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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 June 2015 (04.06.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/078022 A1

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

(21) International Application Number:
PCT/CN2013/088258

(22) International Filing Date:
30 November 2013 (30.11.2013)

(25) Filing Language: English

(26) Publication Language: English

(72) Inventor; and
(71) Applicant : **TANG, Xiaoou** [CN/CN]; HSH 809, Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (CN).

(72) **Inventors**: **QIU, Shi**; Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (CN). **WANG, Xiaogang**; HSH 415, Department of Electrical Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (CN).

(74) **Agent**: **INSIGHT INTELLECTUAL PROPERTY LIMITED**; 19A, Tower A, InDo Building, No. 48A Zhichun Road, Haidian District, Beijing 100098 (CN).

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

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

[Continued on next page]

(54) **Title**: VISUAL SEMANTIC COMPLEX NETWORK AND METHOD FOR FORMING NETWORK

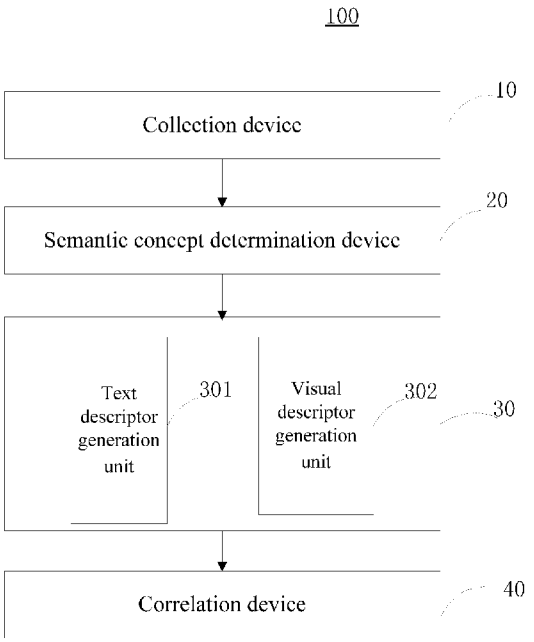


Fig.1

(57) **Abstract**: A visual semantic complex network system and a method for generating the system have been disclosed. The system may comprise a collection device configured to retrieve a plurality of images and a plurality of texts associated with the images in accordance with given query keywords; a semantic concept determination device configured to determine semantic concepts of the retrieved images and retrieved texts for the retrieved images, respectively; a descriptor generation device configured to, from the retrieved images and texts, generate text descriptors and visual descriptors for the determined semantic concepts; and a semantic correlation device configured to determine semantic correlations and visual correlations from the generated text and visual descriptor, respectively, and to combine the determined semantic correlations and the determined visual correlations to generate the visual semantic complex network system.

WO 2015/078022 A1

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

Visual Semantic Complex Network and Method for Forming Network

Technical Field

[0001] The present application refers to a visual semantic complex network system and a method for generating the system.

Background

[0002] The enormous and ever-growing amount of images on the web has inspired many important applications related to web image search, browsing, and clustering. Such applications aim to provide users with easier access to web images. An essential issue facing all these tasks is how to model the relevance of images on the web. This problem is particularly challenging due to the large diversity and complex structures of web images. Most search engines rely on textual information to index web images and measure their relevance. Such an approach has some well known drawbacks. Because of the ambiguous nature of textual description, images indexed by the same keyword may come from irrelevant concepts and exhibit large diversity on visual content. More importantly, some relevant images under different keyword indices such as “palm pixi” and “apple iphone” fail to be connected by this approach. Another approach estimates image relevance by comparing visual features extracted from image contents. Various approximate nearest neighbor (ANN) search algorithms (e.g. hashing) have been used to improve the search efficiency. However, such visual features and ANN algorithms are only effective for images with very similar visual content, i.e. near duplicate, and cannot find relevant images that have the same semantic meaning but moderate difference in visual content.

[0003] Both of the above approaches only allow users to interact with the huge web image collections at a microscopic level, i.e. exploring images within a very small local region either in the textual or visual feature space, which limits the effective access of web images. Although efforts have been made to manually organize portions of web images, it is derived from a human-defined ontology that has inherent discrepancies with dynamic web

images. It is also very expensive to scale.

Summary

[0004] The purpose of this application is to automatically discover and model the visual and semantic structures of web image collections, study their properties at a macroscopic level, and demonstrate the use of such structures and properties through concrete applications. To this end, the present application proposes to model web image collections using the Visual Semantic Complex Network (VSCN), an automatically generated graph structure on which images that are relevant in both semantics and visual content are well connected and organized.

[0005] It shall be noted that images on the web are not distributed randomly, but do tend to form visually and semantically compact clusters. These image clusters can be used as the elementary units for modeling the structures of web image collections. The represent application automatically discovers image clusters with both semantic and visual consistency, and treats them as nodes on the graph.

[0006] In the disclosures of the application, the discovered image clusters is called as semantic concepts, and are associated them with visual and textual descriptors. The semantic concepts are connected with edges based on their visual and semantic correlations. The semantic concepts and their correlations bring structures to web images and allow more accurate modeling of image relevance.

[0007] It will be a better understanding of web image collections at a macroscopic level by studying the structural properties of the VSCN from the perspective of complex network. The present application explores a few of them, including small-world behavior, concept community, hub structures, and isolated concepts, and reveal some interesting findings. Such properties provide valuable information that opens doors for many important applications such as text or content-based web image retrieval, web image browsing, discovering popular web image topics, and defining image similarities based on structural information.

[0008] The application is proposed to address two applications: content-based image retrieval (CBIR) and image browsing. For web-scale CBIR, existing approaches typically

match images with visual features and ANN search algorithms (e.g. hashing). These algorithms often lead only to a small portion of images highly similar to the query (near duplicate). In this work, these detected images are connected to other relevant images that form community structures on the VSCN. Therefore, many more relevant images can be found by exploiting the structural information provided by the VSCN. In the second application, a novel visualization scheme is proposed for web image browsing. Users can explore the web image collections by navigating the VSCN without being limited by query keywords.

