(12) STANDARD PATENT (11) Application No. **AU 2010248862 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Landmarks from digital photo collections

(51) International Patent Classification(s)
G06F 17/30 (2006.01)

(21) Application No: **2010248862** (22) Date of Filing: **2010.05.14**

(87) WIPO No: **WO10/132789**

(30) Priority Data

(31) Number	(32) Date	(33) Country
12/466,880	**2009.05.15**	**US**

(43) Publication Date: **2010.11.18**
(44) Accepted Journal Date: **2016.06.09**

(71) Applicant(s)
Google Inc.

(72) Inventor(s)
Adam, Hartwig;Zhang, Li

(74) Agent / Attorney
Spruson & Ferguson, L 35 St Martins Tower 31 Market St, Sydney, NSW, 2000

(56) Related Art
EP 1921853

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 November 2010 (18.11.2010)

PCT

(10) International Publication Number
WO 2010/132789 A1

(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/US2010/034930

(22) International Filing Date:
14 May 2010 (14.05.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/466,880 15 May 2009 (15.05.2009) US

(71) Applicant *(for all designated States except US)*: **GOOGLE INC.** [US/US]; 1600 Amphitheatre Parkway, Mountain View, CA 94043 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ADAM, Hartwig** [DE/US]; 5404 Packard St., #1, Los Angeles, CA 90019 (US). **ZHANG, Li** [CN/US]; 724 Arastradero Rd., Apt. 315, Palo Alto, CA 94306 (US).

(74) Agents: **MESSINGER, Michael, V.** et al.; Sterne, Kessler, Goldstein & Fox, 1100 New York Avenue, N.W., Washington, DC 20005 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

[Continued on next page]

(54) **Title**: LANDMARKS FROM DIGITAL PHOTO COLLECTIONS

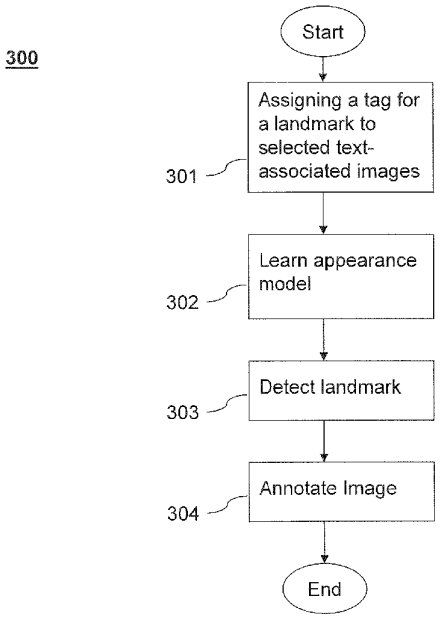

FIG. 3

(57) **Abstract**: Methods and systems for automatic detection of landmarks in digital images and annotation of those images are disclosed. A method for detecting and annotating landmarks in digital images includes the steps of automatically assigning a tag descriptive of a landmark to one or more images in a plurality of text-associated digital images to generate a set of landmark- tagged images,learning an appearance model for the landmark from the set of landmark-tagged images, and detecting the landmark in a new digital image using the appearance model. The method can also include a step of annotating the new image with the tag descriptive of the landmark.

WO 2010/132789 A1

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

LANDMARKS FROM DIGITAL PHOTO COLLECTIONS

BACKGROUND

Technical Field

[0001] The present invention relates generally to digital image collections, and more particularly to identifying popular landmarks in large digital image collections.

Background Art

[0002] With the increased use of digital images, increased digital storage capacity, and interconnectivity offered by digital media such as the Internet, ever larger corpora of digital images are accessible to an increasing number of people. Persons having a range of interests, from various locations spread throughout the world, take photographs of various subjects and make those photographs available for others to view, for instance, on the Internet. For example, digital photographs of various landmarks and tourist sites from across the world may be posted on the web by persons with different levels of skill in taking photographs. Such photographs may show the same landmark from different perspectives, under different conditions, and/or from different distances.

[0003] The vast number of such images available can be useful as an indicator of, or guide to, popular landmarks. To leverage information contained in these large corpora of digital images, it is necessary that the corpora be organized. For example, at digital image web sites such as Picasa Web Albums (from Google Inc., Mountain View, California), starting at a high level menu, one may drill down to a detailed listing of subjects for which photographs are available. Alternatively, one may be able to search one or more sites that have digital photographs. Some tourist information websites, for example, have downloaded images of landmarks associated with published lists of popular tourist sites.

[0004] Most conventional digital photograph organizing systems rely on users to tag photographs. As numerous new photographs are added to these digital image collections, it may not be feasible for users to manually label the photographs in a complete and consistent manner that will increase the usefulness of those digital image collections. A system that can automatically extract information (such as the most popular tourist destinations) from these large collections is described in United States Patent Application

2010248862 10 May 2016

12/119,359 titled "Automatic Discovery of Popular Landmarks," also assigned to Google Inc., California. The system described in Application 12/119,359 uses a processing pipeline comprising a clustering stage based on geo-coding, and a clustering stage based on matching visual features of the images. What is needed, however, are other approaches to automatically discover landmarks and annotate images containing landmarks.

SUMMARY

[0004a] According to a first aspect, the present disclosure provides a method for detecting and annotating landmarks in digital images: (a) automatically assigning, to one or more images in a plurality of text-associated digital images, a tag descriptive of a landmark, to generate a set of landmark- tagged images, wherein images in the set of landmark-tagged Images are algorithmically determined to include the landmark by analyzing text associated with the images of the plurality of images to generate a list of landmark n-grams, wherein the tag descriptive of the landmark is based on at least one landmark n-gram in the list of landmark n-grams; (b) learning an appearance model for the landmark from the set of landmark-tagged images; and (c) detecting the landmark in a new image using the appearance model, wherein said stages (a)-(c) are performed by at least one processor.

[0004b] According to a second aspect, the present disclosure provides a system for automatically detecting and annotating landmarks m digital Images, comprising: at least one collection of text-associated digital images stored in a memory medium; and at least one processor communicatively coupled to said medium, the at least one processor configured to: analyze texts associated with at least one image of the at least one collection to generate a list of landmark n-grams; automatically assign, to one or more images of the at least one collection, a tag descriptive of a landmark, to generate a set of landmark-tagged images, wherein images in the set of landmark-tagged images are algorithmically determined to include the landmark, wherein the tag descriptive of the landmark is based on at least one landmark n-gram in the list of landmarks n-grams; learn an appearance model for the landmark from the set of landmark-tagged images; and detect the landmark in a new image using the appearance model.

[0004c] According to a third aspect, the present disclosure provides a computer program product comprising a computer readable medium having computer program logic recorded thereon for enabling a processor to name images, said computer program logic comprising: a first module configured to enable the processor to assign, to one or more images in a plurality of

2010248862 10 May 2016

text-associated digital images, a tag descriptive of a landmark and based on at least one landmark n-gram selected based on an n-gram score from a list of landmark n-grams, to generate a set of landmark-tagged images, wherein images in the set of landmark-tagged images are algorithmically determined to include the landmark by analysis of text associated with at least one image of the plurality of digital images to generate the list of landmark n-grams; a second module configured to enable the processor to learn an appearance model for the landmark from the set of landmark-tagged images; and a third module configured to enable the processor to detect the landmark in a new image using the appearance model.

[0005] Methods and systems for automatic detection of landmarks in digital images, and annotation of those images, are disclosed. In one embodiment, a method for detecting and annotating landmarks in digital images includes the steps of automatically assigning a tag, descriptive of a landmark, to one or more images in a plurality of text-associated digital images. This generates a set of landmark-tagged images. An appearance model can be learned for the landmark from the set of landmark-tagged images. This allows detection of the landmark in a new digital image using the appearance model. The method can also include a step of annotating the new image with the tag descriptive of the landmark.

