embodiment, all of steps in Figure 4 executed in the processing server 306. However, in alternative embodiments, some or all of these steps could also be performed by the processor of the wireless portable device 301 or at some processing entities in the wireless and wireline data network 304. The division of the workload among 306, 301, and 304, in general is a result of the optimization between minimizing execution times on one hand, and data transmission volume and speed on the other hand.

In step 401, the image undergoes pre-processing designed to optimize the performance of the consecutive steps. Some examples of such image pre-processing 401 are conversion from a color image to a grayscale image, stitching and combining several video frames to create a single larger and higher resolution grayscale image, gamma correction to correct for the gamma response of the digital imaging sensor 302, JPEG artifact removal to correct for the compression artifacts of the communication and image/video compression module 303, missing image/video part marking to correct for missing parts in the image/video due to transmission errors through the wireless and wireline network 304. The exact combination and type of these algorithms depend on the specific device 301, the modules 302 and 303, and may also depend on the wireless network 304. The type and degree of pre-processing conducted depends on the parameters of the input. For example, stitching and combining for video frames is only applied if the original input is a video stream. As another example, the JPEG artifact removal can be applied at different levels

depending on the JPEG compression factor of the image. As yet another example, the gamma correction takes into account the nature and characteristics of the digital imaging sensor 302, since different wireless portable devices 301 with different digital imaging sensors 302 display different gamma responses. The types of decisions and processing
5 executed at 301 are to be contrasted with the prior art described in Figures 1 and 2, in which the software runs on a specific device. Hence, under prior art most of the decisions described above are not made by the software, since prior art software is adapted to the specific hardware on which it runs, and such software is not designed to handle multiple hardware combinations. In essence, prior art software need not be make these decisions, since the
10 device (that is, the combined hardware/software offering in prior art) has no flexibility to make such decisions and has fixed imaging characteristics.

In step 402, the processing is now performed on a single grayscale image. A search is made for “key” alphanumeric characters over a range of values. In this context, a “key” character is one that must be in the given image for the template or templates matching that
15 image, and therefore a character that may be sought out and identified. The search is performed over the whole image for the specific key characters, and the results of the search help identify the location of the alphanumeric strings. An example would be searching for the digits “0” or “1” over the whole image to find locations of a numeric string. The search operation refers to the multiple template matching algorithm described in Figure 5 and in
20 further detail in regards to step 403. Since the algorithm for the search operation detects the existence of a certain specific template of a specific size and orientation, the full search involves iteration over several scales and orientations of the image (since the exact size and orientation of the characters in the image is not known a-priori). The full search may also involve iterations over several “font” templates for a certain character, and/or iterations over
25 several potential “key” characters. For example, the image may be searched for the letter “A” in several fonts, in bold, italics etc. The image may also be searched for other characters

since the existence of the letter "A" in the alphanumeric string is not guaranteed. The search for each "key" character is performed over one or more range of values, in which "range of value" means the ratios of horizontal and vertical size of image pixels between the resized image and the original image. It should be noted that for any character, the ratios for the horizontal and vertical scales need not be the same.

In step 403, the search results of step 402 are compared for the different scales, orientations, fonts and characters so that the actual scale/orientation/font may be determined. This can be done by picking the scale/orientation/font/character combination which has yielded the highest score in the multiple template matching results. An example of such a score function would be the product of the template matching scores for all the different templates at a single pixel. Let us consider a rotated and rescaled version of the original image I after preprocessing 402. This version $I(\alpha, c)$ is rotated by the angle α and rescaled by a factor c . Let us denote by $T_i^A(x, y)$ the value of the normalized cross correlation value of template i of the character "A" at pixel x, y in the image $I(\alpha, c)$. Then a valid score function for $I(\alpha, c)$ would be $\max_{(x, y)} \{\text{Prod}_{i=1..N} T_i^A(x, y)\}$. This score function would yield 1 where the original I contains a version of the character A rotated by $-\alpha$ and scaled by $1/c$. Instead of picking just one likely candidate for α, c based on the maximum score, it is possible to pick several candidates and proceed with all of them to the next steps.

In step 404, the values of α, c , and font have been determined already, and further processing is applied to search for the character line, the line edge, and the line orientation, of consecutive characters or digits in the image. In this context, "line" (also called "character line") is an imaginary line drawn through the centers of the characters in a string, "line edge" is point where a string of characters ends at an extreme character, and "line orientation" is the angle of orientation of a string of characters to a theoretical horizontal line. It is possible to determine the line's edges by characters located at those edges, or by a-priori other knowledge about the expected presence and relative location of specific characters searched

for in the previous steps 402 and 403. For example, a URL could be identified, and its scale and orientation estimated, by locating three consecutive “w” characters. Additionally, the edge of a line could be identified by a sufficiently large area void of characters. A third example would be the letters “ISBN” printed in the proper font which indicate the existence, orientation, size, and edge of an ISBN product code line of text.

Step 404 is accomplished by performing the multi-template search algorithm on the image for multiple characters yet at a fixed scale, orientation, and font. Each pixel in the image is assigned some score function proportional to the probability that this pixel is the center pixel of one of the searched characters. Thus, a new grayscale image J is created where the grayscale value of each pixel is this score function. A sample of such score function for a pixel (x,y) in the image J could be $\max_i \{\prod_{j=1..n} T_j^{C(i)}(x,y)\}$ where i iterates over all characters in the search, c(i) refers to a character, and j iterates over the different templates of the character c(i). A typical result of this stage would be an image which is mostly “dark” (corresponding to low values of the score function for most pixels) and has a row (or more than one row) of bright points (corresponding to high values of the score function for a few pixels). Those bright points on a line would then signify a line of characters. The orientation of this line, as well as the location of the leftmost and rightmost characters in it, are then determined. An example of a method of determining those line parameters would be picking the brightest pixel in the Radon (or Hough) transform of this score-intensity image J. It is important to note that if the number and relative positions of the characters in the line are known in advance (e.g., as in a license plate, an ISBN code, a code printed in advance), then the precise scale of the image c^* could be estimated with greater precision than the original scale c.

In step 405, scale and orientation are corrected. The scale information $\{c, c^*\}$, and the orientation of the line, derived from both steps 403 and 404, are used to re-orient and re-scale

the original image I to create a new image $I^*(\alpha^*, c^*)$. In the new image, the characters of a known font, default size, and orientation, all due to the algorithms previously executed.

The re-scaled and re-oriented image from step 405 is then used for the final string recognition 406, in which every alphanumeric character within a string is recognized. The actual character recognition is performed by searching for the character most like the one in the image at the center point of the character. That is, in contrast with the search over the whole image performed in step 402, here in step 406 the relevant score function is calculated at the “center point” for each character, where this center point is calculated by knowing in advance the character size and assumed spacing. An example of a decision function at this stage would be $C(x,y)=\max_j\{\prod_{i=1..n}T^{(i)}_j(x,y)\}$ where i iterates over all potential characters j over all templates per character. The coordinates (x,y) are estimated based on the line direction and start/end characters estimated in step 405. The knowledge of the character center location allows this stage to reach much higher precision than the previous steps in the task of actual character recognition. The reason is that some characters often resemble parts of other characters. For example the upper part of the digit “9” may yield similar scores to the lower part of the digit “6” or to the digit “0”. However, if one looks for the match around the precise center of the character, then the scores for these different digits will be quite different, and will allow reliable decoding. Another important and novel aspect of an exemplary embodiment of the invention is that at step 406, the relevant score function at each “center point” may be calculated for various different versions of the same character at the same size and at the same font, but under different image distortions typical of the imaging environment of the wireless portable device 301. For example, several different templates of the letter “A” at a given font and at a given size may be compared to the image, where the templates differ in the amount of pre-calculated image smear applied to them or gamma transform applied to them. Thus, if the image indeed contains at this “center point” the letter “A” at the specified font and size, yet the image suffers from smear quantified by a PSF “X”,

then if one of the templates in the comparison represents a similar smear PSF it would yield a high match score, even though the original font's reference character "A" contains no such image smear.

The row or multiple rows of text from step 406 are then decoded into a decoded
5 character string 407 in digitized alphanumeric format.

There are very significant differences between the processing steps outlined in Figure 4, and those of the prior art depicted in Figure 2. For example, prior art relies heavily on binarization 202, whereas in an exemplary embodiment of the present invention the image is converted to gray scale in step 401. Also, whereas in prior art morphological operations 203
10 are applied, in an exemplary embodiment of the current invention characters are located and decoded by the multi-template algorithm in step 402. Also, according to an exemplary embodiment, the present invention searches for key alphanumeric characters 402 over multiple scales, whereas prior art is restricted to one or a very limited number of scales. Also, in the present the scale and orientation correction 405 is executed in reliance, in part, on
15 the search for line, line edge, and line orientation from step 404, a linkage which does not exist in the prior art. These are not the only other differences between prior art and the present invention, there are many others as described herein, but these differences are illustrative of the novelties of the current invention.

Once the string of characters is decoded at the completion of step 407, numerous
20 types of application logic processing 408 become possible. One value of the proposed invention, according to an exemplary embodiment, is that the invention enables fast, easy data entry for the user of the mobile device. This data is human-readable alphanumeric characters, and hence can be read and typed in other ways as well. The logic processing in step 408 will enable the offering of useful applications such as:

25 Product Identification for price comparison/information gathering: The user sees a product (such as a book) in a store with specific codes on it (e.g., the ISBN alphanumeric

code). The user takes a picture/video of the identifying name/code on the product. Based on (e.g., ISBN) code/name of the product, the user receives information on the price of this product, information etc.

URL launching: the user snaps a photo of an http link and later receives a WAP
5 PUSH message for the relevant URL.

Prepaid card loading/Purchased content loading: The user takes a photo of the recently purchased pre-paid card and the credit is charged to his/her account automatically. The operation is equivalent to currently inputting the prepaid digit sequence through an IVR session or via SMS, but the user is spared from actually reading the digits and typing them
10 one by one.

Status inquiry based on printed ticket: The user takes a photo of the lottery ticket, travel ticket, etc., and receives back the relevant information, such as winning status, flight delayed/on time, etc. The alphanumeric information on the ticket is decoded by the system and hence triggers this operation.

15 User authentication for Internet shopping: When the user makes a purchase, a unique code is displayed on the screen and the user snaps a photo, thus verifying his identity via the phone. Since this code is only displayed at this time on this specific screen, it represents a proof of the user's location, which, coupled to the user's phone number, create reliable location-identity authentication.

20 Location Based Coupons: The user is in a real brick and mortar store. Next to each counter, there is a small sign/label with a number/text on it. The user snaps a photo of the label and gets back information, coupons, or discounts relevant to the specific clothes items (jeans, shoes, etc.) he is interested in. The label in the store contains an ID of the store and an ID of the specific display the user is next to. This data is decoded by the server and sent to
25 the store along with the user's phone ID.

Digital signatures for payments, documents, identities: A printed document (such as a ticket, contract, or receipt) is printed together with a digital signature (a number of 20-40 digits) on it. The user snaps a photo of the document and the document is verified by a secure digital signature printed in it. A secure digital signature can be printed in any number of formats, such as, for example, a 40-digit number, or a 20-letter word. This number can be printed by any printer. This signature, once converted again to numerical form, can securely and precisely serve as a standard, legally binding digital signature for any document.