[0009] In one aspect, the present application provides a visual semantic complex network system for Web Images, comprising:

a collection device configured to retrieve a plurality of images and a plurality of texts associated with the images in accordance with given query keywords;

a semantic concept determination device configured to determine semantic concepts and representative images of the retrieved texts and retrieved images , respectively;

a descriptor generation device configured to, from the determined semantic concepts and representative images, generate text descriptors and visual descriptors; and

a correlation device configured to determine semantic correlations and visual correlations from the generated text and visual descriptors, respectively, and to combine the determined semantic correlations and the determined visual correlations to generate the visual semantic complex network system.

[0010] In another aspect, the present application provides a method for forming a visual semantic complex network system for Web images, comprising:

retrieving a plurality of images and a plurality of texts associated with the images in accordance with given query keywords;

determining semantic concepts and representative images of the retrieved texts and retrieved images , respectively;

generating, from the semantic concepts and representative images, text descriptors and visual descriptors; and

determining semantic correlations and visual correlations from the generated text

descriptor and the generated visual descriptor, respectively,

combining the semantic correlations and visual correlations to generate the visual semantic complex network system.

[0011] The above method may be carried out by one or more processor in the computer.

[0012] In another aspect, the present application provides a computer readable storage media comprising:

instructions for retrieving a plurality of images and a plurality of texts associated with the images in accordance with given query keywords;

instructions for determining semantic concepts and representative images of the retrieved texts and retrieved images, respectively;

instructions for generating, from the semantic concepts and representative images, text descriptors and visual descriptors; and

instructions for determining semantic correlations and visual correlations from the generated text descriptor and the generated visual descriptor, respectively,

instructions for combining the semantic correlations and visual correlations to generate the visual semantic complex network system.

[0013] In another aspect, the present application provides a method for searching images with the visual semantic complex network system, comprising:

obtaining a list of images according to a given query image;

determining a group of related concept communities from the obtained list;

determining a group of related semantic concepts from the communities;

gathering, from the determined related semantic concepts, images of a top plurality of concepts; and

forming a re-ranking pool of the gathered images, which are matched with the query image.

[0014] Accordingly, a computer readable storage media is provided and comprises:

instructions for obtaining a list of images according to a given query image;

instructions for determining a group of related concept communities from the obtained list;

instructions for determining a group of related semantic concepts from the communities;

instructions for gathering, from the determined related semantic concepts, images of a top plurality of concepts; and

instructions for forming a re-ranking pool of the gathered images, which are matched with the query image.

[0015] In another aspect, the present application further provides a method for browsing images with semantic concepts. The semantic concepts may be generated in the visual semantic complex network system for Web Images as mentioned in the above. The method may comprises:

entering a query keyword into a display system;

generating a plurality of semantic concepts based on same queries as the entered keyword;

visualizing the generated semantic concepts in a query space of the display system;

switching the query space to a local concept space of the display unit in response to selecting a concept selected from the query space, wherein on the local concept space the selected concept together with its neighbor concepts is shown.

[0016] In addition, the method for browsing images may further comprises:

selecting a centric concept in the local concept space; and

switching back to the query space that the selected concept belongs to.

selecting another concept in the local concept space; and

switching to another local concept space where said another concept and its neighbor concepts are shown.

[0017] The above method may be carried out by one or more processor in the

computer.

Brief Description of Drawings

[0018] Fig. 1 illustrates a block view of the exemplary visual semantic complex network system for Web Images according to one embodiment of the present application.

[0019] Fig. 2 is a flowchart of a method for generating semantic descriptors according to one embodiment of the present application.

[0020] Fig. 3 is a flowchart of a method for generating visual descriptors according to one embodiment of the present application.

[0021] Fig. 4 is a flowchart of a process for forming a visual semantic complex network system for Web images according to an embodiment of the present application;

[0022] Fig. 5 (a)-(f) is illustrates a block view of how to search images with the visual semantic complex network system according to an embodiment of the present application

[0023] Fig. 6 is a flowchart of a method for searching images with the visual semantic complex network system according to an embodiment of the present application.

Detailed Description

[0024] Embodiments of the present application can solve a problem of wasting storage resources or depicting inaccurately during document rendering. Thus, a technical effect of reducing storage space while improving rendering accuracy can be achieved.

[0025] Fig. 1 illustrates a block view of the exemplary visual semantic complex network system 100 for Web Images according to one embodiment of the present application. As shown in Fig. 1, the system 100 comprises a collection device 10, a semantic concept determination device 20, a descriptor generation device 30 and a correlation device 40.

[0026] The collection device 10 is configured to retrieve a plurality of images and texts in accordance with given query keywords. In embodiments of the application, it starts with a plurality of top query keywords of a search engine, and then automatically discovers a larger number of (semantic concepts that are compact image clusters with visual and semantic

consistency. In one instance, take *Bing image search engine* as example, if there is for example, 2,000 keywords of the search engine, there will be about 33,240 semantic concepts to be discovered.

[0027] The semantic concept determination device 20 is configured to determine semantic concepts and representative images of the retrieved texts and retrieved images. In one embodiment of the present application, the semantic concept determination device 20 learns the semantic concepts by discovering keywords that occur frequently in visually similar images. These discovered keywords correlate well with the image content and therefore leads to descriptive concepts. To be specific, for every query q, e.g. "apple", we submit q to an image search engine. With the retrieved collection of images I_q and surrounding texts T_q, their relevant semantic concepts, such as "apple fruit" and "apple iphone", can be automatically discovered. Such concepts have more specific semantic meanings and less visual diversity, and can be viewed as elementary units of web image collections. The learned concepts under query keyword q are denoted as $C_q = \{c_i\}_{i=1}^{M_q}$. The concepts were learned from different queries form the nodes of the VSCN 100.

[0028] The following is a summarized process of the concept discovery.

Algorithm 1 Concept Discovery through Query Expansion

Input: Query q, image collection $\mathcal{I}_q$, surrounding texts $\mathcal{T}_q$.