[0006] Another embodiment is a system for automatically detecting and annotating landmarks in digital images. The system has at least one collection of text-associated digital images stored in a memory medium and at least one processor communicatively coupled to the medium. The processors are configured to automatically assign a tag descriptive of a landmark to one or more images in a plurality of text-associated digital images. This generates a set of landmark-tagged images. An appearance model can be learned for the landmark from the set of landmark-tagged images. This allows detection of the landmark in a new digital image using the appearance model.

[0007] Further features and advantages of the present invention, as well as the structure and operation of various embodiments thereof, are described in detail below with reference to the accompanying drawings. It is noted that the invention is not limited to the specific embodiments described herein. Such embodiments are presented herein for illustrative purposes only. Additional embodiments will be apparent to persons skilled in the relevant art(s) based on the teachings contained herein.

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0008] Reference will be made to the embodiments of the invention, examples of which may be illustrated in the accompanying figures. These figures are intended to be illustrative, not limiting. Although the invention is generally described in the context of these embodiments, it should be understood that it is not intended to limit the scope of the invention to these particular embodiments.

[0009] FIG. 1 shows a system for the automatic detection of landmarks in digital images, according to an embodiment of the present invention.

[0010] FIG. 2 shows more details of a component of the system of FIG. 1, according to an embodiment of the present invention.

[0011] FIG. 3 is a process to automatically detect landmarks in digital images and annotate digital images according to an embodiment of the present invention.

[0012] FIG. 4 is a process for assigning a tag for a landmark to selected text-associated images, according to an embodiment of the present invention.

[0013] FIG. 5 is a process to generate a list of n-grams based on text-associated images, according to an embodiment of the present invention.

[0014] FIG. 6 is a process selecting a set of n-grams from the list of n-grams generated according to the process of FIG. 4, according to an embodiment of the present invention.

DETAILED DESCRIPTION

[0015] While the present invention is described herein with reference to illustrative embodiments for particular applications, it should be understood that the invention is not limited thereto. Those skilled in the art with access to the teachings herein will recognize additional modifications, applications, and embodiments within the scope thereof and additional fields in which the invention would be of significant utility.

Overview

[0016] The present invention includes methods and systems for automatically identifying and classifying objects in digital images. For example, embodiments of the present invention may identify, classify and prioritize most popular tourist landmarks based on digital image collections that are accessible on the Internet. The method and systems of the present invention can enable the efficient maintenance of an up-to-date list and

collections of images for the most popular tourist locations. In some embodiments, the popularity of a tourist location can be approximated based on the number of images of that location posted on the Internet by users.

[0017] Numerous individuals take digital photographs of surroundings in their neighborhoods, locations visited in their day-to-day activities, and sites visited on their touristic travels. The cameras that are used are of various levels of quality and sophistication. The individuals who capture the images are of various skill levels. The images are captured from various angles, in varied levels of lighting, with varied levels of surrounding visual noise, in various weather conditions, etc. Many of these images are then posted on photo sharing websites or made digitally available through other means. Access to a vast collection of digital images, such as digital photographs, is made available through networks such as the Internet.

[0018] Often, users who post images online also annotate the posted images, for example, by adding one or more tags and/or captions. A tag can be used to name an image. Tags can also be assigned to images to assign keywords that relate to an image. For example, an image of the Eiffel Tower may have assigned to it the tags "Eiffel Tower," "Paris," "France," "Europe," "summer," or the name of a person who is shown to be posing in front of the Tower. Tags are valuable as organization tools at various levels of granularity: "France" may be useful in order to classify the image under searches for landmarks in France, while having only "Eiffel Tower" as a tag could exclude the image from searches for landmarks in "Paris" and/or "France." Despite the variation in accuracy and usefulness of the tags of images in determining landmarks contained in those images, the corpora of user-tagged images are a source of valuable information for purposes of building automatic landmark recognition systems.

[0019] Other potential sources of information include various other documents and electronic sources that link text and images. For example, magazine articles about the Eiffel Tower may include a photograph of its subject. Newspaper content, magazine and journal content, articles written and/or posted by individuals including blog postings about various landmarks, etc., often include images that are directly tied to the textual description. Images having recognizable landmark associated text can be referred to as landmark-tagged images.

[0020] Embodiments of the present invention leverage several types of data available regarding images in order to obtain information about popular landmarks. For example, geo-tags, text tags, author information, timestamps (e.g., time or origin), and visual match information, are some of the types of information that are utilized in embodiments of the present invention. Some of this information is available with each image (e.g., in EXIF tags associated with the image). Other information is either user assigned, or algorithmically assigned. When taken individually each of these data types can have substantial weaknesses. For example, geo-location data (e.g., geo-tags) are generally based on the location of the camera and not the landmark that is being photographed. Also, in some cases the geo-location information is based on user provided information such as city name and may therefore not be accurate. Text tags, provided by authors and third parties, may not accurately describe the landmark. Author information for each image can be based on a camera identifier, a person who captures the image, or a person who uploads the image to a web site. Visual match information can also be erroneous in situations such as when several landmarks exist in a small area, when landmarks look alike, and/or when image quality is not sufficient. Embodiments of the present invention, therefore, leverage several types of available information to obtain a high degree of landmark detection and accurate annotation in digital images.

System for Automatic Landmark Recognition and Annotation

[0021] A system 100 for building a database of annotated popular landmark images according to an embodiment of the present invention is shown in FIG. 1. System 100 includes a computer 101, a user interface 102, networks 103 and 104, a text/image document collection 107, an n-gram collection 108, n-gram filters database 109, database of un-annotated images 110, database of appearance models 111, annotated images 112, and text/image sources 105. A person of skill in the art will appreciate that system 100 can include more, less, or different components and modules than those listed above while still being consistent with the present invention.

[0022] Computer 101 may include one or more computers, servers, or like computing devices, that are interconnected by a communication medium. For example, computer 101 can comprise one or more commercially available computing servers that are coupled by one or more local area network, such as an Ethernet network, Gigabit Ethernet network, WIFI network, or the like. A computer 101 includes a processor 121, volatile

memory 122, persistent memory 123, network interface 124, database interface 125, a communication medium 126 to couple modules of computer 101, and an unsupervised image annotator module 127. Processor 121 can include one or more commercially available central processing units (CPU), graphics processor units (GPU), field programmable gate array (FPGA), digital signal processors (DSP), and application specific integrated circuits (ASIC). Processor 121 controls processing within computer 101, receiving inputs and outputting data into or from computer 101. For example, the processing logic of unsupervised image annotator module 127 can be executed on processor 121.

[0023] Volatile memory 122 can include a volatile memory such as dynamic random access memory (DRAM), static random access memory (SRAM), or the like. Volatile memory 122 can be used to store configuration parameters, source data and intermediate results of the processing of module 127. Configuration parameters can include connection information for text/image sources 105, and other parameters that configure the operation of, for example, processing of unsupervised image annotator module 127. Persistent memory 123 can include one or more non-volatile memory devices such as magnetic disk, optical disk, flash memory, read only memory (ROM), or the like. Persistent memory 123 can be used for storing the logic instructions for unsupervised image annotator module 127, configuration parameters, and to store intermediate and other results of the processing in module 127.

[0024] Network interface 124 can include the functionality to communicate with entities connected to computer 101 through networks including network 103, such as text/image sources 105. For example, network interface 124 can include processing components including Internet Protocol (IP) and Hyper-Text Transfer Protocol (HTTP) processing such that enables computer 101 to connect to text/image sources 105 to obtain text and image information. For example, HTTP protocol processing machine software can be implemented as part of network interface 124. Database interface 125 includes the functionality to connect computer 101 to one or more databases used in processing images for landmarks according to embodiments of the present invention. It should be noted that the use of the term "database" does not necessarily refer to a database management system (DBMS), but rather encompasses any collection of data. Database interface 125 therefore can include DBMS functionality to connect to one or more DBMS

systems comprising one or more databases 107-112 or processing logic to communicate with the type of database of each type of database 107-112. Communication medium 126 can connect modules of computer 101, including modules 121-125 and 127. Communication medium 126 can include a communication devices such as a PCI bus, USB, Ethernet, or the like.