Catalog ordering/purchasing: The user is leafing through a catalogue. He snaps a photo of the relevant product with the product code printed next to it, and this is equivalent to an “add to cart operation”. The server decodes the product code and the catalogue ID from the photo, and then sends the information to the catalogue company’s server, along with the user’s phone number.

Business Card exchange: The user snaps a photo of a business card. The details of the business card, possibly in VCF format, are sent back to the user’s phone. The server identifies the phone numbers on the card, and using the carrier database of phone numbers, identifies the contact details of the relevant cellular user. These details are wrapped in the proper “business card” format and sent to the user.

Coupon Verification: A user receives via SMS/MMS/WAP PUSH a coupon to his phone. At the POS terminal (or at the entrance to the business using a POS terminal) he shows the coupon to an authorized clerk with a camera phone, who takes a picture of the user’s phone screen to verify the coupon. The server decodes the number/string displayed on the phone screen and uses the decoded information to verify the coupon.

Figure 5 illustrates graphically some aspects of the multi-template matching algorithm, which is one important algorithm used in an exemplary embodiment of the present invention (in processing steps 402, 404, and 406, for example). The multi-template matching algorithm is based on a well known template matching method for grayscale images called

“Normalized Cross Correlation” (NCC). NCC is currently used in machine vision applications to search for pre-defined objects in images. The main deficiency of NCC is that for images with non-uniform lighting, compression artifacts and/or defocusing issues, the NCC method yields many “false alarms” (that is, incorrect conclusion that a certain status o

5 object appears) and at the same time fails to detect valid objects. The multi-template algorithm extends the traditional NCC by replacing a single template for the NCC operation with a set of N templates, which represent different parts of the object (or character in the present case) that is searched. The templates **505** and **506** represent two potential such templates, representing parts of the digit “1” in a specific font and of a specific size. For each

10 template, the NCC operation is performed over the whole image **501**, yielding the normalized cross correlation images **502** and **503**. The pixels in these images have values between -1 and 1, where a value of 1 for pixel (x,y) indicates a perfect match between a given template and the area in image **501** centered around (x,y). At the right of **502** and **503**, respectively, sample one-dimensional cross sections of those images are shown, showing how a peak of 1

15 is reached exactly at a certain position for each template. A very important point is that even if the image indeed has the object to be searched for centered at some point (x,y), the response peaks for the NCC images for various templates will not necessarily occur at the same point. For example, in the case displayed in Figure 5, there is a certain difference **504** of several pixels in the horizontal direction between the peak for template **505** and the peak

20 for template **506**. These differences can be different for different templates, and are taken into account by the multi-template matching algorithm. Thus, after the correction of these deltas, all the NCC images (such as **502** and **503**) will display a single NCC “peak” at the same (x,y) coordinates which are also the coordinates of the center of the object in the image. For a real life image, the values of those peaks will not reach the theoretical “1.0” value,

25 since the object in the image will not be identical to the template. However, proper score functions and thresholds allow for efficient and reliable detection of the object by judicious

lowering of the detection thresholds for the different NCC images. It should be stressed that the actual templates can be overlapping, partially overlapping or with no overlap. Their size, relative position and shape can be changed for different characters, fonts or environments. Furthermore, masked NCC can be used for these templates to allow for non-rectangular
5 templates.

The system, method, and algorithms, described herein, can be trivially modified and extended to recognize other characters, other fonts or combinations thereof, and other arrangements of text (such as text in two rows, vertical text rather than horizontal, etc.). Nothing in the existing detailed description of the invention makes the invention specific to
10 the recognition of specific fonts or characters or languages/codes.

The system, method, and algorithms described in Figure 4 and 5 enable the reliable detection and decoding of alphanumeric characters in situations where traditional prior art could not perform such decoding. At the same time, potentially other new algorithms could be developed which are extensions of the ones described here or are based on other
15 mechanisms within the contemplation of this invention. Such algorithms could also operate on the system architecture described in Figure 3.

Other variations and modifications of the present invention are possible, given the above description. All variations and modifications which are obvious to those skilled in the art to which the present invention pertains are considered to be within the scope of the
20 protection granted by this Letters patent.

WHAT IS CLAIMED IS:

1. A method for recognizing symbols and identifying users or services, the method comprising:

displaying an image or video clip on a display device in which identification information is embedded in the image or video clip;

5 capturing the image or video clip on an imaging device;

transmitting the image or video clip from the imaging device to a communication network;

transmitting the image or video clip from the communication network to a processing and authentication server;

10 processing the information embedded in the image or video clip by the server to identify logos, alphanumeric characters, or special symbols in the image or video clip, and converting the identified logos or characters or symbols into a digital format to identify the user or location of the user or service provided to the user;

2. The method of claim 1, wherein,

15 the processed information in digital format is used to provide one or more additional services to the user.

3. The method of claim 1, wherein:

the embedded information is a logo;

4. The method of claim 1, wherein:

20 the nature or character of the image or video clip serves as all or part of the identifying information.

5. The method of claim 1, wherein:

the embedded information is signal that is spatially or temporally modulated on the screen of the display device.

6. The method of claim 1, wherein

5 the embedded information is alphanumeric characters;

7. The method of claim 1, wherein:

the embedded information is a bar code.

8. The method of claim 1, wherein:

10 the embedded information is a sequence of signals which are not human readable but which are machine readable.

9. The method of claim 1, wherein:

the communication network is a wireless network.

10. The method of claim 1, wherein:

the communication network is a wireline network.

15 11. The method of claim 1, wherein:

the display device further displays additional information which identifies the type and location of the display device.

12. A system for recognizing symbols and identifying users or services, the system comprising:

a remote server that prepares and transmits an image or video clip to a local node;

a local node that receives the transmission from said server;

a display that presents the image or video clip on either physical or electronic medium;

5 an imaging device for capturing the image or video clip in electronic format;

a communication module for converting the captured image or video clip into digital format and transmitting said digital image or video clip to a communication network;

a communication network that receives the image or video clip transmitted by the communication module, and that transmits such image or video clip to a processing and

10 authentication server; and

a processing and authentication server that receives the transmission from the communication network, and completes the processing to identify the location of the display, the time the display was captured, and the identify of the imaging device.

13. The system of claim **12**, wherein:

15 remote server is one or a plurality of servers or computers.

14. The system of claim **12**, wherein:

the local node is a node selected from the group consisting of a television set, a personal computer running a web browser, an LED display, or an electronic bulletin board.

15. The system of claim **12**, wherein:

20 the display and the imaging device are combined in one unit of hardware.

16. The system of claim **12**, wherein:

there is a communication link between the processing and authentication server and the remote server which allows the execution of additional servers to the user.

17. A method recognizing symbols and identifying users or services, the method comprising:

5 resizing a target image or video clip in order to compared the resized image or a video clip to a pre-existing database of images or video clips;

 determining the best image scale by first searching among all scales where the score is above a pre-defined threshold and then choosing the best image scale among the various image scales tested;

10 repeating all prior procedures for multiple parts of the object image or video clip, to determine the potential locations of different templates representing various parts of the object;

 iterating the combinations of all permutations of the templates for the respective parts of the object in order to determine the permutation with the best match with the object;

15 determining if the best match permutation is sufficiently good to conclude that the object has been correctly identified.

18. The method of claim 17, wherein:

 the best image scale is not determined by applying pre-defined thresholds, but rather by one or more of the techniques of applying other score functions, or weighting the image
20 scales of several scale sets yielding the highest scores, or using a parametric fit to the whole range of scale sets based on their relative scores.

19. The method of claim 17, wherein:

the scale ranges for the various parts of the object during template repetition may be varied for each part in order to determine the optimal image scale for each part.

20. A computer program product, comprising a computer data signal in a carrier wave having computer readable code embodied therein for causing a computer to perform a method comprising:

- 5 displaying an image or video clip on a display device in which identification information is embedded in the image or video clip;
- capturing the image or video clip on an imaging device;
- transmitting the image or video clip from the imaging device to a communication
10 network;
- transmitting the image or video clip from the communication network to a processing and authentication server;
- processing the information embedded in the image or video clip by the server to identify logos, alphanumeric characters, or special symbols in the image or video clip, and
15 converting the identified logos or characters or symbols into a digital format to identify the user or location of the user or service provided to the user;
- using the processed information in digital format to provide one or more of a variety of additional applications.

WHAT IS CLAIMED IS:

1. A method for decoding printed alphanumeric characters from images or video sequences captured by a wireless device, the method comprising:

pre-processing the image or video sequence to optimize processing in all subsequent operations;

5 searching one or more grayscale images for key alphanumeric characters on a range of scales;

comparing the values on said range of scales to a plurality of templates in order to determine the characteristics of the alphanumeric characters;

10 performing additional comparisons to a plurality of templates to determine character lines, line edges, and line orientation;

processing information from prior said pre-processing, said searching, said comparing, and said performing additional comparisons, to determine the corrected scale and orientation of each line;

recognizing the identity of each alphanumeric character in a string of such characters;

15 decoding the entire character string in digitized alphanumeric format.

2. The method of claim 1, wherein:

the pre-processing comprises conversion from a color scale to a grayscale, and the stitching and combining of video frames to create a single larger and higher resolution grayscale image.

20 3. The method of claim 1, wherein:

the pre-processing comprises JPEG artifact removal to correct for compression artifacts of image/video compression executed by the wireless device.

4. The method of claim 1, wherein:

the pre-processing comprises part making of missing image/video data to correct for missing parts in the data due to transmission errors.

5. The method of claim 1, wherein:

5 comparing the key alphanumeric values to a plurality of template in order to determine the characteristics of the alphanumeric characters, comprises executing a modified Normalized Cross Correlation in which multiple parts are identified in the object to be captured from the image or video sequence, each part is compared against one or more templates, and all templates for all parts are cross-correlated to determine the characteristics
10 of each alphanumeric image captured by the wireless device.

6. The method of claim 1, wherein:

the method is conducted in a single session of communication with the wireless communication device.

7. The method of claim 6, further comprising:

15 application logic processing of the decoded character string in digitized alphanumeric format in order to enable additional applications.

8. The method of claim 1, wherein:

the method is conducted in multiple sessions of communication with the wireless communication device.

20 9. The method of claim 8, further comprising:

application logic processing of the decoded character string in digitized alphanumeric format in order to enable additional applications.

10. A system for decoding printed alphanumeric characters from images or video sequences captured by a wireless device, the system comprising:

5 an object to be imaged or to be captured by video sequence, that contains within it alphanumeric characters;

a wireless portable device for capturing the image video sequence, and transmitting the captured image or video sequence to a data network;

10 a data network for receiving the image or video sequence transmitted by the wireless portable device, and for retransmitting it to a storage server;

a storage receiver for receiving the retransmitted image or video sequence, for storing the complete image or video sequence before processing, and for retransmitting the stored image or video sequence to a processing server;

15 a processing server for decoding the printed alphanumeric characters from the image or video sequence, and for transmitting the decoded characters to an additional server.

11. The system of claim 10 wherein:

20 the wireless portable device is any device that transmits and receives on any radio communication network, that has a means for photographically capturing an image or video sequence, and that is of sufficiently small dimensions and weight that it may be transported by an unaided human being.

12. The system of claim 10 wherein:

the wireless portable device is a wireless telephone with built-in camera capability.

