

A System And Method For Image Retrieval

5 TECHNICAL FIELD

The present disclosure relates broadly to a system for image retrieval and to a method of image retrieval.

10

BACKGROUND

With the progression of imaging technology such as integration of high definition cameras with portable devices, large amounts/numbers of visual data are being
15 collected e.g. by mobile and/or wearable cameras. It has been recognized that concept-based image retrieval is desired to make use of the visual data. One major challenge for such retrieval has been the so-called semantic gap, where images are characterized by visual features and the queries/collections are defined in terms of semantic concepts.

20 Technologies such as deep learning have become available to extract semantics from images. However, it has been recognized by the inventors that such semantics are relatively primitive as compared to possible query topics (which may be more sophisticated). It has been recognized by the inventors that a level of expertise and a tedious process (e.g. use of trial-and-error) is typically needed to translate query topics
25 into more primitive semantic concepts for efficient image retrieval.

As such, there exists a significant challenge to apply logic to a large collection of data, in particular, image data. One fundamental problem, in attempting to use query topics at a higher level than to translate the query topics to be more primitive, is
30 understanding contents of image data or images so as to support deeper insights. Current approaches may typically resort to deep learning technologies (e.g. usage of available databases (such as ImageNet, Place365, etc.) or convolutional neural

networks (such as, AlexNet, ResNet, GoogleNet, Faster-RCNN, etc.) to annotate images based on training data, according to a pre-defined set of semantic descriptors. With such tags, it may then be possible to answer relatively simple questions/queries e.g. whether an image contains a specific object (e.g. a laptop, a dog, a tree etc.).

5

However, it has been recognized by the inventors that there also exists an even more significant challenge in that there are queries or query topics of a higher level and therefore, it is challenging to retrieve information at higher levels, e.g. images relating to driving a car to some place, or playing tennis with somebody, or doing grocery shopping at a certain place, etc. These higher-level information/queries are typically considered as events or activities as opposed to relatively simple object identification. Such events/activities are recognized to be significantly more difficult to retrieve from image data because they involve e.g. multiple concepts and relationships.

10

15

It has been recognized that conventional technologies are not able to address the one or more problems identified above adequately. Hence, in view of the above, there exists a need for a system for image retrieval and a method of image retrieval that seek to address at least one of the problems discussed above.

20

SUMMARY

In accordance with an aspect of the present disclosure, there is provided a system for image retrieval, the system comprising an extraction module coupled to a dataset of one or more image data; the extraction module being configured to extract one or more semantic descriptors from the one or more image data; a mapping module coupled to the extraction module; a query topics module coupled to the mapping module, the query topics module being arranged to provide one or more query topics to the mapping module; wherein the mapping module is configured to receive the one or more semantic descriptors from the extraction module and to map the one or more semantic descriptors to the one or more query topics from the query topics module; further wherein the mapping module is configured to map the one or more semantic descriptors to the one or more query topics

25

30

from the query topics module based on a correlation of at least one of the one or more semantic descriptors to the one or more query topics.

5 The mapping module may be configured to receive ground truth data based on a ground truth subset of the one or more image data and to establish an initial semantic relevance map using the ground truth data and at least one of the one or more query topics.

10 The mapping module may be configured to use a training subset of the ground truth data to obtain a relevance rate of each semantic descriptor of the training subset based on the one or more query topics.

15 The relevance rate may comprise a determination of a first number of image data tagged to the each semantic descriptor determined to be positively relevant to the one or more query topics and of a second number of image data tagged to the each semantic descriptor determined to be negatively relevant to the one or more query topics.

20 The mapping module may be configured to obtain a set of node weights of aspect clusters, the aspect clusters being based on semantic descriptors of the training subset of the ground truth data.

25 The mapping module may be configured to apply the relevance rate to the rest of the ground truth data outside of the training subset, said rest of the ground truth data outside of the training subset being termed a verification set; and the application of the relevance rate to the verification set may be is to perform image data retrieval from the verification set.

30 The mapping module may be configured to determine, based on the one or more query topics, at least one of whether there is any image data from the verification set that is not retrieved and whether there is any image data from the verification set that is incorrectly retrieved.

The mapping module may be configured to expand the training subset based on the determination of at least one of whether there is any image data from the verification set that

is not retrieved and whether there is any image data from the verification set that is incorrectly retrieved.

5 The mapping module may be configured to obtain a revised relevance rate of each semantic descriptor of the expanded training subset based on the one or more query topics.

10 In accordance with another aspect of the present disclosure, there is provided a method of image retrieval, the method comprising providing a dataset of one or more image data; extracting one or more semantic descriptors from the one or more image data using an extraction module; providing one or more query topics; mapping using a mapping module the one or more semantic descriptors to the one or more query topics based on a correlation of at least one of the one or more semantic descriptors to the one or more query topics.

15 The method may further comprise receiving at the mapping module ground truth data based on a ground truth subset of the one or more image data and establishing an initial semantic relevance map using the ground truth data and at least one of the one or more query topics.

20 The method may further comprise using at the mapping module a training subset of the ground truth data to obtain a relevance rate of each semantic descriptor of the training subset based on the one or more query topics.

25 The method may further comprise determining a first number of image data tagged to the each semantic descriptor determined to be positively relevant to the one or more query topics and determining a second number of image data tagged to the each semantic descriptor determined to be negatively relevant to the one or more query topics.

30 The method may further comprise obtaining at the mapping module a set of node weights of aspect clusters, the aspect clusters being based on semantic descriptors of the training subset of the ground truth data.

The method may further comprise applying at the mapping module the relevance rate to the rest of the ground truth data outside of the training subset, said rest of the ground truth data outside of the training subset being termed a verification set; and the application of the relevance rate to the verification set being to perform image data retrieval from the verification set.

The method may further comprise determining at the mapping module, based on the one or more query topics, at least one of whether there is any image data from the verification set that is not retrieved and whether there is any image data from the verification set that is incorrectly retrieved.

The method may further comprise expanding the training subset based on the determination of at least one of whether there is any image data from the verification set that is not retrieved and whether there is any image data from the verification set that is incorrectly retrieved.

The method may further comprise obtaining at the mapping module a revised relevance rate of each semantic descriptor of the expanded training subset based on the one or more query topics.

In accordance with yet another aspect of the present disclosure, there is provided a non-transitory tangible computer readable storage medium having stored thereon software instructions that, when executed by a computer processor of a system for image retrieval, cause the computer processor to perform a method of image retrieval, by executing the steps comprising providing a dataset of one or more image data; extracting one or more semantic descriptors from the one or more image data using an extraction module; providing one or more query topics; mapping using a mapping module the one or more semantic descriptors to the one or more query topics based on a correlation of at least one of the one or more semantic descriptors to the one or more query topics.

BRIEF DESCRIPTION OF THE DRAWINGS

Exemplary embodiments of the present disclosure will be better understood and readily apparent to one of ordinary skill in the art from the following written description, by way of example only, and in conjunction with the drawings, in which:

FIG. 1 is a schematic diagram of a system for image retrieval in an exemplary embodiment.

FIG. 2 is a schematic diagram of a system framework in another exemplary embodiment.

FIG. 3 is a schematic flowchart broadly illustrating a method of constructing/updating a semantic relevance map in an exemplary embodiment.

FIG. 4 is a schematic flowchart for illustrating a method of image retrieval in an exemplary embodiment.

FIG. 5 is a schematic drawing of a computer system suitable for implementing an exemplary embodiment.

FIG. 6 is a schematic drawing of a wireless communication device suitable for implementing an exemplary embodiment.

FIG. 7 is a schematic block diagram for illustrating a system for image retrieval in an exemplary embodiment.

DETAILED DESCRIPTION

The exemplary embodiments described herein may provide a system for image retrieval and a method of image retrieval or of retrieving one or more images. One or

more query topics and one or more semantic descriptors from one or more images may be obtained. The one or more semantic descriptors may be tagged to each of the one or more images. The one or more semantic descriptors may be mapped to the one or more query topics. The mapping may be based on relevance/correlation of at least one of the one or more semantic descriptors to the one or more query topics. Retrieval of the one or more images may be based on the tagging of the one or more semantic descriptors and based on the relevance of the tagged semantic descriptors to the one or more query topics.

In the description herein, it will be appreciated that term image broadly encompasses still images, moving images (such as videos) etc.

FIG. 1 is a schematic diagram of a system for image retrieval in an exemplary embodiment.

In the exemplary embodiment, the system 100 comprises a database/dataset 102 of one or more images, videos and accompanying metadata. Such content may be termed broadly as image data. The database 102 may be provided by, for example but is not limited to, a database connected to the system 100. Such a database may also be provided via an input module into the system 100.

In the exemplary embodiment, the system 100 further comprises an extraction module 104 coupled to the database 102. In the exemplary embodiment, the extraction module 104 is configured to extract/identify features of the image data of the database 102. The features may comprise one or more semantic concepts/semantic descriptors/semantic attributes. The extraction module 104 may tag each constituent of the image data with the one or more semantic concepts/semantic descriptors/semantic attributes. For example, each image may be tagged with one or more semantic descriptors. In the exemplary embodiment, the tagging may be provided based on a suite of deep learning methods e.g. implemented within the extraction module 104. In some exemplary embodiments, the extraction module 104 may be provided in the form of a dedicated extractor device and/or in the form of a processing module.