[0025] Unsupervised image annotator module 127 includes the functionality to identify landmarks, generate appearance models for selected landmarks, and to annotate images according to an embodiment of the present invention. Landmarks contained in images can be identified based on explicit tags already associated with the image, or through algorithmic means as described below. The functionality of unsupervised image annotator module 127 can be implemented in software, firmware, hardware, or any combination thereof. In one embodiment, processing logic for the functionality of unsupervised image annotator module 127 can be implemented in a computer programming language or script language such as C, C++, Assembly, Java, JavaScript, Perl, or the like.

[0026] Networks 103 can include a means of connecting computer 101 to one or more text/image sources 105. Network 104 can include a means of connecting computer 101 to one or more databases 107-112. Networks 103 and 104 can include one or more network mediums including peripheral connections such as USB, FireWire, or local area networks such as Ethernet, WIFI, or wide area networks such as a PSTN or Internet. In one embodiment, network 103 includes the Internet, and network 104 includes an Ethernet based local area network.

[0027] User interface 102 can be connected to one or more computers 101 using any one or a combination of interconnection mechanisms such as PCI bus, IEEE 1394 Firewire interface, Ethernet interface, an IEEE 802.11 interface, or the like. User interface 102 allows a user or other external entity to interact with computer 101. In some embodiments, one or more databases 107-112 can also be interacted with through user interface 102. One or more of a graphical user interface, a web interface, and application programming interface may be included in user interface 130.

[0028] Text/image sources 105 can include various types of digital document collections that include images of landmarks and associated text (e.g., landmark-tagged images). In one embodiment, text/image sources 105 include one or more photo collections that have

photos associated with captions and tags. Captions, as used herein, refer to a title assigned to a photo. Tags, as used herein, refer to one or more words or phrases assigned to a photo. Often, captions as well as the tags are assigned by the author (e.g., originator of the photograph, or person uploading the photograph to a photosharing website) of the photograph. However, captions and tags can also be assigned to a photograph by a third party, or an automated tool. Unless each is identified separately, the term "tag" in the following description includes tags as well as captions.

[0029] Text/image sources 105 can also include collections of hypertext documents that hyperlink images to documents (and vice versa), and can also include newspaper corpora, magazine and journal corpora, blog archives, digital libraries having digitized books, third-party annotated photo depositories, and personal and business web sites. For example, tourism and/or travel-related websites, digital travel guides, city websites, etc. are some resources that generally includes images of landmarks and descriptions of those landmarks. However, any collection of digital data where a correlation between one or more images and associated text can be drawn can be included in text/image sources 105.

[0030] Text/image collection 107 is a database where, in some embodiments, local copies and/or modified versions of text/image data originally accessed in remote text/image sources 105 are saved, for example, for more convenient and reliable access for processing by unsupervised image annotator 127. For example, because accessing data and images in text/image sources 105 over network 103, which can be a wide area network such as the Internet, may involve long latencies, there may be a process (not shown) in computer 101 that makes copies of such data and images in a local or locally attached network location such as in text/image collection 107. Text/image collection 107 can also include collections of images that are already tagged, for example, user photo collections in Picasa Web Albums and/or image collections already processed according to teachings in the present invention. In some embodiments, text/image collection 107 can include a data structure corresponding to each image, in which the data structure includes one or more pointers to images and/or documents in text/image sources 105, for example, to avoid having to create a separate copy of the image and/or documents from text/image sources 105.

[0031] N-gram collection 108 is a database that includes a collection of n-grams. N-grams can be extracted from captions, tags, or text documents associated with images in,

for example, text/image collection 107 or text/image sources 105. As used herein an n-gram is a sequence of one or more words. The selection of n-grams can be done using methods similar to one or more of several techniques used, for example, in text analysis. The selection and extraction of n-grams, according to embodiments of this invention, is further described below.

[0032] N-gram filters database 109 includes one or more lists of n-grams to be filtered out of n-gram collection 108, and/or one or more filtering rules to be applied to n-gram collection 108. For example, one list in n-gram filters database 109 can be a "bad words list" where the n-grams appearing in the bad words list are not extracted from text/image collections 107, or text/image sources 105, and are removed from n-gram collection 108 if they are found to be present. Another list may be a list of n-grams that occur too frequently in image associated text, and therefore are of little value as landmark identifiers. Words such as "the" and "of" can be considered in this category. Another list can be a list of phrases, that are known to appear too frequently and are therefore not sufficiently useful as discriminatory landmark identifiers.

[0033] Unannotated images database 110 includes images that are yet to be annotated (e.g., tagged) according to embodiments of the present invention. For example, unannotated images database 110 can include the untagged digital images uploaded by one or more users in order to be processed using an embodiment of the present invention.

[0034] Appearance models database 111 holds the recognition models, herein referred to as appearance models, that are derived in order to recognize landmarks in images, for example, images in unannotated images database 110.

[0035] Annotated images database 112 contains the images that are annotated according to embodiments of the present invention. For example, images from unannotated images database 110 are stored in annotated images database 112, after they are processed by unsupervised image annotator 127 according to an embodiment of the present invention. A person of skill in the art will recognize that although databases 107-112 are described as separate databases above, databases 107-112 can be arranged and/or implemented in various ways consistent with the present invention.

[0036] FIG. 2 shows more details of unsupervised image annotator module 127, according to an embodiment of the present invention. In this embodiment, unsupervised image annotator module 127 includes three processing modules: a landmark identifier

201, an appearance model generator 202, and an image annotator 203. Modules 201, 202, and 203 can be implemented in software, firmware, hardware, or a combination thereof. In one embodiment, modules 201-203 are implemented in software using the C++ programming language. In one embodiment, a computer program product may have logic including the computer program logic of modules 201-203 recorded on a computer readable medium such as a hard disk, flash disk, or other form of storage.

[0037] Landmark identifier module 201 includes the functionality to identify landmarks in text/image collections 107 and/or text/image sources 105. Landmark identifier module 201 can, in one embodiment, use images and associated text from text/image sources 105 as input, and copy such images and associated text to text/image collection 107. Landmark identifier module 201 can also analyze the text in text/image sources 105 while using and updating n-grams collection 108. N-gram filters database 109 can also be used in the processing within landmark identifier module 201.

[0038] Appearance model generator 202 includes the functionality to generate one or more appearance models for each landmark that is, for example, identified by landmark identifier module 201. In one example, appearance model generator 202 can take as input the images and identified landmarks in text/image collection 107, and generate one or more appearance models for each of the landmarks. The generated appearance models can be written to appearance models database 111.

[0039] An appearance model, as used herein, is a template to be used in the automatic recognition of certain common features in images. In one embodiment of the present invention, an appearance model used for the recognition of a landmark can include a feature vector comprising numerical scores for a set of predetermined image features. Methods of object recognition in images and of generating feature vectors are well known in the art. For example, methods of object recognition in images are described in David G. Lowe, "Object recognition from local scale-invariant features," *International Conference on Computer Vision*, Corfu, Greece (September 1999), pp. 1150-1157. In addition to the visual recognition components, an appearance model can also include information such as geo-location information for the corresponding landmark. For example, the geo-location information in the appearance model for a particular landmark can specify a geographic point and/or a geographic area. Specifying a geographic area

can reduce uncertainties created due to the variance in accuracy of the geo-location information of images.

[0040] Image annotator module 203 includes the functionality to automatically recognize landmarks in images and appropriately annotate such images with information identifying the one or more corresponding landmarks. In one embodiment, image annotator module 203 can use appearance models from appearance models database 111 to automatically recognize landmarks in images from unannotated images database 110. The images can then be annotated, for example by associating one or more tags, according to the recognized landmarks in each image and the annotated images can be written to annotated images database 112.

Method for Automatic Landmark Recognition and Annotation

[0041] FIG. 3 shows a process 300 that annotates an image that includes one or more popular landmarks, according to an embodiment of the present invention. Process 300 can be implemented, for example, in unsupervised image annotator module 127. Steps 301-304 of process 300 can be implemented in landmark identifier module 201, appearance model generator module 202, and image annotator module 203, as appropriate. A person of skill in the art will understand that the functionality described herein with respect to process 300 can be implemented using modules 201-203 in ways other than that described below. For example, in one embodiment, landmark identifier module 201, appearance model generator module 202, and image annotator module 203, can each be separate processes that together implement process 300. In another embodiment, landmark identifier module 201, appearance model generator module 202, and image annotator module 203, can each be a separate thread that together implement process 300. In yet another embodiment, landmark identifier module 201, appearance model generator module 202, and image annotator module 203, can all be implemented as a single process implementing process 300.