13. The system of claim 10, wherein:

the wireless portable device comprises a digital imaging sensor, and a communication and image/video compression module.

14. The system of claim 10, wherein:

5 the additional server is a wireless messaging server for receiving the decoded characters transmitted by the processing server, and for retransmitting the decoded characters to a data network.

15. The system of claim 14, further comprising:

10 a content/information server for receiving the decoded characters from the processing server, for further processing the decoded characters by adding additional information as necessary, for retrieving content based on the decoded characters and the additional information, and for transmitting the processed decoded characters and additional information back to the processing server.

16. A processing server within a telecommunication system for decoding printed

15 alphanumeric characters from images or video sequences captured by a wireless device, the processing server comprising:

a server for interacting with a plurality of storage servers, a plurality of content/information servers, and a plurality of wireless messaging servers, within the telecommunication system for decoding printed alphanumeric characters from images;

20 the server accessing image or video sequence data sent from a data network via a storage server, the server converting the image or video sequence data into a digital sequence of decoded alphanumeric characters, and the server communicating such digital sequence to an additional server.

17. The processing server of claim 16, wherein:
the additional server is a content/information server.

18. The processing server of claim 16, wherein:
the additional server is a wireless messaging server.

5 19. A computer program product, comprising a computer data signal in a carrier wave having computer readable code embodied therein for causing a computer to perform a method comprising:

pre-processing an alphanumeric image or video sequence;

10 searching on a range of scales for key alphanumeric characters in the image or sequence;

determining appropriate image scales;

searching for character lines, line edges, and line orientations;

correcting for the scale and orientation;

recognizing the strings of alphanumeric characters;

15 decoding the character strings.

20. The computer program product of claim 19, further comprising:

processing application logic in order to execute various applications on the decoded character string.

WHAT IS CLAIMED IS:

1. A method for imaging a document, and using a reference document to place pieces of the document in their correct relative position and resize such pieces in order to generate a single unified image, the method comprising:

electronically capturing a document with one or multiple images using an imaging
5 device;

performing pre-processing of said images to optimize the results of subsequent image recognition, enhancement, and decoding;

comparing said images against a database of reference documents to determine the most closely fitting reference document; and

10 applying knowledge from said closely fitting reference document to adjust geometrically orientation, shape, and size of said electronically captured images so that said images correspond as closely as possibly to said reference document.

2. The method of claim 1, wherein the method further comprises:

after completion of processing, routing the document to one or a multiplicity of
15 electronic or physical locations.

3. The method of claim 1, wherein the method further comprises:

applying metadata from said database of reference documents to selectively and optimally process the data from each area of said document as such area has been identified by said geometric adjustment of said captured electronic images.

20 4. The method of claim 3, wherein the method further comprises:

after completion of processing, routing the document to at least one of electronic and physical locations.

5. The method of claim 3, wherein the method further comprises:

applying an optical recognition technique decoding information on said imaged document by comparison to known optical symbols.

6. The method of claim 5, wherein:

5 said optical recognition technique is Optical Character Recognition.

7. The method of claim 5, wherein:

said optical recognition technique is Optical Mark Recognition.

8. The method of claim 6, wherein the method further comprises:

10 after completion of processing, routing the document to at least one of electronic and physical locations.

9. The method of claim 7, where in the method further comprises:

after completion of processing, routing the document to at least one of electronic and physical locations.

10. The method of claim 1, wherein the method further comprises:

15 identification of symbols within said document by said comparison of said images and said geometric adjustment of said images; and
decoding of said symbols.

11. The method of claim 8, wherein the imaging device captures photographic images of the document.

12. The method of claim 8, wherein the imaging device captures video images of the document.

13. The method of claim 9, wherein the imaging device captures video photographic images of the document.

5 14. The method of claim 10, wherein the imaging device captures video images of the document.

15. The method of claim 1, wherein:

said imaging device captures at least two images of said document;

said at least two images are of at least two different parts of the document;

10 said at least two images are recognized as processed so that they are recognized as said at least two different parts of a reference document; and

based on said recognition, forming a unified image of a higher photographic quality than at least one of said at least two images.

16. A system for imaging a document, and using a reference document to place
15 pieces of the document in their correct relative position and resize such pieces in order to generate a single unified image, the system comprising:

at least one document to be electronically captured;

a portable imaging device for electronically capturing said document with at least one
image;

20 a network for pre-processing said at least one image to optimize the results of subsequent image recognition, enhancement, and decoding;

a database comprising reference documents for comparing against said at least one pre-processed image; and

at least one server for receiving said at least one pre-processed image from the network, storing said at least one image, performing final processing, comparing said at least one image against at least one reference document, and routing the processed images to one or more recipients.

17. The system of claim 16, wherein:

said imaging device captures at least two images of said document;

said at least two images are of at least two different parts of the document;

10 said at least two images are recognized as processed so that they are recognized as two different parts of a reference document; and

based on a result of said recognition, forming a unified image of a higher photographic quality than at least one of said at least two images.

18. The system of claim 16, wherein:

15 said portable imaging device is configured to electronically capture at least one of photographic images and video clips of said document.

19. The system of claim 16, wherein:

said portable imaging device is configured to electronically capture photographic images of said document, and cannot electronically capture video clips of said document.

20 20. A computer program product stored on a computer readable medium for causing a computer medium to perform a method comprising:

electronically capturing a document with at least one image using an imaging device;

performing pre-processing of said at least one image to optimize results of subsequent image recognition, enhancement, and decoding;

comparing said at least one image against reference documents stored in a database, to determine most closely fitting reference document;

- 5 applying knowledge from said closely fitting reference document to adjust geometrically orientation, shape, and size of said electronically captured images so that said at least one image corresponds as closely as possibly to said reference document.

