An information-providing system is disclosed. The information-providing system can include an image-receiving module that receives an image from a device, an item-selection module that identifies an item based on the received image, an information-retrieving module that retrieves information relating to the item, and a data transmitting module that transmits the retrieved information to the device, wherein the item is identified by matching one or more features of the received image with features identified from a training image associated with the item.
FIG. 1

Server Obtains Information Relating to the Identified Item

Server Identifies Feature IDs Based on Image Feature IDs

Server Transmits Information to Requesting Device

Requesting Device Captures an Image of an Item

Requesting Device Transmits Image to Server

Server Calculates Feature IDs from Image

FIG. 2
Obtain Training Images Depicting Various Items

Identify Keypoints of a Training Image

Generate a Descriptor for Each Keypoint

Quantify the Description to Generate a Feature ID

Store Feature ID with Corresponding Item Associated with the Training Image

FIG. 4
### Table 5.1

<table>
<thead>
<tr>
<th>Feature ID</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Harry Potter &amp; Chamber of Secrets</td>
</tr>
<tr>
<td></td>
<td>Harry Potter &amp; Goblet of Fire</td>
</tr>
<tr>
<td>2</td>
<td>Da Vinci Code</td>
</tr>
<tr>
<td>3</td>
<td>Harry Potter &amp; Goblet of Fire</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,000,000</td>
</tr>
</tbody>
</table>

**FIG. 5**

**FIG. 7**
Receive Scanned Image from User Device

Identity Keypoints of the Scanned Image

Generate a Descriptor for each Keypoint

Quantify the Descriptor to Generate a Feature ID

Select Items Corresponding to the Feature IDs Based on Information in the Database

Determine a Total number of Hits for Each of the Selected Items

Select the Item with the Most Hits to be the Item Associated with the Scanned Image

Perform Geometric Verification of Scanned Image

FIG. 6
FIG. 8

FIG. 10
FIG. 9
SYSTEMS AND METHODS FOR OBTAINING INFORMATION BASED ON AN IMAGE

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims the priority benefit of Chinese Patent Application No. 201210235833.5, filed on Apr. 25, 2012, the contents of which are incorporated by reference herein in their entirety for all purposes.

FIELD

[0002] The present disclosure generally relates to image-based information retrieval, and more particularly, to methods and systems for identifying one or more items from a scanned image and retrieving information relating to the identified one or more items.

BACKGROUND

[0003] Very often in everyday life do people encounter items in the physical world that they would like to have more information about. For example, someone looking at a movie poster on a wall may want to find out more about the director and actors involved with the movie, such as their previous work. He may also want to see a preview of the movie and, if he likes the preview, find a nearby theater and buy a ticket to see the movie. Similarly, a person browsing in a book store may want to read reviews of a particular book or cross-shop the book at online book stores.

[0004] There are a number of existing ways to obtain information relating to items such as the movie poster or book. One way is to conduct a manual search using, for example, a browser or application-based search engine on a PC or mobile device. This process is usually tedious and slow because it requires the user to manually enter a descriptive search string. Also, it may only work well for text-based searches. It is usually difficult to run a search for an image without specialized software.

[0005] Another existing mechanism for retrieving information regarding an item is to scan a barcode (linear or matrix) associated with the item. The barcode can usually be found on or in close proximity of the item. It can be scanned using, for example, a dedicated scanner, such as a common barcode scanner, or a mobile device equipped with a camera and the required scanning application. However, there are certain limitations with scanning barcodes. For example, the amount of information retrievable from a barcode is usually limited. Scanning the barcode on a product in a supermarket may only provide the name and price of the product. More advanced barcodes, such as Quick Response (QR) codes, can provide a Web link, name, contact information such as an address, phone number, email address, and/or some other similar data type when scanned. Nevertheless, the information retrievable from these barcodes is typically limited to the information available in the corresponding backend system/database, such as an inventory management system of a supermarket. Such system/database may not have all the information desired by the person interested in the item.

[0006] Radio Frequency Identification (RFID) technology is another mechanism for automatically identifying and tracking tags attached to an item. RFID technology relies on radio-frequency electromagnetic fields to transfer data in a non-contacting fashion. An RFID system typically requires RFID tags to be attached to the item and a reader for reading data associated with a particular item from the corresponding tag. The reader can transmit the data to a computer system to be further processed. Nevertheless, RFID technology has the same shortcomings as barcodes in that only a relatively limited amount of information can be retrieved from reading the RFID tags. Furthermore, the fact that it requires special tags and readers makes it a less desirable solution for retrieving information since most people do not carry a RFID reader on them.

[0007] Accordingly, information retrieval systems and methods that can provide a simpler and more user-friendly experience and have access to a large information repository for providing information relating to a wide range of items are highly desirable.

SUMMARY

[0008] This generally relates to systems and methods for retrieving information relating to an item based on a scanned image of the item. In particular, the systems and methods can involve using a device, such as a smartphone, to capture a 2-dimensional image of an item and transmit the captured image to a server. The server can analyze the image against pre-stored data to determine a corresponding item associated with the image and obtain information relating to the item from a data repository such as the Internet. The information can then be transmitted from the server to the device.

[0009] In one embodiment, an information-providing system is disclosed. The information-providing system can include an image-receiving module that receives an image from a device, an item-selection module that identifies an item based on the received image, an information-retrieving module that retrieves information relating to the item, and a data transmitting module that transmits the retrieved information to the device, wherein the item is identified by matching one or more features of the received image with features identified from a training image associated with the item.

