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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 June 2009 (18.06.2009)

PCT

(10) International Publication Number
WO 2009/075976 A2

(51) International Patent Classification:
G06F 17/00 (2006.01) *G06F 17/30* (2006.01)

(21) International Application Number:
PCT/US2008/082942

(22) International Filing Date:
10 November 2008 (10.11.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/954,246 12 December 2007 (12.12.2007) US

(71) Applicant *(for all designated States except US)*: **MICROSOFT CORPORATION** [US/US]; Attn: Sharon Rydberg, LCA, International Patents Department, One Microsoft Way, Redmond, WA 98052-6399 (US).

(72) Inventors: **TAN, Desney S.**; c/o Microsoft Corporation, LCA, International Patents Department, One Microsoft Way, Redmond, Washington 98052-6399 (US). **KAPOOR, Ashish**; c/o Microsoft Corporation, LCA, International Patents Department, One Microsoft Way, Redmond, Washington 98052-6399 (US). **WINDER, Simon A.J.**; c/o Microsoft Corporation, LCA, International Patents Department, One Microsoft Way, Redmond, Washington 98052-6399 (US). **FOGARTY, James A.**; c/o Microsoft Corporation, LCA, International Patents Department, One Microsoft Way, Redmond, Washington 98052-6399 (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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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,

[Continued on next page]

(54) Title: INTERACTIVE CONCEPT LEARNING IN IMAGE SEARCH

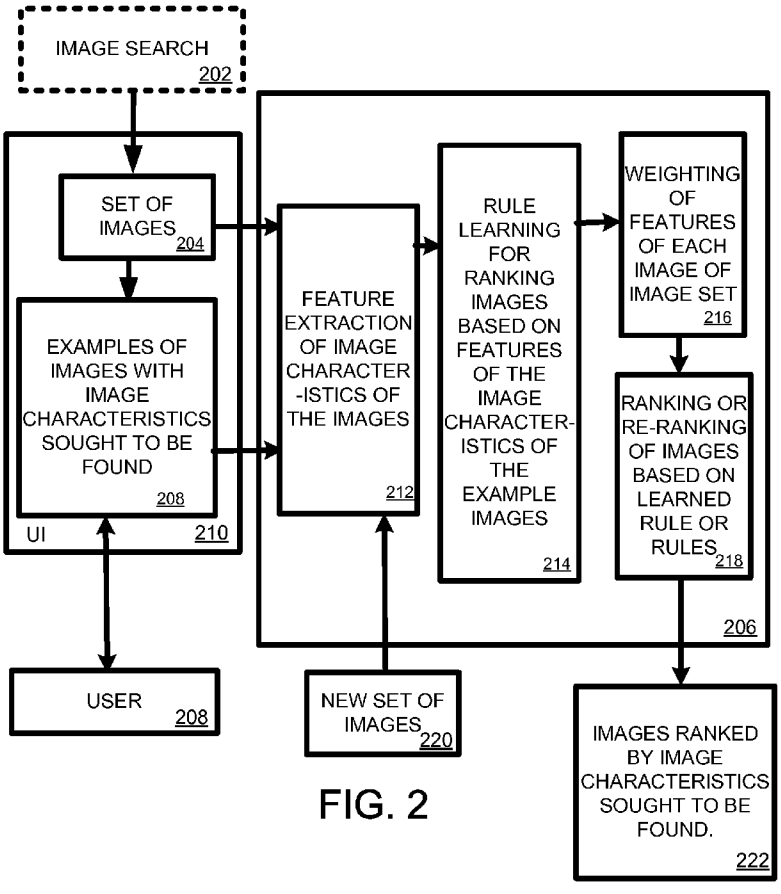


FIG. 2

(57) Abstract: An interactive concept learning image search technique that allows end-users to quickly create their own rules for re-ranking images based on the image characteristics of the images. The image characteristics can include visual characteristics as well as semantic features or characteristics, or may include a combination of both. End-users can then rank or re-rank any current or future image search results according to their rule or rules. End-users provide examples of images each rule should match and examples of images the rule should reject. The technique learns the common image characteristics of the examples, and any current or future image search results can then be ranked or re-ranked according to the learned rules.

WO 2009/075976 A2

GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

INTERACTIVE CONCEPT LEARNING

IN IMAGE SEARCH

BACKGROUND

[0001] There is growing interest in image retrieval such as would be encountered, for example, in an image search on the World Wide Web because of the limitations inherent in metadata-based systems. Presently most image searches are conducted by searching using keywords. This requires that images be located by matching the keywords in a search query to metadata associated with an image. This type of a search requires mechanisms (such as human labeling) that can describe every image via metadata which is impractical for very large databases, or for images that are generated automatically, such as, for example, from surveillance cameras.

[0002] Another reason image search, such as searching for images on the World Wide Web, is difficult is in part because a handful of keywords are generally insufficient for characterizing the visual properties of an image, such as, for example, that an image with vertical lines in it is sought. Popular search engines have begun to provide a set of automatically generated meta tags based on simple characteristics of images (such as tags for black and white images or images that contain a face), but such approaches are limited by the fact that it is unclear what tags end-users want to be able to use in examining Web image search results.

SUMMARY

[0003] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter.

[0004] The interactive concept learning image search technique described herein allows end-users to quickly create their own rules, based on the image characteristics of images for ranking or re-ranking images returned in response to a search query. These image characteristics can include visual as well as semantic features or characteristics, for example, or a combination of both. Once one or more rules have been created end-users can then rank or re-rank any future image search results according to their rule or any combination of rules. In one embodiment of the technique end-users provide example images each rule should match and examples of images the rule should reject. The technique learns the common image characteristics of the example images as a rule or concept and can then rank or re-rank any current or future image search, image list or image database according to the learned concept.

[0005] In the following description of embodiments of the disclosure, reference is made to the accompanying drawings which form a part hereof, and in which are shown, by way of illustration, specific embodiments in which the technique may be practiced. It is understood that other embodiments may be utilized and structural changes may be made without departing from the scope of the disclosure.

DESCRIPTION OF THE DRAWINGS

[0006] The specific features, aspects, and advantages of the disclosure will become better understood with regard to the following description, appended claims, and accompanying drawings where:

[0007] FIG. 1 provides an overview of one possible environment in which image searches are typically carried out.

[0008] FIG. 2 is a diagram depicting one exemplary architecture in which one embodiment of the interactive concept learning in image search technique can be employed.

[0010] FIG. 3 is a flow diagram depicting a generalized exemplary embodiment of a process employing one embodiment of the interactive concept learning in image search technique.

[0011] FIG. 4 is a flow diagram depicting another exemplary embodiment of a process employing one embodiment of the interactive concept learning in image search technique.

[0012] FIG. 5 is a schematic of an exemplary computing device in which the present interactive concept learning in image search technique can be practiced.

DETAILED DESCRIPTION

[0013] In the following description of the interactive concept learning in image search technique, reference is made to the accompanying drawings, which form a part thereof, and which is shown by way of illustration examples by which the interactive concept learning in image search technique may be practiced. It is to be understood that other embodiments may be utilized and structural changes may be made without departing from the scope of the claimed subject matter.

1.0 Interactive Concept Learning in Image Search

[0014] The interactive concept learning image search technique provides an image search application that allows end-users to quickly create their own rules for ranking or re-ranking images based on the image characteristics of the images. These image characteristics may include visual characteristics or semantic features or characteristics, or a combination of both. Any current or future image search results can then be ranked or re-ranked according to the created rule or rules. End-users provide examples of images each rule should match and examples of images the rule should reject. The technique learns the common image characteristics of the examples. Search results of any current or future image search can then be ranked or re-ranked according to one or more of the learned rules.