In the exemplary embodiment, the extraction module 104 is further coupled to a mapping module 106 which is in turn coupled to a query topics module 108. In the exemplary embodiment, the query topics module 108 may be provided for a user to input one or more query topics for image retrieval from the image data of the database 102. The query topics module 108 may also be used to input one or more pre-determined training query topics for use by the mapping module 106, e.g. for training purposes. In some exemplary embodiments, the mapping module 106 may be provided in the form of a dedicated mapper device and/or in the form of a processing module. In some exemplary embodiments, the query topics module 108 may be provided in the form of an input device and/or coupled to a database of pre-determined query topics.

In the exemplary embodiment, the mapping module 106 is configured to provide semantic relevance mapping 110 wherein the mapping module 106 constructs a relevance map and updates the relevance map with the relevance map indicating a correlation/relevance between the one or more semantic concepts/semantic descriptors/semantic attributes extracted/identified by the extraction module 104 with each (or a next) query topic received at the query topics module 108.

In the exemplary embodiment, the mapping module 106 may be further configured to provide feature weighting 112 wherein the one or more semantic concepts/semantic descriptors/semantic attributes are collectively assigned weights to predict the extent/level to which an image is related (or relevant) to each query topic received at the query topics module 108. For example, different categories or groups of semantic descriptors may be collectively assigned weights for the predicted relevancy.

In the exemplary embodiment, the mapping module 106 may be further configured to provide fine-tuning and/or smoothing (see numeral 114) wherein the predicted relevance level of the constituents of the image data to each query topic is further filtered/improved in a temporal smoothing process to enhance semantic coherence. The determined relevance level ("Search Result" in FIG. 1) may be used to update the correlation established during semantic relevance mapping 110 and to update the constructed relevance map.

In the exemplary embodiment, based on the above, a systematic image retrieval procedure/method may be provided that integrates visual analytics methods comprising, for example, three main steps that may be implemented with the system 100.

In the exemplary embodiment, the procedure/method may comprise, as a first main step, semantic extraction (implemented via the extraction module 104) wherein images (e.g. lifelog images) are tagged with respect to a set of semantic descriptors based on a suite of deep learning methods.

In the exemplary embodiment, the procedure/method may further comprise, as a second main step, relevance mapping (implemented via the mapping module 106) wherein a correlation table is built e.g. by learning positive and negative cases of correlation using ground truth data. The correlation table may specify which semantic descriptors are relevant to query topics.

In the exemplary embodiment, the procedure/method may further comprise, as a third main step, feature weighing (implemented via the mapping module 106) wherein categories of semantic descriptors are collectively assigned weights to predict how much (or the extent/level) an image is related/relevant to a query topic. The predicted relevance level of an image may be further filtered in a temporal smoothing process to enhance semantic coherence (also implemented via the mapping module 106).

The technical features of the above image retrieval method may comprise an integrated image search process model that combines deep learning and semantic mapping and linear regression for concept-based image retrieval. FIG. 2 illustrates an exemplary system framework in this regard.

FIG. 2 is a schematic diagram of a system framework in another exemplary embodiment. In this exemplary embodiment, the system framework provides a system 200 that functions substantially similarly to the system 100 described with reference to FIG. 1.

In the exemplary embodiment, semantic concepts, descriptors and attributes may be terms that are used interchangeably.

In the exemplary embodiment, the system 200 comprises a database 202 to provide image data to the system 200. In the exemplary embodiment, the image data is, for example but is not limited to, lifelog data comprising one or more lifelog images and accompanying lifelog metadata. In the exemplary embodiment, the lifelog data stored in the database 202 may be represented by a tuple (m_i, n_i) , where m_i is an image and n_i is the metadata, for example as indicated in FIG. 2.

In the exemplary embodiment, the system 200 further comprises a semantics extraction module 204 coupled to the database 202. In the exemplary embodiment, the semantics extraction module 204 is configured to extract/identify features of the lifelog data stored in the database 202. The semantics extraction module 204 may tag the lifelog data with one or more semantic descriptors.

In the exemplary embodiment, based on offline analysis using deep learning and data cleaning, each constituent or image of the lifelog data may be described by an activation vector p_{ij} with respect to a set of semantic descriptors d_j . The descriptors may be divided into sub-categories $(D_1, D_2, \dots, D_N)$ according to the deep learning models used to extract the features. For example, as indicated in FIG. 2, a semantic descriptor D_1 may be an object; D_2 may be a particular place; D_3 may be a human; D_4 may be an indicator of image quality; D_{N-2} may be metadata for time associated with an image; D_{N-1} may be metadata for location associated with an image; D_N may be metadata for an activity associated with an image etc.

In the exemplary embodiment, two types of semantic descriptors may be extracted from two sources. It will be appreciated that the exemplary embodiment is not limited as such and may be expanded to include other types of possible descriptors.

In the exemplary embodiment, a first type of semantic descriptors is directly obtained from lifelog metadata (see numeral 203). Depending on the format and the quality of the

data, as well as the task requirements, the original metadata from the database 202 may be cleaned before extraction of semantic information/descriptors such as time stamps (e.g. in hours etc.), location (e.g. work, home, church, etc.), activity (e.g. walk, run, transport, etc.). In the exemplary embodiment, these descriptors may be denoted as 0/1 vectors in-line with
5 the lifelog images.

In the exemplary embodiment, a second type of semantic descriptors is obtained via image-based retrieval, where lifelog images are either tagged with semantic concepts based on deep learning (e.g. object, place, human, etc.) or characterized by low level features (e.g.
10 image quality). See numeral 205.

In the exemplary embodiment, Convolutional Neural Networks (CNN)-based classifiers and detectors are adopted/used to translate each lifelog image into a set of feature vectors, with each element representing the probability that, for example, an
15 object/scene exists in the image. Prediction and detection may be used with the classifiers and detectors. The CNN models may include, but are not limited to, an object-centric classifier pre-trained on for example available image databases such as ImageNet1K, a scene-centric classifier pre-trained on for example available scene recognition image databases such as Places365, an object detector pre-trained on for example Microsoft
20 Common Objects in Context (MSCOCO), and any suitable hybrid classifiers fine-tuned on additional visual concepts.

For CNN classifiers, CNN models may be used to predict objects and places depicted in lifelog images. In the exemplary embodiment, the CNN models may be pre-
25 trained on image datasets, such as ImageNet1K and Places365. At the time of writing, ImageNet1K is a dataset with 1.2 million images, each annotated according to 1000 object classes, and Places365 comprises 1.8 million images, each tagged against 365 place categories.

In the exemplary embodiment, a residual neural network is further used. For
30 example, a ResNet with 152 layers may be pre-trained on each of the above datasets ImageNet1K and Places365 respectively, which are referred to as ResNet152-ImageNet1K

and ResNet152-Places365. In the exemplary embodiment, if a lifelog image is passed through ResNet152-ImageNet1K, the image results in obtaining a one thousand dimensional probability vector. If a lifelog image is passed through ResNet152-Place365, a 365 dimensional probability vector may be extracted to predict place information. Both vectors are extracted from the last layer of the network. In the exemplary embodiment, data augmentation may be performed to generate scaled and rotated versions for each lifelog image. Further, instead of average operation, the maximum activation value may be chosen for each class.

Apart from CNN classifiers, for CNN detectors, object detection may be performed to locate objects in lifelog images. In the exemplary embodiment, a region-CNN such as a Faster R-CNN with a feature/pattern recognition network such as Inception-ResNet may be adopted as the base CNN architecture, and which is pre-trained on a MSCOCO training set. In the exemplary embodiment, when Faster R-CNN is used, a lifelog image may be annotated with the top 20 detections based on the maximum probability for each category. It will be appreciated that other values are also possible. For example, based on the current feature space of MSCOCO, it is possible to annotate with 1 to 80 detections.

In the exemplary embodiment, it may be useful to further enhance the semantic description by including task-specific classifiers. In the exemplary embodiment, for task-specific classifiers, some datasets may provide task-specific classifiers where a set of concepts (or semantic descriptors) is annotated for part or all the lifelog images.

In the exemplary embodiment, still referring to numeral 205, for image quality assessment (e.g. for tagging with semantic concepts characterized by low level features), a modified laplacian method and variance of laplacian may be used for blurriness assessment.

In the exemplary embodiment, the semantics extraction module 204 is coupled to a mapping module 206. The mapping module 206 implements relevance mapping (shown at dotted box 207). The relevance mapping may be implemented with feature weighting (shown at dotted box 209).

The mapping module 206 is configured to provide semantics relevance mapping wherein the mapping module 206 constructs and updates a semantics relevance map R_{jk} at box 207. The semantics relevance map R_{jk} may indicate a correlation between the one or more semantic descriptors extracted/identified by the semantics extraction module 204 with one or more query topics received by the system 200. In the exemplary embodiment, the mapping module 206 may be configured to receive and use a set of ground truth G_{ik} to construct/update the semantics relevance map R_{jk} (see provided ground truth 210). In some exemplary embodiments, the ground truth 210 may be provided via an annotation module or device usable by a user to annotate data. The annotation module may be used in an offline mode or an online mode.