[0010] In another embodiment, the system can also include a training image processing module that identifies one or more features from at least one training image. In another embodiment, the training image processing module can further include: a keypoint-identifying module that identifies at least one keypoint of the received image, a descriptor-generating module that generates a descriptor for each of the at least one keypoint, a feature ID generating module that quantifies a descriptor to generate at least one feature ID, and a database-access module that stores the at least one feature ID and at least one of its corresponding item in a database. In another embodiment, the system can also include a database for storing the features identified from the training image. In another embodiment, the database can store the features and one or more items associated with each of the features. In another embodiment, the information is retrieved from the Internet. In another embodiment, the identified item is a book and the received image includes a book cover of the book.

[0011] In yet another embodiment, the item-selection module can further include: a keypoint-identifying module that identifies at least one keypoint of the received image, a descriptor-generating module that generates a descriptor for each of the at least one keypoint, a feature ID generating module that quantifies a descriptor to generate at least one feature ID, an item-selecting module that selects at least one item corresponding to each of the at least one feature ID, a hit-counting module that determines a total number of hits for each of the selected items, and a top item selection module.
that selects one of the selected items that best matches with the received image. In yet another embodiment, the item-selection module includes: a threshold module that determines whether the number of hits for an item exceeds a predetermined threshold, and an item-eliminating module that eliminates an item if the number of hits for the item does not exceed the predetermined threshold. In yet another embodiment, the item-selection module includes a geometric verification module that performs geometric verification on the received image and the training image associated with the best-matching item.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] FIG. 1 is a block diagram illustrating the exemplary components of an information-retrieval system, according to an embodiment of the disclosure.

[0013] FIG. 2 is a flowchart illustrating the exemplary steps in an image-based information retrieval process, according to an embodiment of the disclosure.

[0014] FIGS. 3A-3C are screen shots on the requesting device illustrating exemplary user interfaces for retrieving information based on a scanned image, according to an embodiment of the disclosure.

[0015] FIG. 4 is a flowchart illustrating the exemplary steps of an image-based information retrieval process, according to an embodiment of the disclosure.

[0016] FIG. 5 illustrates an exemplary database table for storing features IDs and items, according to an embodiment of the disclosure.

[0017] FIG. 6 illustrates exemplary steps in the process of determining an item based on a scanned image received from a user device, according to an embodiment of the disclosure.

[0018] FIG. 7 is a block diagram illustrating exemplary modules of the server for providing information in response to receiving a scanned image from the user device, according to an embodiment of the disclosure.

[0019] FIG. 8 is a block diagram illustrating the exemplary modules of the training image processing module of FIG. 7, according to an embodiment of the disclosure.

[0020] FIG. 9 is a block diagram illustrating the exemplary modules of the item-selection module of FIG. 7, according to an embodiment of the disclosure.

[0021] FIG. 10 illustrates exemplary common hardware components of a server, according to an embodiment of the disclosure.

DETAILED DESCRIPTION

[0022] In the following description of preferred embodiments, reference is made to the accompanying drawings which form a part hereof, and in which it is shown by way of illustration specific embodiments in which the disclosure can be practiced. It is to be understood that other embodiments can be used and structural changes can be made without departing from the scope of the embodiments of this disclosure.

[0023] This generally relates to systems and methods for retrieving information relating to an item based on a scanned image of the item. In particular, the systems and methods can involve using a device, such as a smartphone, to capture a 2-dimensional image of an item and transmit the captured image to a server. The server can analyze the image against pre-stored data to determine a corresponding item associated with the image and obtain information relating to the item from a data repository such as the Internet. The information can then be transmitted from the server to the device.

[0024] FIG. 1 is a block diagram illustrating the exemplary components of an information-retrieval system, according to an embodiment of the disclosure. As illustrated, a device 100 can be connected to a server 102, which in turn can be connected to the Internet 104. The device 100 can be any electronic device with image-capturing capability. In particular, the device 100 can include an image-capturing component such as a camera or a webcam. Although the device 100 in FIG. 1 is shown to be a smartphone, it can also be other devices including, for example, a personal computer (PC), Mac, desktop computer, laptop computer, tablet, PC, e-reader, camera, camcorder, in-car communication device, and other consumer electronic devices. The device 100 can use its camera to capture an image of an item of interest and send the captured image to the server 102. Accordingly, the device 100 can also include communication component for communicating with other devices including the server 102. For example, the device 100 can be connected to the server via a wired or a wireless connection/network including, but not limited to, the Internet, local area network (LAN), wide area network (WAN), cellular network, Wi-Fi network, virtual private network (VPN), and Bluetooth connection. In this embodiment, the image can be transmitted as part of a request requesting the server 102 to identify one or more items in the image and return additional information relating to the identified one or more items to the device 100. Accordingly, the device 100 can also be referred to as a “requesting device” hereinafter.

[0025] Although only one device 100 is shown to be connected to the server 102, it should be understood that additional devices can also be connected to the server 102 and request information in the same fashion. The server 102 can be any suitable computing device or devices capable of receiving image data from one or more devices 100, identifying an item based on the image data, retrieving additional data regarding the identified item from internal and/or external sources such as the Internet 104, and transmitting the retrieved data back to the requesting device(s). Various methods can be employed by the server to extract information (e.g., features such as the color, brightness, and/or relative position of the pixels) from the image and identify, based on the extracted information, one or more items associated with the image (e.g., an item depicted in the image). The information extracted from an image can identify features that, alone or in combination, can be used to identify the image among a collection of images.