Output: Learned concept set $\mathcal{C}_q = \{c_i\}_{i=1}^{M_q}$.

1: **Initialization:** $\mathcal{C}_q := \emptyset$, $r_I(w) := 0$.
2: **for all** images $I_k \in \mathcal{I}_q$ **do**
3: Find the top K visual neighbors, denote as $\mathcal{N}(I_k)$
4: Let $W(I_k) = \{w_{I_k}^i\}_{i=1}^{T}$ be the T most frequent words in the surrounding texts of $\mathcal{N}(I_k)$.
5: **for all** words $w_{I_k}^i \in W(I_k)$ **do**
6: $r_I(w_{I_k}^i) := r_I(w_{I_k}^i) + (T - i)$.
7: **end for**
8: **end for**
9: Combine q and the M_q words with largest $r_I(w)$ to form $\mathcal{C}_q$.

[0029] The descriptors generation device 30 is configured to, from the retrieved images, the retrieved texts and elementary units, generate a text descriptor and a visual descriptor for the determined semantic concepts. As the number of concepts is very large

(for example, 33,240 in the embodiment, and potentially even larger if we expand the VSCN), two efficient methods to compute semantic and visual correlations will be described below.

[0030] In particular, the descriptor generation device 30 comprises a text descriptor generation unit 301 and a visual descriptor generation unit 302.

[0031] In one embodiment, the text descriptor generation unit 301 operates to collect the text snippets corresponding to the semantic concepts, compute/determine the term frequency (TF) vector of the collected snippets to keep a first plurality of terms in the vector with the highest term frequency (that is, the other terms in the vector will be cancelled), and thus the computed the term frequency vector is truncated. And then the text descriptor generation unit 301 operates to normalize the truncated vectors and determine the semantic correlation from the truncated vectors. For example, the text descriptor generation unit 301 operates to L_2-normalize the truncated vectors.

[0032] To be specific, for each concept $c_i \in \mathrm{C}$, the text descriptor generation unit 301 may operate to carry out the following steps as shown in Fig. 2. At step s201, the text descriptor generation unit 301 utilizes c_i as a query input on the web search (for example, Google web search), and collect the top K (for example, K=50) searched snippets, denoted as $S(c_i)$. At step s202, the text descriptor generation unit 301 computes/determines the term frequency (TF) vector of $S(c_i)$ and keeps, for example, the top M (for example, K=100) terms with highest TFs, that is, the TF vector is truncated. At step s203, the text descriptor generation unit 301 normalizes the truncated vector to form text descriptor.

[0033] The visual descriptor generation unit 302 is configured to encode each of the retrieved images by a hashing function H so as to generate a binary vector for each retrieved image, accumulate the generate binary vectors and quantize the accumulated vector back to binary vector such that a visual correlation (visual descriptor) between each two of the concept will be formed from the quantized binary vector.

[0034] To be specific, the visual descriptor generation unit 302 may operate to carry out the following steps as shown in Fig. 3. At step S301, for a concept $c_i \in \mathrm{C}$, its exemplar

image set by I_{ci}. $I_k \in \mathrm{I}_{ci}$ is encoded in an M-dimensional binary vector $H(I_k)$ using an M-bit base hashing function H (Here we represent each bit with ± 1). At step S302, the visual descriptor generation unit 302 operates to accumulate the binary vectors as $A = \sum H(I_k)$. At step 303, the visual descriptor generation unit 302 operates to quantize the accumulated vectors back to binary vector, which is donated as visual descriptor $simhash(c_i) = sign(A)$.

[0035] The correlation device 40 is configured to determine the semantic correlation from the generated text and visual descriptor so as to combine the semantic correlation and visual correlation to generate a K-nearest-neighbor (K-NN) graph network system.

[0036] The semantic correlation may be determined by using the conventional means. For example, for a short text x, a set of snippets $S(x)$ is obtained from the web search. A snippet is a short text summary generated by the search engine for each search result item with query c. The text descriptor generation unit 301 collects the snippets of the top N search result items, which provide rich semantic context for x. And then the text descriptor generation unit 301 operates to determine the similarity between two texts x_1 and x_2 by computing the textual similarity between $S(x_1)$ and $S(x_2)$ using the term vector model and cosine similarity.

[0037] After the result vector $ntf(c_i)$ as the text descriptors is determined as shown in Fig.2, the correlation device 40 operates to determine the semantic correlation between c_i and c_j by rule of:

$$S\_Cor = Cosine(ntf(c_i), ntf(c_j)). \quad (1)$$

[0038] As to the visual correlation, it may be measured by the visual similarity between their corresponding exemplar image sets. For each concept, its exemplar image set consists of the top K (for example, 300) images retrieved from the search engine by using the concept as query keyword. This exemplar image set is further represented as a binary code by

the conventional sim-hashing algorithm. This sim-hashing code can be viewed as a visual signature of the original exemplar image set. The visual similarity between any pair of exemplar image sets can then be approximated by the negative of hamming distance between their sim-hashing codes. To be specific, once the visual descriptor $simhash(c_i) = sign(A)$ is determined by unit 302 at step S303, the correlation device 40 operates to determine the visual correlation between c_i and c_j by rule of,

$$V\_Cor = 1 - \frac{1}{M} HamDist(simhash(c_i), simhash(c_j)) \quad . \quad (2)$$

[0039] And then, the correlation device 40 operates to combine the semantic correlation and visual correlation by $Cor = S\_cor + V\_cor$. Finally, the system 100 build the VSCN as a K-nearest-neighbor (K-NN) graph by connecting each node to its top K neighbors with the largest correlations.

[0040] Hereinafter, the present application also proposes a process 400 for forming a visual semantic complex network system for Web images. As shown in Fig. 4, in step S401, the process retrieve a plurality of images and a plurality of texts associated with the images in accordance with given query keywords.

[0041] In step S402, the process determines semantic concepts and representative images of the retrieved texts and retrieved images..

[0042] In step S403, the process generates, from the determined semantic concepts and representative image, text descriptors and visual descriptors for the determined semantic concepts. The step S403 may comprise the step of determining semantic correlations and visual correlations from the generated text descriptor and the generated visual descriptor as discussed in reference to Figs. 2 and 3 above.