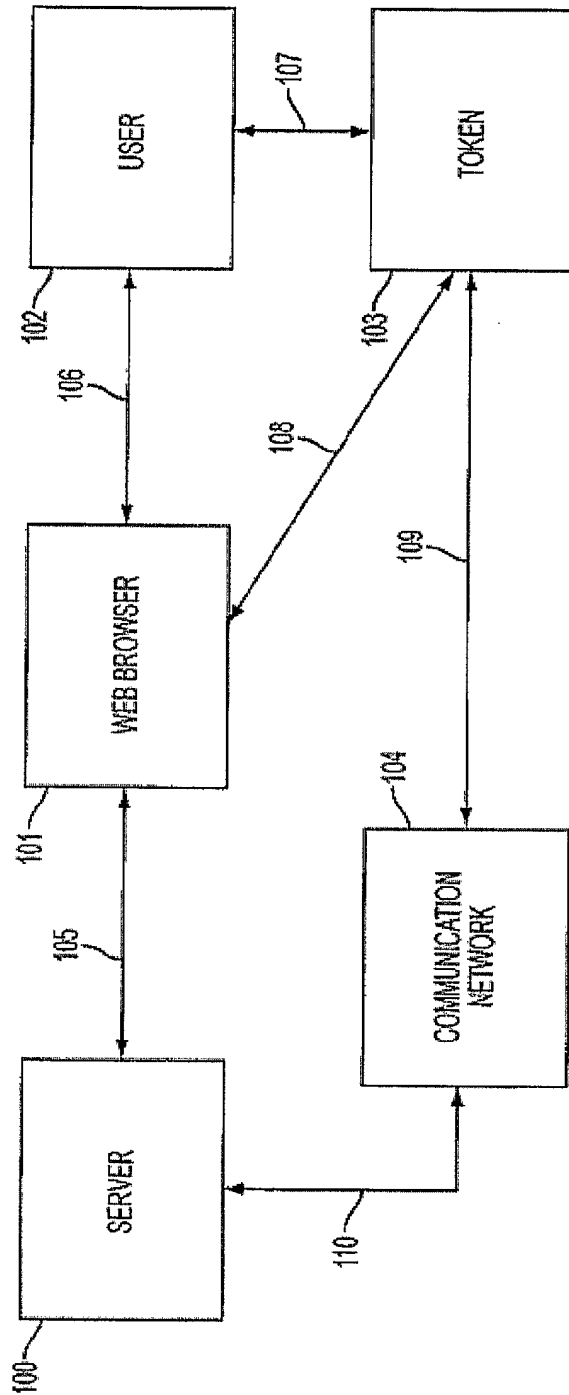


FIG. 1
PRIOR ART

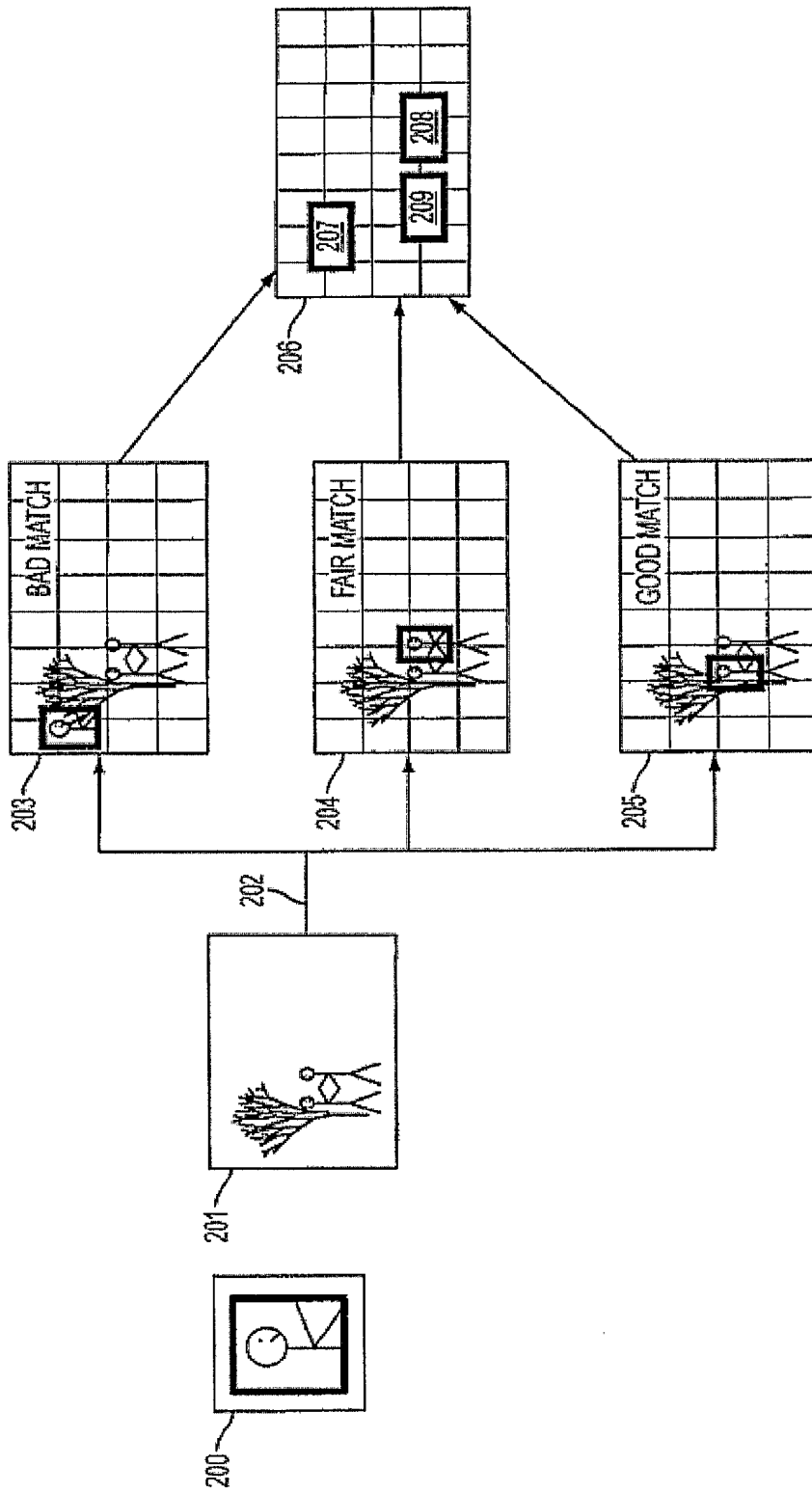


FIG. 2
PRIOR ART

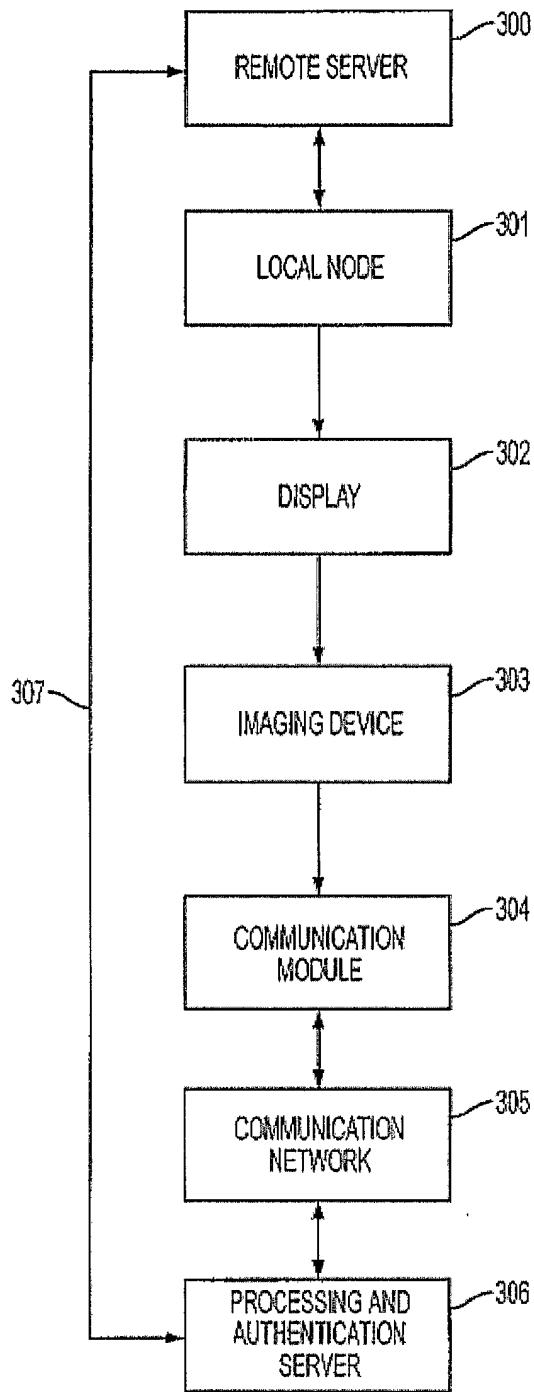


FIG. 3

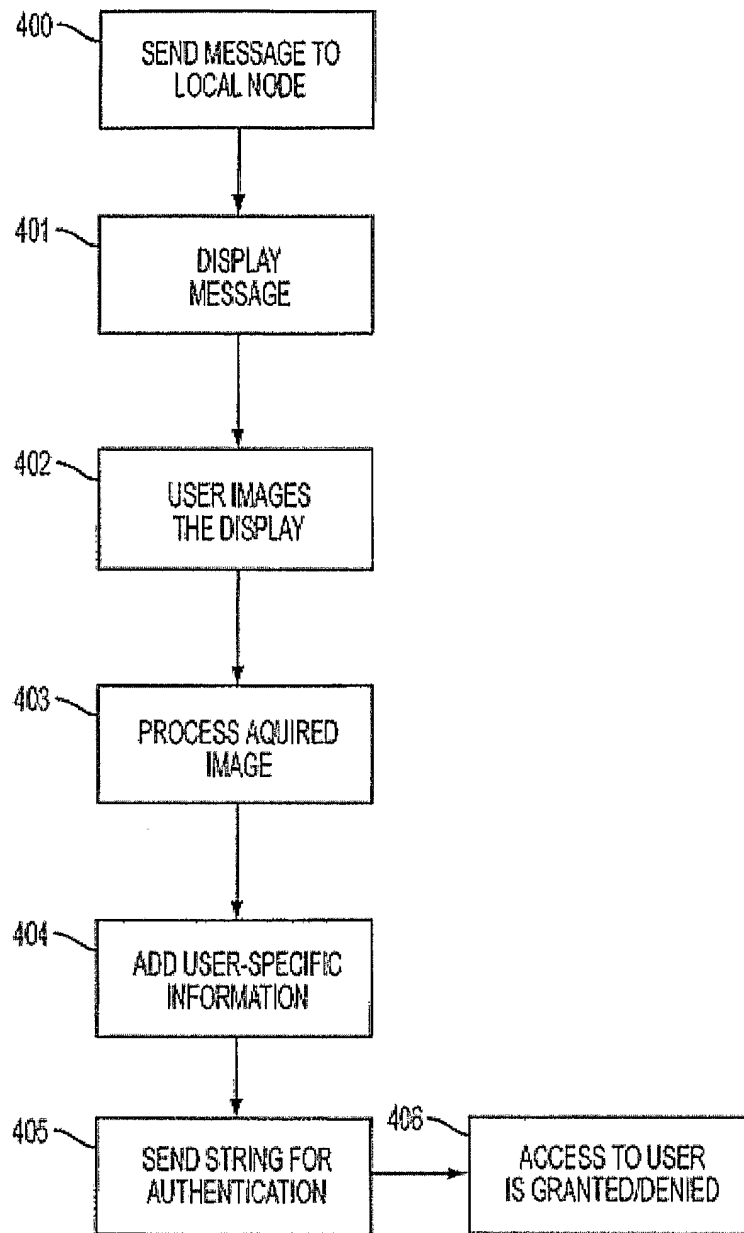


FIG. 4

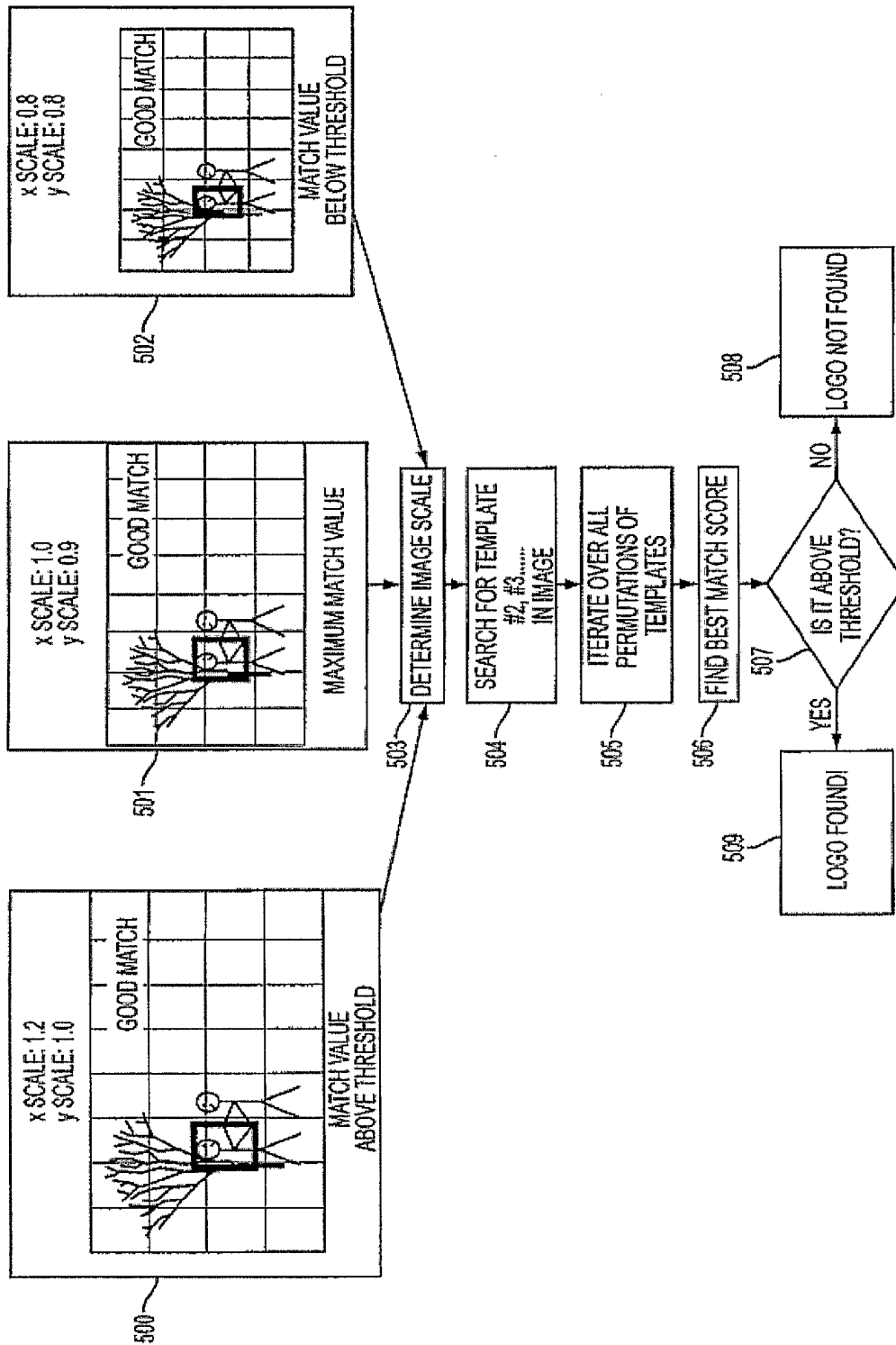


FIG. 5

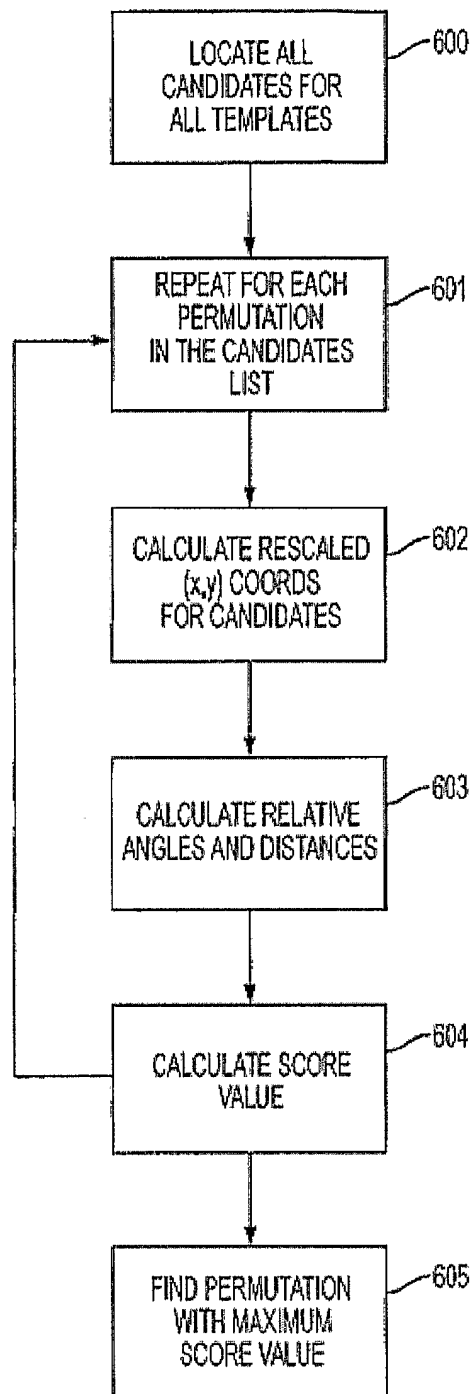


FIG. 6

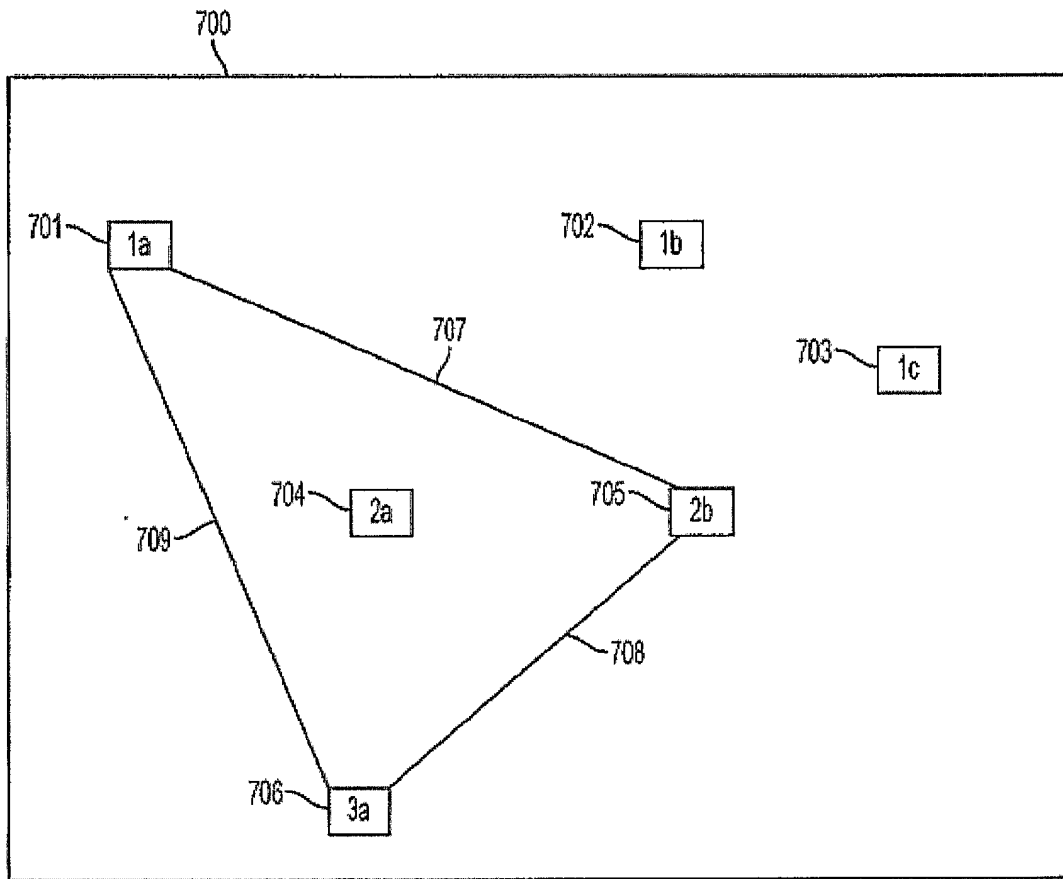


FIG. 7

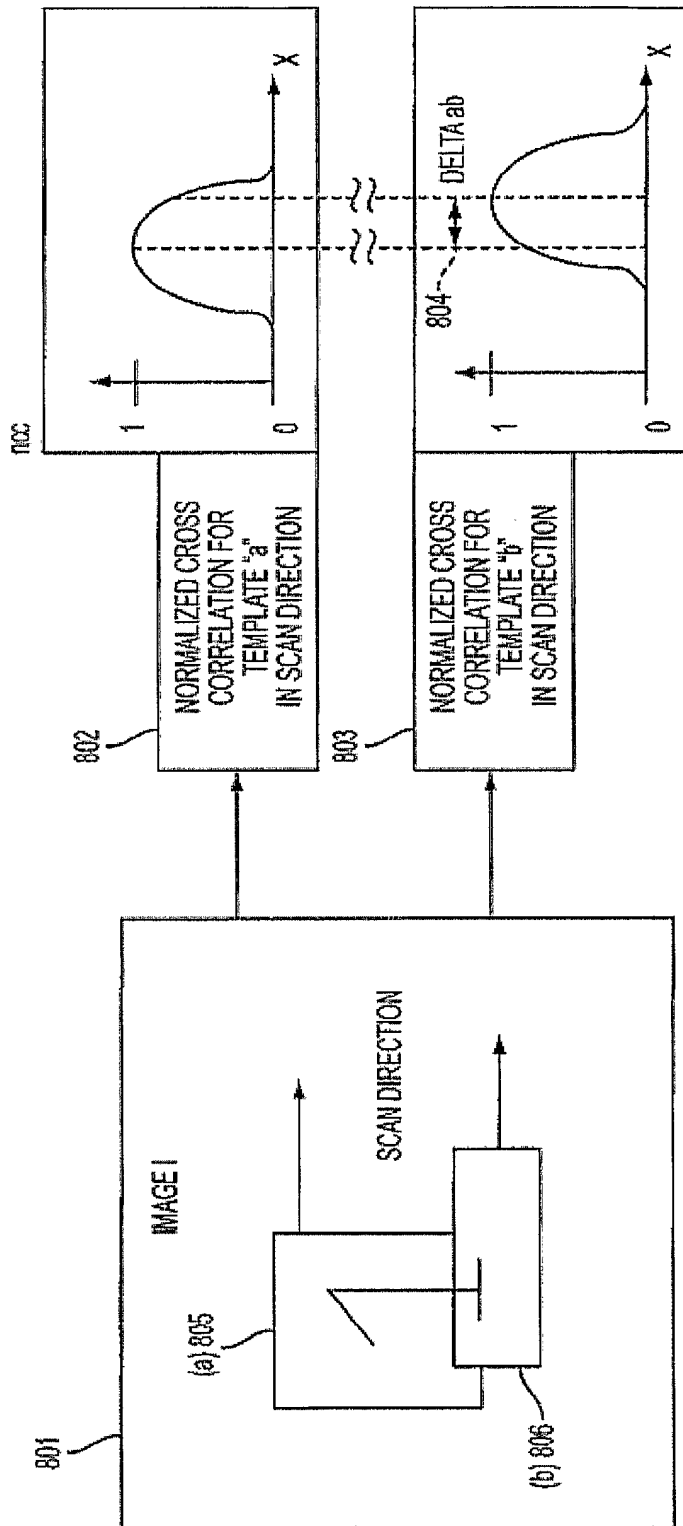


FIG. 8