[0015] The interactive concept learning in image search technique can be deployed in terms of a stand alone application on a personal computer or similar computing device, within a Web page or be integrated in a Web browser.

[0016] The various embodiments of the interactive concept learning in image search technique provide for many unexpected results and advantages. For example, a user can create their own image search rules based on the image characteristics of an image they are looking for. This image search can be totally independent of any search words or metadata associated with images on the Web, in a database or on the user's computing device or an associated server. Alternately the search terms can be combined with the learned rules to create hybrid embodiments of the technique.

[0017] The interactive concept learning in image search technique can be used to search any image repository such as the World Wide Web or an image database resident on a user's computing device.

1.1 Search Environment

[0018] FIG. 1 provides an overview of an exemplary environment in which image searches such as, for example, image searches on the Web or other network, may be carried out. Typically, a user searches for information on a topic on the Internet or on a Local Area Network (LAN) (e.g., inside a business).

[0019] The Internet is a collection of millions of computers linked together and in communication on a computer network. A personal home computer 102 may be linked to the Internet or Web using a phone line and a modem, a digital subscriber line (DSL), a wireless connection, or a cable modem 104 that talks to an Internet service provider (ISP) 106. A computer in a larger entity such as a business will usually have a network interface card (NIC) that directly connects it to a local area network (LAN) 110 inside the business. The business can then connect its LAN 210 to an ISP 106 using a high-speed phone line like a T1 line 112. ISPs then connect to larger ISPs 114, and the largest ISPs 116 maintain fiber optic "backbones" for an entire nation or region. In this way, every computer on the Internet can be connected to every other computer on the Internet.

[0020] The World Wide Web (referred sometimes as the Web herein) is a system of interlinked hypertext documents accessed via the Internet. There are millions of pages of information and images available on the World Wide Web. When a person conducting a search seeks to find information on a particular subject or an image of a certain type they typically visit an Internet search engine to find this information on other Web sites via a browser. Although there are differences in the ways different search engines work, they typically crawl the Web (or other networks or databases), inspect the content they find, keep an index of the words they find and where they find them, and allow users to query or search for words or combinations of words in that index. Searching through the index to find information involves a user building a search query and submitting it through the search engine via a browser or client-side application. Text and images on a Web page returned in response to a query can contain hyperlinks to other Web pages at the same or different Web site.

[0021] It should be noted that while the interactive concept learning in image search technique may be used to search images over a LAN 110 or on the Internet as discussed above, the technique can equally well be used to search for images on a user's computing device, such as is later discussed with respect to FIG. 5.

1.2 Exemplary Architecture

[0022] One exemplary architecture 200 (residing on a computing device 500 such as discussed later with respect to FIG. 5) in which the interactive concept learning in image search technique can be employed is shown in FIG. 2. To acquire a set of images 202, an image search 202 can be employed, which processes search queries that are typically input into a search engine via a browser. The Web browser is a software application that enables a user to display and interact with text, images, and other information typically located on a Web page at a Web site on the World Wide Web or other network. The search query typically returns a set of search results which are displayed to a user and manipulated by the user via a User Interface 210.

[0023] The interactive concept in learning image search technique module 206 interfaces with the user interface (UI) 210, which interfaces with a browser and search engine and displays to the user a set of images returned from the search or other image database 202 and which also allows the user to provide examples of images with the image characteristics sought to be found (or examples of images that do not have the image characteristics sought to be found) 208. The interactive concept learning in image search module 206 of the technique learns rules or concepts regarding the type of images a user is trying to find based on the image characteristics of the example images and rearranges the images returned in a search (either the images returned in response to the present search or a new set of images 220 of a future search) in accordance with the learned rules. The rule or concept learning involves extracting features of the image characteristics of the images of the image set and the example images in a feature extraction module 212. A rule learning module 214 attempts to learn a rule by ranking

images of the image set based on the features of the image characteristics of the example images. This involves weighting the features of each image of the image set in a feature weighting module 216. The images of the image set are then ranked or re-ranked based on the learned rule or rules in a ranking module 218. The ranked images may be stored in a re-ranked image database 222.

1.3 Exemplary Process Employing the Interactive Concept Learning in Image Search Technique.

[0024] A general exemplary process employing the interactive concept learning in image search technique is shown in FIG. 3. In a very general sense, this embodiment of the technique operates as follows. A set of images is obtained (box 302). This set of images can be obtained, for example, by searching for them by keyword on a network, or by searching for them on a user's computing device. Examples of images with image characteristics sought to be found are provided (box 304). For example, these example images can be selected from the set of images returned in response to the aforementioned search by a user. Rules or concepts that specify which type of images are to be found (which images are to be ranked more highly in a set of images) are learned via a machine learning technique based on the provided example images (box 306). Various machine learning techniques may be used, such as, for example, Support Vector Machines (SVMs), regression, Bayesian networks, or nearest neighbor classifiers. The learned rules or concepts can then be employed for ranking or re-ranking images, such as those found in an image search, based on their image characteristics (box 308). Images with image characteristics sought to be found are ranked higher than images without the image characteristics sought to be found. The learned rule or concepts can also be applied to another set of images (for example, images returned in a new search or in a different image database) based on the sought image characteristics (box 310).

[0025] Another more detailed exemplary process employing the interactive concept learning image search technique is shown in FIG. 4. A set of images is obtained (box 402). A number of features, such as, for example, visual features

are computed for each of the set of images (box 404). A user then selects an example of an image, either a positive example having features of the type of image that they are looking for, or a negative example not having features of the type of image they are looking for (box 406). (It should be noted that user can continue to provide positive or negative examples to the rule even if they apply the rule to other search queries. In fact, continuous training like this is sometimes important in making a rule as robust as possible.) A concept or rule is learned by computing a weighting of features of each image of the set of images based on the example image (or images) provided to determine the most relevant of the images of the image set (box 408). The images of the image set are then re-ranked according to the computed weighting and shown to the user (box 410). The user evaluates the re-ranked images (box 412) and if not satisfied with the re-ranked images (box 414) provides an additional example image (box 406). This continues until the user is satisfied that the most relevant images of the re-ranked images contain one or more images that the user is looking for. It should be noted that it is also possible to provide the user with a selected set of example images for the user to specify which image is more like the one they are looking in order to speed the process of determining the most relevant images (e.g., those images with the sought after image characteristics (box 416) (e.g., active learning as will be discussed in detail later). Additionally, it is also possible to apply the learned rule or rules to a totally different set of images if desired (box 418).

1.4 Exemplary Embodiments

[0026] Various alternate embodiments of the interactive concept learning in image search technique can be implemented.

[0027] One exemplary embodiment of the interactive concept learning in image search technique can be implemented as an application that retrieves images from a keyword-based Web image search engine, although the technique can equally well be applied to perform image searches of images stored on the desktop computer or other computing device. The interactive concept learning in image search technique enables the ranking or re-ranking of image search results

or images in an image database according to user-created rules based on the image characteristics of the images.

[0028] In one embodiment of the interactive concept learning in image search technique, each rule is defined as a nearest-neighbor classifier, computing a score that indicates how similar an image is to the examples that were used to train that rule. The training of such rules requires learning a distance function from the examples provided by an end-user. In order to help end-users provide informative examples that help the interactive concept learning in image search technique determine what rule the user is creating, in one embodiment, the interactive concept learning in image search technique implements active learning criteria to help a user provide good examples for learning the rules. Details of these and other implementations of the interactive concept learning in image search technique are discussed in greater detail below.