In the exemplary embodiment, the system 200 further comprises a query topics module 208 configured to provide one or more query topics as an input. In the exemplary embodiment, query topics may be defined as purpose-driven high-level semantic concepts (e.g., "Waiting in an airport lounge.").

In the exemplary embodiment, the mapping module 206 may be further configured to provide/perform feature weighting wherein the one or more semantic concepts/semantic descriptors/semantic attributes, e.g. under different categories, are collectively assigned weights to predict the extent/level to which a lifelog image is related to the query topic q_k . The mapping module 206 may be further configured to provide fine-tuning and/or smoothing wherein the predicted relevance level of the one or more lifelog images to the query topic q_k is further filtered/improved in a temporal smoothing process to enhance semantic coherence. The final/determined relevance level may be used to update the correlation established during semantic relevance mapping.

In the exemplary embodiment, the feature weighting (at box 209) may be performed because the inventors recognise that different types of features may contribute differently e.g. to information/image retrieval. In the exemplary embodiment, feature importance/relevance may be learned with statistical modelling such as using a Conditional Random Field (CRF) model to weigh the contributions of different types of features for individual event queries (or query topics q_k). Examples of the types of features may include

ImageNet1K, Places365, MSCOCO, location, time and the number of people. For each of these feature types, there may be available relevant concepts “R”, (e.g. *food* in *Eating Lunch*), and concepts to be avoided, “Avoid”, (e.g. *kitchen* in *Hiking*).

5 For an exemplary formulation, there is one node per feature. The unary is defined as $\phi_U(s_i) = \text{mean}(\text{score}_{rel}[G_{ik} = s_{ij}])$ for $s_i = \{0, 1\}$, where G_{ik} is an annotation for whether each image corresponds to the task. Broadly, there is provided a mean relevance score to $G_{ik}=1$ for a relevant correspondence to a concept or to $G_{ik}=0$ for a non-correspondence to a concept. The pairwise potentials are defined to enforce that the nodes activation values be
10 positive.

In the exemplary embodiment, since the similarity/relevance of an image to an event may be derived without considering adjacent image frames, temporal smoothing may be desirably incorporated into the system framework (i.e. of system 200) to refine the similarity,
15 with an assumption that adjacent images are with semantic coherence. The similarity between an image and an event may be smoothed using a triangular window of size W , which may be adaptive to event topics. In the exemplary embodiment, a search such as a greedy search may be performed to find the sub-optimal value of W , by testing the retrieval performances on the manually annotated lifelog images e.g. ground truth 210. As an
20 example, the relevance of an identified image may be adjusted based on its adjacent images for semantic coherence, with the number of adjacent images being determined based on the sub-optimal value of W .

In the exemplary embodiment, the system 200 may implement image/information
25 retrieval, for example, for lifelog data or lifelog images.

In the exemplary embodiment, to further map the symbol-level semantic attributes to higher level activities and events, the mapping module 206 may be configured to implement an automatic semantic relevance mapping (ASRM) process that may usefully provide
30 information mapping between semantic concepts/semantic descriptors/semantic attributes and events/search topics (or query topics).

As described above, a lifelog image may be tagged with one or more semantic descriptors. It has been recognised by the inventors that some semantic concepts/descriptors/attributes are more relevant to a query topic/event than others. For example, a "cup" may be relevant to the event "drinking coffee" whereas a "tree" may not be.

5 The inventors also recognise that relevance may be negative, meaning that the activation of a semantic concept/semantic descriptor/semantic attribute is negatively related or unfavorable to an event. This may happen when several semantically related attributes are co-activated, and some of the attributes contribute negatively to a respective event. For example, the query "shopping for a TV set" may return results contaminated by laptop
10 computers. That is, there may be false positives attributed to "laptop". In the exemplary embodiment, it is useful to exclude those images that have high activation of "laptop". As another example, the topic "working on the computer" may be closely related to "screen". However, other attributes that may appear to be relevant, e.g. "television", may be activated because "television" is closely related to "screen", leading to false alarms/false positives in
15 the information/image retrieval. In this example, the activation of attribute "television" is recognised to negatively affect the retrieval result. In the exemplary embodiment, semantic-aversion may be initiated explicitly by a user, e.g., the user may configure the system 200 (via configuring the mapping module 206) to retrieve photos/images related to "lecturing in the classroom", while excluding those showing the "projector screen".

20 In the exemplary embodiment, the mapping module 206 may be configured so that a constructed semantic relevance map may be used to specify a relationship/correlation in the form of a two-dimensional matrix. In the exemplary embodiment, the two-dimensional matrix may be referred to as an ASRM matrix which represents the relationship/correlation between
25 generic concepts/descriptors/attributes and purpose-driven query topics/events. The elements of the ASRM matrix may be a numeric value between -1 and 1, where 1 indicates a positive relation, -1 indicates a negative relation, and 0 indicates no relation. The value may be further discretized to {-1,0,1} with a pre-defined threshold. Thus, the function of the ASRM matrix may be an analogy or be similar to a human cognitive process in retrieval of
30 memory. For example, there are three cues for goal-driven retrieval of memory, namely, relevant, non-relevant and inhibitive cues. Relevant cues enhance the recall of past experiences, non-relevant cues typically have less effect on the recall while inhibitive cues

induce a suppression on the recall.

In this exemplary embodiment, the ASRM process/method (which may be implemented via the mapping module 206) is capable of programmatically searching relevant concepts related to given events/topics.

An exemplary method of constructing/updating the semantic relevance map R_{jk} (at box 207) is described with reference to FIG. 3.

FIG. 3 is a schematic flowchart broadly illustrating a method/concept of constructing/updating a semantic relevance map R_{jk} in an exemplary embodiment. In the exemplary embodiment, the mapping module 206 of the system 200 described with reference to FIG. 2 may implement the method 300. Like numerals are used for exemplary implementations of modules/elements as described in FIG. 2.

In the exemplary embodiment, at step 302, a subset of lifelog images are annotated with event labels, serving as ground truth to validate the effectiveness of relevant concepts searching. Ground truth is understood to be data that is confirmed or validated to be correct.

In the exemplary embodiment, the ground truth may not have to be made available prior to the search (or offline mode). Rather, the ground truth may be established in an interactive process e.g. having a human user label "Yes" (i.e. image retrieved is relevant to query topic/event) or "No" (i.e. image retrieved is not relevant to query topic/event) to the retrieved subset of images (or online mode). As such, the ground truth (i.e. whether an image is relevant to a query topic) for a subset of the lifelog data may be obtained through manual annotation, and is denoted as $G_{ik} \in \{0,1\}$, where 1 indicates m_i is related to a query topic q_k , and 0 otherwise. Compare ground truth 210 of FIG. 2.

In the exemplary embodiment, concepts (or semantic descriptors) with higher activation to an event (or topic) may have greater impact (positively or negatively) on the retrieval of images/information. For example, certain descriptors may be positively relevant to a topic while other descriptors may be negatively relevant to that topic. Therefore, in the

exemplary embodiment, a sub-set of the descriptors with high activation levels may be considered to be main factors. For each semantic descriptor, lifelog images may be retrieved that are annotated with a given event (true positives) (see step 304A), and the average activation of the corresponding descriptors of the images may be computed as p_rel (see step 306A). On the other hand, if images irrelevant to a topic are retrieved (false positives) (see step 304B), the average activation of the corresponding descriptors are computed as n_rel (see step 306B).

At step 308, the positive and negative activation are combined to compute the relevance/correlation of a descriptor to a topic. This may enable the semantics relevance map to be constructed/updated. The above steps 304A, 304B, 306A, 306B are repeated for a next topic for each descriptor.

In the exemplary embodiment, the numeric relevance value of step 308 may be discretised to $\{-1,0,1\}$ with a pre-defined threshold. For example, the threshold may be notated as θ_+ and θ_- . The sub-optimal threshold may be determined in a search process (e.g. by a greedy search), by testing the retrieval performance over the annotated training set (see step 302). In various exemplary embodiments, different thresholding strategies may be adopted which may lead to different computational loads and performance. The different thresholding strategies may include setting a (1) fixed threshold, (2) threshold adapted to a dataset (e.g. for a particular user), and/or (3) threshold tailored for a dataset and one or more query topics.

It is recognised that a database or dataset of image data (e.g. at numerals 102 of FIG. 1 and 202 of FIG. 2) may comprise a plurality of smaller datasets and these may respond to the different thresholding strategies described above.

An exemplary implementation is further described.

Lifelog information/data retrieval may begin from a query topic q_k , which in the exemplary implementation is defined as a purpose-driven high-level semantic concept (e.g., "Waiting in an airport lounge.").