[0026] In some embodiments, this can be done by the server querying a database containing pre-stored feature IDs and corresponding item(s). The feature IDs in the database can be generated from a collection of training images of various items. As referred to hereinafter, a training image can be any existing image depicting one or more items. One or more features can be identified from a training image by using a feature extracting mechanism, an example of which will be discussed in detail below. A unique ID (e.g., feature ID) for each of these features can be stored in the database along with the names or IDs of one or more items associated with the training image from which the features were identified. Essentially, the database can store pairing of various features and items.

[0027] When a request (or a scanned image) from a requesting device is received by the server, the server can perform the same feature extracting process on the scanned image to
generate one or more feature IDs from the scanned image. These feature IDs can then be used to query the database to find the matching item(s). The item with the best matching score can be identified as the item associated with the particular scanned image.

[0028] Referring again to FIG. 1, the server 102 can be network-enabled to communicate with one or more requesting devices. In addition, the server 102 can also have access to internal and/or external information repositories to search and retrieve information relating to items identified to be associated with a scanned image. In the embodiment illustrated in FIG. 1, the exemplary external repository can be the Internet 104. After the server identifies the item associated with the scanned image received from the requesting device, the server can gather information relevant to the item from the Internet and forward some or all of the gathered information to the requesting device. In other embodiments, the repository can be one or more internal or external database storing information regarding a collection of items. The type of information stored in the database can be customized based on the categories (e.g., movie, book) of items to be covered by the image-based information look-up service defined by the provider. In one embodiment, the server can include any information relating to any items in the physical world.

[0029] FIG. 2 is a flowchart illustrating the exemplary steps in an image-based information retrieval process, according to an embodiment of the disclosure. First, a requesting device can capture an image of an item (step 201). In one embodiment, this can be done by, for example, scanning a 2-dimensional item, such as a book, using the camera on a smartphone.

[0030] After the image is captured, the requesting device can transmit the image to a server to identify the item depicted in or associated with the image (step 202). For example, the scanned book cover can be sent to the server to retrieve information regarding the particular book. The image-transmitting step can take place automatically after the requesting device determines that the scanning operation was successful. Additionally or alternatively, the device may perform one or more quality-assurance steps to ensure that the captured image meets certain criteria in terms of clarity, resolution, size, brightness, etc. so that it can be properly analyzed by the server. If the image does not meet one or more of the criteria, the device can prompt the user to scan the image again. In some embodiments, the user has to manually transmit the image to the server. In some embodiments, the user can also enter additional information, such as keywords specifying the type of information to be returned by the server, to be transmitted with the image to the server. In the book cover example, the scanned image of the book cover can be transmitted to the server. The user may optionally enter keywords, such as “author” and/or “cover designer,” to be transmitted with the image. These keywords may direct the server to search for information specifically relating to the author and/or cover designer of this particular book. After the server identifies the book from the scanned image, the server can then search the Internet (or another data repository) for information relating to the item (step 203). In the book cover example, the server can determine that the cover image corresponds to the cover of one of the Harry Potter books. The server can then use a search on the Internet for information, such as title and plot summary of all Harry Potter books, readers’ reviews of the book, information relating to the physical appearance of the book and, online book stores offering this particular book for sale. Essentially, any information regarding the book that is available on the Internet can be found by the server and made available to the user device. In the embodiments where the scanned image is transmitted to the server along with keywords entered by the user, the search can incorporate these keywords to provide additional results tailored to the user’s interest.

[0031] After receiving the image transmitted from the requesting device, the server can identify one or more features of the image (step 203). In one embodiment, the one or more features can be represented by one or more feature IDs. For example, the image of the book cover may be processed by the server to extract certain features defined by, for example, color, brightness, and/or relative position of the pixels of the image. Each of these features can then be quantified as a unique feature ID.

[0032] Next, the server can identify an item based on the feature IDs calculated from the image (step 204). This can involve looking up items corresponding to each of the feature IDs in a database and ranking these items based on the number of feature IDs to which they correspond. The item ranked the highest can then be identified as the best match for the scanned image (e.g., most likely to be the item depicted in the image). Steps 203 and 204 will be described in further detail below.

[0033] After the item is determined, the server can then search the Internet (or another data repository) for information relating to the item (step 205). In the book cover example, the server can determine that the cover image corresponds to the cover of one of the Harry Potter books. The server can then use a search on the Internet for information, such as title and plot summary of all Harry Potter books, readers’ reviews of the book, information relating to the physical appearance of the book and, online book stores offering this particular book for sale. Essentially, any information regarding the book that is available on the Internet can be found by the server and made available to the user device. In the embodiments where the scanned image is transmitted to the server along with keywords entered by the user, the search can incorporate these keywords to provide additional results tailored to the user’s interest.

[0034] The server can then transmit the search results to the requesting device (step 206). The results can be displayed on the screen of the device for user browsing. They can also include web links to other websites where additional information can be available to the user. For example, the user may follow one of the links to the online book store to make a copy of the Harry Potter book. Additionally or alternatively, the user may also purchase and download an electronic copy of the book onto his device so that he can start reading right away.

[0035] The exemplary embodiments discussed above can provide a much simpler and effective way for retrieving information than any existing mechanisms including those described in the Background section. First, the disclosed systems and methods do not require any customized hardware, such as a barcode scanner. Any device with a camera can be used to scan an image of an item and receive information regarding the particular item from the server. Furthermore, the item of interest does not have to come with any barcode, QR code, or any other type of code to enable the information retrieval process. All it takes is for the user to scan, or capture in another way, a two-dimensional image of the item using the camera on his mobile phone to have access to potentially all kinds of information relating to the item. This process cannot be any more straightforward from the user end. Another advantage of the disclosed systems and methods is that the backend server can have access to the whole Internet to find information relating to the item. This can overcome the limitations of existing systems where only a limited amount of information (e.g., the information available in a closed system) can be returned in response to an inquiry based on, for example, a QR code.