[0029] It should be noted that many alternative embodiments to the discussed embodiments are possible, and that steps and elements discussed herein may be changed, added, or eliminated, depending on the particular embodiment. These alternative embodiments include alternative steps and alternative elements that may be used, and structural changes that may be made, without departing from the scope of the disclosure.

1.4.1 Image Queries to Obtain a Set of Images

[0030] In one exemplary embodiment, as shown in FIG. 2, the interactive concept learning in image search technique programmatically retrieves images using queries to conduct an image search 202 using any of a number of search engines. In one embodiment of the technique, a format parameter in the query indicates that the engine should return its results list in an XML format, and the interactive concept learning in image search technique downloads the thumbnail images for each image (although full sized images could also be used with the technique). The image characteristics of each image returned in response to the search query are analyzed 212 as they are downloaded, and the resulting images

are ranked according to any active learned rules 214, 216. Alternatively, the image characteristics of images can be analyzed when the web crawler first finds and indexes the image, and these characteristics need only be downloaded, rather than computed, when the search results are downloaded.

1.4.2 Ranking Image Results in accordance with Learned End-User Rules

[0031] As discussed previously, images are ranked or re-ranked by applying a set of end-user created rules. In one exemplary embodiment of the technique, users can enable and disable the learned rules or concept (for example, in one exemplary user interface by moving them between Active Rule and Available Rule panes on the display of the user interface 210). In one embodiment the user 208 can also change the relative weight of one rule of a set of rules. For example, in one exemplary user interface 210 a slider control on each active rule allows the user to control the relative weighting of multiple rules. In this embodiment, every active rule computes a score for each image, and scores are multiplied by a weighting, such as, for example, between -1 and 1. Images are thus ranked by a weighted sum of scores. Mathematically the image score of an image i can be expressed as:

$$ImageScore(i) = \sum_{r \in ActiveRules} weight_r * score_r(i).$$

where $weight_r$ is the weight of the rule and $score_r(i)$ is the score of the image i for the rule r.

1.4.3 Applying Concept Rules via Nearest-Neighbor Classifier.

[0032] In one embodiment of the interactive concept learning in image search technique each rule is learned via a nearest-neighbor classifier. Each rule is defined as a set of positive examples (images illustrating what the rule should match), a set of negative examples (images illustrating what the rule should reject), and a distance metric. Given these parameters, a rule scores a new image

of an image set by computing the distance between that image and each positive or negative example image, then dividing the distance to the nearest positive example by the sum of the distance to the nearest positive and nearest negative example. Mathematically the score of an image for a given rule can be expressed as:

$$score_r(i) = 1 - \frac{mindist_p}{mindist_p + mindist_N}$$

where $mindist_p$ is the minimum distance between the image and a positive example image and $mindist_N$ is the minimum distance between the image and a negative example image. Note that $score_r(i)$ ranges between 0 and 1, approaching 1 when i is near a positive example and far from negative examples, having value .5 when i is equally close or far from the nearest positive and negative examples, and approaching 0 when i is near a negative example and far from positive examples. For its distance metric, the interactive concept learning in image search technique uses a weighted sum of a set of several component distance metrics:

$$Distance(i, j) = \sum_{m \in metrics} weight_m * distance_m(i, j)$$

where the overall distance between two images is described by the sum of distances of each metric weighted by their relative importance. The weights are the factor that is learned (as described below) and the set of these uniquely defines each rule.

[0033] Given this approach, the core of this embodiment of the technique's ability to re-rank images according to their image characteristics lies in a set of component distance metrics and the interactive concept learning in image search technique's ability to learn how to weight those different metrics.

1.4.3.1 Distance Metrics

[0034] In one embodiment, the interactive concept learning in image search technique implements image distance metrics, for example, based on histograms of hue, saturation, and luminosity of pixels, an edge histogram, a global shape histogram, and a texture histogram. The technique computes and stores these histograms for each image, using them to efficiently compute distances between images.

[0035] In this embodiment, the hue, saturation, luminosity, and edge histograms are computed over the pixels in each thumbnail image and normalized to account for thumbnails of varying size (or full-sized images if full-sized images are used). In this embodiment, two distance metrics are defined for each histogram. The first distance metric is the quadratic distance between two histograms, a measure of histogram distance that accounts for the similarity between different bins in the histogram. In the case of the luminosity histogram, for example, an image entirely of luminance 1.0 is considered more similar to an image entirely of luminance 0.8 than it is to an image entirely of luminance 0.4 (the simpler Euclidean comparison would treat the two images as equally dissimilar from the first, provided the three luminosity values are in different histogram bins). The second metric for each image's histograms is the difference in histogram entropy.

[0036] In this embodiment, a histogram is also computed representing the overall structure of each image by applying a shape descriptor to the entire image. This descriptor sums local image gradients into bins over a log-polar target-shaped region covering the entire image, normalizing the resulting histogram. Similar histograms (using Euclidean distance) correspond to images with similar overall structure, and the descriptor offers a degree of invariance to illumination, translation, scale, and rotation. For example, this distance metric will generally indicate that two frontal close-ups of a face are similar and that two driver-side views of a car are similar. It will also generally indicate that a close-up of a face is different from a view of a car. The shape descriptor does not consider color, and so it complements the color histogram metrics.

[0037] Finally, a texture histogram is computed that preserves less geometric information than the global shape histogram but allows discrimination between the distribution of structures present in an image without regard to their arrangement. In one embodiment a bag-of-words approach is applied that samples a number of patches from the image. For efficiency, samples on a regular grid of partially overlapping blocks and computes a descriptor for each block. An offline recursive clustering analysis of a large image database is used to learn a set of discriminative textures, and a histogram is computed at runtime by resolving each sampled block to a bin based on the identified discriminative textures. Less formally, this metric considers images similar if they contain similar patches. It might consider images of two different city skylines to be similar, while the previously discussed global shape descriptor might consider the two skylines different.

[0038] In other embodiments of the technique, non-visual distance metrics, such as keywords can also be included as distance metrics. Many search engines index and return multiple keywords with each piece of content. Hence, if for example "car" were the keyword used to retrieve a set of images, each of those images will also have other keywords associated with it, such as "blue, 4-wheel drive, porsche" for one image or "yellow, convertible, fast" for another. In practice, while the focus is mostly on image characteristics in this disclosure, any characteristics that are associated with images or content can be used.

1.4.3.2 Learning over Distance Metrics

[0039] As discussed previously, the interactive concept learning in image search technique learns rules from positive and negative examples of images that rule should match or reject. Given a set of positive examples, there are many concepts a person might be attempting to specify. In most applications of nearest-neighbor algorithms, the developer of a system carefully tunes a distance function based on their knowledge of the problem being solved. However, with the current technique it is not known beforehand what notion of similarity will be appropriate for an end-user's rule. If an attempt is made to treat all distance metrics equally, dimensionality guarantees a very large number of images will be

required in order to specify even the simplest rules. As previously discussed with respect to FIG. 2, the interactive concept learning in image search technique therefore, in one embodiment, defines a concept or rule learning problem as a matter of learning a set of weights 216 based on which distance metrics best correspond to the provided examples. The interactive concept learning in image search technique can learn, for example, whether a set of images are similar because of their color histogram, their global shape descriptor, or a combination of the two. One method for learning these weights is described next.