[0043] In step S404, the process determines semantic correlations and visual correlations from the generated text descriptor and the generated visual descriptor, respectively. Specifically, a semantic correlation between each two of the text concepts may be generated by collecting a plurality of text snippets corresponding to the semantic concepts, determining a term frequency vector of the collected snippets; truncating the computed vector

such that a plurality of terms in the vector with the highest term frequency is maintained; and normalizing the truncated vectors to generate said text descriptors, such that the visual correlation between each two of the text concepts are generated from the quantized binary vector. The visual correlation may be generated by encoding each of the retrieved images by a hashing function so as to generate a binary vector for each retrieved image, accumulating the generate binary vectors; quantizing the accumulated vector back to a binary vector as said visual descriptor; and determining the visual correlation from the truncated vectors. The generations of the semantic correlation and the visual correlation have been discussed in the above, and thus the detailed descriptions thereof are omitted.

[0044] In step S405, the process 400 combines the semantic correlations and visual correlations to generate the visual semantic complex network system.

[0045] As well known in the art, the complex networks have many important properties, some of which are explored with the proposed VSCN 100. The study of these properties not only yields a better understanding of web image collections at a macroscopic level, but also provides valuable information that assists in important tasks including CBIR and image browsing, as will be discussed later.

1) Small-World Behavior

[0046] The small-world behavior exists in many complex networks such as social networks and the World Wide Web. It means that most nodes can be reached from the others in a small number of hops. It is of great interest to study whether this phenomenon also exists in our VSCN 100. The small-world behavior has important implications in some applications such as image browsing by navigating the VSCN 100.

[0047] As the VSCN 100 is constructed locally, it is interesting to know how it is globally connected. It finds that even for a small neighborhood size ($K = 5$), there already emerges a dominant connected component that includes more than half of the nodes on the VSCN, as shown in Figure 3 (a). The largest connected component grows quickly with K and covers 96% of the VSCN when $K = 20$. Thus, the VSCN is a well connected network.

[0048] The average shortest path length is determined by

$$L = \frac{1}{|V|(|V|-1)} \sum_{v_i, v_j \in V, v_i \neq v_j} d(v_i, v_j). \quad (3)$$

[0049] V is defined as the largest connected component to avoid divergence of L. Figure 3 (a) shows L as a function of K. L drops quickly at the beginning. For K>20, the average separation between two nodes on the largest connected components is only about six hops. The existence of a dominant connected component and its small separation between nodes suggest it is possible to navigate the VSCN 100 by following its edges, which inspires the novel image browsing scheme as will be discussed below. Hereinafter, K will be fixed at 20 for purpose of description, but the present application is not limited thereto.

2) In-degree Distribution

[0050] In-degree is an important measurement in complex networks. On the VSCN 100, the nodes have identical out-degree (for example, K=20), but their in-degrees differ widely from 0 to 500. Only 1% of nodes have in-degrees larger than 100. In general, representative and popular concepts that are neighbors of many other concepts have high in-degrees, and form hub structures. Isolated concepts have zero in-degree. They are typically uncommon concepts such as “geodesic dome” and “ant grasshopper”, or the failures of concept detection such as “dscn jpg” which does not have semantic meanings. Figure 5 shows part of the VSCN, with concepts of large in-degrees. We can identify several semantic regions formed by these concepts, including traveling, entertainments, wallpapers, and automobile, which correspond to the green, yellow, dark blue, and light blue regions, respectively.

[0051] Hereinafter, a method 600 for searching images with the visual semantic complex network system 100 will be discussed.

[0052] Generally, given a query image (Fig. 5 (a)), its nearest neighbors in the database are retrieved with a baseline method or any other available method. Based on the initial retrieval result, the semantic meaning of the query image is estimated using a small set of relevant semantic concepts on the VSCN100. Images under these semantic concepts are then gathered to form a re-ranking pool. Images inside the pool are ranked based on their

visual similarity to the query image, and the ranking list is returned (Fig. 5 (f)). The VSCN brings two key benefits: (1) as the search space is greatly reduced, the re-ranking pool contains significantly less noise than the entire database, leading to superior retrieval result. (2) The re-ranking pool contains a more manageable number of images than the entire database (a few thousand v.s. millions). It allows the use of more powerful features and similarity measures, further promoting the performance.

[0053] To be specific, the method 600 for searching images with the visual semantic complex network system according to an embodiment of the present application will be discussed in referring to Fig. 6. At step 601, a list of images is obtained according to a given image by using any conventional means in the art.

[0054] At step S602, a group of close related concept communities will be determined from the list returned from step S601.

[0055] The semantic regions suggest the existence of community structures on the VSCN. In the literature of complex networks, a community is referred to as a subgraph with tightly connected nodes. On the VSCN, it corresponds to a group of (for example, closely) related semantic concepts, called a concept community. To find such communities, the inventors adopt the graph-based agglomerative algorithm in the art due to its good performance and high efficiency. The algorithm starts by treating each single node as a cluster, and iteratively merges clusters with largest affinity, measured via the product of in-degrees and out-degrees between the two clusters.

[0056] The inventors observe a few interesting facts from the clustering results. First, the size of clusters approximately follows a power-laws distribution, and 10% of the clusters are with size larger than 10. They cover 52% nodes on the VSCN. Second, these clusters correspond to various semantic topics, such as cars, food, plants, and animals.

[0057] At step S603, a group of close related semantic concepts will be determined from the communities as determined in step S602.

[0058] A key step of our approach is to estimate the semantic meaning of the query image, which is done at two levels. At the community level, it estimates the query image's semantic meaning using a set of concept communities discovered in the above. As concept communities group similar concepts, estimating the relevant communities is more reliable

than estimating individual concepts. Then, at the concept level, a smaller set of relevant concepts are further identified from the previously identified communities. Both levels fully exploit the structural information of the VSCN, which makes our approach more robust.