[0036] FIGS. 3a-3c are exemplary screen shots of the requesting device illustrating user interfaces for retrieving information based on a scanned image, according to embodiments of the disclosure. In particular, FIG. 3a illustrates an exemplary interface 300 for initiating an image-scanning
operation. This interface can be an interface of an application (e.g., a book cover scanning application for retrieving information on a book by scanning the cover) installed on the device. In some embodiments, the interface can be launched from the home screen of the device directly and/or the camera application of the device. It can be launched by a softkey, hardkey, or voice control of the device. As illustrated in FIG. 3a, a menu including one or more options 304, 306, 308, 310 can be superimposed on top of the camera view, which may be darken to indicate that no active scanning is taking place. One of the options can be an “Image Scan” option 308 that can allow a user to scan an image of an item or object (e.g., a book cover) that is of interest to the user. The interface 300 can include additional menu options, such as “2D Code Scan” for scanning QR codes and “Translate” for translating text captured by the camera. Although three menu options are illustrated in the screen shot of FIG. 3a, it should be understood that the menu options can vary depending on the functions provided by the device and/or a particular application. In this embodiment, the menu can also include a “Cancel” softkey to leave this interface 300. Optionally, a frame 302 indicating the scanning area can be displayed to allow the user visually control the subsequent image scanning operation.

If the user hits the “Image Scan” softkey on the interface 300 of FIG. 3a, the scanning operation can be initiated and a corresponding “Image Scan” interface, such as the one illustrated in FIG. 3b, 36, can be displayed on the user device. As illustrated in FIG. 3b, the superimposed option menu is no longer displayed on the interface 312 of FIG. 3b. Instead, the image frame 302 can expand to occupy most of the display area on the interface 312. In this example, the device can be positioned so that the cover of a Harry Potter book can be viewed within the frame 302 through the camera lens of the device. As with any other camera-based operations, the user can move the device around to ensure that the item can be fully captured by the camera. The device can automatically start scanning when it determines that the item is within the frame 302. In this embodiment, the scan can produce a 2-dimensional image of the item. Optionally, as illustrated in FIG. 3b, the size 316 of the image file can also be displayed on the interface 312. The image can be automatically transmitted to the server to start a query to retrieve information relating to the item in the image. In some embodiment, one or more optional fields can be displayed either on the “Image Scan” interface 312 or a subsequent interface to allow the user to specify additional search and/or filtering criteria, such as keywords. For example, the user can input “author” and/or “price” to request specific information regarding the particular book.

The processes being performed by the server will be discussed in detail in the paragraphs after. After the server finds relevant information regarding the item associated with the scanned image, it can transmit this information in a specific format to the user device. FIG. 3c illustrates an exemplary “Scan Result” screen 318 displaying the information returned from the server. As illustrated, the “Scan Result” screen 318 can include a thumbnail of the original scanned image 320, the title 322, author 324, and edition 326 of the book. Each of the title 322, author 324, and edition 326 can also be a link to retrieve additional information regarding the respective field. In this embodiment, the screen 318 can also display information relating to at least one online book store (e.g., Amazon.com) 328 that has the book in stock. Clicking on the name of the store, for example, can allow the user to be redirected to the online store where he may purchase a copy of the book. Additionally or alternatively, the “Scan Result” screen 318 can display any other information identified by server as relating to the item in the scanned image. Depending on the type of item captured in the image, the “Scan Result” screen can be customized to display different kinds of information. For example, if the scanned image includes a movie poster, the scan results can include, for example, a clip of the preview of the movie, a list of other movies involving the same director or actors, list of theaters and show times, and/or a link to a movie ticket website where the user can purchase a ticket to see the movie.

As apparent from the exemplary user interfaces 300, 312, 318 of FIGS. 3a-c, embodiments of the disclosure may only require minimum user input, namely, pointing the camera of the user’s device at an item of interest, to obtain potentially all sorts of information regarding the item. From the user’s point of view, the whole process can be carried out in a seamless and extremely straightforward fashion with minimal amount of user input.

At mentioned above, when the server receives the scanned image from the requesting device, the server can identify various features of the image. The server can include a database of features extracted from a collection of known images (i.e., training images) of items. The server can then match the features extracted from the scanned image with the features from the collection of training images to identify a best-matching training image from the collection of training images. Because each training image can be associated with at least one known item, the server can identify one or more items relating to the scanned image if features from the scanned image are found to match with features from multiple training images.

First, the processing of training images by the server to generate and store a list of features and their associated items is discussed. FIG. 4 is a flow chart illustrating the exemplary steps of such a process, according to an embodiment of the disclosure. First, the server can collect images of various items (step 401). For example, if the server is to provide information relating to books and movies based on scanned book covers and movie posters received from the users, the server can have a collection of images of book covers, movie posters, and/or DVD covers. These images are referred to as training images and can be stored on the server as references. The process of generating feature IDs representing the features of a training image is discussed in the following steps 402-404. It should be understood that the process described in these steps is one of many different methods that can be used for generating a list of feature IDs from an image. Other suitable methods can also be employed to accomplish the same results.