Given a vector of semantic attributes/ descriptors $d = [d_j]_{j=1}^M$ and a query topic q_k , the ASRM process (e.g. implemented at mapping module 206 of FIG. 2) may specify the relationship/relevance between d and q_k in the form of a vector, $r = [r_j]_{j=1}^M$, where $r_j \in [-1, 1]$ denotes whether an attribute d_j is relevant to the current topic, i.e., 1 denoting relevant, -1 denoting inhibitive (negative), and 0 denoting non-relevant (i.e. the relevance between semantic descriptors and a query topic q_k is specified by the relevance map $R = r_{jk}$, where $-1 \leq r_{jk} \leq 1$). In the exemplary implementation, the ASRM vector may be expressed by a discretized r_j on a pair of thresholds (i.e. discretized by applying a threshold), which is a tuple denoted as $\theta = (\theta_+, \theta_-)$, whereby

$$r_{ik}^* = \begin{cases} 1, & \text{if } r_{ij} \geq \theta_+ \\ -1, & \text{if } r_{ij} \leq \theta_- \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

In the exemplary implementation, for lifelog applications, a set of query topics, denoted as $q = [q_k]_{k=1}^K$ may be established. By extending the relevance vector for all the K topics, a discretized ASRM matrix (or the discretized semantic relevance map) may be established and represented/denoted as $R^* = [r]_{jk}^*$.

In the exemplary implementation, the dataset of lifelog data is passed through a suite of semantic extraction modules, including but not limited to, deep learning methods or networks. For example, eight different types of networks may be used, leading to eight different semantic aspects or categories/groups for the lifelog data. In addition, it will be appreciated that any number of networks may be used.

The aforementioned semantic aspects $\{\delta_s\}_{s=1}^8$ (where “8” represents the exemplary total number of semantic aspects in the exemplary implementation) may have varying contributions to the query topics, e.g. some queries may be place-sensitive, some may be time-sensitive, and some may be human-sensitive etc. it is recognised that the different networks used for the semantic aspects may provide such varying contributions.

In the exemplary implementation, feature aspect weighting may account for such sensitivity by assigning different weights to the attribute aspects or semantics aspects/categories. With the ASRM vector r_k^* (see equation 1), the semantic attributes are associated with query topic q_k . In the implementation, r_k^* may be separated into two parts r_k^+ and r_k^- , which represent the ASRM rates for relevant and inhibitive (negative) attributes, respectively. The semantic attributes of each aspect cluster or semantics aspects/categories may thus be divided into a relevant group (δ_s^+) and an inhibitive group (δ_s^-) (and non-relevant attributes may be dropped). Using such formulation, there are two nodes for each aspect cluster, one for relevant, and the other for inhibitive attributes. The node activation levels of an aspect cluster may thus be computed as:

$$\begin{cases} \Delta_s^+(k) = \frac{1}{D_s} \sum_{d_j \in \delta_s^+} (r_{jk}^+) \\ \Delta_s^-(k) = \frac{1}{D_s} \sum_{d_j \in \delta_s^-} (r_{jk}^-) \end{cases} \quad (2)$$

where D_s represents the dimension of the aspect cluster

In the exemplary implementation, to characterize the contributions of semantic aspects to the association with a query topic, a statistical modelling method such as a CRF method is applied to learn the weights for aspect clusters for each query topic. In the exemplary implementation, the weights for each node of the semantic aspects are obtained by using the MAXFLOW software tool (e.g. for computation of the so-called mincut/maxflow algorithm) on training samples. The input to the software tool is the node activation levels of aspect clusters and the correct query topic label of each image in a training set e.g. obtained from ground truth. The output is the node weights of the aspect clusters for each query topic, i.e., $\{w_s^+(k), w_s^-(k)\}$ for $\{\delta_1, \dots, \delta_S\}$. Compare box 209 of FIG. 2.

In the exemplary implementation, to learn/update the ASRM matrix R^* , a subset of lifelog images are randomly selected and their relevance to provided query topics are annotated, for example by manual annotation. These images are used to form a Learning

Set, where each image is relevant to just one topic. Compare the Learning Set to ground truth as described above.

The activation of semantic attributes may be expressed as a $M \times N$ matrix $A = [a_{ij}]_{i=1}^N$, which may be called attribute activation matrix. In the exemplary implementation, an iterative learning procedure is applied on the Learning Set. For a topic q_k , the Learning Set is randomly split into two subsets: a training set I_t and a verification set I_v . The training set is as referred to above with regard to the node weights of the aspect clusters.

The two subsets I_t and I_v are further divided into positive sets and negative sets: $I_t = I_t^+ \cup I_t^-$ and $I_v = I_v^+ \cup I_v^-$, where I_t^+ and I_v^+ are composed of samples relevant to q_k , and I_t^- and I_v^- are composed of samples not relevant to q_k .

In each iteration, for each semantic attribute d_j , its relevance rate to topic q_k is computed as

$$r_{jk} = \alpha_1 \left(\frac{1}{P_k} \sum_{n=1}^{P_k} a_j^{(n)} \right) + \alpha_2 \left(\frac{1}{N_k} \sum_{n=1}^{N_k} a_j^{(n)} \right) \quad (3)$$

Where P_k is the number of positive samples in I_t^+ , N_k is the number of negative samples in I_t^- , α_1 and α_2 are pre-defined weights of positive and negative samples, respectively. For example, one may set $\alpha_1 = \alpha_2$ and $\alpha_1 + \alpha_2 = 1$.

As described, the computation of r_{jk} is based on the training set I_t . Each semantic attribute d_j is independent of the dataset, i.e. the same set of semantic descriptors d_j is used for both the training set I_t and the verification set I_v .

Next, the relevance rate r_{jk} is transformed into a discrete value r^* according to equation (1). In various exemplary embodiments, both θ_+ and θ_- may be initialized arbitrarily and fine-tuned based on e.g. a greedy search. Generally, it is considered the smaller the

range of the threshold between θ_+ and θ_- , the wider the results may be. For example, comparing equation (1), r_{ij} becomes relevant when larger than or equal to θ_+ or smaller than or equal to θ_- .

5 In this exemplary implementation, a threshold of $\theta_+ = \theta_- = 0.05$ is adopted. This results in an ASRM vector for topic q_k on the training set $I_t = I_t^+ \cup I_t^-$. Next, the node weights of aspect clusters $\{w_s^+(k), w_s^-(k)\}$ are obtained on the training set $I_t = I_t^+ \cup I_t^-$ and the ASRM vector/relevance map may be updated. Subsequently, the learned/updated ASRM vector is applied to the samples in the verification set $I_v = I_v^+ \cup I_v^-$ to perform lifelog
10 image retrieval. If there are images in I_v^+ which are not retrieved i.e. missed, some of the missed images are moved from I_v^+ to I_t^+ , and if there are images in I_v^- which are falsely retrieved i.e. false positives, some of those images are moved from I_v^- to I_t^- . In certain limited scenarios, it is possible that all the missed and/or false positive images are moved. In the exemplary implementation, the learning iteration is applied again on the extended
15 training set I_t .

In the exemplary implementation, the learning process/algorithm is terminated when there is no more missed or false positive sample(s), or if there is no performance gain by the above iterating updating process. After performing the above process for all the topics, an
20 ASRM matrix $R^* = [r^*]$ and final weights for aspect clusters $W^* = [w^+(k), w^-(k)]$ is then obtained. Compare box 209.

A procedure of image retrieval with the ASRM process is shown in an example Algorithm 1 below. One main step is to find and retain activated descriptors where correct
25 retrieval is rewarded and false positives are penalized. In the example, the procedure is iterated with different thresholds of activation to achieve overall suboptimal retrieval performance. The alphabets A to M appended to the steps shown in the Algorithm is meant for notation only and are not part of the Algorithm.

Algorithm 1: Automatic semantic relevance mapping**Input:** Feature vector d_j , Ground truth G_{ik} , Query topic q_k --- (A)**Output:** Discretized relevance map R^* , Feature weight w ,
Retrieval result $Arg Max_K(p_{ik}) \geq \zeta$ --- (B)**Initialize:** From ground truth, retrieval samples of images
for a query task, assign mean activation value to the
relevance map: $p_{rel} = mean(p_{ik} | m_{ik} = G_{ik} = 1)$ --- (C) $R(t=0) = \{r_{jk}^0\} = \alpha_1 \cdot p_{rel}$ --- (D)Discretize: R^* according to Eq. (1) --- (E)for $\theta \leftarrow \theta_1$ to θ_T dofor $k \leftarrow 1$ to N do

$$\sum D_k^+ = \sum_{d_j \in \Sigma D_k^+} (r_{ik} \cdot p_{ik})$$

$$\sum D_k^- = \sum_{d_j \in \Sigma D_k^-} (r_{ik} \cdot p_{ik})$$
 --- (F)

Perform feature weighting using CRF, get $w(t)$. --- (G)

Perform temporal smoothing. --- (H)

Retrieve images: $p_{ik} \geq \zeta$; Compute precision score

$$P^t, P_{max} = Max\{P_{max}, P^t\}$$
 --- (I)

If $\exists(p_{ik} < \zeta)$ AND $m_{ik} = G_{ik} = 1$ then // Exists missed positivesRandomly select X samples from positives thatwere not retrieved $p_{ik}^X < \zeta$.

$$p_{rel} = mean(p_{ik}^X | m_{ik} = G_{ik} = 1)$$
 --- (J)

If $\exists(p_{ik} < \zeta)$ AND $m_{ik} = G_{ik} = 0$ then // Exists false
positivesGet top X false positives $p_{ik}^X > \zeta$.

$$n_{rel} = mean(p_{ik}^X | m_{ik} = G_{ik} = 0)$$
 --- (K)

$$Update R(t): r_{jk}^t = \alpha_1 \cdot p_{rel} - \alpha_2 \cdot n_{rel}$$
 --- (L)

Update $R(t)^*$ according to Eq. (1) with θ_t --- (M)

Algorithm 1 is explained below.