1.1 Community-level Estimation

[0059] The detected concept communities is referred by $\{T_i\}_{i=1}^{K_T}$. Given a query image I_q, a list of top-ranked images and their distances to I_q are returned by a baseline retrieval algorithm (e.g. ITQ hashing). From the truncated list $\{(I_k, d_k)\}_{k=1}^{N_I}$, we calculate a relevance score for each T_i as:

$$s(T_i) = \sum_{k=1}^{N_I} exp(\frac{-d_k}{\sigma}) \chi[c(I_k), T_i]. \quad (2)$$

$c(I_k)$ is the concept to which the database image I_k belongs. $\chi[c(I_k), T_i]$ is 1 if $c(I_k) \in T_i$ and 0 otherwise. $\sigma = \frac{1}{N_I} \sum_{k=1}^{N_I} d_k$. After calculating relevance scores for all the communities, we keep the top N_T with the largest relevance scores. The concepts included in these concept communities are aggregated and denoted by $C' = \{c_i'\}_{i=1}^{N_{C'}}$.

1.2 Concept-level Estimation

[0060] The results of community-level estimation enable us to focus on a small subset of concepts C'. In order to best identify the most relevant concepts out of C', we jointly leverage two sources of information. The first source is the relevance score derived from the ranking list returned by the baseline retrieval algorithm. Similar to Section 5.1, we compute the initial relevance score for each concept $c_i' \in C'$ as:

$$s(c_i') = \sum_{k=1}^{N_I} exp(\frac{-d_k}{\sigma}) \mathbf{1}[c(I_k) = c_i'], \quad (3)$$

Where $\mathbf{1}[\cdot]$ is the indicator function, and σ is the same as that in Equation 3. As $s(c_i')$ is not sufficiently reliable, we introduce the second source of information---correlations between semantic concepts---to refine the noisy relevance score. To this end, we further construct a

graph $G'(V',E',W')$ by extracting a subgraph from the VSCN, where V' are nodes corresponding to C', E' are edges with both nodes in V', and W' are the weights associated with E'. To integrate the two information sources, we conduct a Random Walk with Restart (RWR) on G', characterized by

$$p^{n+1} = \alpha \mathbf{P}^T p^n + (1-\alpha)\pi, \qquad (4)$$

where p^n is the walker's probability distribution over V' at step n. $\mathbf{P}$ is the transition matrix derived from W' and $\pi(i) = s(c_i') / \sum_i s(c_i')$. The physical meaning of Equation 5 can be interpreted as, at each step, the random walker either walks, with probability α, along the E' according to the transition matrix $\mathbf{P}$ or restarts, with probability $1-\alpha$, from a fixed probability distribution π. Therefore, the two information sources, incorporated into the two terms on the r.h.s. of Equation 5, respectively, are combined by RWR up to the balance factor α.

The equilibrium distribution p of the RWR is known as the personalized PageRank vector, which has the following analytical solution:

$$p = (1-\alpha)(\mathbf{I} - \alpha \mathbf{P}^T)^{-1}\pi, \qquad (5)$$

where a larger probability in p indicates higher relevance of the corresponding node. We rank the semantic concepts according to their probability values in p, and take the top N_C to represent the semantic meaning of the query image.

[0061] At Step s604, images of the top N_C concepts are gathered and formed a re-ranking pool of the gathered images, which are matched with the query image.

[0062] In one aspect of the present application, there is disclosed a new browsing scheme that helps users explore the VSCN 100 and find images of interest is proposed. The user starts browsing by entering a query keyword to the system. Since the size of the VSCN is huge, it provides local views. This scheme allows users to browse two spaces---the query space and the local concept space---each of which only presents a small subgraph of the entire

VSCN 100. A query space visualizes semantic concepts generated by the same query. For example, the query space of "apple" contains concepts such as "apple fruit", "apple iphone", "apple pie", and their corresponding images. A local concept space visualizes a centric concept (e.g., "apple iphone") together with its neighbor concepts (e.g. "htc diamond" and "palm pixi"), which may come from different query keywords. In this way, it bridges images of most related concepts and helps users access more images of interest without being limited by their initial queries.

[0063] In the browsing process, users can freely switch between the two spaces. A user who chooses a particular concept in the query space enters into the local concept space and the chosen concept becomes the centric concept. The user can then move to a new concept space by choosing a neighboring concept. If the user chooses the centric concept in a local concept space, he will move back to the query space the centric concept belongs to. In this way, users can navigate over the VSCN and search for target images.

[0064] The embodiments of the present invention may be implemented using certain hardware, software, or a combination thereof. In addition, the embodiments of the present invention may be adapted to a computer program product embodied on one or more computer readable storage media (comprising but not limited to disk storage, CD-ROM, optical memory and the like) containing computer program codes.

[0065] In the foregoing descriptions, various aspects, steps, or components are grouped together in a single embodiment for purposes of illustrations. The disclosure is not to be interpreted as requiring all of the disclosed variations for the claimed subject matter. The following claims are incorporated into this Description of the Exemplary Embodiments, with each claim standing on its own as a separate embodiment of the disclosure.

[0066] Moreover, it will be apparent to those skilled in the art from consideration of the specification and practice of the present disclosure that various modifications and variations can be made to the disclosed systems and methods without departing from the scope of the disclosure, as claimed. Thus, it is intended that the specification and examples be considered as exemplary only, with a true scope of the present disclosure being indicated by the following claims and their equivalents.

What is claimed is:

1. A visual semantic complex network system, comprising:

a collection device configured to retrieve a plurality of images and a plurality of texts associated with the images in accordance with given query keywords;

a semantic concept determination device configured to determine semantic concepts and representative images of the retrieved texts and retrieved images , respectively;

a descriptor generation device configured to, from the determined semantic concepts and representative images, generate text descriptors and visual descriptors; and

a correlation device configured to determine semantic correlations and visual correlations from the generated text and visual descriptors, respectively, and to combine the determined semantic correlations and the determined visual correlations to generate the visual semantic complex network system.

2. A system according to claim 1, wherein the system comprises a K -nearest-neighbor graph network system, K is an integer.