In this embodiment, to extract one or more features from a training image, the server can first identify a number of keypoints of the image (step 402). First, scale-invariant feature transform (SIFT) features can be extracted from the training image. SIFT features can be invariant with respect to, for example, rotation, scaling, and illumination changes of the image. They can also be relatively stable with respect to, for example, the changing of the viewing angle, affine transformation, noise, and other factors that may affect an image. In one embodiment, SIFT feature extraction can be carried out as follows.
First, scale space extrema can be detected. To effectively extract stable keypoints, the training image can be convolved with Difference of Gaussians (DoGs) that occur at multiple scales.

\[ D(x,y) = (G(x,y,a) - G(x,y,b)) \ast I(x,y) \]

This can be achieved by generating Gaussian image pyramids. The image pyramids can be in P groups and each group can include S layers. The layers of the first group can be generated by convolving the original image with DoGs that occur at multiple scales (adjacent layers can have a scale difference of factor k). The next group can be generated by downsampling the previous group of images. A DoG pyramid can be generated from the differences between the adjacent Gaussian image pyramids.

To locate the scale space extrema, each sampling point (e.g., pixel) in the DoG pyramid can be compared to its eight adjacent pixels at the same scale and nine upper and nine lower neighboring pixels in each of the neighboring scales (a total of 3 + 2 × 9 = 26 pixels). If the value of the pixel is lesser or greater than the value of the 26 neighboring pixels, the pixel can be determined as a local extremum (i.e., a keypoint).

Next, an accurate location of each keypoint can be determined. Specifically, this can be done by fitting a 3-dimensional quadratic function to accurately determine the location and scale of each keypoint. At the same time, low contrast candidate points and edge response points along an edge can be discarded to improve consistency in the later feature-matching processes and also increase noise-rejection capability. To accurately locating a keypoint can include determining a main orientation of the keypoint and generating a descriptor of the keypoint.

To determine the orientation of a keypoint, the keypoint can be used as the center of the neighboring window for sampling. An orientation histogram can be used for determining the gradient orientation of the neighboring pixels. An orientation histogram with 36 bins can be formed, with each bin covering 10 degrees for a total range of 0-360 degrees. The peaks in this histogram can correspond to the dominant orientations of the neighboring gradient of the keypoint, and thus can be used as the dominant orientations of the keypoint.

In the gradient orientation histogram, the orientations corresponding to the peaks that are within 80% of the highest peaks can be the supplemental direction of the keypoint.

Referring to FIG. 4, next, a descriptor for each keypoint can be generated (step 403). In this embodiment, first, the zero degree direction of the axes can be rotated to match to the dominant orientation of the keypoint to achieve rotational invariance. Then, in a 16 × 16 region around the keypoint, a set of orientation histograms can be created on 4 × 4 pixel neighborhoods each with 8 bins. A sum of each gradient orientation can then be determined. Each keypoint can be described by the 4 × 4 = 16 sums. Since there are 4 × 4 × 16 histograms, each with 8 bins, 128 values can be generated for each keypoint to form a 128-dimensional SIFT feature vector. The SIFT vector can already have the effects from geometric distortion factors such as scaling or rotating of the training image removed. This vector can then be normalized to unit length in order to enhance invariance to affine changes in illumination.

The 128-dimensional SIFT feature vector can then be quantified as a feature ID (e.g., a number from 1-1,000,000) (step 404). That is, each SIFT feature vector representing a feature of the training image can have a corresponding numeric feature ID. Typically, more than one feature can be identified from a training image. Accordingly, each training image may be associated with multiple feature IDs. Because each training image can be associated with an item (e.g., the item depicted in the image), the item can also be associated with the multiple feature IDs.

Similarly, the same feature may be found in different training images. For example, the book cover of “Harry Potter and the Chamber of Secrets” may share some of the same features with that of “Harry Potter and the Goblet of Fire” (e.g., both covers may include an image of the text “Harry Potter”). Accordingly, each feature ID may be associated with multiple items. The relationship between the features (as identified by their respective feature ID) and the items can be captured and stored in a database accessible to the server (step 405). In the various embodiments, the database can be any suitable data storage program/format including, but not limited to, a list, text file, spreadsheet, relational database, and/or object-oriented database. FIG. 5 illustrates an exemplary database table 500 for storing features IDs 502 and their corresponding items 504. It should be understood that the format of the database and the structure of the database can vary and do not have to conform to the two-column format shown in FIG. 5. In this example, the feature IDs 502 can be unique incremental numbers from, for example, 1 to 1 million. Each feature ID can be associated with a feature from one of the training images and generated using the process described in FIG. 4. The number of features IDs can depend on the number of features ascertainable from the training images and does not have to be limited to 1 million as shown in table 500.

As shown in the table 500, each feature ID can correspond to one or more items. As previously discussed, when the same feature is found in two different images, the corresponding feature ID can be associated with two different items. For example, as shown in the table of FIG. 5, both “Harry Potter and the Chamber of Secrets” and “Harry Potter and the Goblet of Fire” can be associated with the feature ID “1.” However, “Harry Potter and the Goblet of Fire” may also have a second feature (identified by feature ID “3”), that is not associated with “Harry Potter and the Chamber of Secrets.” In this example, feature ID “2” can correspond to another book, “Da Vinci Code.” Although book names are listed under the “item” column 504 in table 500, it should be understood that the book names can be replaced by item IDs in other embodiments, where each book can be associated with a unique item ID.

When the server receives a scanned image from a user device as an request for information relating to the item in the image, the server can process the scanned image to extract features IDs representing various features of the scanned image and look up the corresponding item(s) from the database (e.g., the table of FIG. 5) to determine an item associated with the scanned image.