5 For Algorithm 1, at (A), the inputs used are a feature/descriptor/attribute vector d_j , ground truth G_{ik} and query topic q_k . The outputs from implementing Algorithm 1 are the discretized relevance map R^* , feature weight w , and the retrieval result $Arg Max_K(p_{ik}) \geq \zeta$ (see (B)).

10 In Algorithm 1, for initialisation, a sample of images are retrieved from ground truth for a query task and a mean activation value is assigned to a relevance map (see (C)). The sample of images make up a training set and a verification set. At (D), the relevance map is established and discretized according to equation (1) (see (E)). Compare also FIG. 3.

15 At (F), the dimensions for the feature vector are calculated based on initial threshold values θ . Compare equation (2) for a scenario of an aspect cluster. Subsequently, at (G),

feature weighting is performed using CRF to obtain feature weights and temporal smoothing is performed thereafter (see (H)).

At (I), images are retrieved from the combined training and verification set, and a precision score is computed. Images that are retrieved are compared to the verification set. See above description of $I_V = I_V^+ \cup I_V^-$.

At (J), it is determined if there are any missed positives (i.e. images that are relevant to the query topic that were not retrieved). If there exists missed positives, an arbitrary number of images that are missed positives is randomly selected and the mean activation value to the relevance map is adjusted accordingly (see (J)).

At (K), it is determined if there are any false positives (i.e. images that are not relevant to the query topic but were retrieved). If there exists false positives, an arbitrary number of images that are false positives are selected and the mean activation value to the relevance map is adjusted accordingly (see (K)). To select the false positives, one may first rank the false positives according to their activation levels, and select an arbitrary top few (X) ones for adjusting the relevance map. Random selection may also be implemented to enhance diversity.

Based on the results obtained at (J) and at (K), the relevance map (of (D)) is updated (see (L)). Compare also FIG. 3.

At (M), the discretized relevance map is iteratively updated according to equation (1) with varying threshold values θ and across all query topics (i.e. k from 1 to N).

The method described in the exemplary embodiments discussed above was evaluated using a dataset distributed by the National Institute of Informatics (NII), Japan, under the NTCIR-13 Liflog-2 conference. The dataset was contributed by two users (denoted as u1 and u2). u1 had 91,044 lifelog images and u2 had 20,471 lifelog images, both sets of images accompanied with metadata. To test the effectiveness of the method of image retrieval, two sets of retrieval tasks were attempted.

The first set (called LIT), had 10 query topics (see Table 1 below), and the performance using the described method was compared against a baseline where a semantic relevance matrix was constructed manually. The second set (called LSAT) had 20 query topics (see Table 2 below). For the LSAT, not all topics listed are applicable to both users. The semantic relevance mapping was automatically generated using the ASRM process and different strategies were applied to set the thresholds $\theta = (\theta_+, \theta_-)$, namely, (1) fixed (θ does not change), (2) adaptive to user (θ is different for two users), and (3) adaptive to user and query topics (θ is set for each user and each query topic).

Table 1: Query topics in LIT set.

Event Topics	Description (both users)
Eating	User is consuming food
Walking	User is walking
Running	User is running
Hiking	User is hiking
Gym/Yoga	User is exercising in a gym or practicing yoga
Socializing	User is interacting with others
Bus	User is taking a bus
Car/Taxi	User is driving or taking taxi
Train	User is taking train
Plane	User is in an aero plane

Table 2: Event topics in LSAT set.

Event Topics	User 1	User 2
Eating Lunch	Y	Y
Gardening	Y	
Castle at Night	Y	

Coffee	Y	
Sunset	Y	
Graveyard	Y	
Presenting/Lecturing	Y	
Grocery Shopping	Y	Y
Working Late	Y	Y
On the Computer	Y	Y
Cooking	Y	Y
Flying	Y	Y
Fruit or Vegetable Juice	Y	
Photo of the Sea	Y	
Having Beers in a Bar	Y	
Television Recording	Y	
Working in a Coffee Shop	Y	
Painting Walls	Y	
Exercises at home	Y	
Benbulbin Mountain	Y	

To identify relevant concepts, feature weights and temporal smoothing parameters for event topics, a subset of lifelog images were manually annotated with these topics. Specifically, 22,304 lifelog images (18,209 for user 1, and 4,095 for user 2) were sampled, which is approximately 20% of the entire dataset. Through random sampling, about half of the relevant images were selected for training and the remainder were selected for testing/verification.

The results of the LIT retrieval is shown in Table 3 below. The evaluation metric is mean Average Precision (mAP) across the tasks/topics. The results in Table 3 show that the ASRM process/method outperformed the manual method by an increase of approximately 11% for u1 and approximately 33% for u2.

Table 3: Performance in simple tasks: comparing manual and ASRM method in setting semantic relevance map.

	User 1	User 2
Manual	0.260	0.309
ASRM	0.291	0.411

5

For the 20 LSAT query topics, relevance mapping was not constructed manually since it is too time consuming to do so. Only the ASRM process was used. In this evaluation, the effect of thresholds for relevant concepts searching using ASRM was explored. Three thresholding configurations ((1) fixed (θ does not change), (2) adaptive to user (θ is different for two users), and (3) adaptive to user and query topics (θ is set for each user and each query topic)) were tested.

As shown in the Table 4 below, both configurations (2) and (3) outperformed than using fixed thresholds, i.e. configuration (1). Further, as shown in Table 4, configuration (3) outperformed configuration (2) by a large margin.

15

Table 4: Effect of thresholds for relevant concepts searching

	User 1	User 2
Fixed	0.502	0.748
Adaptive to User	0.528	0.761
Adaptive to User and Event	0.654	0.826

20

It is noted that the above performance evaluation was conducted on the annotated data (serving as ground truth), which is recognised to be relatively small.

A formal evaluation was performed by the NTCIR-13 organizer on the entire dataset for the 20 tasks in LSAT. The official evaluation evaluated the number of events detected in

a given day (compared to the ground truth) as well as the accuracy of the event-detection process (given a sliding five minute window, e.g. over the set of images which were to be detected). The used metrics were precision and recall, and the official score was based on the mean of precision over the topics. It was evaluated that the described method achieved
 5 an official score of 57.6%, which ranked at first place in the benchmarking.

FIG. 4 is a schematic flowchart for illustrating a method of image retrieval in an exemplary embodiment.

10 At step 402, a dataset of one or more image data is provided. At step 404, one or more semantic descriptors is extracted from the one or more image data e.g. using an extraction module. At step 406, one or more query topics is provided. At step 408, the one or more semantic descriptors is mapped, e.g. using a mapping module, to the one or more query topics based on a correlation of at least one of the one or more semantic descriptors
 15 to the one or more query topics.

FIG. 7 is a schematic block diagram for illustrating a system for image retrieval in an exemplary embodiment.

20 The system 700 comprises an extraction module 702 coupled to a dataset of one or more image data, the extraction module being configured to extract one or more semantic descriptors from the one or more image data. For example, the one or more semantic descriptors are provided as $d = [d_j]_{j=1}^M$. For example, the dataset of one or more image data may be provided by an external database 703. The system 700 also comprises a mapping
 25 module 704 coupled to the extraction module 702. The system 700 further comprises a query topics module 706 coupled to the mapping module 704, the query topics module 706 being arranged to provide one or more query topics to the mapping module 704. For example, the one or more query topics are provided as $q = [q_k]_{k=1}^K$. The mapping module 704 is configured to receive the one or more semantic descriptors from the extraction
 30 module 702 and to map the one or more semantic descriptors to the one or more query topics from the query topics module 706. The mapping module 704 is configured to map the one or more semantic descriptors to the one or more query topics from the query topics

module 706 based on a correlation of at least one of the one or more semantic descriptors to the one or more query topics. For example, the correlation is in terms of the vector $r = [r_j]_{j=1}^M$. Across all K topics, the vector may be discretised as $R^* = [r]_{jk}^*$. In the description, the phrase “one or more” or “at least one” is intended to cover both the singular and a plurality. For example, a semantic descriptor may be mapped to a query topic or to a plurality (two or more) of query topics. For example, a plurality of semantic descriptors may be mapped to a query topic or to a plurality of query topics.

In the exemplary embodiment, the mapping module may receive ground truth data based on a ground truth subset of the one or more image data and may establish an initial semantic relevance map using the ground truth data and at least one of the one or more query topics. For example, compare FIG. 3. For example, the ground truth data may be provided by an annotation module 708.

The mapping module may be configured to use a training subset of the ground truth data to obtain a relevance rate of each semantic descriptor of the training subset based on the one or more query topics. The relevance rate may comprise a determination of a first number of image data tagged to the each semantic descriptor determined to be positively relevant to the one or more query topics and of a second number of image data tagged to the each semantic descriptor determined to be negatively relevant to the one or more query topics. For example, compare equation (3).