[0040] Given a set of positive and negative examples, one embodiment of the interactive concept learning in image search technique learns a set of distance metric weights such that the distance between two images with the same label (positive or negative) is minimized and the distance between two images with different labels is maximized. Specifically, an objective function that separates the two classes as much as possible is minimized while keeping examples of the same class close together. Mathematically this can be described as:

$$f(weights) = \sum_{i,j \in Pos} D(i,j) + \sum_{i,j \in Neg} D(i,j) + \sum_{i \in All} \ln \sum_{j \in all} e^{-D(i,j)}$$

The first two terms of the previous equation correspond to within-class distances, thus the minimization of the function favors weights that minimize the distance between data of the same class. The third term considers all examples and favors maximum separation. The combination of terms thus favors weights that collapse each of the classes while maximizing the distance between data with different labels. The function is convex, and the unique global minimum is efficiently found using standard non-linear optimization techniques.

[0041] Less formally, the interactive concept learning in image search technique learns what notions of distance are relevant based on the examples that a user has provided. For example, if all of the provided positive examples are mostly yellow, and they have no other common characteristics, and the negative examples are not mostly yellow, the interactive concept learning in image search technique will learn that hue histogram similarity is the relevant distance (giving it

a large weight and other distance metrics small weights). The resulting rule will give high scores to images with hue histograms similar to those of the positive examples used to train the rule. In a situation where the positive examples have multiple characteristics in common, those characteristics will each receive some weighting.

1.4.4 Active Learning to Provide Better Example Images

[0042] As discussed previously, for example with respect to FIG. 4, because it might sometimes be difficult to determine what images an end-user should provide as examples in order to help the interactive concept learning in image search technique learn the correct concept or rule, the interactive concept learning in image search technique uses active learning techniques 416 to identify images from the set of images returned from the current query that provide the most information about the rule a person is creating. In general, in this aspect of the technique, in one embodiment, two or more images about which the technique is most confused are provided to a user to select which of the images most readily represents the rule or concept the user is attempting to establish.

[0043] In one embodiment, the interactive concept learning in image search technique's first active learning procedure finds images of the image set that, given the current set of learned distance weights, are closest to the boundary between positive and negative. These are the images about which the technique is most uncertain, so labeling them provides the most information within the space defined by the current distance weights. In one embodiment, mathematically, the interactive concept learning in image search technique selects images with the smallest value:

$$uncertain(i) = abs(mindist_p - mindist_N)$$

[0044] The interactive concept learning in image search technique's second active learning procedure finds images of the image set that will result in the exploration of new weightings of the distance metrics. For one embodiment of the

interactive concept learning in image search technique, $activedistance(i)$, a heuristic based on data density and uncertainty, selecting images with the smallest value, is applied:

$$activedistance(i) = (mindist_p + mindist_N) * uncertain(i)$$

The intuition behind this heuristic is to find images that are in dense portions of the space (near other labeled examples) but are still very uncertain. The first term captures density, as the distance to positive and negative examples will be lower in dense portions of the space. The second term captures uncertainty, as discussed in the previous paragraph. Images with low scores are selected, and labeling those images gives the interactive concept learning in image search technique new information to use in finding a weighting of distance metrics that pushes positive and negative examples away from each other.

1.4.4.1 User Interface Condition Considerations

[0045] There are various considerations on how the technique should present data to the user. When a user edits a concept or rule by providing new example images, how the user interface presents the effect of a rule on the current image set is a consideration. There are two dimensions of one embodiment of the rule editing user interface that deserve careful attention. The first dimension, herein called Editing Presentation, is how the interface presents the effect of a rule on the current image set while that rule is being edited. The second, herein called Active Learning Presentation, is how the interface presents the examples identified by the active learning procedures. Crossing these two dimensions yields six possible user interface embodiments of the interactive concept learning in image search technique.

1.4.4.1.1 Editing Presentation: Displaying the Ranked Images

[0046] As a rule or concept is being edited, the technique needs to present what rule has been learned. Presenting the positive and negative examples that form the basis for the rule is straightforward.

[0047] The previously discussed learned distance metric weights can be illustrated by showing the images in the current query results or image set ranked according to the rule being edited. Editing Presentation considers whether the technique should show the entire set of images as they are ranked by the rule that is being edited (Single), or show only a small subset of the images, those that rank at the very top and those that rank at the very bottom (Split).

[0048] The Single approach to displaying the ranked images provides the end-user with access to the entire set of images from the current query, so they have more images to choose from in training the rule. But the rule is unlikely to ever be completely perfect, so a person may become overly focused on the noisy boundary between positive and negative images, continuing to provide training examples that are no longer noticeably improving a rule.

The Split approach to displaying the ranked images avoids the possibility of continuing to provide training examples that are no longer noticeably improving a rule, as end-users can provide training examples only until the small subset of images displayed from the top of the ranking match the desired concept and the small subset of images displayed from the bottom of the ranking are examples of images that have been correctly rejected. This comes at the cost that fewer images are available to choose from when providing training examples.

1.4.4.1.2 Active Learning Presentation: Displaying Active Learning Examples

[0049] A similar tradeoff is explored in considering how to present images selected by the technique's active learning procedures.

[0050] These images, by definition, will exist in the most uncertain regions of the image query space. They may therefore either help a participant quickly find effective examples, or their presence may lead a participant to continue providing examples even after they are no longer noticeably improving a rule. Active Learning Presentation considers whether the interactive concept learning in image search technique should place active learning images in a separate pane (Explicit), randomly distribute active learning images near the top and bottom of a set of results in the hope of them being visually distinct from the nearby images may lead active learning images to be selected for use as a positive or negative example (Embedded), or rely only upon the ranked query results (None).

1.4.4.1.3 User Interface Conditions for Optimum Rule Editing and Active Learning Image Presentation

[0051] In one embodiment of the interactive concept learning in image search technique, crossing these two dimensions of Active Learning Presentation and Editing Presentation discussed above yields six user interface conditions as discussed below which should be considered in designing a user interface for various embodiments of the technique.

[0052] *Split-Explicit.* In the case where only the top and bottom ranked images are displayed and active learning images are explicitly displayed, in one embodiment of the technique the split-explicit condition uses three scroll panes to display this data to the user. In one exemplary embodiment, these three panes present the 1) 50 top-ranked images, 2) 10 images selected by the technique's active learning procedures, and 3) the 50 bottom-ranked image results.

[0053] *Split-Embedded.* In the case where only the top and bottom ranked images of the image set are displayed and the active learning images are embedded images, in one embodiment the split-embedded condition user interface uses two scroll panes to display the data. In one embodiment, the first pane presents the 50 top-ranked results, as well as 5 randomly-seeded active

learning images. The second pane presents the 50 bottom-ranked results, as well as 5 randomly-seeded active learning results.

[0054] 3) *Split-None*. In one embodiment, in the case where only the top and bottom ranked images are displayed and no active learning images are shown, the split-none technique uses two scroll panes. In one embodiment these panes present the 50 top-ranked results and the 50 bottom-ranked results.

[0055] 4) *Single-Explicit.* In one embodiment in the case where the technique provides the end-user with access to the entire set of images so that they have more images to choose from in training the rule and the active learning images are explicitly displayed, the split-explicit condition uses two scroll panes. Ten active learning results are displayed in one scroll pane. The other pane displays the remainder of the ranked images of the image set.

[0056] 5) *Single-Embedded.* In one embodiment in the case where the technique provides the end-user with access to the entire set of images so that they have more images to choose from in training the rule and the active learning images are embedded, the single-embedded condition uses one scroll pane. This condition seeds the top and bottom of the rankings with active learning results using the same procedure as the Split-Embedded condition and then displays the entire modified ranking in one scroll pane.

[0057] 6) *Single-None.* In one embodiment in the case where the technique provides the end-user with access to the entire set of images so that they have more images to choose from in training the rule and no active learning images are shown, the single-none condition uses one scroll pane. It displays all of the ranked query images.

