



- (51) **International Patent Classification:**
G06K 9/46 (2006.01) *G06K 9/62* (2006.01)
- (21) **International Application Number:**
PCT/IB2011/001459
- (22) **International Filing Date:**
18 May 2011 (18.05.2011)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant** (for all designated States except US): **LTU TECHNOLOGIES** [FR/FR]; 132 rue de Rivoli, F-75001 Paris (FR).
- (72) **Inventors; and**
- (75) **Inventors/Applicants** (for US only): **DOLLE, Simon** [FR/FR]; 1 rue Minard, F-92130 Issy Les Moulineaux (FR). **MASSOUDI, Ayoub** [FR/FR]; 16 rue de Louvain, F-92400 Courbevoie (FR). **ALLUSSE, Yannick** [FR/FR]; 32 rue la Fontaine, F-75016 Paris (FR). **JAHARD, Frédéric** [FR/FR]; 20 rue du Midi, F-94300 Vincennes (FR). **WINTER, Alexandre** [US/US]; 865 President Street, Brooklyn, New York 11215 (US).
- (74) **Agent:** **MAILLET, Alain**; 5 place Newquay, B.P. 70250, F-35802 Dinard Cedex (FR).

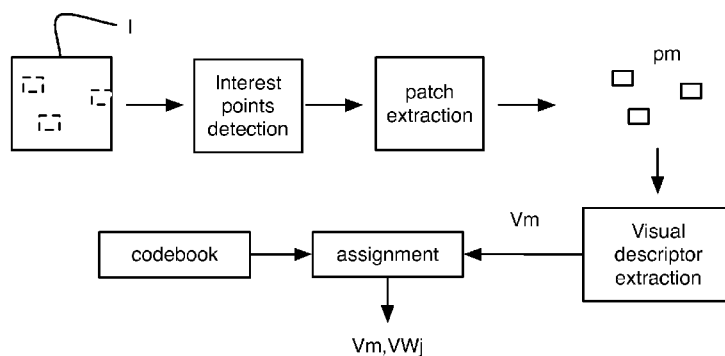
(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE,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

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

Published:

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

(54) **Title:** METHOD AND APPARATUS FOR DETECTING VISUAL WORDS WHICH ARE REPRESENTATIVE OF A SPECIFIC IMAGE CATEGORY



(57) **Abstract:** The present invention concerns a method for detecting visual words which are representative of a specifier image category from a reference codebook obtained from visual local descriptors extracted from training images, said method comprising the following steps: - forming (1) a general training set (S) which gathers images belonging to multiple image categories, - forming (2) a specific category training set (D) which gathers images belonging to the specifier image category to detect, - obtaining (3) a reference codebook from visual local descriptors extracted from images of said general training sets (S), - Computing (4) a distribution of visual words of the reference codebook over the specifier category training set and a distribution of visual words of the reference codebook over the general training set (S), and - determining (5) which visual words of the reference codebook are representative of the specific image category by comparing the two distributions of visual words. The present invention concerns also an apparatus, a computer program and an information storage means.

Method and apparatus for detecting visual words which are representative of a specific image category

The present invention concerns a method for detecting visual words which are representative of a specific image category from a visual codebook obtained from visual local descriptors extracted from training images.

5 Because of the rapid growth of networks and the fast evolution of content sharing solutions, people are getting more and more easy access to a variety of numerical documents in an 'Anytime, Anywhere'. This fact raises the challenge of data traceability and authentication.

10 Indeed, a document has cultural and commercial value at the same time. On one hand, from the cultural point of view, the traceability of document can be of great interest in database search. This can be useful in history, human science or any other research domain. It could be interesting, for example, to analyze viewing statistics of a given political speech or documentary film... Thus traceability is an important feature that contributes to the enrichment of documents by making its indexing and retrieval
15 more efficient and exploitable. On the other hand, digital content producers, distributors as well as copyright holders are careful to guarantee the traceability of

their content and protect their commercial rights on distributed content. This legal aspect of documents traceability is fundamental.