In the exemplary embodiment, the mapping module may obtain a set of node weights of aspect clusters, the aspect clusters being based on semantic descriptors of the training subset of the ground truth data. For example, the set of node weights may be provided by a feature weighting module 710.

In the exemplary embodiment, the mapping module may be configured to apply the relevance rate to the rest of the ground truth data outside of the training subset, said rest of the ground truth data outside of the training subset being termed a verification set; and the application of the relevance rate to the verification set being to perform image data retrieval

from the verification set. For example, the relevance rate may be discretised to r^* according to equation (1) and applied on a verification set I_v .

In the exemplary embodiment, the mapping module may be configured to determine, based on the one or more query topics, at least one of (1) whether there is any image data from the verification set that is not retrieved and (2) whether there is any image data from the verification set that is incorrectly retrieved. Based on the determination, the mapping module may be configured to expand the training subset. For example, if there are images not retrieved in I_v^+ , they are moved from I_v^+ to I_t^+ and if there are images in I_v^- which are incorrectly or falsely retrieved, they are moved from I_v^- to I_t^- .

In the exemplary embodiment, the mapping module may be configured to obtain a revised relevance rate of each semantic descriptor of the expanded training subset I_t based on the one or more query topics. For example, iterations are performed and after completion of the iterations, an ASRM matrix $R^* = [r^*]$ and weights for aspect clusters $W^* = [w^+(k), w^-(k)]$ may be obtained.

In another exemplary embodiment, a non-transitory tangible computer readable storage medium having stored thereon software instructions may be provided. The instructions, when executed by a computer processor of a system for image retrieval, may cause the computer processor to perform a method of image retrieval, by executing the steps comprising providing a dataset of one or more image data; extracting one or more semantic descriptors from the one or more image data using an extraction module; providing one or more query topics; mapping using a mapping module the one or more semantic descriptors to the one or more query topics based on a correlation of at least one of the one or more semantic descriptors to the one or more query topics. The computer processor may also execute any steps that have been described in the present disclosure, and these steps may be stored as software instructions or code on the storage medium.

In yet another exemplary embodiment, it may be provided that a system for image retrieval be provided at a remote location. A user may use a query module/component such as, but not limited to, via a Graphical User Interface on a wireless device to provide one or

more query topics to the system. A dataset of one or more image data may be provided to the system at the remote location, for example from another location or from the user via the wireless device. The system may perform the mapping based on the dataset of one or more image data and the one or more query topics. The system may then provide a retrieval, such as an output of one or more retrieved image data, to the user e.g. via the wireless device.

In the described exemplary embodiments, with an established semantic relevance map R^* , it is possible to provide a value of the threshold and obtain/retrieve a set of image data based on the relevance/correlation of the semantic descriptors to the provided threshold value and the activation vector p_{ij} of each image to the semantic descriptors. For example, one or more semantic descriptors may have been extracted from a dataset of image data and training may be performed (e.g. using a subset of the image data) using predetermined query topics that may include "grocery", "shopping", a location "place A" etc. The relevance/correlation value/score of the semantic descriptors may be taken into account with the predetermined query topics (e.g. negatively relevant, positively relevant), and R^* is established. Based on provision of a desired threshold pair of $\theta = (\theta_+, \theta_-)$, image retrieval may be performed according to a query topic "doing grocery shopping at a certain place A". For example, based on the relevance or value/score, images (and their surrounding images and/or associated aspects/categories) that are deemed not relevant to grocery but relevant to shopping and/or place A are not retrieved; images (and their surrounding images and/or associated aspects/categories) that are deemed relevant to grocery and shopping but not relevant to place A are not retrieved etc. Therefore, the described exemplary embodiments may take into account multiple concepts and relationships and may resolve higher level information/queries/events/activities.

In the described exemplary embodiments, an automated semantic relevance mapping (ASRM) process that generates a semantic relevance map in an iterative process is disclosed. In the described exemplary embodiments, the relevance map may specify the correlation between query topics and semantic descriptors. In the described exemplary embodiments, ground truth data may be used to learn positive and negative contributions of semantic descriptors. In the described exemplary embodiments, a mechanism may be

developed to make the model adaptable to different datasets and query topics, and thus, may provide flexibility and adaptability to variances in data and tasks.

In the described exemplary embodiments, image retrieval may be performed, in particular, image retrieval from large data collection, e.g. lifelogged visual data collected by wearable devices. It is appreciated that although lifelog data is described for some exemplary implementations or embodiments, the exemplary embodiments are applicable to any other forms of image data such as moving image data, still image data etc.

The described exemplary embodiments may provide a method to enhance image retrieval for given query topics about events. In particular, the described exemplary embodiments may desirably provide a model that leverages on deep learning technologies and linear regression to bridge the semantic gap. The described exemplary embodiments may provide a novel process/algorithm called automatic semantic relevance mapping (ASRM) which may automatically generate a relevance map to connect/map atomic semantic descriptors with query topics. The inventors have recognised that manual mapping processes are typically tedious and depend heavily on human expertise. The described exemplary embodiments may provide an efficient process, may automate a mapping process and achieve better retrieval performance.

The inventors have also recognised that basic approaches for image retrieval may employ a relevance rate within $[0,1]$, which is only able to model the effects of relevant and non-relevant cues. In described exemplary embodiments, the relevance rate may be extended to $\{+1,0,-1\}$, to characterize the effects of cues that are relevant, non-relevant and inhibitive, respectively. The inventors have recognised that while an ontology-based semantic parsing process is relevant in this context, there are no existing semantic parsing methods (or systems) that are able to sufficiently deal with the issue.

The described exemplary embodiments may address a problem of concept-based image retrieval, e.g. to retrieve images from a large digital photo dataset given a semantically defined query topic. Some components in the described exemplary embodiments in this regard may include, but are not limited to, utilising various CNNs to

describe lifelog images using a set of semantic descriptors (e.g. object and scene features), automated identification of relevant semantic descriptors for a query topic, and optimization of the retrieval result based on feature weighting adapted to events and temporal smoothing to incorporate semantic coherence.

5

The inventors have recognised that it may often require a comprehensive understanding of key information such as when, where, and what to address the problem mentioned above. The inventors have also recognised that conventional technologies have not addressed the problem adequately.

10

The inventors have recognised that there may be many possible usage scenarios for the described exemplary embodiments. For example, in the healthcare industry, the described exemplary embodiments may provide a means to find a moment when an individual takes medication X. As another example, in manufacturing, the described exemplary embodiments may provide a means to find a moment when Y machine was inspected (e.g. for quality checks). The described exemplary embodiments may be useful in various other industries such as in the construction industry (e.g. for inspection and/or quality checks via wearable devices), aerospace industry (e.g. for maintenance, repair and overhauls of aircrafts via wearable devices of ground staff) and in the policing industry (e.g. for law enforcement and security via wearable devices of security staff).

20

The described exemplary embodiments may further provide a means in the context of lifelogging. Lifelogging is an emerging application domain with technologies embracing latest advances in wearable computing and data analytics. The inventors have recognised that lifelogging may have a great potential to promote digital health and digital lifestyle.

25

Different exemplary embodiments can be implemented in the context of data structure, program modules, program and computer instructions executed in a computer implemented environment. A general purpose computing environment is briefly disclosed herein. One or more exemplary embodiments may be embodied in one or more computer systems, such as is schematically illustrated in FIG. 5.

30

One or more exemplary embodiments may be implemented as software, such as a computer program being executed within a computer system 500, and instructing the computer system 500 to conduct a method of an exemplary embodiment.

5 The computer system 500 comprises a computer unit 502, input modules such as a keyboard 504 and a pointing device 506 and a plurality of output devices such as a display 508, and printer 510. A user can interact with the computer unit 502 using the above devices. The pointing device can be implemented with a mouse, track ball, pen device or any similar device. One or more other input devices (not shown) such as a joystick, game
10 pad, satellite dish, scanner, touch sensitive screen or the like can also be connected to the computer unit 502. The display 508 may include a cathode ray tube (CRT), liquid crystal display (LCD), field emission display (FED), plasma display or any other device that produces an image that is viewable by the user.

15 The computer unit 502 can be connected to a computer network 512 via a suitable transceiver device 514, to enable access to e.g. the Internet or other network systems such as Local Area Network (LAN) or Wide Area Network (WAN) or a personal network. The network 512 can comprise a server, a router, a network personal computer, a peer device or other common network node, a wireless telephone or wireless personal digital assistant.
20 Networking environments may be found in offices, enterprise-wide computer networks and home computer systems etc. The transceiver device 514 can be a modem/router unit located within or external to the computer unit 502, and may be any type of modem/router such as a cable modem or a satellite modem.

25 It will be appreciated that network connections shown are exemplary and other ways of establishing a communications link between computers can be used. The existence of any of various protocols, such as TCP/IP, Frame Relay, Ethernet, FTP, HTTP and the like, is presumed, and the computer unit 502 can be operated in a client-server configuration to permit a user to retrieve web pages from a web-based server. Furthermore, any of various
30 web browsers can be used to display and manipulate data on web pages.