[0058] Additionally, in one embodiment, users are shown the actual distance metrics and the weights associated with them (i.e. which ones the system has selected as important). Users can then manually edit these, for example, indicating that they did not mean for “color” to be an important metric, but rather “shape”.

2.0 The Computing Environment

[0059] The interactive concept learning in image search technique is designed to operate in a computing environment. The following description is intended to provide a brief, general description of a suitable computing environment in which the interactive concept learning in image search technique can be implemented. The technique is operational with numerous general purpose or special purpose computing system environments or configurations. Examples of well known computing systems, environments, and/or configurations that may be suitable include, but are not limited to, personal computers, server computers, hand-held or laptop devices (for example, media players, notebook computers, cellular phones, personal data assistants, voice recorders), multiprocessor systems, microprocessor-based systems, set top boxes, programmable consumer electronics, network PCs, minicomputers, mainframe computers, distributed computing environments that include any of the above systems or devices, and the like.

[0060] FIG. 5 illustrates an example of a suitable computing system environment. The computing system environment is only one example of a suitable computing environment and is not intended to suggest any limitation as to the scope of use or functionality of the present technique. Neither should the computing environment be interpreted as having any dependency or requirement relating to any one or combination of components illustrated in the exemplary operating environment. With reference to FIG. 5, an exemplary system for implementing the interactive concept learning in image search technique includes a computing device, such as computing device 500. In its most basic configuration, computing device 500 typically includes at least one processing unit 502 and memory 504. Depending on the exact configuration and type of computing device, memory 504 may be volatile (such as RAM), non-volatile (such as ROM, flash memory, etc.) or some combination of the two. This most basic configuration is illustrated in FIG. 5 by dashed line 506. Additionally, device 500 may also have additional features/functionality. For example, device 500 may

also include additional storage (removable and/or non-removable) including, but not limited to, magnetic or optical disks or tape. Such additional storage is illustrated in FIG. 5 by removable storage 508 and non-removable storage 510. Computer storage media includes volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules or other data. Memory 504, removable storage 508 and non-removable storage 510 are all examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can accessed by device 500. Any such computer storage media may be part of device 500.

[0061] Device 500 has a display 518, and may also contain communications connection(s) 512 that allow the device to communicate with other devices. Communications connection(s) 512 is an example of communication media. Communication media typically embodies computer readable instructions, data structures, program modules or other data in a modulated data signal such as a carrier wave or other transport mechanism and includes any information delivery media. The term "modulated data signal" means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal, thereby changing the configuration or state of the receiving device of the signal. By way of example, and not limitation, communication media includes wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, RF, infrared and other wireless media. The term computer readable media as used herein includes both storage media and communication media.

[0062] Device 500 may have various input device(s) 514 such as a keyboard, mouse, pen, camera, touch input device, and so on. Output device(s) 516 such as speakers, a printer, and so on may also be included. All of these devices are well known in the art and need not be discussed at length here.

[0063] The interactive concept learning in image search technique may be described in the general context of computer-executable instructions, such as program modules, being executed by a computing device. Generally, program modules include routines, programs, objects, components, data structures, and so on, that perform particular tasks or implement particular abstract data types. The interactive concept learning in image search technique may be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote computer storage media including memory storage devices.

[0064] It should also be noted that any or all of the aforementioned alternate embodiments described herein may be used in any combination desired to form additional hybrid embodiments. Although the subject matter has been described in language specific to structural features and/or methodological acts, it is to be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or acts described above. For example, even though some of the described embodiments use the visual characteristics to learn the rule or rules to rank the images found in a search, the visual characteristics could also be combined with the keywords of the search to learn the rule or rules to rank the images of an image set to create a hybrid embodiment to rank the images found in a search. The specific features and acts described above are disclosed as example forms of implementing the claims.

Wherefore, what is claimed is:

1. A computer-implemented process for finding one or more images of desired image characteristics, comprising:

obtaining a set of images (302);

providing example images with image characteristics sought to be found (304);

learning one or more rules for ranking images based on the image characteristics of the example images via a machine learning technique (306); and

ranking the set of images based on the one or more learned rules (308).

2. The computer-implemented process of Claim 1 further comprising applying the one or more learned rules to a second set of images.

3. The computer-implemented process of Claim 1 further comprising combining two or more learned rules before ranking the set of images.

4. The computer-implemented process of Claim 1 further comprising obtaining the set of images in response to a search query submitted over a computer network.

5. The computer-implemented process of Claim 1 further comprising obtaining the set of images by locally searching for images on a computing device.

6. The computer-implemented process of Claim 1, further comprising repeating the actions of:

providing examples of images with image characteristics sought to be found;

learning one or more rules for ranking images based on the image characteristics of the example images via a machine learning technique; and

ranking the set of images based on the one or more learned rules.

until the images of the set of images are ranked with the images with image characteristics sought to be found being ranked highest.

7. The computer-implemented process of Claim 1 wherein each rule is based on learning a distance metric which computes a score for each image returned in the set of images that indicates how similar each image returned in the set of images is to the examples of images with image characteristics sought to be found.

8. The computer-implemented process of Claim 7, further comprising

computing a set of image histograms computed over the pixels in each image in the set of images;

computing two distance metrics for each histogram of the set of image histograms;

computing a structure histogram that represents the overall structure of each image in the set of images by applying a shape descriptor to the whole image;

computing a texture histogram for each image in the set of images; and

using the computed image histograms, the distance metrics for each of the computed image histograms, the structure histogram and the texture histogram for each image to compute a score for each image in the set of images that indicates how similar each image in the set of images is to the example images with image characteristics sought to be found and example images not having image characteristics sought to be found.

9. The computer-implemented process of Claim 1 further comprising learning the one or more rules for ranking images based on a set of distance metrics that are used to represent the distance of any given image to any other given image.

10. The computer-implemented process of Claim 8 wherein the two distance metrics comprise:

the quadratic distance between the image histograms of two images; and

the distance in histogram entropy between two images.

11. The computer-implemented process of Claim 9 wherein the set of distance metrics are based on visual characteristics or semantic characteristics.

12. A computer-implemented process for searching for an image of desired image characteristics, comprising:

obtaining a set of images (402);

computing visual features for each of the set of images (404);

selecting an example image, either a positive example image having features of a type of image being sought, or a negative example image not having features of the type of image being sought (406);

learning a rule by computing a weighting of features of each of the set of images based on one or more example images provided to determine the most relevant of the images of the image set (408); and

ranking the images of the image set according to the computed weighting (410).

13. The computer-implemented process of Claim 11 further comprising;

showing the ranked images to a user;

the user evaluating the ranked images and if not satisfied with the ranked images providing one or more example images until the user is satisfied that the most relevant images of the ranked images contain one or more images that the user is seeking.

14. The computer-implemented process of Claim 12 further comprising providing the user with example images for the user to specify at least one image is more like the image the user is seeking prior to the user providing new example images.

15. The computer-implemented process of Claim 12 further comprising applying one or more of the learned rules to a new set of images to rank the new set of images.

16. A system for ranking images retrieved in response to an image search, comprising:

a general purpose computing device (500);

a computer program comprising program modules executable by the general purpose computing device, wherein the computing device is directed by the program modules of the computer program to,

receive a set of images in response to an image search (204);

specify examples of images with image characteristics sought to be found (210);

extract features based on image characteristics of each of the set of images received in response to the image search and the specified example images (212);

learn a rule based on the features of the image characteristics of the example images (214);

weight the features of each of the images of the set of images (218);

rank the images of the image set based on one or more learned rules (218); and

display the ranked images of the image set (222) to a user (208).