[0042] In step 301, images and text associated with those images are analyzed to identify landmarks, particularly popular landmarks. Popular landmarks, in general, are those landmarks that appear most frequently in the analyzed image/text sources, such as text/image sources 105. The input to the processing in step 301, in one embodiment, is one or more image/text sources accessible to the one or more computers on which process 300 is being executed. For example, process 300 can be executing on computer 101 and

can have accessibility to text/image sources 105 over a network 103. The output from step 301, according to one embodiment, can be a selected set of images, identified landmarks in those images, and associated text and n-grams. For example, the output of step 301 can be written into text/image collection 107. Step 301 is further described with respect to FIGs. 4-6 below.

[0043] In step 302, one or more appearance models are derived or learned for landmarks identified in step 301. A person skilled in the art will recognize that one of many methods may be used to learn an appearance model from the landmark-tagged images obtained as a result of step 301. According to one embodiment, the appearance model for a particular landmark comprises a feature vector that numerically quantifies one or more visual aspects of one or more images considered to contain the particular landmark. As described earlier feature vector generation is well known in the art and an approach for feature vector generation, such as that can be used in the present invention, is described in David G. Lowe, "Object recognition from local scale-invariant features," cited above. The feature vector, for example, ideally includes a substantial number of features that are relatively invariant to the numerous varying conditions such as camera distance, camera angle, image quality, lighting conditions, etc. In some embodiments of the present invention, the one or more appearance models corresponding to a particular image can also include non-visual aspects of an image such as geo-location information. An appearance model can include any information, including visual characteristics of the particular landmark and geo-location information, that can be used in automatically recognizing the existence of that landmark in images.

[0044] In step 303, the one or more appearance models obtained in step 302 are used to detect a corresponding landmark in images. In one embodiment, one or more appearance models in appearance model database 111 are used in the detection of a corresponding landmark in unannotated images 110 database. For example, feature vectors of an appearance model from appearance models database 111 can be compared to feature vectors generated for the image from unannotated images database 110 that is being considered. If the feature vectors match beyond a predetermined threshold level, the image being considered is recognized to include the landmark that corresponds to the matched appearance model. Object recognition technology, such as that can be used in step 303 in an embodiment of the present invention, is generally well known. One

approach to object recognition that can be used in the present invention is described in Lowe, "Object recognition from local scale-invariant features," cited above.

[0045] In step 304, the image being analyzed can be annotated if it is determined to have within it, a particular landmark corresponding to the one or more appearance models that were used in the detection, for example, in step 303. Annotated images and the respective annotations can be written to annotated images database 112. The annotation associated with an annotated image can include text associated with each one of the appearance models that were found to have a match in that annotated image. It is also contemplated that the annotations associated with the annotated image can include text or phrases based on additional processing of the text associated with the corresponding appearance models. For example, in an embodiment in which the text associated with the corresponding appearance models are of the form of simple tags such as "Statue of David," and "Rome," step 304 may include additional processing to generate a sentence such as "Statute of David in Rome, Italy," "Statue of David in Palacio Veccio, Rome, Italy," or the like.

[0046] In FIG. 4, processing involved in step 301 is shown in further detail. The functionality of step 301 includes steps 401-403. In step 401, an n-gram set of words or phrases descriptive of landmarks is generated and/or an existing n-gram set is updated. For example, step 401 can take as input text/image sources 105 and produce as output n-grams in n-gram collection 108. A more detailed description of step 401, as in how one or more n-grams descriptive of landmarks are generated is provided below in relation to FIG. 5.

[0047] In step 402, a set of n-grams that are preliminarily considered as being useful for landmark determination is scored. For example, the initial set of n-grams considered in step 402, can be the set of n-grams derived from text/image sources 105 in step 401. The processing of step 402 can create a list of n-grams in n-gram collection 108. The n-grams are filtered according to various criteria including having each n-gram scored and keeping only a predetermined number of n-grams with the highest scores. An n-gram score $S(k)$ is assigned to each of the n-grams $N(k)$ in n-gram collection 108. A method of determining $S(k)$ is described below. Processing of step 402, is further described with respect to FIG. 6 below.

[0048] In step 403, images are assigned tags from n-gram collection 108. For example, for each pair of image and n-gram combination, a pairing-score can be assigned. The

pairing-score can be defined such that the higher valued pairing-scores imply strongly related image and n-gram pairs. In one example, the pairing formed by image *I(i)* from image/text collection 107 and the n-gram *N(k)* from n-gram collection 108, can be assigned a pairing-score defined by the product of the strength *L(i,k)* of the link between *I(i)* and *N(k)* and the n-gram score of *N(k)*, i.e., *L(i,k)*S(k)*. A method of determining *L(i,k)* is described below. A list of candidate n-grams can be generated by focusing on the n-grams with high pairing-scores, and truncating the list appropriately. In one instance, the list can be truncated when the pairing-score falls lower than half of the highest pairing-score in the list. In this manner, each image can be assigned the most relevant n-grams.

[0049] FIG. 5 shows processing steps 501-504 in the generation of the set of n-grams according to step 401 described above. In step 501, one or more text/image sources 105 is accessed, for example, by landmark identifier module 201. Accessing of text/image sources 105 can include connecting to such sources either over a local network, or a wide area network such as the Internet. The text/image sources 105 that are selected to be processed can be identified based on various methods such as input from users or operators, automatic identification and classification of web sited by program components (e.g., identification of photo repository web sites by web bots), or a list of websites or other repositories that are monitored for content. Methods of connecting to sources such as text/image sources 105 are well known. Where necessary, an implementation of the present invention should also consider aspects of copyrights, privacy, etc., that may be involved in the use of images owned by various parties.

[0050] In step 502, a list of potential landmark descriptor n-grams are retrieved from text associated with images in text/image sources 105. The extraction of n-grams from photo repositories where photos are associated with tags and/or captions can include the collection of the set of tags and/or captions associated with photos of the photo repositories of text/image source 105. When image/text sources include other documents and/or content that associates images with corresponding text, one or more of numerous text analysis methods can be used to extract terms (tags) that potentially correspond to landmarks. For example, a text associated with an image in a tourism website can be automatically analyzed using a method well-known in the art such as term-frequency-inverse document frequency (TF-IDF) over the text available to identify potential tags. In

one embodiment, TF-IDF is applied to the tags associated with photos in a photo repository from a text/image source 105.

[0051] Predetermined rules can be applied to determine a narrowed and/or filtered set of tags that refer to landmarks from the potentially large number of available tags. For example, in step 503, one or more filtering rules or criteria can be applied to the set of n-grams of potential landmark descriptors collected in step 502. One filter that can be applied to the list of potential landmark descriptor n-grams is a bad words filter. The bad words filter includes a list of n-grams and phrases that are predetermined as bad and/or unhelpful to discriminate among landmarks. Another filter that is applied can be a stop word list. The stop word list can include n-grams that are expected to occur so frequently in tags and/or descriptors that they are unlikely to be helpful as landmark descriptors. Words such as "of," "the," and "and" are example n-grams that can be included in a stop word list. Another filter that can be applied is a minimum reliability measure, such as a minimum number of authors filter. The minimum number of authors filter can be used to remove any n-grams from the list of potential landmark descriptor n-grams that have less than a predetermined number of unique authors using those n-grams in their tags. For example, it may be predetermined that for any n-gram to be included in n-gram collection 108, the n-gram should be detected in the tags used by three or more unique authors.