The computer unit 502 in the example comprises a processor 518, a Random Access Memory (RAM) 520 and a Read Only Memory (ROM) 522. The ROM 522 can be a system memory storing basic input/ output system (BIOS) information. The RAM 520 can store one or more program modules such as operating systems, application programs and program data.

The computer unit 502 further comprises a number of Input/Output (I/O) interface units, for example I/O interface unit 524 to the display 508, and I/O interface unit 526 to the keyboard 504. The components of the computer unit 502 typically communicate and interface/couple connectedly via an interconnected system bus 528 and in a manner known to the person skilled in the relevant art. The bus 528 can be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures.

It will be appreciated that other devices can also be connected to the system bus 528. For example, a universal serial bus (USB) interface can be used for coupling a video or digital camera to the system bus 528. An IEEE 1394 interface may be used to couple additional devices to the computer unit 502. Other manufacturer interfaces are also possible such as FireWire developed by Apple Computer and i.Link developed by Sony. Coupling of devices to the system bus 528 can also be via a parallel port, a game port, a PCI board or any other interface used to couple an input device to a computer. It will also be appreciated that, while the components are not shown in the figure, sound/audio can be recorded and reproduced with a microphone and a speaker. A sound card may be used to couple a microphone and a speaker to the system bus 528. It will be appreciated that several peripheral devices can be coupled to the system bus 528 via alternative interfaces simultaneously.

An application program can be supplied to the user of the computer system 500 being encoded/stored on a data storage medium such as a CD-ROM or flash memory carrier. The application program can be read using a corresponding data storage medium drive of a data storage device 530. The data storage medium is not limited to being portable and can include instances of being embedded in the computer unit 502. The data storage

device 530 can comprise a hard disk interface unit and/or a removable memory interface unit (both not shown in detail) respectively coupling a hard disk drive and/or a removable memory drive to the system bus 528. This can enable reading/writing of data. Examples of removable memory drives include magnetic disk drives and optical disk drives. The drives and their associated computer-readable media, such as a floppy disk provide nonvolatile storage of computer readable instructions, data structures, program modules and other data for the computer unit 502. It will be appreciated that the computer unit 502 may include several of such drives. Furthermore, the computer unit 502 may include drives for interfacing with other types of computer readable media.

The application program is read and controlled in its execution by the processor 518. Intermediate storage of program data may be accomplished using RAM 520. The method(s) of the exemplary embodiments can be implemented as computer readable instructions, computer executable components, or software modules. One or more software modules may alternatively be used. These can include an executable program, a data link library, a configuration file, a database, a graphical image, a binary data file, a text data file, an object file, a source code file, or the like. When one or more computer processors execute one or more of the software modules, the software modules interact to cause one or more computer systems to perform according to the teachings herein.

The operation of the computer unit 502 can be controlled by a variety of different program modules. Examples of program modules are routines, programs, objects, components, data structures, libraries, etc. that perform particular tasks or implement particular abstract data types. The exemplary embodiments may also be practiced with other computer system configurations, including handheld devices, multiprocessor systems, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, personal digital assistants, mobile telephones and the like. Furthermore, the exemplary embodiments may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a wireless or wired communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

Different exemplary embodiments can be implemented in the context of data structure, program modules, program and computer instructions executed in a communication device. An exemplary communication device is briefly disclosed herein. One
5 or more exemplary embodiments may be embodied in one or more communication devices e.g. 600, such as is schematically illustrated in FIG. 6.

One or more exemplary embodiments may be implemented as software, such as a computer program being executed within a communication device 600, and instructing the
10 communication device 600 to conduct a method of an exemplary embodiment.

The communication device 600 comprises a processor module 602, an input module such as a touchscreen interface or a keypad 604 and an output module such as a display 606 on a touchscreen.

15 The processor module 602 is coupled to a first communication unit 608 for communication with a cellular network 610. The first communication unit 608 can include, but is not limited to, a subscriber identity module (SIM) card loading bay. The cellular network 610 can, for example, be a 3G or 4G network.

20 The processor module 602 is further coupled to a second communication unit 612 for connection to a network 614. For example, the second communication unit 612 can enable access to e.g. the Internet or other network systems such as Local Area Network (LAN) or Wide Area Network (WAN) or a personal network. The network 614 can comprise a server, a
25 router, a network personal computer, a peer device or other common network node, a wireless telephone or wireless personal digital assistant. Networking environments may be found in offices, enterprise-wide computer networks and home computer systems etc. The second communication unit 612 can include, but is not limited to, a wireless network card or an ethernet network cable port. The second communication unit 612 can also be a
30 modem/router unit and may be any type of modem/router such as a cable-type modem or a satellite-type modem.

It will be appreciated that network connections shown are exemplary and other ways of establishing a communications link between computers can be used. The existence of any of various protocols, such as TCP/IP, Frame Relay, Ethernet, FTP, HTTP and the like, is presumed, and the communication device 600 can be operated in a client-server
5 configuration to permit a user to retrieve web pages from a web-based server. Furthermore, any of various web browsers can be used to display and manipulate data on web pages.

The processor module 602 in the example includes a processor 616, a Random Access Memory (RAM) 618 and a Read Only Memory (ROM) 620. The ROM 620 can be a
10 system memory storing basic input/ output system (BIOS) information. The RAM 618 can store one or more program modules such as operating systems, application programs and program data.

The processor module 602 also includes a number of Input/Output (I/O) interfaces,
15 for example I/O interface 622 to the display 606, and I/O interface 624 to the keypad 604.

The components of the processor module 602 typically communicate and interface/couple connectedly via an interconnected bus 626 and in a manner known to the person skilled in the relevant art. The bus 626 can be any of several types of bus structures
20 including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures.

It will be appreciated that other devices can also be connected to the system bus 626. For example, a universal serial bus (USB) interface can be used for coupling an
25 accessory of the communication device, such as a card reader, to the system bus 626.

The application program is typically supplied to the user of the communication device 600 encoded on a data storage medium such as a flash memory module or memory card/stick and read utilising a corresponding memory reader-writer of a data storage device
30 628. The data storage medium is not limited to being portable and can include instances of being embedded in the communication device 600.

The application program is read and controlled in its execution by the processor 616. Intermediate storage of program data may be accomplished using RAM 618. The method(s) of the exemplary embodiments can be implemented as computer readable instructions, computer executable components, or software modules. One or more software modules may alternatively be used. These can include an executable program, a data link library, a configuration file, a database, a graphical image, a binary data file, a text data file, an object file, a source code file, or the like. When one or more processor modules execute one or more of the software modules, the software modules interact to cause one or more processor modules to perform according to the teachings herein.

The operation of the communication device 600 can be controlled by a variety of different program modules. Examples of program modules are routines, programs, objects, components, data structures, libraries, etc. that perform particular tasks or implement particular abstract data types.

The exemplary embodiments may also be practiced with other computer system configurations, including handheld devices, multiprocessor systems/servers, microprocessor-based or programmable consumer electronics, network PCs, minicomputers, mainframe computers, personal digital assistants, mobile telephones and the like. Furthermore, the exemplary embodiments may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a wireless or wired communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

The terms "coupled" or "connected" as used in this description are intended to cover both directly connected or connected through one or more intermediate means, unless otherwise stated.

The description herein may be, in certain portions, explicitly or implicitly described as algorithms and/or functional operations that operate on data within a computer memory or an electronic circuit. These algorithmic descriptions and/or functional operations are usually

used by those skilled in the information/data processing arts for efficient description. An algorithm is generally relating to a self-consistent sequence of steps leading to a desired result. The algorithmic steps can include physical manipulations of physical quantities, such as electrical, magnetic or optical signals capable of being stored, transmitted, transferred, combined, compared, and otherwise manipulated.

Further, unless specifically stated otherwise, and would ordinarily be apparent from the following, a person skilled in the art will appreciate that throughout the present specification, discussions utilizing terms such as “scanning”, “calculating”, “determining”, “replacing”, “generating”, “initializing”, “outputting”, and the like, refer to action and processes of an instructing processor/computer system, or similar electronic circuit/device/component, that manipulates/processes and transforms data represented as physical quantities within the described system into other data similarly represented as physical quantities within the system or other information storage, transmission or display devices etc.

The description also discloses relevant device/apparatus for performing the steps of the described methods. Such apparatus may be specifically constructed for the purposes of the methods, or may comprise a general purpose computer/processor or other device selectively activated or reconfigured by a computer program stored in a storage member. The algorithms and displays described herein are not inherently related to any particular computer or other apparatus. It is understood that general purpose devices/machines may be used in accordance with the teachings herein. Alternatively, the construction of a specialized device/apparatus to perform the method steps may be desired.

In addition, it is submitted that the description also implicitly covers a computer program, in that it would be clear that the steps of the methods described herein may be put into effect by computer code. It will be appreciated that a large variety of programming languages and coding can be used to implement the teachings of the description herein. Moreover, the computer program if applicable is not limited to any particular control flow and can use different control flows without departing from the scope of the invention.