17. The system of Claim 16 further comprising a module to apply one or more learned rules to rank a new set of images.

18. The system of Claim 16 further comprising a module to use key words to retrieve the set of images in the image search.

19. The system of Claim 18 further comprising using the key words combined with the features of the image characteristics sought to be found to learn one or more rules to rank the set of images.

20. The system of Claim 16 further comprising a user interface which displays the ranked images of the image set and provides positive and negative example images for the user to choose to speed the learning of the rule.

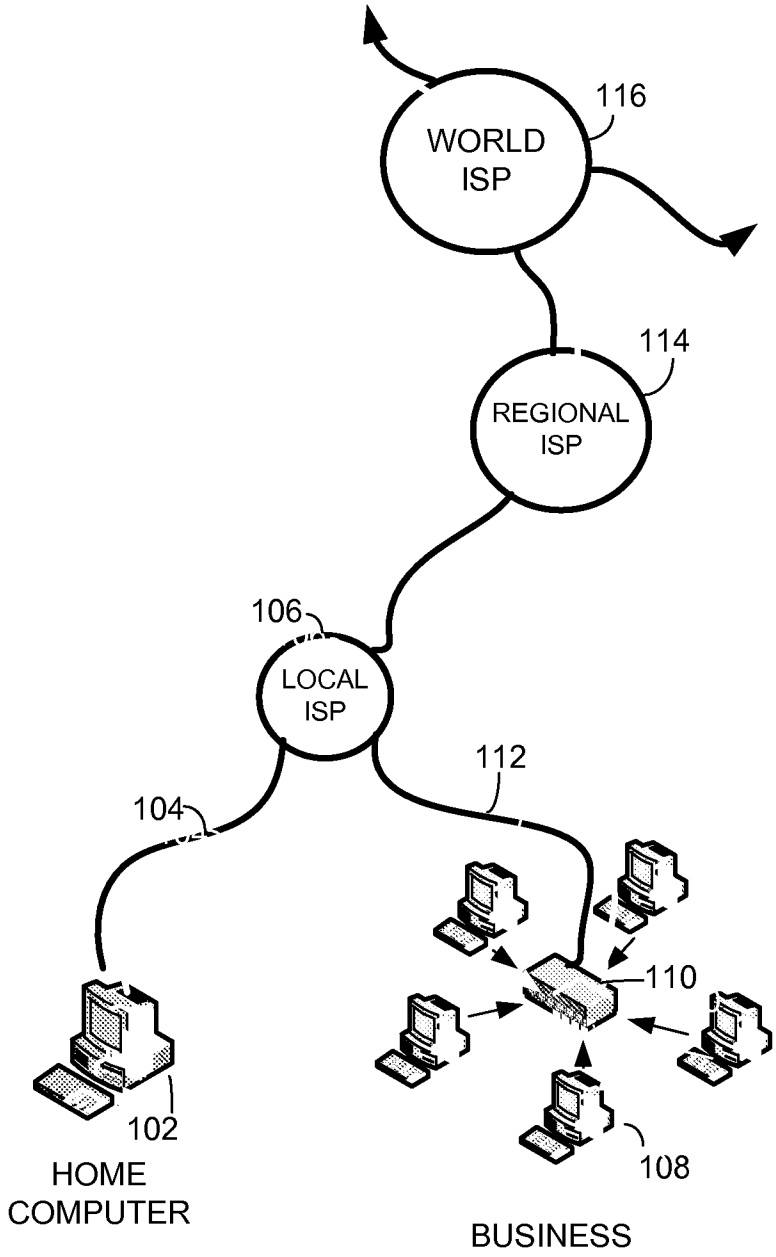

WORLD ISP
116
REGIONAL ISP
114
LOCAL ISP
106
104
112
110
102
HOME COMPUTER
108
BUSINESS

FIG. 1

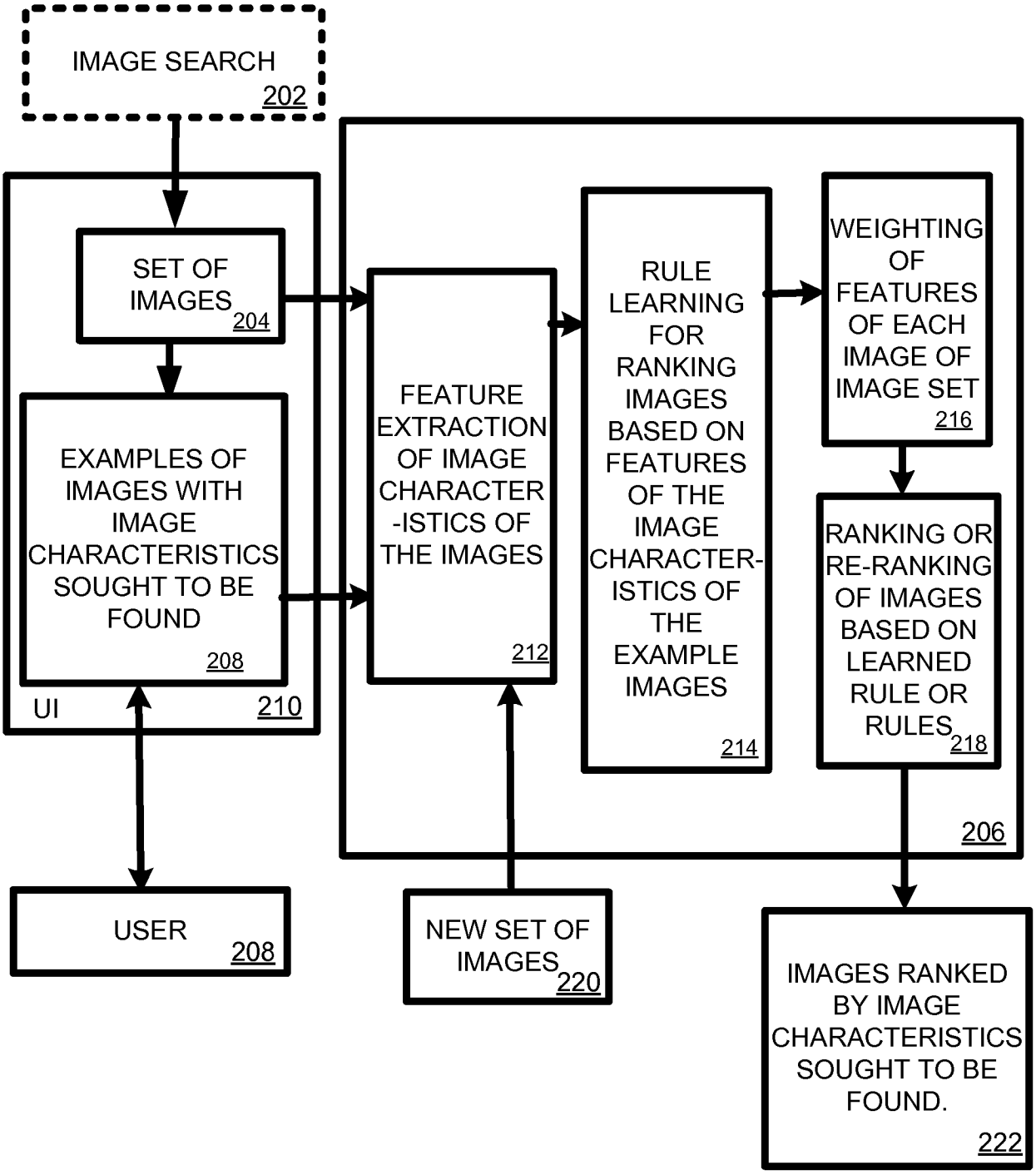

IMAGE SEARCH
202
SET OF IMAGES
204
EXAMPLES OF IMAGES WITH IMAGE CHARACTERISTICS SOUGHT TO BE FOUND
208
UI
210
FEATURE EXTRACTION OF IMAGE CHARACTER-ISTICS OF THE IMAGES
212
RULE LEARNING FOR RANKING IMAGES BASED ON FEATURES OF THE IMAGE CHARACTER-ISTICS OF THE EXAMPLE IMAGES
214
WEIGHTING OF FEATURES OF EACH IMAGE OF IMAGE SET
216
RANKING OR RE-RANKING OF IMAGES BASED ON LEARNED RULE OR RULES
218
206
USER
208
NEW SET OF IMAGES
220
IMAGES RANKED BY IMAGE CHARACTERISTICS SOUGHT TO BE FOUND.
222