[0052] In step 504, the list of potential landmark descriptor n-grams remaining after the one or more rules and/or filters are applied in step 503, can be written in n-gram collection 108. The set of n-grams from n-gram collection 108 used by subsequent processing steps, such as processing step 402, is a set of n-grams that have been filtered according to several filters as described above, and would therefore include only n-grams that are substantially descriptive of landmarks

[0053] FIG. 6 shows steps 601-608 illustrating the processing involved in step 402, according to one embodiment. In step 601, the images associated with the n-grams selected in step 401, are assigned correlation-weights. In one embodiment, the images associated with the n-grams selected in step 401 are copied into text/image collection 107 and the weight assignment and additional processing is performed upon those images. The correlation-weight $W(i)$ of an image $I(i)$ is an inverse measure of the level of correlation of the image $I(i)$ to other images in text/image collection 107. For example, if image $I(i)$ is not correlated with any other images in text/image collection 107, then

image *I(i)* is assigned a correlation-weight of 1; if image *I(i)* is correlated to 2 other images in text/image collection 107, then image *I(i)* and each of its two correlated images is assigned a correlation-weight of 1/3. A predetermined set of rules or criteria can be used to determine if two images are correlated. For example, two images can be considered correlated when they are taken by the same author and at very close geo-location (e.g., within 1/4 miles from each other).

[0054] In step 602, a matching images graph is created from images in, for example, text/image collection 107. Nodes in the matching images graph represents images in text/image collection 107. Each edge in the matching images graph represents the extent to which the images corresponding to the two connected nodes match. For example, the matching score *M(i,j)* assigned to the edge between images *I(i)* and *I(j)*, can be a numeric value that is derived based on the match between the feature vector of image *I(i)* and the feature vector of image *I(j)*. Individual features in the feature vectors may be assigned configurable weights, and the matching score *M(i,j)* can be the summation of such the weights of the matching features.

[0055] In step 603, links (referred to as image-name links) are formed between each of the n-grams in n-gram collection 108 and images in text/image collection 107. The image-name links can be a binary variable set to 1 if the n-gram is contained by the tags of the images and 0 otherwise. However, in order to increase the robustness of the results, the output is smoothed by averaging over a set of images that are visually similar rather than considering single images. For example, image-name link between image *I(i)* and n-gram *k*, *L(i,k)*, can be defined as:

[0056]
$$L(i,k) = \frac{\sum_{\text{for all images } j \text{ with } n-\text{gram } k} M(i,j) * W(j)}{\sum_{\text{for all images } j} M(i,j) * W(j)}$$

[0057] where, as noted above, *M(i,j)* is the matching-score between images *I(i)* and *I(j)* in the image-matching graph, and *W(j)* is the correlation weight of image *I(j)*.

[0058] In step 604, the geo-reliability of each image in text/image collection 107 is estimated. The geo-reliability of image *I(i)*, *G(i)*, is an estimation of the accuracy of the image's geo-location information, based on a comparison of the visual consistency of images with geo-location coordinates within a predetermined distance to each other. For example,

[0059] $$G(i) = \frac{\sum_{\text{for } n \text{ nearest images } j \text{ to image } i} M(i,j) * W(j)}{\sum_{\text{for } n \text{ nearest images } j \text{ to image } i} W(j)}$$

[0060] where, n can be a configurable parameter.

[0061] In step 605, a geo-variance can optionally be computed for each n-gram $N(k)$. For example, geo-variance $V(k)$ of $N(k)$ can be expressed as:

[0062] $$V(k) = EW[(loc(i) - EW(loc(i))^2]$$

[0063] where $loc(i)$ represents the geo-location of image $I(i)$, and EW is the weighted expectation. The weighted expectation is helpful in capturing the variance of the most significant location points for the n-gram. Weights can be computed as $L(i,k)*W(i)*G(i)$, i.e., the product of image-name link, image weight and image's geo-reliability. Subsequently, n-grams with $V(k)$ larger than a threshold geo-variance can be filtered out from the n-gram collection 108.

[0064] In step 606, the n-gram score $S(k)$ of each n-gram $N(k)$ in text/image collection 107 is determined using a measure that is designed to capture the internal link strength between images that have n-gram $N(k)$ in its tags, and the external link strength between images that have n-gram $N(k)$ in its tags and images that do not have n-gram $N(k)$ in its tags. For example, $S(k)$ can be expressed as:

[0065] $$S(k) = \frac{\sum_{\text{for all image pairs}(i,j)} W(i) * L(i,k) * M(i,j) * L(j,k) * W(j)}{\sum_{\text{for all image pairs}(i,j)} W(i) * L(i,k) * M(i,j) * (1 - L(j,k)) * W(j)}$$

[0066] The larger the $S(k)$, the more likely that n-gram $N(k)$ refers to a meaningful, visually distinguishable entity, and therefore more likely to be a landmark name.

[0067] In step 607, after the n-grams are scored, a further filtering can optionally be implemented to identify the most popular landmark n-grams. For example, the n-gram scores of a predetermined number of n-grams having the highest n-gram scores can be averaged to determine a threshold-average score. Thereafter, all n-grams other than those n-grams having a score higher than the threshold-average score can be removed from n-gram collection 108.

[0068] In step 608, n-grams that are considered to refer to the same landmark location are merged. Although the scoring step, and the subsequent filtering based on scores, generally leaves a list of n-grams that meaningfully refer to landmarks, many n-grams

referring to the same landmark can still remain in n-gram collection 108. Multiple n-grams referring to the same landmark can exist because of several reasons including different names for the same landmark, different formulations of the same name, and substring truncation. It would be desirable to merge such duplicate n-grams together in a meaningful manner. To address this, in one example, if two n-grams *N(k)* and *N(l)* have their scores within a predetermined distance from each other, and if the images they are linked to are substantially overlapped, then the two n-grams *N(k)* and *N(l)* are merged. The substantial overlap of images can be determined, for example, by considering the Bhattacharya distance of *L(i,k)* for each image *I(i)* and n-gram *N(k)* pair, and determining whether the Bhattacharya distance is above a predetermined threshold. The computation of Bhattacharya distance is well-known in the art.

Conclusion

[0069] The processing functionality of module 127 and/or modules 201-203, can be achieved in software, hardware, or a combination thereof. For example, modules 201 and 203 may be implemented entirely as software modules, or some of the functionality of the appearance model generator module 202 may be implemented using hardware such as a field programmable gate array (FPGA). It will be understood by a person of skill in the art that unsupervised image annotator module 127 and or computer 101 may include additional components and modules that facilitate the functions of the present invention.

[0070] It is to be appreciated that the Detailed Description section, and not the Summary and Abstract sections, is intended to be used to interpret the claims. The Summary and Abstract sections may set forth one or more but not all exemplary embodiments of the present invention as contemplated by the inventor(s), and thus, are not intended to limit the present invention and the appended claims in any way.

[0071] The present invention has been described above with the aid of functional building blocks illustrating the implementation of specified functions and relationships thereof. The boundaries of these functional building blocks have been arbitrarily defined herein for the convenience of the description. Alternate boundaries can be defined so long as the specified functions and relationships thereof are appropriately performed.

[0072] The foregoing description of the specific embodiments will so fully reveal the general nature of the invention that others can, by applying knowledge within the skill of the art, readily modify and/or adapt for various applications such specific embodiments,

without undue experimentation, without departing from the general concept of the present invention. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed embodiments, based on the teaching and guidance presented herein. It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance.

[0073] The breadth and scope of the present invention should not be limited by any of the above-described exemplary embodiments, but should be defined only in accordance with the following claims and their equivalents.

2010248862 10 May 2016

CLAIMS

1. A method for detecting and annotating landmarks in digital images:

(a) automatically assigning, to one or more images in a plurality of text-associated digital images, a tag descriptive of a landmark, to generate a set of landmark-tagged images, wherein images in the set of landmark-tagged images are algorithmically determined to include the landmark by analyzing text associated with the images of the plurality of digital images to generate a list of landmark n-grams, wherein the tag descriptive of the landmark is based on at least one landmark n-gram in the list of landmark n-grams;

(b) learning an appearance model for the landmark from the set of landmark-tagged images; and

(c) detecting the landmark in a new image using the appearance model, wherein said stages (a)-(c) are performed by at least one processor.