3. A system according to claim 2, wherein the correlation device is configured to generate a a semantic and a visual correlation between each two of the concepts, respectively.

4. A system according to claim 3, wherein the descriptor generation device comprises a text descriptor generation unit configured to,

collect a plurality of text snippets corresponding to the semantic concepts,

determine a term frequency vector of the collected snippets;

truncate the determined vector such that a plurality of terms in the vector with the highest term frequency is maintained; and

normalize the truncated vectors to generate said text descriptors, such that the semantic correlation between each two of the semantic concepts are generated from the normalized vector.

.

5. A system according to claim 3, wherein the descriptor generation device comprises a visual descriptor generation unit configured to,

encode each of the representative images by a hashing function so as to generate a binary vector for each retrieved image,

accumulate the generate binary vectors; and

quantize the accumulated vector back to a binary vector as said visual descriptor, such that the visual correlation between each two of the concepts are generated from the binary vector.

6. A system according to claim 5, wherein

the semantic correlations and the visual correlations are combined to generate the visual semantic complex network system.

7. A method for forming a visual semantic complex network system for Web images, comprising:

retrieving a plurality of images and a plurality of texts associated with the images in accordance with given query keywords;

determining semantic concepts and representative images of the retrieved texts and retrieved images , respectively;

generating, from the semantic concepts and the representative images, text descriptors and visual descriptors; and

determining semantic correlations and visual correlations from the generated text descriptor and the generated visual descriptor, respectively,

combining the semantic correlations and visual correlations to generate the visual semantic complex network system.

8. A method according to claim 7, wherein the step of determining semantic correlations and visual correlations from the generated text descriptor and the generated visual descriptor comprises:

generate a semantic and a visual correlation between each two of the concepts.

9. A method according to claim 8, wherein the step of generating a semantic correlation between each two of the semantic concepts comprises:

collecting a plurality of text snippets corresponding to the semantic concepts,

determining a term frequency vector of the collected snippets;

truncating the computed vector such that a plurality of terms in the vector with the highest term frequency is maintained; and

normalizing the truncated vectors to generate said text descriptors, such that the semantic correlation between each two of the semantic concepts are generated from the normalized vectors.

.

10. A method according to claim 8, wherein the step of determining the semantic correlations comprises:

encoding each of the retrieved images by a hashing function so as to generate a binary vector for each retrieved image,

accumulating the generate binary vectors;

quantizing the accumulated vector back to a binary vector as said visual descriptor; and

determining the semantic correlation from the truncated vectors.

11. A method according to claim 8, wherein the step of determining the visual correlations comprises:

encoding each of the representative images by a hashing function so as to generate a binary vector for each retrieved image,

accumulating the generate binary vectors;

quantizing the accumulated vector back to a binary vector as said visual

descriptor; and

determining the s visual correlation from the truncated vectors.

12. A method according to claim 11, wherein the step of combining further comprises:

combining the semantic correlations and visual correlations to generate the visual semantic complex network system.

13. A method for searching images with the visual semantic complex network system, comprising:

obtaining a list of images according to a given query image;

determining a group of related concept communities from the obtained list;

determining a group of related semantic concepts from the determined communities;

gathering, from the determined related semantic concepts, images of a top plurality of concepts; and

forming a re-ranking pool of the gathered images, which are matched with the query image.

14. A method for browsing images with semantic concepts, comprising:

entering a query keyword into a display system;

generating a plurality of semantic concepts based on same queries as the entered keyword;

visualizing the generated semantic concepts in a query space of the display system; and

switching the query space to a local concept space of the display unit in response to selecting a concept selected from the query space, wherein on the local concept space the selected concept together with its neighbor concepts are shown.

15. A method according to claim 14, further comprising:

selecting a centric concept in the local concept space; and

switching back to the query space that the selected concept belongs to.

selecting another concept in the local concept space; and

switching to another local concept space where said another concept and its neighbor concepts are shown.