FIG. 2

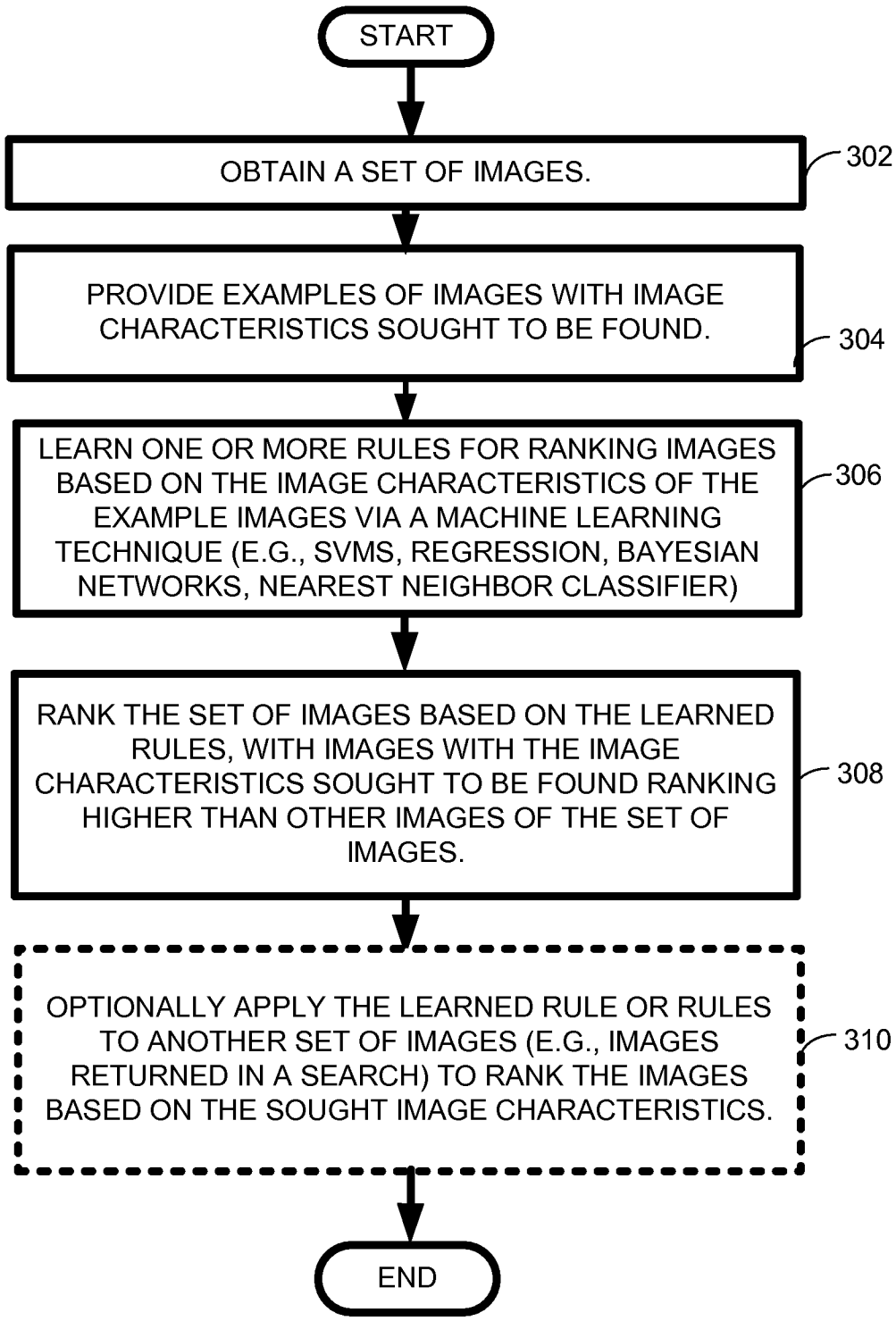

START
OBTAIN A SET OF IMAGES.
302
PROVIDE EXAMPLES OF IMAGES WITH IMAGE CHARACTERISTICS SOUGHT TO BE FOUND.
304
LEARN ONE OR MORE RULES FOR RANKING IMAGES BASED ON THE IMAGE CHARACTERISTICS OF THE EXAMPLE IMAGES VIA A MACHINE LEARNING TECHNIQUE (E.G., SVMS, REGRESSION, BAYESIAN NETWORKS, NEAREST NEIGHBOR CLASSIFIER)
306
RANK THE SET OF IMAGES BASED ON THE LEARNED RULES, WITH IMAGES WITH THE IMAGE CHARACTERISTICS SOUGHT TO BE FOUND RANKING HIGHER THAN OTHER IMAGES OF THE SET OF IMAGES.
308
OPTIONALLY APPLY THE LEARNED RULE OR RULES TO ANOTHER SET OF IMAGES (E.G., IMAGES RETURNED IN A SEARCH) TO RANK THE IMAGES BASED ON THE SOUGHT IMAGE CHARACTERISTICS.
310
END