2. The method of claim 1, further comprising:

(d) annotating the new image with the tag descriptive of the landmark.

3. The method of claim 1, wherein stage (a) comprises computing an n-gram score for each landmark n-gram in an n-gram set, wherein the n-gram set is a subset of the list of landmark n-grams.

4. The method of claim 1, wherein stage (a) comprises:

electronically accessing the plurality of text-associated digital images; and

retrieving at least one of said landmark n-grams from a text associated with an image in the plurality of text-associated digital images.

5. The method of claim 4, wherein stage (a) further comprises:

choosing said landmark n-grams having at least a minimum reliability measure.

6. The method of claim 5, wherein the reliability measure is based on a number of unique authors.

7. The method of claim 3, wherein stage (a)(i) comprises:

assigning correlation-weights to the plurality of text-associated digital images, wherein the correlation-weights are based on correlation of metadata of images in the plurality of text-associated digital images;

2010248862 10 May 2016

generating a matching-images graph from the plurality of text-associated digital images; and

linking said landmark n-grams to images in the plurality of text-associated digital images to generate links between landmark n-grams and images in the plurality of text-associated digital images.

8. The method of claim 7, wherein stage (a)(i) further comprises:

estimating a geo-reliability score for each image of the plurality of text-associated digital images using the matching-images graph.

9. The method of claim 7, wherein the n-gram score is based on the matching-images graph.

10. The method of claim 9, wherein the n-gram score is computed as a ratio of strength of internal edges of said matching-images graph and strength of external edges of said matching images graph, wherein an internal edge exists between images having at least one common landmark n-gram, and wherein an external edge exists between images not having at least one common landmark n-gram.

11. The method of claim 8, wherein stage (a)(i) further comprises:

computing a variance of geo-location for a landmark n-gram of said n-gram set, wherein the variance is based on geo-locations of images having said landmark n-gram in their n-gram set in said matching-images graph; and

removing from said n-gram set any landmark n-grams having a variance of geo-location exceeding a predetermined threshold.

12. The method of claim 7, wherein stage (a) further comprises:

merging two or more landmark n-grams in said n-gram set.

13. The method of claim 12, wherein the merging is based at least on one of, a similarity of the score of said two or more landmark n-grams, and an overlap of images having said two or more landmark n-grams in linked landmark n-grams.

14. The method of claim 7, wherein the metadata includes information relating to at least one of,

an author,

a geo-location, and

a time of origin.

2010248862 10 May 2016

15. The method of claim 7, wherein each link in the matching-images graph represents matching feature descriptors between two images of the plurality of text-associated digital images.

16. A system for automatically detecting and annotating landmarks in digital images, comprising:

at least one collection of text-associated digital images stored in a memory medium; and

at least one processor communicatively coupled to said medium, the at least one processor configured to:

analyze text associated with at least one image of the at least one collection to generate a list of landmark n-grams;

automatically assign, to one or more images of the at least one collection, a tag descriptive of a landmark, to generate a set of landmark-tagged images, wherein images in the set of landmark-tagged images are algorithmically determined to include the landmark, wherein the tag descriptive of the landmark is based on at least one landmark n-gram in the list of landmarks n-grams;

learn an appearance model for the landmark from the set of landmark-tagged images; and

detect the landmark in a new image using the appearance model.

17. The system of claim 16, wherein the at least one processor is further configured to:

annotate the new image with the tag descriptive of the landmark.

18. The system of claim 16, wherein the at least one processor is further configured to

compute an n-gram score for each landmark n-gram in an n-gram set, wherein the n-gram set is a subset of the list of landmark n-grams.

19. The system of claim 18, wherein the at least one processor is further configured to:

assign correlation-weights to the images of the at least one collection, wherein the correlation-weights are based on correlation of metadata of images of the at least one collection;

generate a matching-images graph from the images of the at least one collection; and

link said landmark n-grams to images in the images of the at least one collection to generate links between landmark n-grams and images in the images of the at least one collection.

2010248862 10 May 2016

20. A computer program product comprising a computer readable medium having computer program logic recorded thereon for enabling a processor to name images, said computer program logic comprising:

a first module configured to enable the processor to assign, to one or more images in a plurality of text-associated digital images, a tag descriptive of a landmark and based on at least one landmark n-gram selected based on an n-gram score from a list of landmark n-grams, to generate a set of landmark-tagged images, wherein images in the set of landmark-tagged images are algorithmically determined to include the landmark by analysis of text associated with at least one image of the plurality of digital images to generate the list of landmark n-grams;

a second module configured to enable the processor to learn an appearance model for the landmark from the set of landmark-tagged images; and

a third module configured to enable the processor to detect the landmark in a new image using the appearance model.

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

a fourth module configured to enable the processor to annotate the new image with the tag descriptive of the landmark.

22. The computer program product of claim 20, wherein the first module is further configured to compute the n-gram score for each landmark n-gram in an n-gram set, wherein the n-gram set is a subset of the list of landmark n-grams.

23. The computer program product of claim 22, wherein the first module is further configured to:

assign correlation-weights to the plurality of text-associated digital images, wherein the correlation-weights are based on correlation of metadata of images in the plurality of text-associated digital images;

generate a matching-images graph from the plurality of text-associated digital images; and

link said landmark n-grams to images in the plurality of text-associated digital images to generate links between landmark n-grams and images in the plurality of text-associated digital images.

Google Inc.

Patent Attorneys for the Applicant/Nominated Person

SPRUSON & FERGUSON

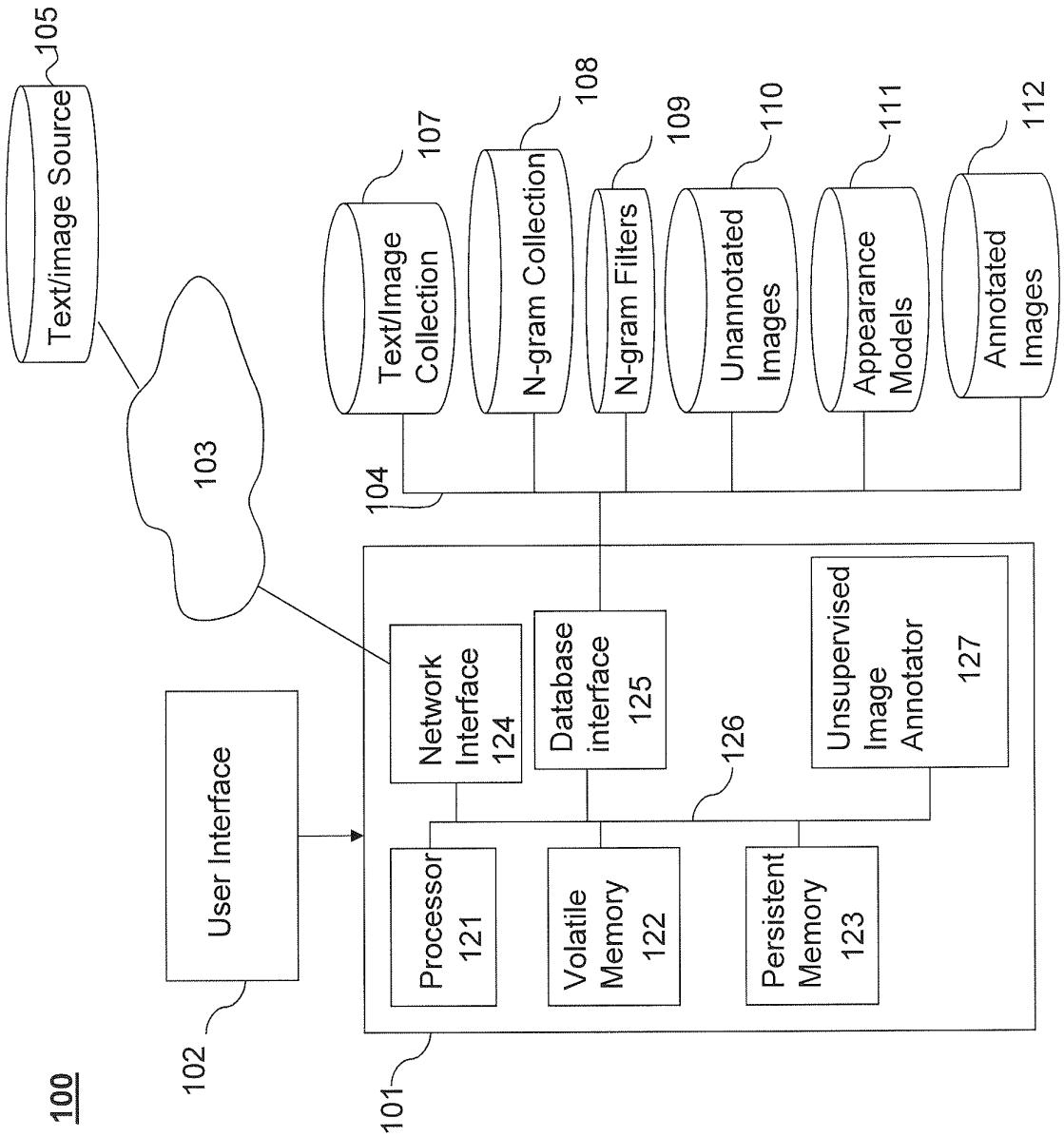
100
101
102
103
104
105
Text/image Source
User Interface
Network Interface 124
Database interface 125
126
Unsupervised Image Annotator 127
Processor 121
Volatile Memory 122
Persistent Memory 123
Text/Image Collection
107
N-gram Collection
108
N-gram Filters
109
Unannotated Images
110
Appearance Models
111
Annotated Images
112