FIG. 6 illustrates exemplary steps in the process of determining an item based on a scanned image received from a user device, according to embodiments of the disclosure. First, the server can receive a scanned image from the user device (step 601). The scanned image can be transmitted over a network. In some embodiments, other information associated with the scanned image can also be received by the server. Such information may include an ID of the user device from which the scanned image was transmitted and/or keywords specified by the user. The scanned image can then be
processed to generate one or more feature IDs that identify the various features of the image. The same process discussed above for generating feature IDs from training images can be applied on the scanned image. In particular, keypoints of the scanned image can be identified (step 602). A descriptor for each identified keypoint can be generated (step 603). The descriptors can then be quantified to generate feature IDs (step 604). Steps 602-604 can correspond to steps 402-404 of FIG. 4. As such, details of exemplary implementations of each of these steps are not repeated here.

With the feature IDs determined, corresponding item(s) for each feature ID can be looked up from a database (e.g., table of FIG. 5) listing all feature IDs generated from the training images and the items corresponding to each of these feature IDs. That is, the items corresponding to the feature IDs of the scanned image can be selected from the database (step 605). Using the table of FIG. 5 as an example, if the feature IDs from the scanned image include “1” and “3,” both “Harry Potter and the Chamber of Secrets” and “Harry Potter and the Goblet of Fire” can be selected because each of these items corresponds to at least one of feature IDs “1” and “3.” In contrast, “Da Vinci Code” is not selected because the scanned image did not generate feature ID “2.”

Next, the total number of hits for each of the selected items can be determined (step 606). For example, “Harry Potter and the Chamber of Secrets” can have a total of one hit while “Harry Potter and the Goblet of Fire” can have a total of two hits based on the information in the table of FIG. 5. Typically, the higher the total number of hits is for an item, the better match it can be with regard to the scanned image. Optionally, the total number of hits for each item can be compared to a predetermined threshold value to eliminate feature IDs with relatively low number of hits (step 607). In one embodiment, for example, the threshold number can be 20. If the total hit of an item does not meet the threshold, it can be eliminated from consideration. Among the items that exceed the threshold, the item with the most hits can be selected to be the corresponding item for the particular scanned image (step 608).

In some embodiment, a geometric verification step can be performed to further verify that the scanner image matches with the training image associated with the candidate item (step 610) before an item is determined to be the best match for the scanned image. In particular, geometric verification can involve matching the individual pixels or features from the scanned image with those from the training image of the item selected through the process described above. This can be done by, for example, measuring and comparing the relative distances between two or more pixels in each of the two images. Based on how well the relative distances between the pixels match in the two images, it can be determined whether the training image is a top match for the scanned image. If the geometric verification is successful, the item associated with the training image can be confirmed to be the item associated with the scanned image.

After an item is determined to be the item corresponding to the scanned image, the server can search for information relating to the item in one or more data repositories. For example, if “Harry Potter and the Goblet of Fire” is determined to be the item associated with the scanned image received from the user device, the server can conduct a search for information relating to this particular book. The results from the search can then be transmitted back to the user device for display, as shown in the screen shot of FIG. 3c.

Although the above embodiments describe identifying books and movies from images of book covers and movie posters, respectively, the same methods and systems can be applied to obtain information relating to any item, as long as an image of the item can be captured by scanning or other mechanisms and the item in the image can be recognized based on the information available (e.g., information extracted from the training images) to the server. In various embodiments, the item can also include, for example, the logo of a product, a screen shot from another device, a work of art such as a painting, or a 3-dimensional object such as a building. It should also be understood that the processes for extracting information such as feature IDs from an image are not limited to those described in the embodiments above.

Without departing from the spirit of the disclosure, other suitable processes for recognizing text, graphics, facial expressions, geographic locations, 1D and 2D codes, etc. can also be used for identifying a particular item for the purpose of providing information relating to the item. Examples of other types of image processing systems and methods are described in, for example, Chinese Patent Application No. 201210123853.5, filed Apr. 26, 2012, the content of which is incorporated by reference herein in its entirety.

The above-described exemplary processes including, but not limited to, generating a list of feature IDs from training images, storing these feature IDs with their corresponding items in the database, determining a best-matching item for a scanned image using the information stored in the database, and obtaining information relating to the best-matching item can be implemented using various combinations of software, firmware, and hardware technologies on the server (or a cluster of servers). The server may include one or more modules for facilitating the various tasks of these processes. FIGS. 7-9 illustrate exemplary modules in an exemplary server for performing these tasks. In some embodiments, these modules can be implemented mostly in software.

FIG. 7 illustrates exemplary modules of the server for providing information in response to receiving a scanned image from the user device, according to embodiments of the disclosure. The server 700 can include, for example, a training-image processing module 702, an image-receiving module 704, an item-selection module 706, an information retrieval module 708, a data-transmission module 710, and a database 712. The image-receiving module 704 can receive one or more scanned images from user devices. The received scanned images can be processed by the item-selection module 706 to determine an item that is the best match to the scanned image. In some embodiments, the item-selection module 706 can perform one or more of steps illustrated in FIG. 6. The information retrieval module 708 can be connected to the Internet for retrieving information relating to the selected item. The information can then be transmitted back to the requesting device by the data-transmission module 710.