FIG. 3

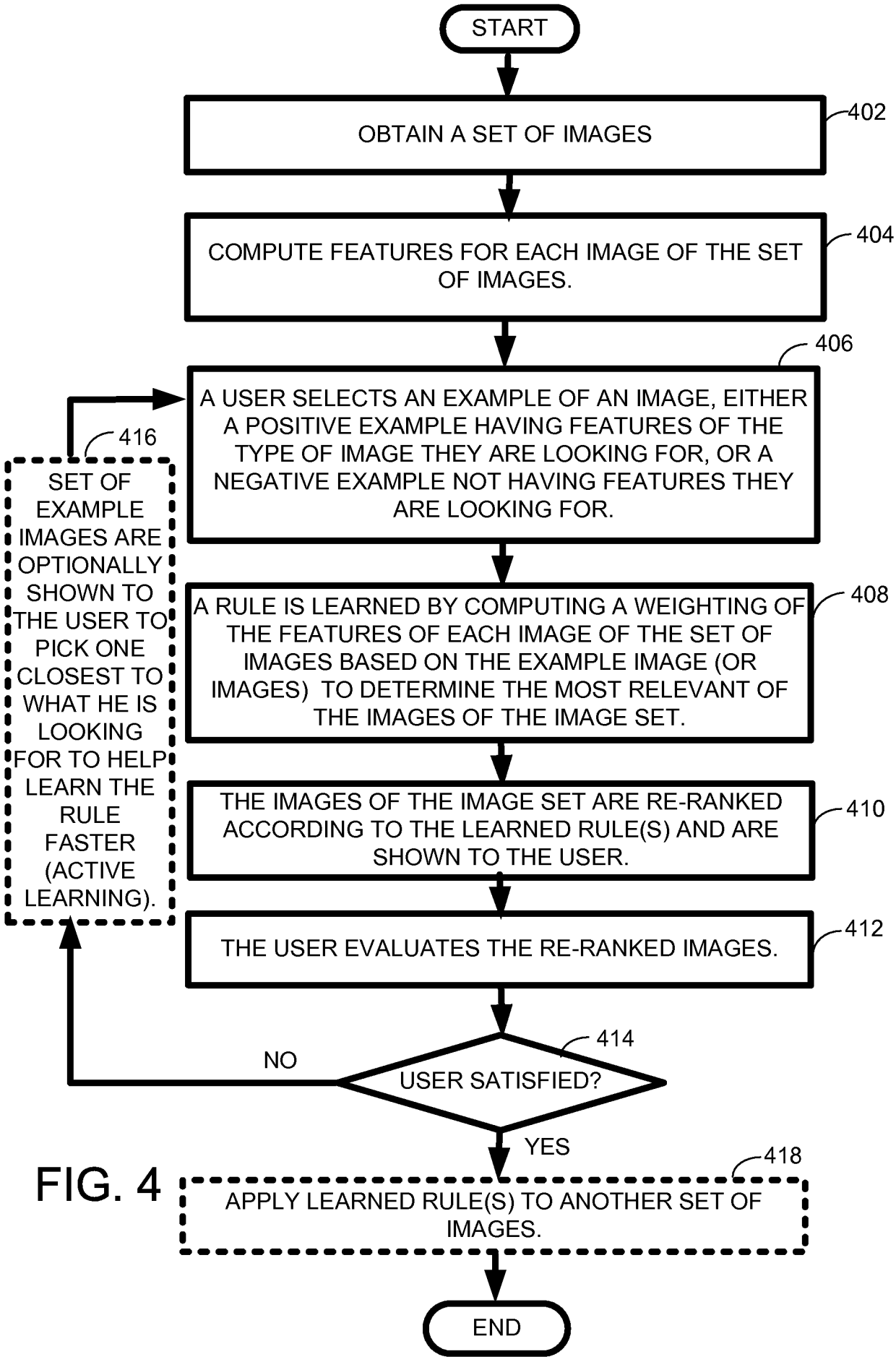

START
OBTAIN A SET OF IMAGES
402
COMPUTE FEATURES FOR EACH IMAGE OF THE SET OF IMAGES.
404
406
A USER SELECTS AN EXAMPLE OF AN IMAGE, EITHER A POSITIVE EXAMPLE HAVING FEATURES OF THE TYPE OF IMAGE THEY ARE LOOKING FOR, OR A NEGATIVE EXAMPLE NOT HAVING FEATURES THEY ARE LOOKING FOR.
416
SET OF EXAMPLE IMAGES ARE OPTIONALLY SHOWN TO THE USER TO PICK ONE CLOSEST TO WHAT HE IS LOOKING FOR TO HELP LEARN THE RULE FASTER (ACTIVE LEARNING).
A RULE IS LEARNED BY COMPUTING A WEIGHTING OF THE FEATURES OF EACH IMAGE OF THE SET OF IMAGES BASED ON THE EXAMPLE IMAGE (OR IMAGES) TO DETERMINE THE MOST RELEVANT OF THE IMAGES OF THE IMAGE SET.
408
THE IMAGES OF THE IMAGE SET ARE RE-RANKED ACCORDING TO THE LEARNED RULE(S) AND ARE SHOWN TO THE USER.
410
THE USER EVALUATES THE RE-RANKED IMAGES.
412
414
NO
USER SATISFIED?
YES
418
APPLY LEARNED RULE(S) TO ANOTHER SET OF IMAGES.
END

FIG. 4

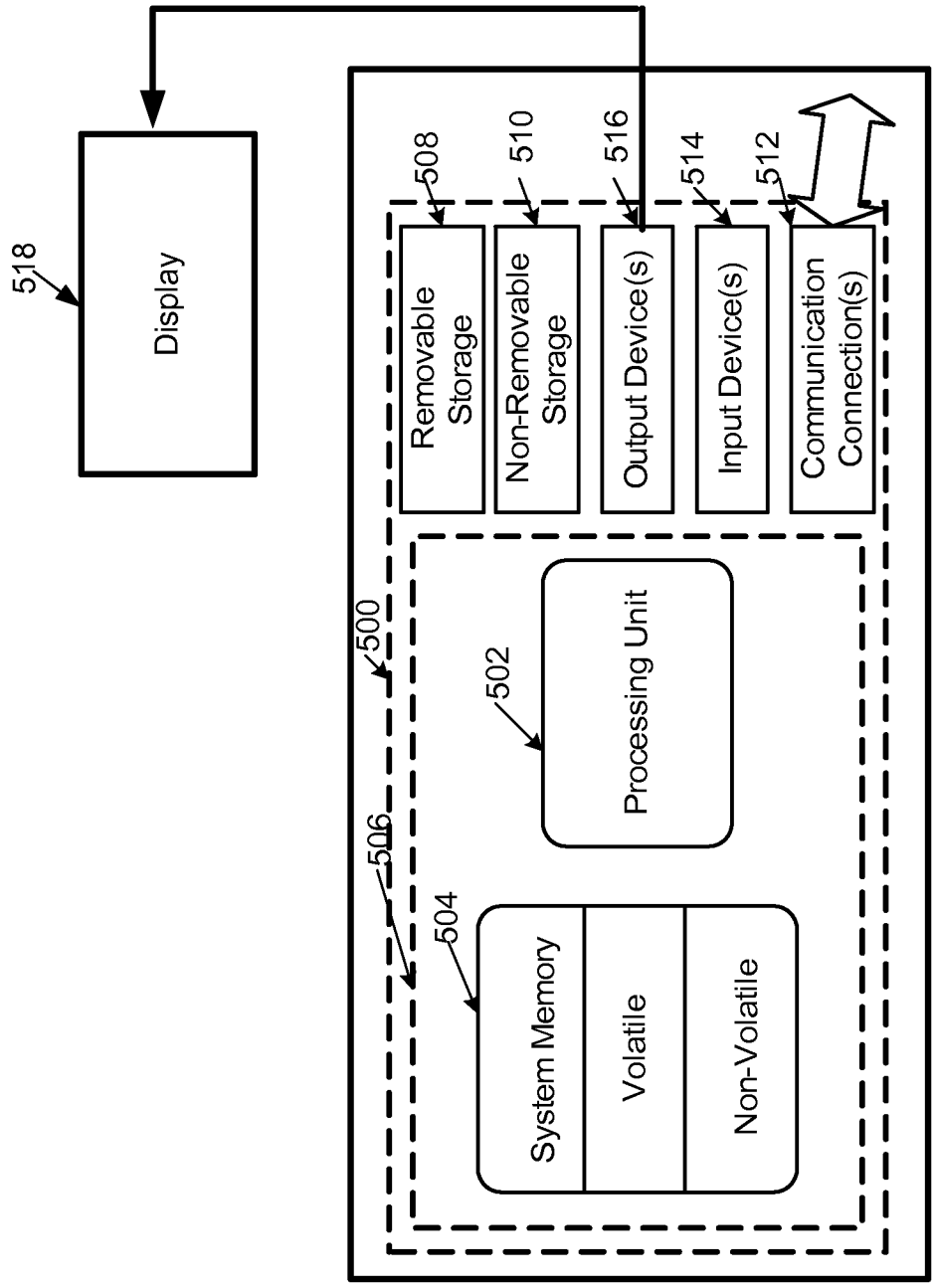

518
Display
500
508
Removable Storage
510
Non-Removable Storage
516
Output Device(s)
514
Input Device(s)
512
Communication Connection(s)
502
Processing Unit
506
504
System Memory
Volatile
Non-Volatile

FIG. 5