Furthermore, one or more of the steps of the computer program if applicable may be performed in parallel and/or sequentially. Such a computer program if applicable may be stored on any computer readable medium. The computer readable medium may include storage devices such as magnetic or optical disks, memory chips, or other storage devices suitable for interfacing with a suitable reader/general purpose computer. In such instances, the computer readable storage medium is non-transitory. Such storage medium also covers all computer-readable media e.g. medium that stores data only for short periods of time and/or only in the presence of power, such as register memory, processor cache and Random Access Memory (RAM) and the like. The computer readable medium may even include a wired medium such as exemplified in the Internet system, or wireless medium such as exemplified in bluetooth technology. The computer program when loaded and executed on a suitable reader effectively results in an apparatus that can implement the steps of the described methods.

The exemplary embodiments may also be implemented as hardware modules. A module is a functional hardware unit designed for use with other components or modules. For example, a module may be implemented using digital or discrete electronic components, or it can form a portion of an entire electronic circuit such as an Application Specific Integrated Circuit (ASIC). A person skilled in the art will understand that the exemplary embodiments can also be implemented as a combination of hardware and software modules.

Additionally, when describing some embodiments, the disclosure may have disclosed a method and/or process as a particular sequence of steps. However, unless otherwise required, it will be appreciated the method or process should not be limited to the particular sequence of steps disclosed. Other sequences of steps may be possible. The particular order of the steps disclosed herein should not be construed as undue limitations. Unless otherwise required, a method and/or process disclosed herein should not be limited to the steps being carried out in the order written. The sequence of steps may be varied and still remain within the scope of the disclosure.

Further, in the description herein, the word "substantially" whenever used is understood to include, but not restricted to, "entirely" or "completely" and the like. In addition, terms such as "comprising", "comprise", and the like whenever used, are intended to be non-restricting descriptive language in that they broadly include elements/components recited after such terms, in addition to other components not explicitly recited. For an example, when "comprising" is used, reference to a "one" feature is also intended to be a reference to "at least one" of that feature. Terms such as "consisting", "consist", and the like, may, in the appropriate context, be considered as a subset of terms such as "comprising", "comprise", and the like. Therefore, in embodiments disclosed herein using the terms such as "comprising", "comprise", and the like, it will be appreciated that these embodiments provide teaching for corresponding embodiments using terms such as "consisting", "consist", and the like. Further, terms such as "about", "approximately" and the like whenever used, typically means a reasonable variation, for example a variation of +/- 5% of the disclosed value, or a variance of 4% of the disclosed value, or a variance of 3% of the disclosed value, a variance of 2% of the disclosed value or a variance of 1% of the disclosed value.

Furthermore, in the description herein, certain values may be disclosed in a range. The values showing the end points of a range are intended to illustrate a preferred range. Whenever a range has been described, it is intended that the range covers and teaches all possible sub-ranges as well as individual numerical values within that range. That is, the end points of a range should not be interpreted as inflexible limitations. For example, a description of a range of 1% to 5% is intended to have specifically disclosed sub-ranges 1% to 2%, 1% to 3%, 1% to 4%, 2% to 3% etc., as well as individually, values within that range such as 1%, 2%, 3%, 4% and 5%. The intention of the above specific disclosure is applicable to any depth/breadth of a range.

In the described exemplary embodiments, it will be appreciated that retrieval of one or more images may broadly encompass a mere identification of the one or more images, a display of the one or more images etc., and is not limited to an actual output or transmission of the one or more images.

In the described exemplary embodiments, it will be appreciated that the concept of semantics aspects and feature weighting is not limited to usage of deep learning networks used for semantics descriptors extraction. Rather, such aspects may also be due to e.g. recognition of certain groups of descriptors being more relevant to a particular topic etc.

5

It will be appreciated by a person skilled in the art that other variations and/or modifications may be made to the specific embodiments without departing from the scope of the invention as broadly described. For example, in the description herein, features of different exemplary embodiments may be mixed, combined, interchanged, incorporated, adopted, modified, included etc. or the like across different exemplary embodiments. The present embodiments are, therefore, to be considered in all respects to be illustrative and not restrictive.

10

15

CLAIMS

1. A system for image retrieval, the system comprising,
an extraction module coupled to a dataset of one or more image data; the extraction
5 module being configured to extract one or more semantic descriptors from the one or more
image data;

a mapping module coupled to the extraction module;

a query topics module coupled to the mapping module, the query topics module
being arranged to provide one or more query topics to the mapping module;

10 wherein the mapping module is configured to receive the one or more semantic
descriptors from the extraction module and to map the one or more semantic descriptors to
the one or more query topics from the query topics module;

further wherein the mapping module is configured to map the one or more semantic
descriptors to the one or more query topics from the query topics module based on a
15 correlation of at least one of the one or more semantic descriptors to the one or more query
topics;

further wherein the mapping module is configured to receive ground truth data based
on a ground truth subset of the one or more image data and to establish an initial semantic
relevance map using the ground truth data and at least one of the one or more query topics;

20 further wherein the mapping module is configured to use a training subset of the
ground truth data to obtain a relevance rate of each semantic descriptor of the training
subset based on the one or more query topics; and

wherein the relevance rate comprises a determination of a first number of image data
tagged to the each semantic descriptor determined to be positively relevant to the one or
25 more query topics and of a second number of image data tagged to the each semantic
descriptor determined to be negatively relevant to the one or more query topics.

2. The system as claimed in claim 1, further comprising the mapping module
being configured to obtain a set of node weights of aspect clusters, the aspect clusters being
30 based on semantic descriptors of the training subset of the ground truth data.

3. The system as claimed in any one of claims 1 or 2, further comprising the mapping module being configured to apply the relevance rate to the rest of the ground truth data outside of the training subset, said rest of the ground truth data outside of the training subset being termed a verification set; and the application of the relevance rate to the verification set being to perform image data retrieval from the verification set.

4. The system as claimed in claim 3, further comprising the mapping module being configured to determine, based on the one or more query topics, at least one of whether there is any image data from the verification set that is not retrieved and whether there is any image data from the verification set that is incorrectly retrieved.

5. The system as claimed in claim 4, further comprising the mapping module being configured to expand the training subset based on the determination of at least one of whether there is any image data from the verification set that is not retrieved and whether there is any image data from the verification set that is incorrectly retrieved.

6. The system as claimed in claim 5, further comprising the mapping module being configured to obtain a revised relevance rate of each semantic descriptor of the expanded training subset based on the one or more query topics.

7. A method of image retrieval, the method comprising,
providing a dataset of one or more image data;
extracting one or more semantic descriptors from the one or more image data using an extraction module;
providing one or more query topics;
mapping using a mapping module the one or more semantic descriptors to the one or more query topics based on a correlation of at least one of the one or more semantic descriptors to the one or more query topics;
receiving at the mapping module ground truth data based on a ground truth subset of the one or more image data and establishing an initial semantic relevance map using the ground truth data and at least one of the one or more query topics;

using at the mapping module a training subset of the ground truth data to obtain a relevance rate of each semantic descriptor of the training subset based on the one or more query topics; and

5 determining a first number of image data tagged to the each semantic descriptor determined to be positively relevant to the one or more query topics and determining a second number of image data tagged to the each semantic descriptor determined to be negatively relevant to the one or more query topics.

8. The method as claimed in claim 7, further comprising obtaining at the
10 mapping module a set of node weights of aspect clusters, the aspect clusters being based on semantic descriptors of the training subset of the ground truth data.

9. The method as claimed in any one of claims 7 or 8, further comprising
15 applying at the mapping module the relevance rate to the rest of the ground truth data outside of the training subset, said rest of the ground truth data outside of the training subset being termed a verification set; and the application of the relevance rate to the verification set being to perform image data retrieval from the verification set.

10. The method as claimed in claim 9, further comprising determining at the
20 mapping module, based on the one or more query topics, at least one of whether there is any image data from the verification set that is not retrieved and whether there is any image data from the verification set that is incorrectly retrieved.

11. The method as claimed in claim 10, further comprising expanding the training
25 subset based on the determination of at least one of whether there is any image data from the verification set that is not retrieved and whether there is any image data from the verification set that is incorrectly retrieved.

12. The method as claimed in claim 11, further comprising obtaining at the
30 mapping module a revised relevance rate of each semantic descriptor of the expanded training subset based on the one or more query topics.

13. A non-transitory tangible computer readable storage medium having stored thereon software instructions that, when executed by a computer processor of a system for image retrieval, cause the computer processor to perform a method of image retrieval, by executing the steps comprising,

5 providing a dataset of one or more image data;

 extracting one or more semantic descriptors from the one or more image data using an extraction module;

 providing one or more query topics;

10 mapping using a mapping module the one or more semantic descriptors to the one or more query topics based on a correlation of at least one of the one or more semantic descriptors to the one or more query topics;

 receiving at the mapping module ground truth data based on a ground truth subset of the one or more image data and establishing an initial semantic relevance map using the ground truth data and at least one of the one or more query topics;

15 using at the mapping module a training subset of the ground truth data to obtain a relevance rate of each semantic descriptor of the training subset based on the one or more query topics; and

 determining a first number of image data tagged to the each semantic descriptor determined to be positively relevant to the one or more query topics and determining a
20 second number of image data tagged to the each semantic descriptor determined to be negatively relevant to the one or more query topics.