The server can also include a training image processing module 702 for collecting and processing training images to identify various feature IDs and their corresponding items. In some embodiments, the training image processing module 702 can perform one or more steps illustrated in FIG. 4. The corresponding feature IDs and items identified by the training image processing module 702 can be stored in the database 712, which can be accessed by the item-selection module 706 during the process of selecting an item based on a received scanned image.
FIG. 8 is a block diagram illustrating the exemplary modules of the training image processing module of FIG. 7, according to embodiments of the disclosure. As illustrated, the training image processing module 800 can include, for example, a training image obtaining module 801, a keypoint-identifying module 802, a descriptor-generating module 803, a feature ID generating module 804, and a database-access module 805. The training image obtaining module 801 can obtain training images (e.g., performing step 401 in FIG. 4) to be processed by the other modules of the training image processing module 800. The keypoint-identifying module 802 can identify keypoints of a training image (e.g., performing step 402 in FIG. 4). The descriptor-generating module 804 can generate a descriptor for each identified keypoint (e.g., performing step 403 in FIG. 4). The feature ID generating module 804 can quantify the descriptors to generate corresponding feature IDs (e.g., performing step 404 in FIG. 4). The database-access module 805 can read and write to a database. In particular, the database-access module 805 can store feature IDs with the names of their corresponding items in the database (e.g., performing step 405 in FIG. 4).

FIG. 9 is a block diagram illustrating the exemplary modules of the item-selection module of FIG. 7, according to embodiments of the disclosure. The item-selection module 900 can include one or more of the following sub-modules: a scanned image receiving module 901 for receiving scanned images from one or more user devices (e.g., performing step 601 in FIG. 6); a keypoint-identifying module 902 for identifying keypoints of a scanned image (e.g., performing step 602 in FIG. 6); a descriptor-generating module 903 for generating a descriptor for each keypoint (e.g., performing step 603 in FIG. 6); a feature ID generating module 904 for quantifying a descriptor to generate a feature ID (e.g., performing step 604 in FIG. 6); an item-selecting module 905 for selecting items corresponding to one or more feature IDs (e.g., performing step 605 in FIG. 6); a hit-counting module 906 for determining a total number of hits for each of the selected items (e.g., performing step 606 in FIG. 6); a threshold module 907 for determining whether the number of hits for an item exceeds a predetermined threshold (e.g., performing step 607 in FIG. 6); a top item selecting module 908 for selecting the item as the item that best matches the scanned image (e.g., performing step 608 in FIG. 6); an item-eliminating module 909 for eliminating an item if the number of hits for the item does not exceed a predetermined threshold (e.g., performing step 609 in FIG. 6); and a geometric verification module 910 for performing geometric verification on a scanned image based on the training image associated with the best-matching item (e.g., performing step 610 in FIG. 6).

In some embodiments, the training image processing module 800 and the item-selection module 900 can share one or more of the keypoint-identifying module, descriptor-generating module, and feature ID generating module 904.

In some embodiments, one or more of these modules on the server can be stored and/or transported within any non-transitory computer-readable storage medium for use by or in connection with an instruction execution system, apparatus, or device, such as a computer-based system, processor-containing system, or other system that can fetch the instructions from the instruction execution system, apparatus, or device and execute the instructions. In the context of this document, a "non-transitory computer-readable storage medium" can be any medium that can contain or store the program for use by or in connection with the instruction execution system, apparatus, or device. The non-transitory computer-readable storage medium can include, but is not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus or device, a portable computer diskette (magnetic), a random access memory (RAM) (magnetic), a read-only memory (ROM) (magnetic), an erasable programmable read-only memory (EPROM) (magnetic), a portable optical disc such as a CD, CD-R, CD-RW, DVD, DVD-R, or DVD-RW, or flash memory such as compact flash cards, secured digital cards, USB memory devices, memory sticks, and the like.

The non-transitory computer-readable storage medium can be a part of a computing system serving as the server. FIG. 10 illustrates exemplary computer components of a system such as a computing system. As illustrated, the system 1000 can include a central processing unit (CPU) 1002, I/O components 1004 including, but not limited to one or more of display, keypad, touch screen, speaker, and microphone, storage medium 1006 such as the ones listed in the last paragraph, and network interface 1008, all of which can be connected to each other via a system bus 1010. The storage medium 1006 can include one or more of the modules of FIGS. 7-9.

Although the modules illustrated in FIGS. 7-9 are described to be modules on the server, it should be understood that, in other embodiments, one or more of these modules can be part of the requesting device. That is, at least part of the above-described processes of, for example, generating feature IDs from training images and scanned images, identifying one or more items based on the feature IDs, and searching for information relating to the identified items can be performed by the requesting device without involving the server.

For example, referring again to FIG. 7, one or more of the image-receiving module 704, training image processing module 702, item-selection module 706, information retrieval module 708, data transition module 710, and database 712 can be a part of the requesting device. The image-receiving module 704 can be connected to the camera of the device for receiving images captured by the camera. It can also be connected to, for example, a communication module for receiving an image from an email program, messaging application, the Internet, and/or removable storage modules such as SIM cards and USB drives. The requesting device can also process these images locally using one or more of the other modules shown in FIG. 7. For example, it can perform one or more steps illustrated in FIG. 4 to identify feature IDs from training images and store them in a local database. Alternatively or additionally, the requesting device can also perform one or more steps illustrated in FIG. 6 to select an item based on a scanned image. Alternatively or additionally, the requesting device can retrieve information relating to the selected item from the Internet or other internal or external data repositories. Depending on the networked by the requesting device, one or more modules illustrated in FIGS. 8 and 9 can also reside on the requesting device.

In one embodiment, one of the requesting devices can process image-based information retrieval request from one or other requesting devices. As such, a dedicated server is necessary to carry out the processes of methods discussed above. In other embodiments, the various steps and tasks involved in the processes described in view of FIGS. 4-6 can be divided between the server and one or more requesting devices. For example, the server may handle the processing of
training images and the requesting device can handle the processing of scanned images locally and transmit only the feature IDs generated from the scanned image to the server for identifying an item and retrieving information from the Internet. In another example, the server can process the training images to obtain a list of feature IDs with corresponding items (e.g., the table of FIG. 5) and transmit the list to one or more requesting devices. The requesting device(s) can store the list locally and, after processing a scanned image, identify an item based on information from the scanned image and the list. The requesting device(s) can then search for information relating to the item by either directly connecting to the Internet and perform a search or passing along the item to the server and ask the server to perform the search and return the search results.