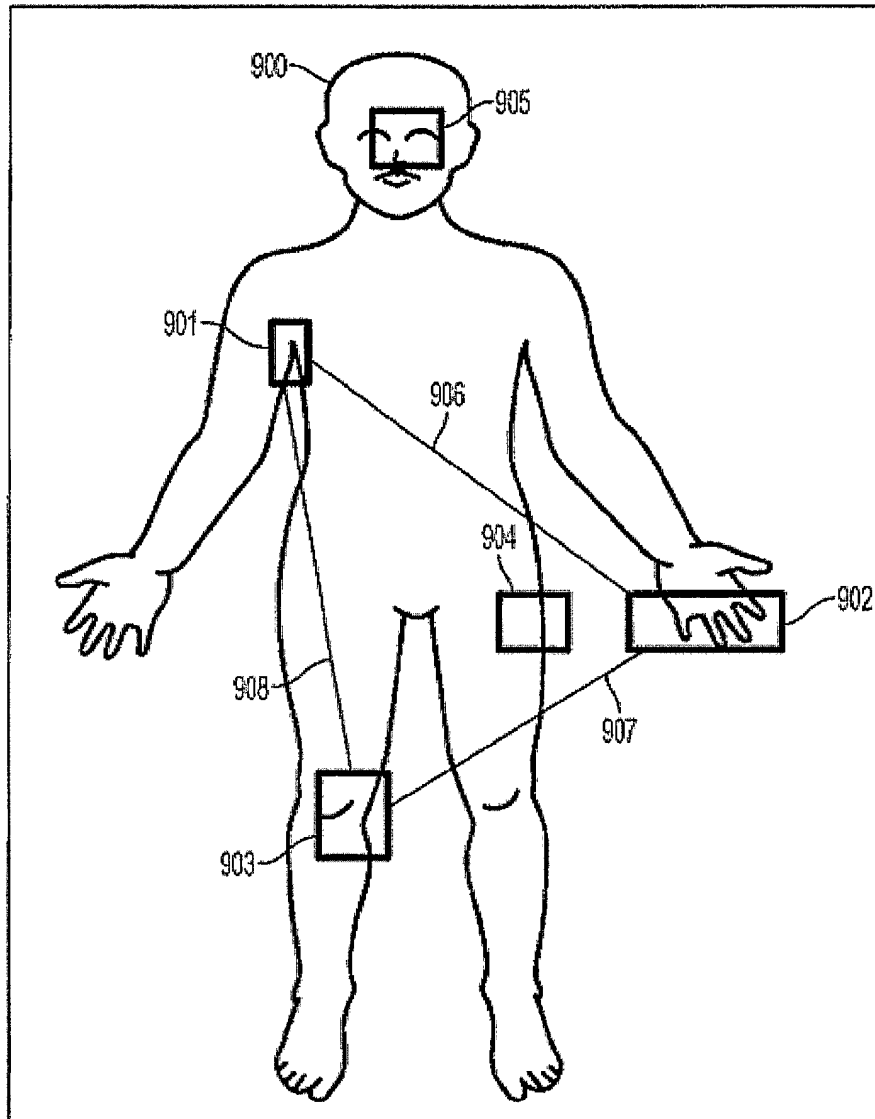


FIG. 9

END OF ADDENDUM A

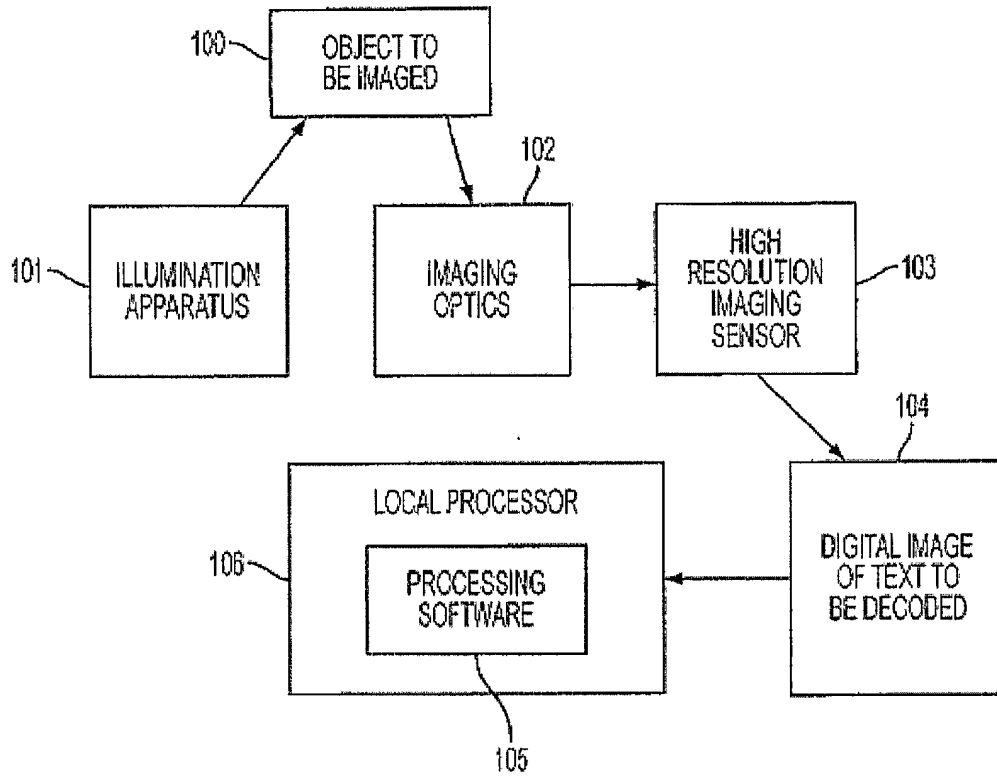


FIG. 1

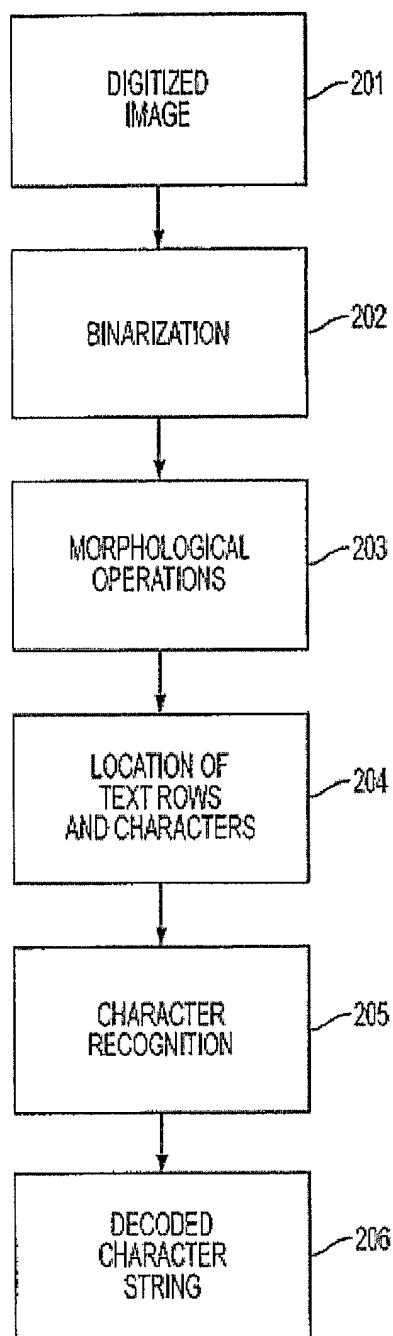


FIG. 2

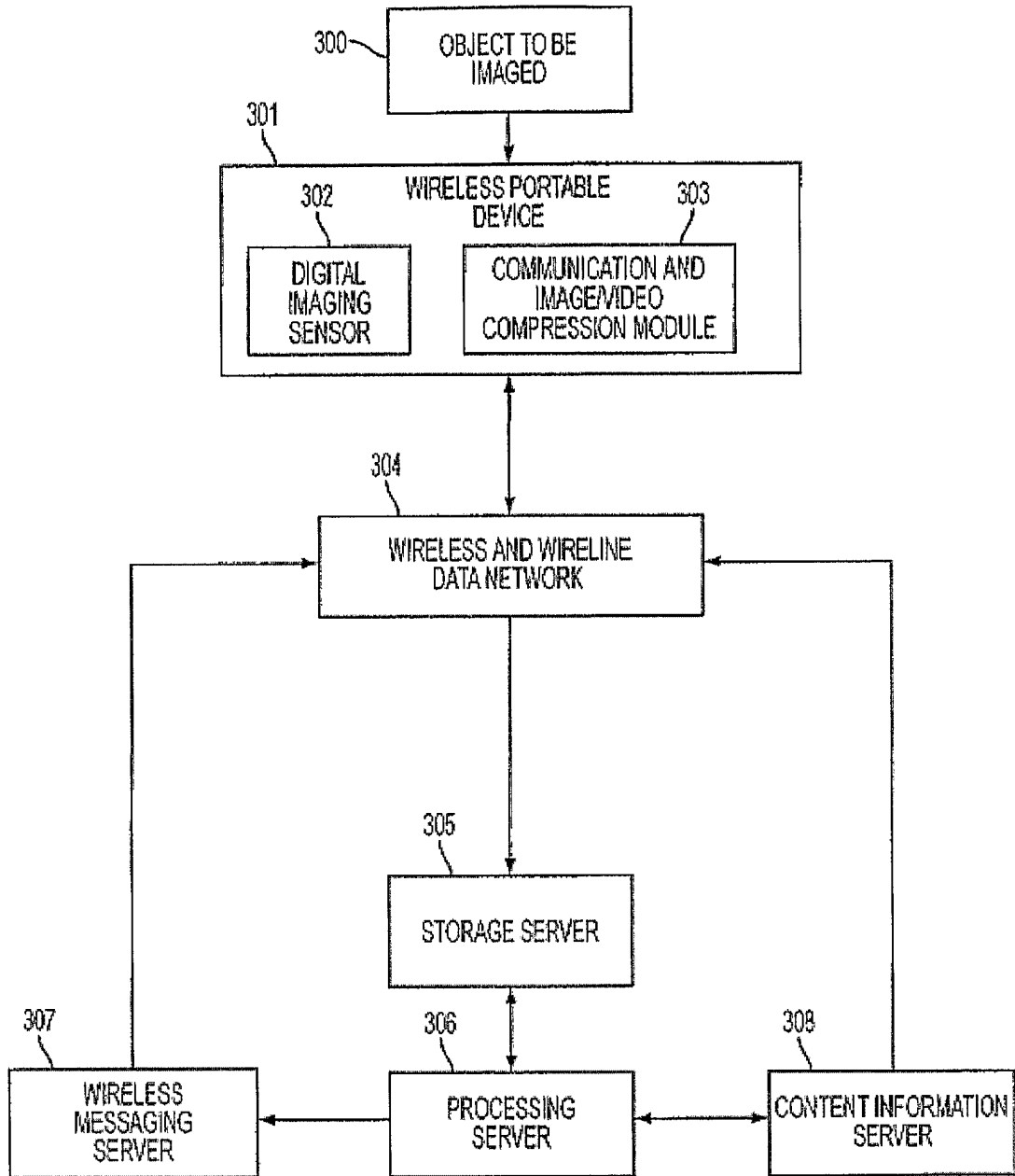


FIG. 3

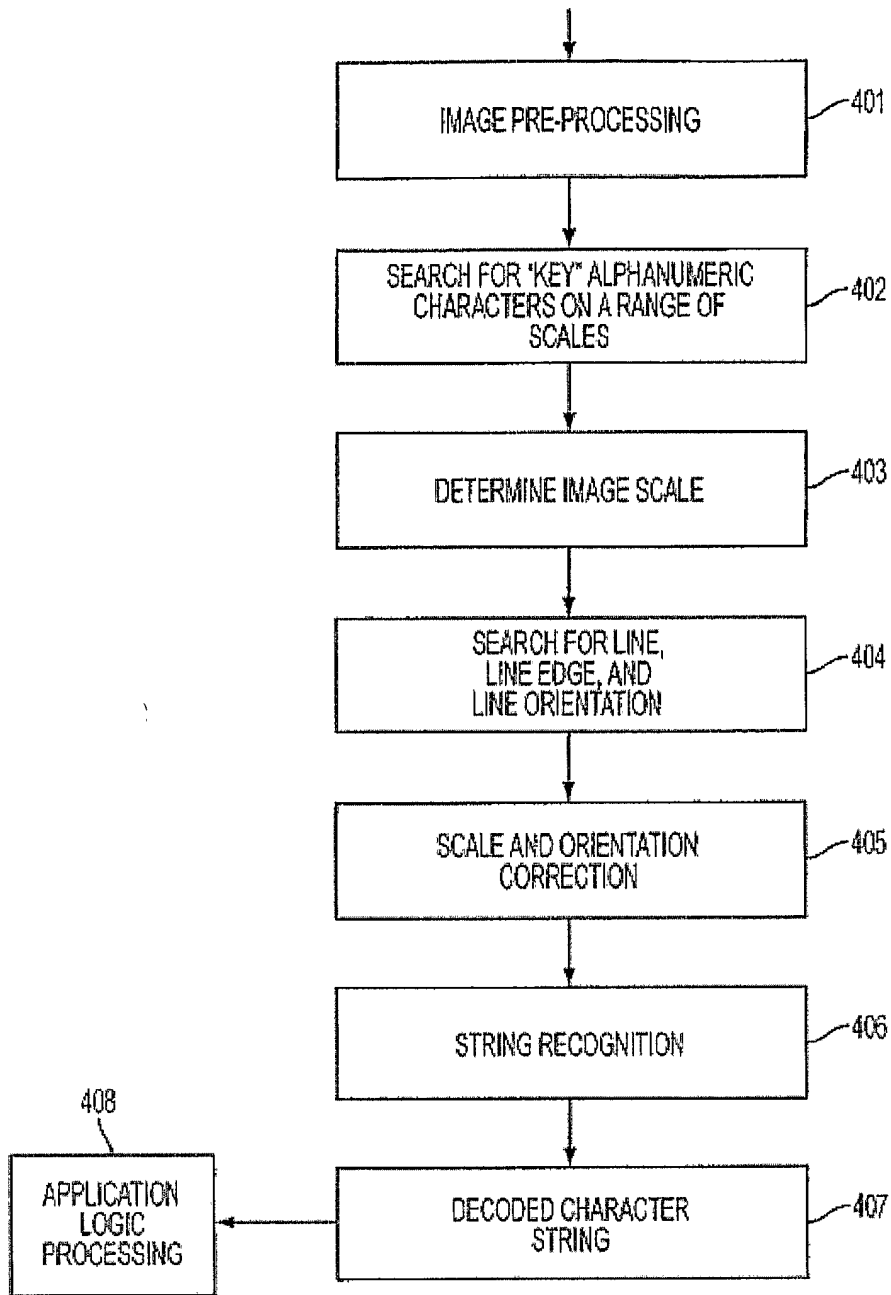


FIG. 4

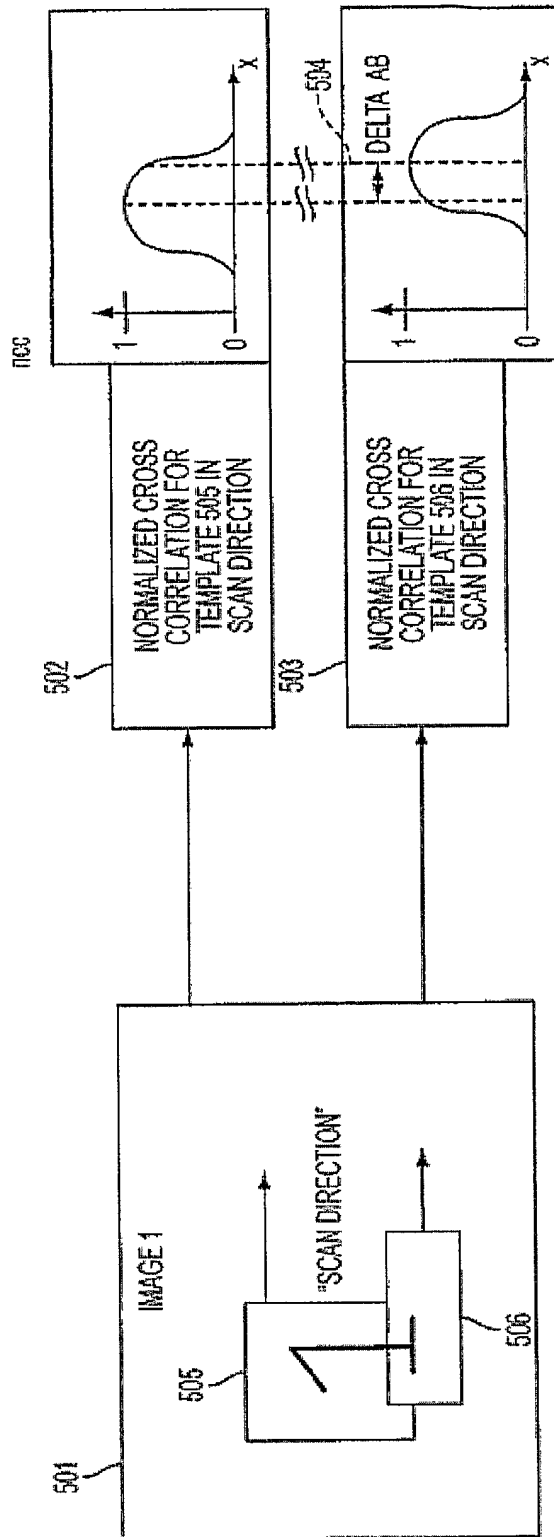


FIG. 5

END OF ADDENDUM B

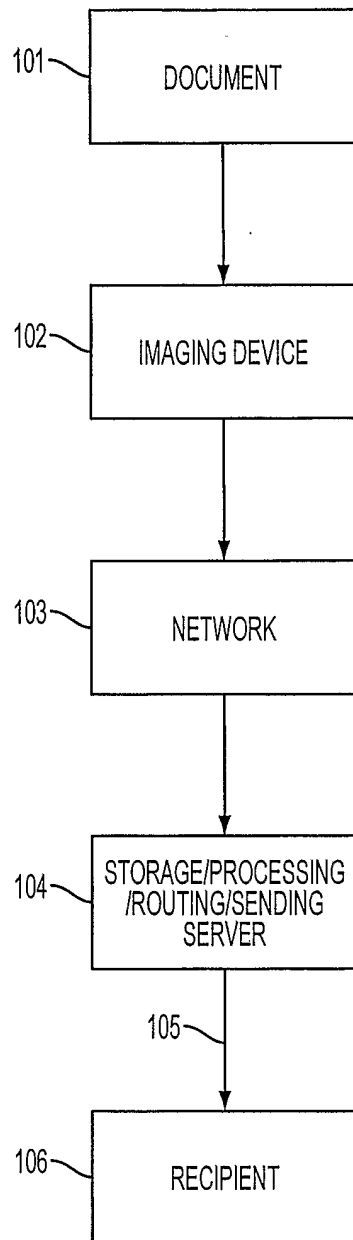


FIG. 1
RELATED ART