FIG. 1

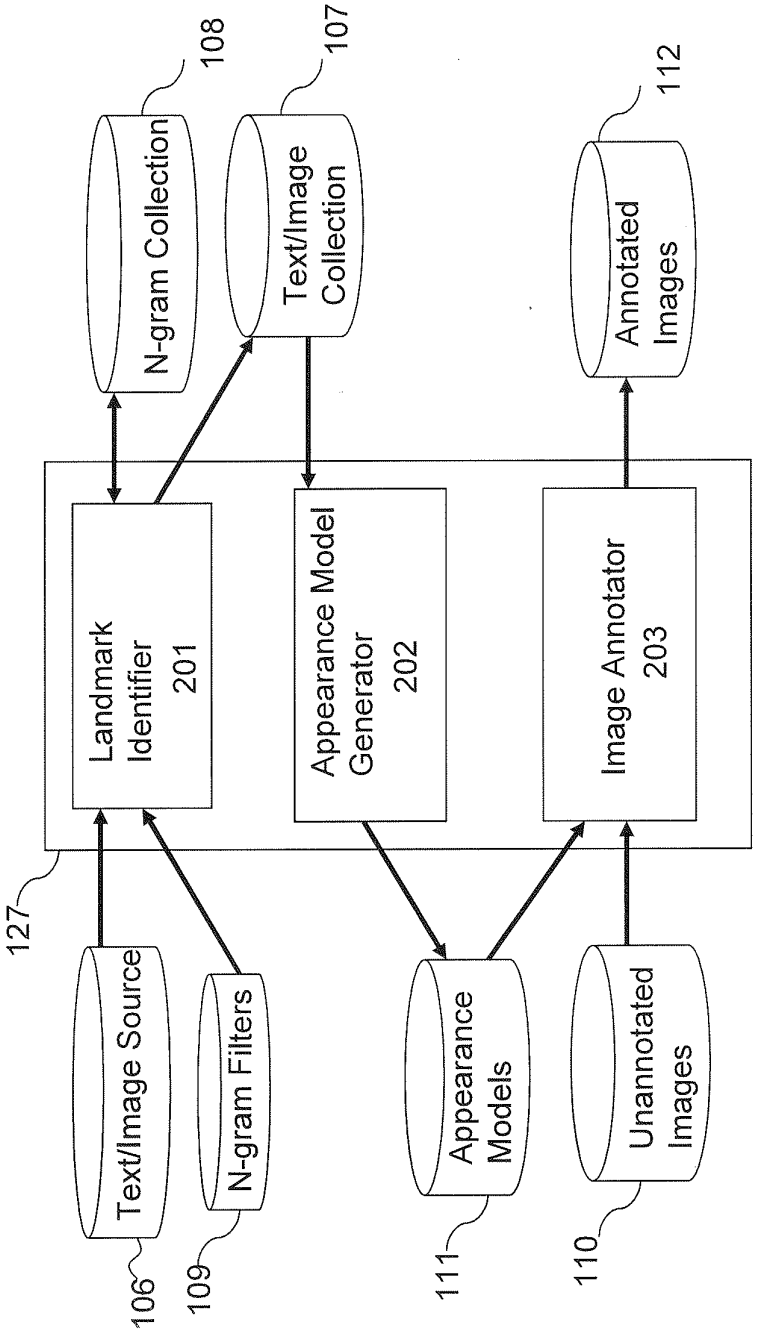
N-gram Collection
108
Text/Image Collection
107
Annotated Images
112
Landmark Identifier 201
Appearance Model Generator 202
Image Annotator 203
127
Text/Image Source
106
N-gram Filters
109
Appearance Models
111
Unannotated Images
110

FIG. 2

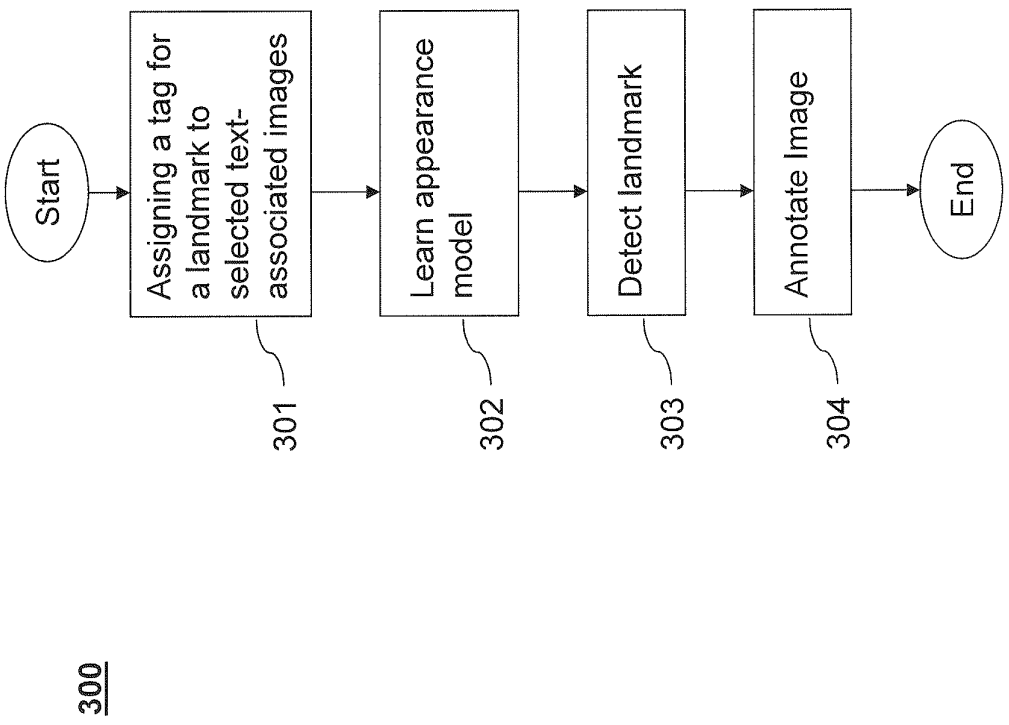
300
Start
Assigning a tag for a landmark to selected text-associated images
Learn appearance model
Detect landmark
Annotate Image
End
301
302
303
304