[0071] Essentially, embodiments of the disclosure provides methods and systems that can allow a user to scan a 2-dimensional image of any item of his/her interest, for example, his/her smartphone, and provide him with all sorts of information relating to the item that can be ascertained from, for example, the Internet and/or other existing data repositories. This can provide a simple, low-cost, but also user-friendly and effective way of looking up information relating to anything that can be captured in an image.

[0072] Although embodiments of this disclosure have been fully described with reference to the accompanying drawings, it is to be noted that various changes and modifications will become apparent to those skilled in the art. Such changes and modifications are to be understood as being included within the scope of embodiments of this disclosure as defined by the appended claims.

1. An image analysis system comprising:
   - an image-receiving module that receives an image from an external source,
   - a keypoint-identifying module that identifies at least one keypoint from the image,
   - a descriptor-generating module that generates a descriptor for each of the at least one identified keypoint,
   - a feature ID-generating module that generates a feature ID to each descriptor,
   - an item-selecting module that selects at least one item based on at least one feature ID generated by the feature ID-generating module,
   - a top item selecting module that selects a best-matched item from the selected at least one item, and
   - an information retrieving module that retrieves information relating to the best-matched item.

2. The image analysis system of claim 1, wherein the external source comprises a network.

3. The image analysis system of claim 1, wherein the keypoint-identifying module identifies the at least one keypoint from the image extracts at least one Scale-Invariant feature transform (SIFT) feature from the image.

4. The image analysis system of claim 1, wherein the keypoint-identifying module identifying the at least one keypoint from the image comprises determining a location and orientation associated with the at least one identified keypoint.

5. The image analysis system of claim 1, wherein the descriptor-generating module generating a descriptor for each of the at least one identified keypoint comprises generating a SIFT vector for each of the at least one identified keypoint.

6. The image analysis system of claim 5, wherein the feature ID-generating module generating a feature ID to each descriptor comprises assigning a unique feature ID to each SIFT vector.

7. The image analysis system of claim 1, wherein the item-selecting module selecting at least one item based on at least one feature ID comprises:
   - querying a database containing multiple feature IDs and items corresponding to each of the feature IDs to obtain at least one item corresponding to the at least one item, and
   - the top item selecting module selecting a best-matched item comprises:
     - determining an item with a most number of hits from the obtained at least one item as the best-matched item.

8-10. (canceled)

11. An information-providing system comprising:
   - an image-receiving module that receives an image from a device,
   - an item-selection module that identifies an item based on the received image,
   - an information-retrieving module that retrieves information relating to the item, and
   - a data transmitting module that transmits the retrieved information to the device,
   wherein the item is identified by matching one or more features of the received image with features extracted from a training image associated with the item.

12. The information-providing system of claim 11, comprising:
   - a training image processing module that identifies one or more features from at least one training image.

13. (canceled)

14. The information-providing system of claim 11, comprising:
   - a database for storing the features identified from the training image.

15. (canceled)

16. The information-providing system of claim 11, wherein the information is retrieved from the Internet.

17. The information-providing system of claim 11, wherein the identified item is a book and the received image comprises a book cover of the book.

18. The information-providing system of claim 11, wherein the item-selection module comprises:
   - a keypoint-identifying module that identifies at least one keypoint of the received image,
   - a descriptor-generating module that generates a descriptor for each of the at least one keypoint,
   - a feature ID generating module that quantifies a descriptor to generate at least one feature ID,
   - an item-selecting module that selects at least one item corresponding to each of the at least one feature ID,
   - a hit-counting module that determines a total number of hits for each of the at least one selected item, and
   - a top item selection module that selects one of the at least one selected item that best matches with the received image.

19-20. (canceled)

21. An information-providing method comprising:
   - receiving an image from a device,
   - identifying an item based on the received image,
   - retrieving information relating to the item, and
   - transmitting the retrieved information to the device,
wherein the item is identified by matching one or more features of the received image with features identified from a training image associated with the item.

22. The information-providing method of claim 21, comprising:
identifying one or more features from at least one training image.

23. (canceled)

24. The information-providing method of claim 21, comprising:
storing the features identified from the training image.

25. (canceled)

26. The information-providing method of claim 21, wherein retrieving the information relating to the item comprises retrieving the information related to the item from the Internet.

27. The information-providing method of claim 21, wherein the identified item is a book and the received image comprises a book cover of the book.

28. The information-providing method of claim 21, wherein identifying an item based on the received image comprises:
identifying at least one keypoint of the received image,
generating a descriptor for each of the at least one keypoint,
quantifying a descriptor to generate at least one feature ID,
selecting at least one item corresponding to each of the at least one feature ID,
determining a total number of hits for each of the at least one selected item, and
selecting one of the at least one selected item that best matches with the received image.

29-31. (canceled)

32. An information-requesting device comprising:
a camera that captures an image of an item,
a keypoint-identifying module that identifies at least one keypoint from the image,
a descriptor-generating module that generates a descriptor for each of the at least one identified keypoint,
a feature ID-generating module that generates a feature ID to each descriptor,
a data-receiving module that receives a list of corresponding feature IDs and items from an external source,
an item lookup module that identifies an item from the received list based on the at least one generated feature ID, and
an information retrieving module that retrieves information relating to the identified item.

33-35. (canceled)