POSTAL SERVICE™	Date of Mailing	<input type="checkbox"/> Remit COD Charges to Sender via Express Mail	EMCA No.	
	FROM: Dear my Dime Dav, Now this is my style - to get to scribble my words and draw ☺☹☺☹☹.			TO:
	Delivered By	Date Delivered	Recoi	
	Check Number	MO Number	Date P	



PROCESSED IMAGE 202

<input type="checkbox"/> Registered Mail	<input type="checkbox"/> Express Mail	<input type="checkbox"/> Form 3 Request
Date of Mailing	<input type="checkbox"/> Remit COD Charges to Sender via Express Mail	EMCA No.
FROM: Dear my Dime Dav, Now this is my style - to get to scribble my words and draw ☺☹☺☹☹.		
Delivered By	Date Delivered	

FIG. 2
RELATED ART

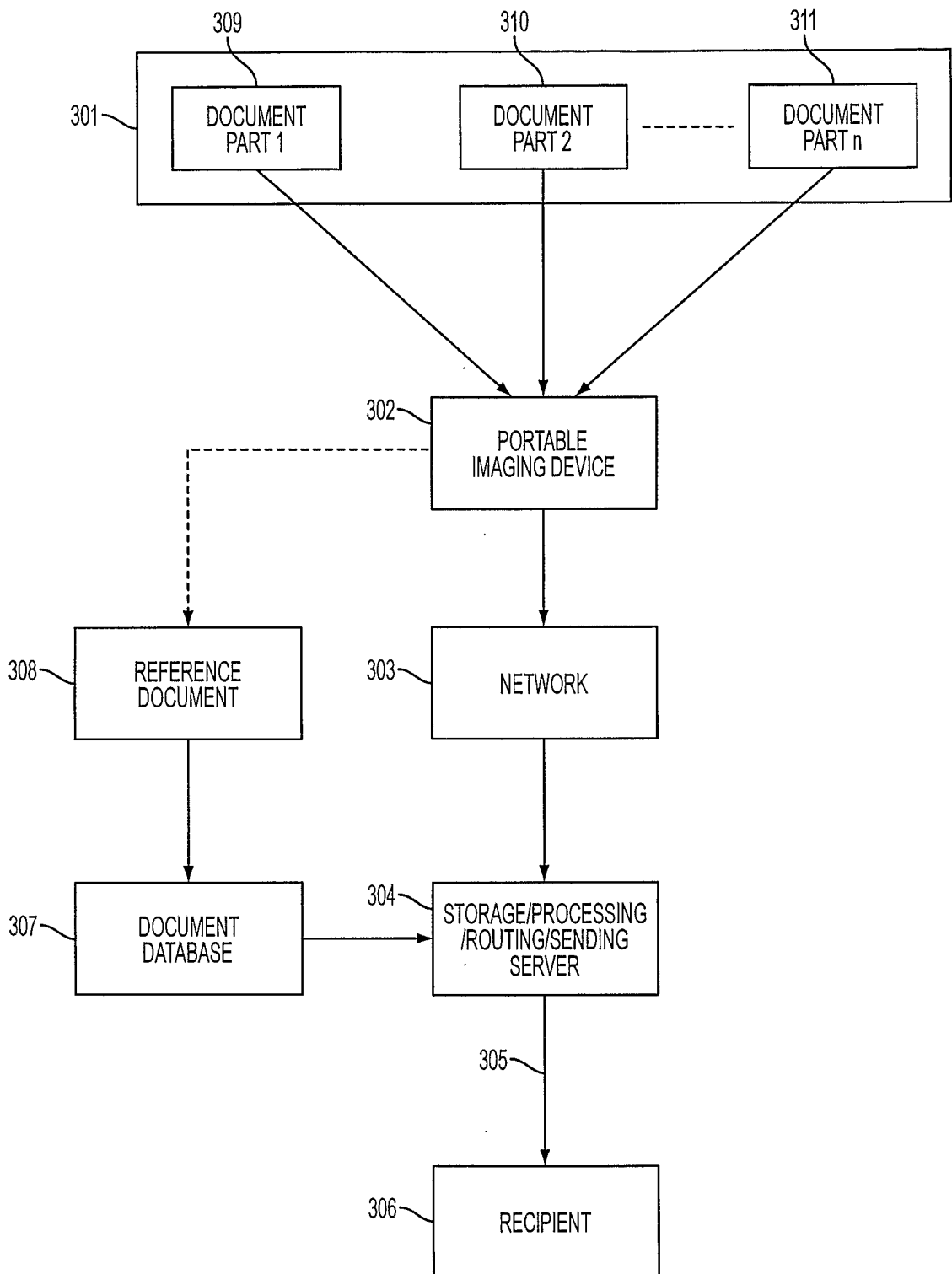


FIG. 3

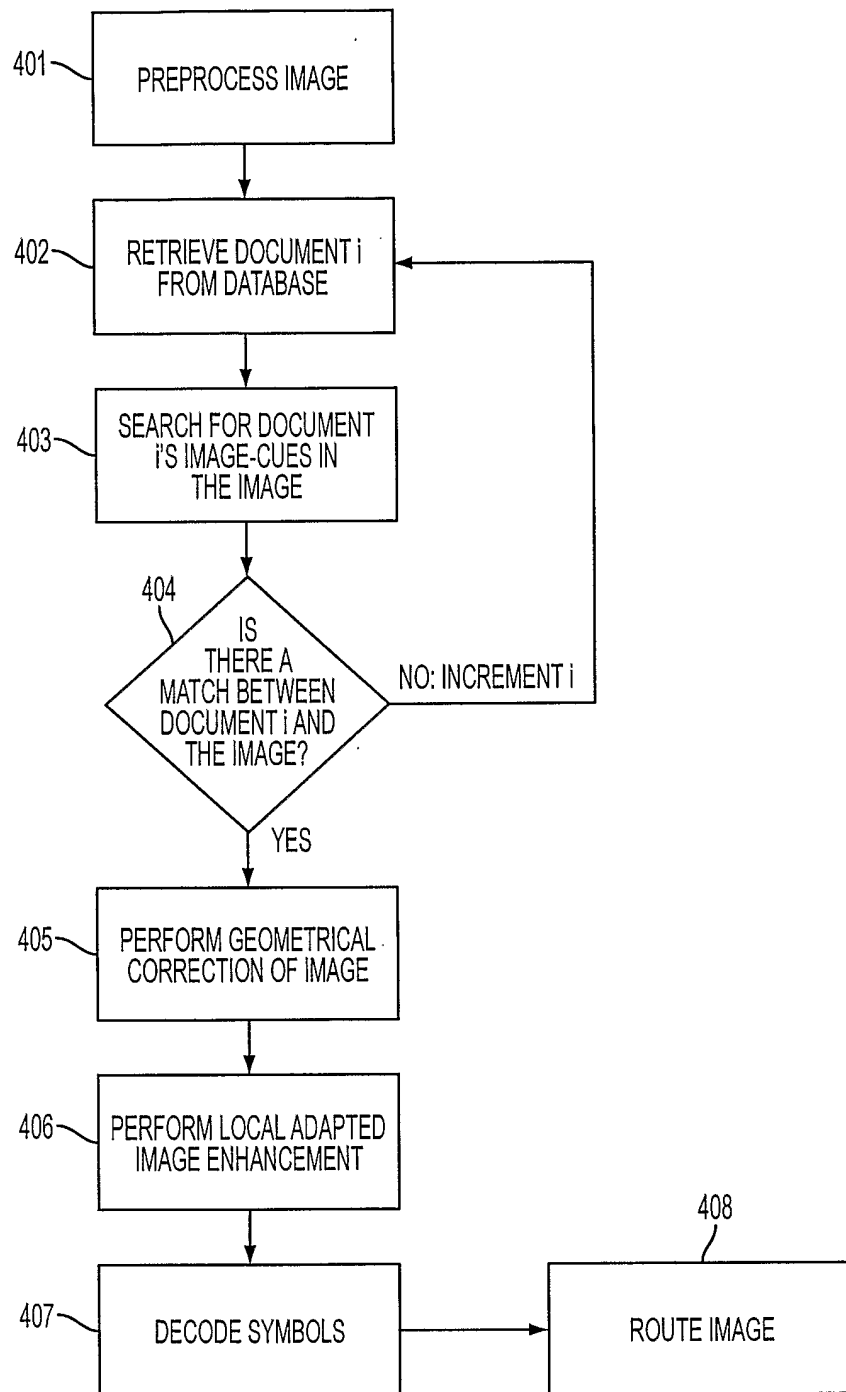


FIG. 4

500

United States Patent and Trademark Office
Credit Card Payment Form
Please Read Instructions before Completing this Form