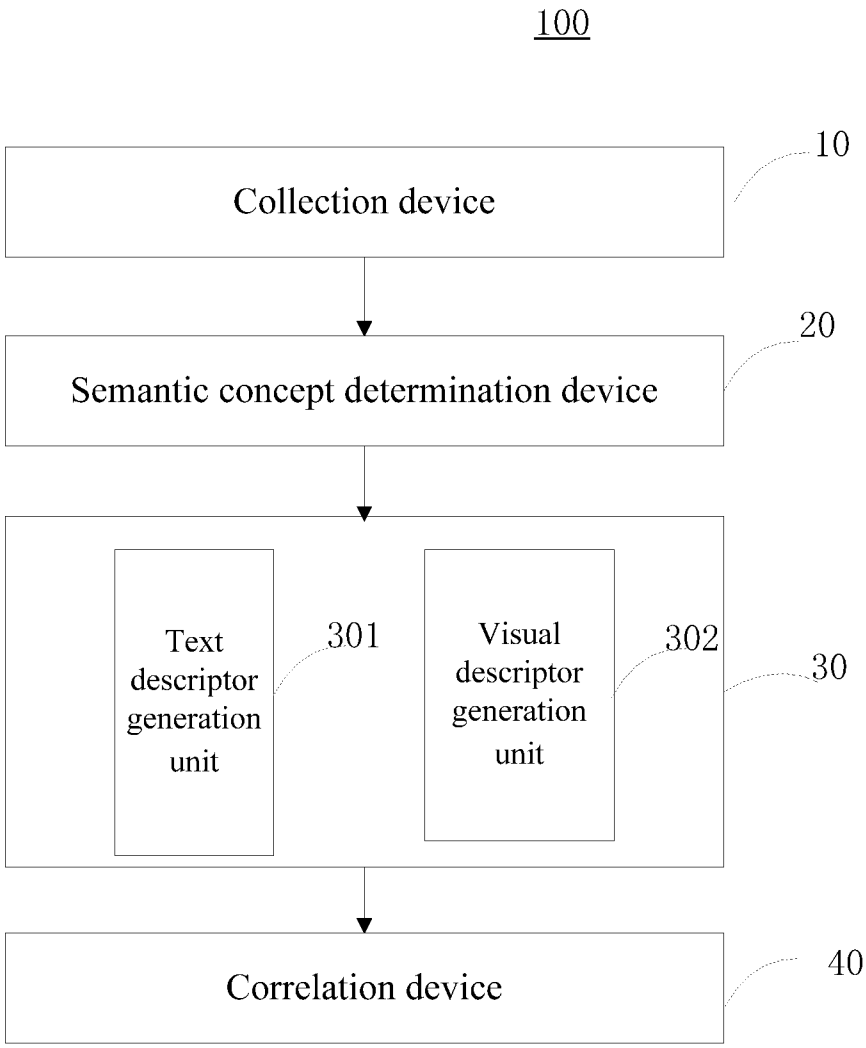

100
Collection device
10
Semantic concept determination device
20
Text descriptor generation unit
301
Visual descriptor generation unit
302
30
Correlation device
40

Fig.1

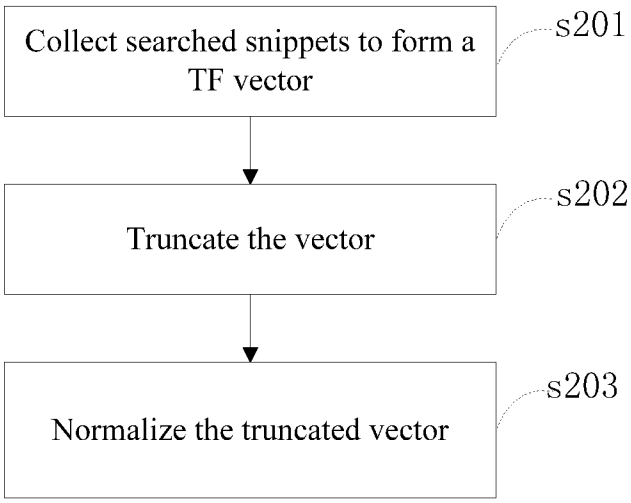

Collect searched snippets to form a TF vector
s201
Truncate the vector
s202
Normalize the truncated vector
s203

Fig. 2

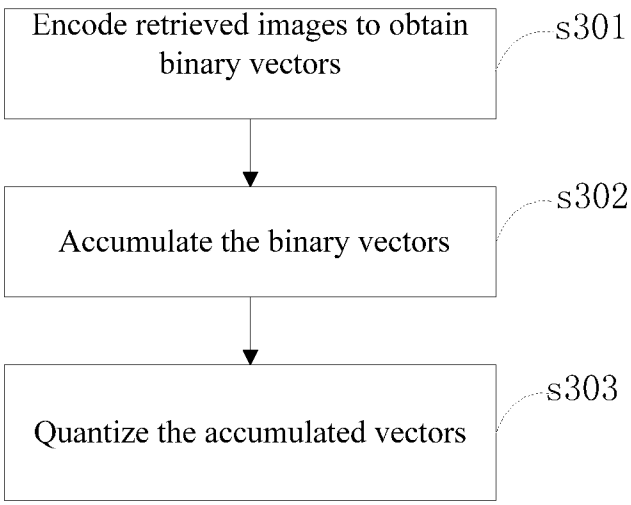
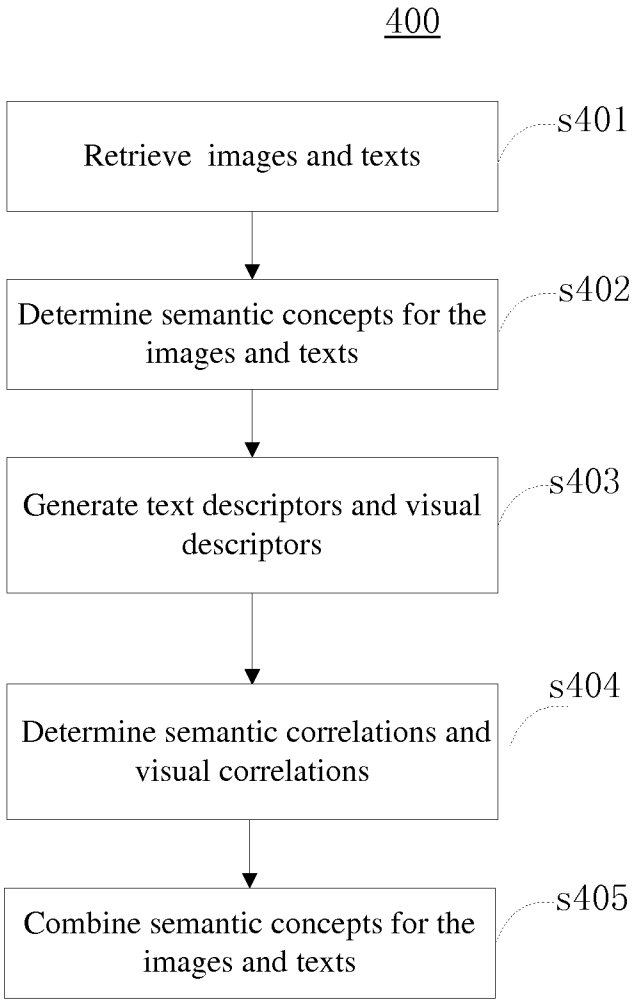

Encode retrieved images to obtain binary vectors
s301
Accumulate the binary vectors
s302
Quantize the accumulated vectors
s303