FIG. 3

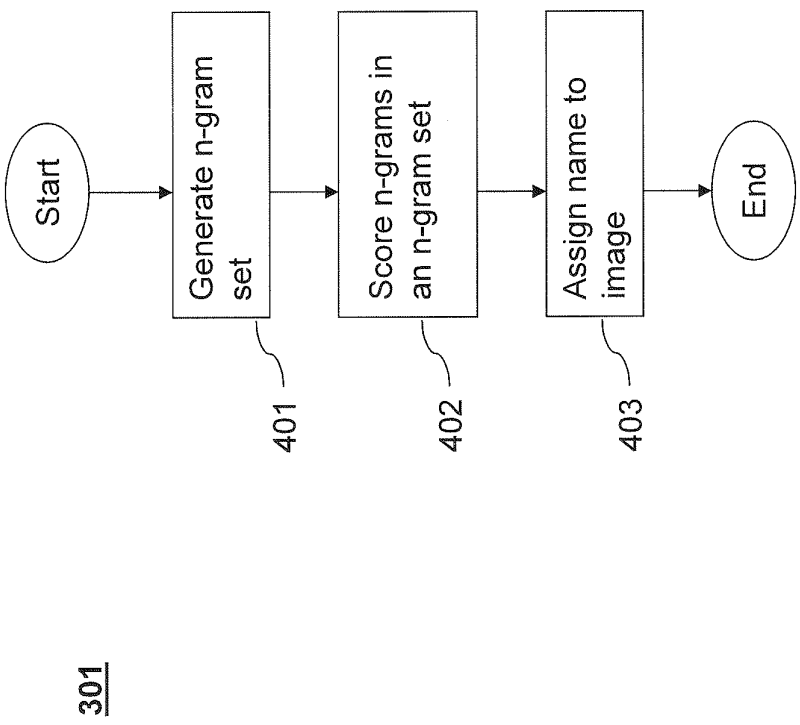
301
Start
Generate n-gram set
Score n-grams in an n-gram set
Assign name to image
End
401
402
403

FIG. 4

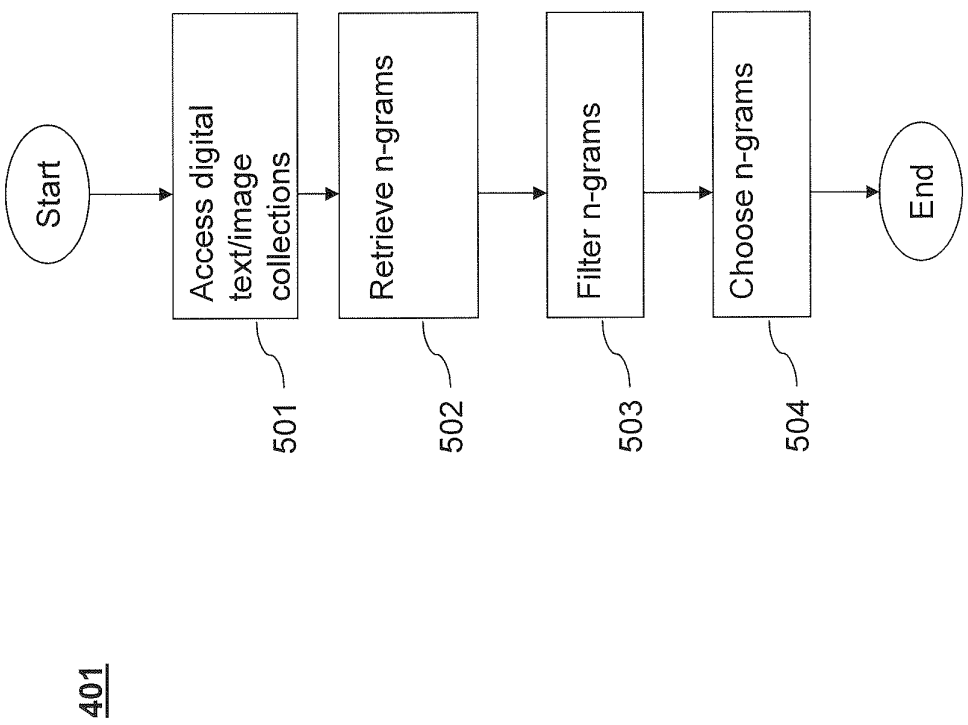
401
Start
Access digital text/image collections
Retrieve n-grams
Filter n-grams
Choose n-grams
End
501
502
503
504

FIG. 5

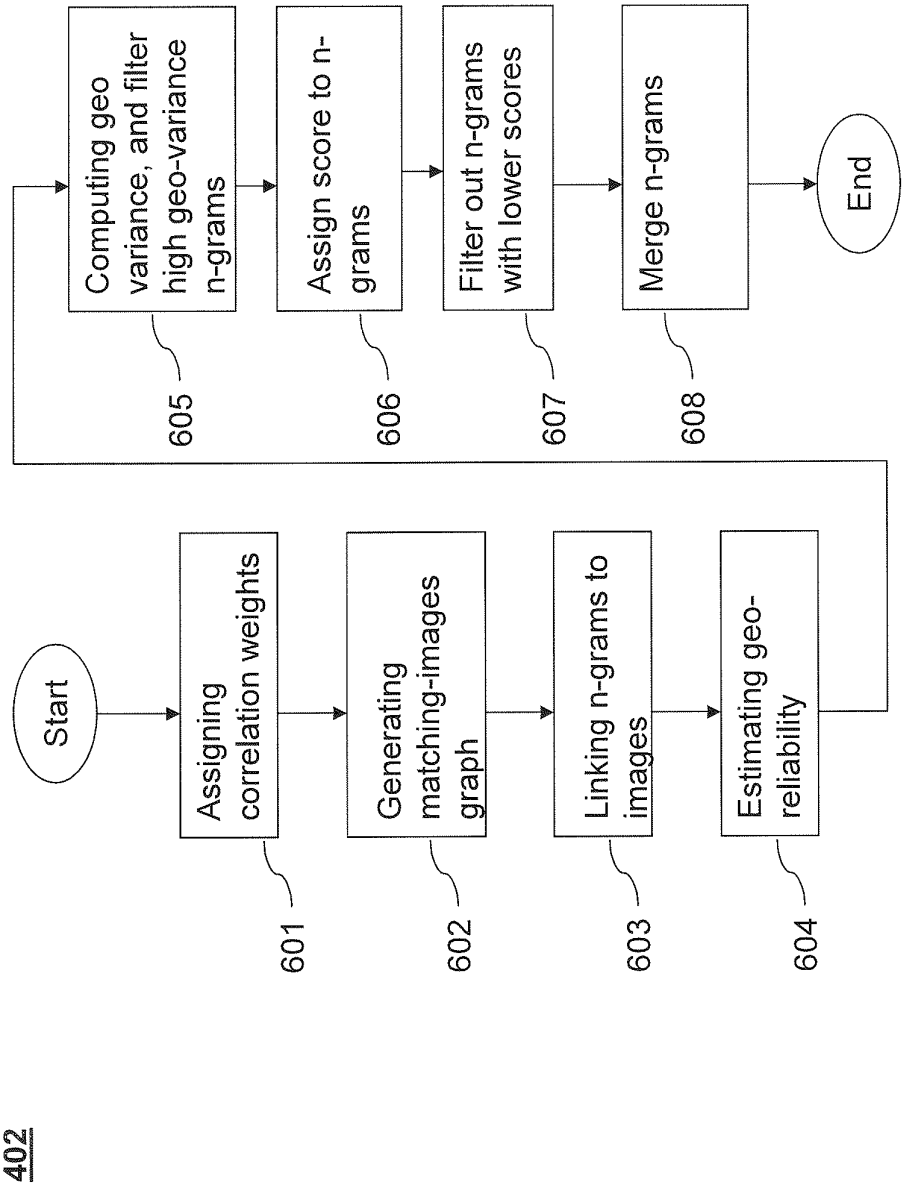
402
Start
Assigning correlation weights
601
Generating matching-images graph
602
Linking n-grams to images
603
Estimating geo-reliability
604
Computing geo variance, and filter high geo-variance n-grams
605
Assign score to n-grams
606
Filter out n-grams with lower scores
607
Merge n-grams
608
End

FIG. 6