Credit Card Information 502

Credit Card Account # **501** Type Visa MasterCard American Express Discover

Credit Card Expiration Date:

Name as it Appears on Credit Card:

Payment Amount (US Dollars):

Cardholder Signature: _____ Date: _____

Credit Card Billing Address 503

Street Address 1: _____

Street Address 2: _____

City: _____

State/Province: _____ Zip/Postal Code: _____

Country: _____

Daytime Phone # _____ Fax # _____

Request and Payment Information 504

Description of Request and Payment Information: _____

<input type="checkbox"/> Patent Fee	<input type="checkbox"/> Patent Maintenance Fee	<input type="checkbox"/> Trademark Fee	<input type="checkbox"/> Other Fee
Application No.:	Application No.:	Application No.:	Application No.:
Priority No.:	Priority No.:	Priority No.:	
Attorney/Agent No.:		Identify Trademark Mark:	

If the cardholder includes a credit card number on any form or document other than this Credit Card Payment Form, the United States Patent and Trademark Office will not be liable in the event that the credit card number becomes public knowledge.

FIG. 5A

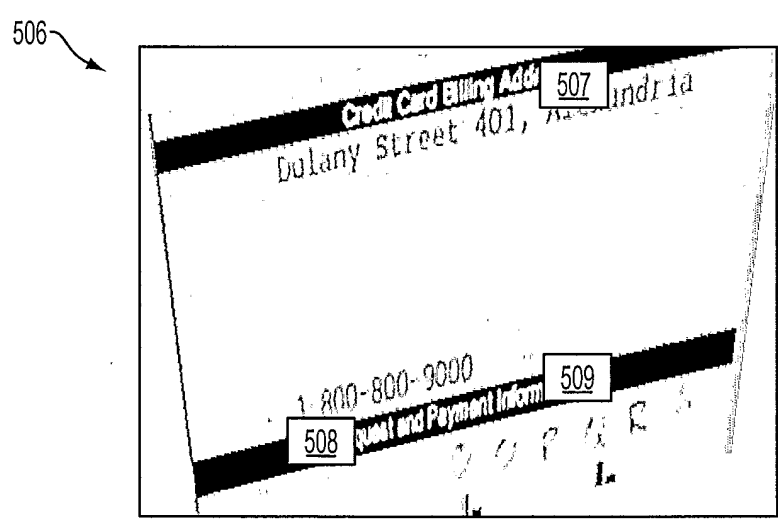


FIG. 5B

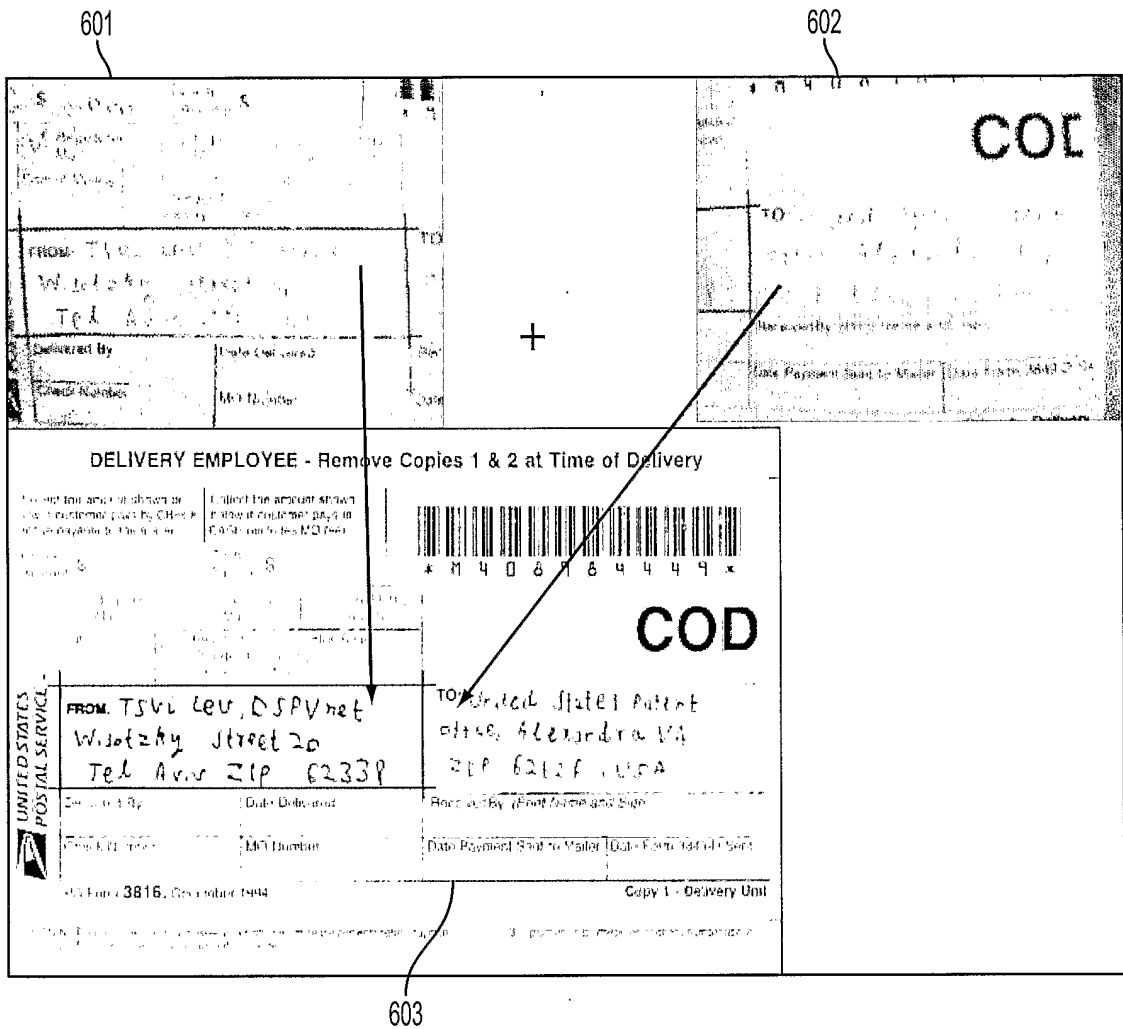


FIG. 6

701

PTO/SB/16 (10-05)

Approved for use through 07/31/2003. OMB 0651-0032

U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE

Persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET - Page 1 of 2

PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

INVENTOR(S)		
Family Name or Surname	Residence (City and either State or Foreign Country)	
<i>Additional inventors are being named on the _____ separately.</i>		
TITLE OF THE INVENTION (500 characters max)		

702

FIG. 7

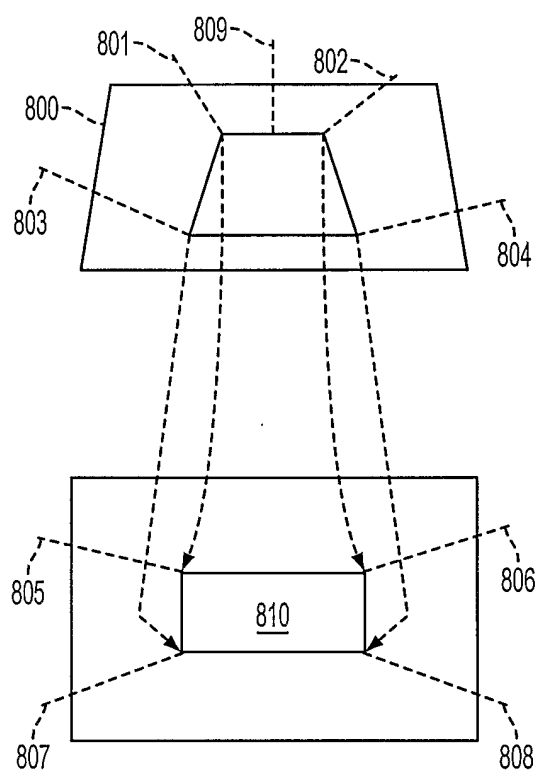


FIG. 8

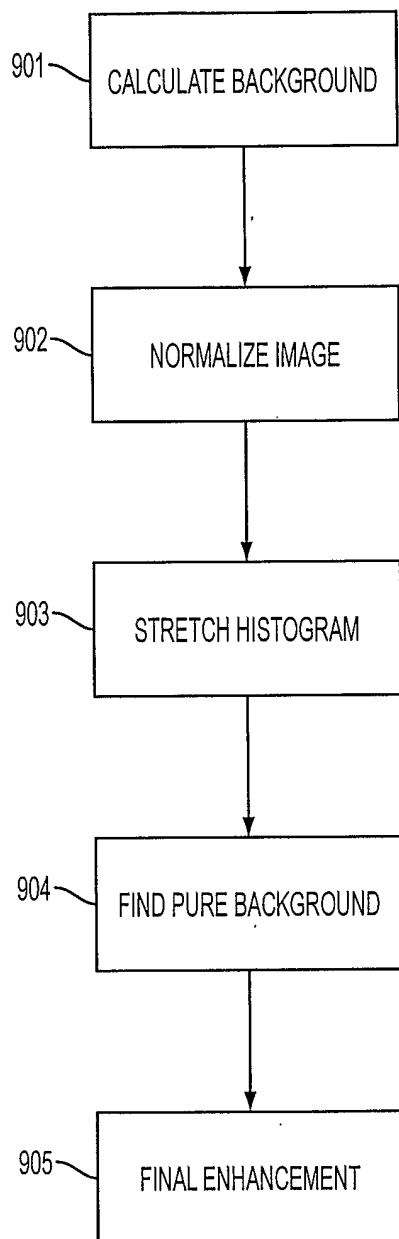


FIG. 9