Fig. 3


400
Retrieve images and texts
s401
Determine semantic concepts for the images and texts
s402
Generate text descriptors and visual descriptors
s403
Determine semantic correlations and visual correlations
s404
Combine semantic concepts for the images and texts
s405

c

Fig. 4

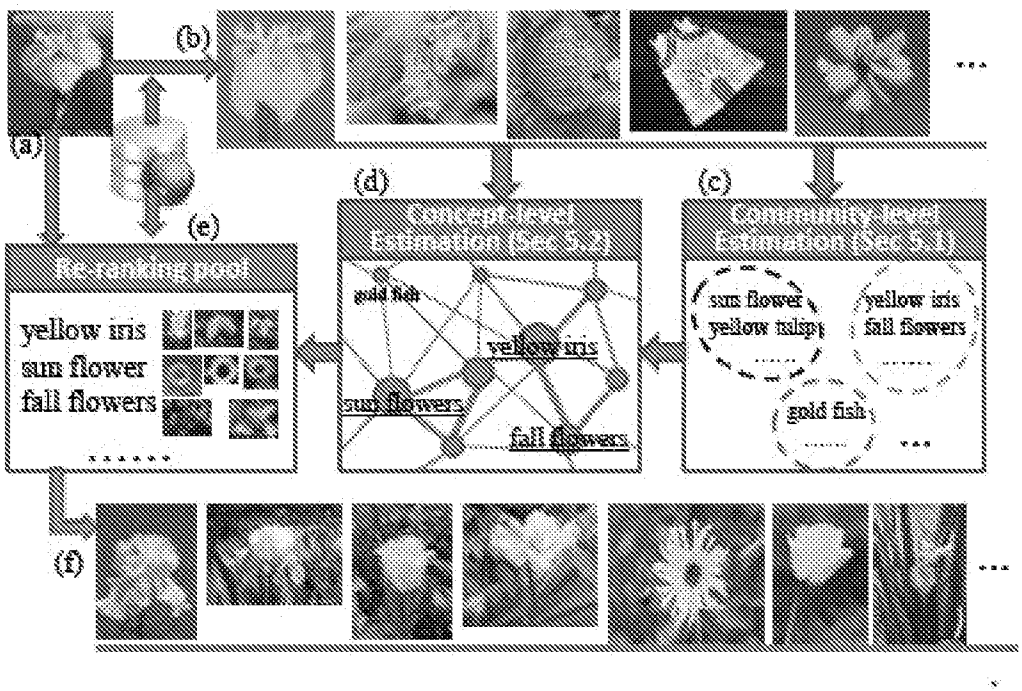
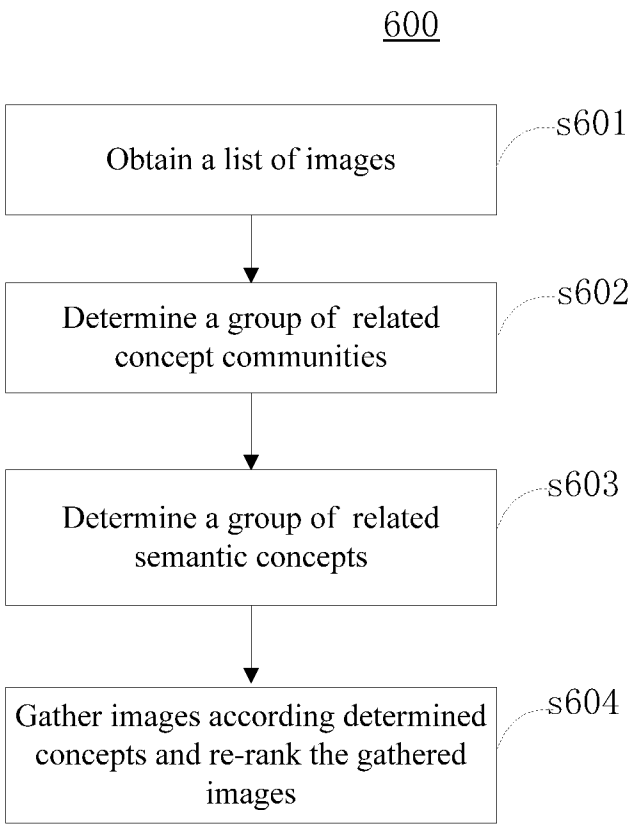

(a)
(b)
(c)
(d)
(e)
(f)
Concept-level Estimation (Sec. 5.2)
Community-level Estimation (Sec. 5.1)
Re-ranking pool
yellow iris
sun flower
fall flowers
gold fish
yellow iris
sun flowers
fall flowers
sun flower
yellow tulip
yellow iris
fall flowers
gold fish

Fig.5


600
Obtain a list of images
s601
Determine a group of related concept communities
s602
Determine a group of related semantic concepts
s603
Gather images according determined concepts and re-rank the gathered images
s604

c

Fig.6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/088258

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: semantic, image?, text?, concept?, descriptor?, correlat+, visual, search+, rank+, re-rank+, query, space

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Xiaogang Wang et.al. "Query-specific visual semantic spaces for web image re-ranking" *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on,* 20 June 2011 (2011-06-20), section 2. Approach Overview, figure 2	13-15
Y	Xiaogang Wang et.al. "Query-specific visual semantic spaces for web image re-ranking" *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on,* 20 June 2011 (2011-06-20), section 2. Approach Overview, figure 2	1-12
Y	CN 101751447 A (CHINESE ACAD SCI AUTOMATION INST) 23 June 2010 (2010-06-23) description, paragraphs [0006]-[0011]	1-12
A	CN 101751447 A (CHINESE ACAD SCI AUTOMATION INST) 23 June 2010 (2010-06-23) the whole document	13-15
A	CN 102902821 A (UNIV BEIJING POSTS&TELECOM) 30 January 2013 (2013-01-30) the whole document	1-15

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 August 2014	**24 September 2014**
Name and mailing address of the ISA/ **STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA(ISA/CN)** **6,Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088 China** Facsimile No. **(86-10)62019451**	Authorized officer **SUN,Weiwei** Telephone No. **(86-10)62412076**

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011072048 A1 (MICROSOFT CORP) 24 March 2011 (2011-03-24) the whole document	1-15
A	US 7996762 B2 (MICROSOFT CORP) 09 August 2011 (2011-08-09) the whole document	1-15

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/088258

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 101751447 A	23 June 2010	None	
CN 102902821 A	30 January 2013	None	
US 2011072048 A1	24 March 2011	US 8392430 B2	05 March 2013
US 7996762 B2	09 August 2011	US 2009083010 A1	26 March 2009