Without loss of generality, the aim of document traceability (or authentication) can be:

- 5 • Copyright enforcement: copyright holders need to know where and how their content is used.
- Monitoring: it consists in supervising a multimedia source (web, magazines, newspapers, TV...) and detect any usage of a given document.
- Copy detection: copy detection consists in identifying the original content. The
- 10 main difference between monitoring and copy detection is the nature of transformations that are managed. Indeed, monitoring usually handles soft distortions whereas copy detection handles strong attacks. Thus, copy detection is more general than monitoring.

15 A document may comprise text and/or images and/or video and/or any other multimedia content.

 The traceability of document which includes text means detecting text documents which belong to a specific category, i.e. documents which talk about a same subject-matter. This is usually carried out using text-based matching in which a list of words which are representative of the specific category is built. Such a list is

20 usually called list of stop words in the state of the art.

 For example, L. Hao ("Automatic Identification of Stop Words in Chinese Text Classification," 2008 International Conference on Computer Science and Software Engineering, 2008, pp. 718–722) builds a list of stop words by gathering words which are both frequent in the text of a document and not correlated with existing document

25 categories. The method implies the collection of large sets for various categories and is limited to build a list of stop words for characterising a specific category of documents from text analysis.

 W.J. Wilbur and K. Sirotkin ("The automatic identification of stop words," Journal of information science, vol. 18, 1992, p. 45) identify stop words as words

30 having the same probability of occurring in a pair of related text documents as in random pairs. They collect pairs of related text documents and random pairs of documents and identify words that are almost as frequent in both these two collections. As they consider random pair of documents to detect stop words, their

method is unable to detect words which characterise a specific category. It is only able to detect general stop words.

Traceability means traceability of images but also traceability of any other documents which include, for example, text. Traceability is usually carried out using visual-word-based indexing and matching.

A visual word is an index of a reference codebook which gathers a finite set of visual words previously computed from a set of training images.

A set of visual words is standalone, i.e. it may be used independently of the documents from which this set has been computed. Visuals words may thus be used for different purposes such as false positives removal in image or video retrieval, image or video retrieval recall improvement, image or video classification, image or video annotation, image or video segmentation and image or video clustering. This list of applications is not restrictive but is given only to illustrate the broad scope of the present invention in image and more generally in document retrieval.

J. Sivic and A. Zisserman ("Video Google: A Text Retrieval Approach to Object Matching in Videos," Computer Vision, IEEE International Conference on, Los Alamitos, CA, USA: IEEE Computer Society, 2003, p. 1470) propose a visual-word-based indexing and matching techniques, called the Inverse Document Frequency weighting (IDF), where an histogram is built, each bin of which being relative to a visual word $\mathbf{i}$ computed from a video and weighted by a coefficient $\mathbf{c}(\mathbf{i})$. The coefficient $\mathbf{c}(\mathbf{i})$ is designed to be high if the visual word $\mathbf{i}$ is rare in a reference codebook and low if the visual word $\mathbf{i}$ is frequent in the reference codebook. This approach is useful to handle visual words that are common in all the images of a video but is inefficient to characterise a specific image category. To make an analogy with text retrieval, the IDF weighting scheme will handle correctly words like "the", "has", "like" but category specific words such as "gene" or "cell" will still have a high weight because these words are rare in a text. This will be problematic if a document, similar to another document, already categorised as belonging to biological domain, is looking for because all the documents, which would contain rare words, would be very similar .

H. Jégou, M. Douze, and C. Schmid, ("On the burstiness of visual elements," Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 1169–1176) try to overcome the problems of recurrent visual words (burstiness) with a more sophisticated weighting. They modify the scoring scheme to

take burstiness into account. Their modification takes place at two different levels: image to image comparison and image to base comparison. In both cases, they are unable to capture the specificity of one image category and they never explicitly build a list of visual words which is relative to a specific image category.

- 5 The “Bag of Features” indexation (BoF) is another visual-word-based indexing and matching technique which transposes techniques that have been successfully employed in text retrieval to image retrieval.

 The BoF indexation of an image I , as illustrated in **Fig. 1**, is a process that takes as input the image I and computes a set of visual words VW_j it contains, their position
10 in the image I and eventually other relevant information. It implies the detection of interest points on the image, the extraction of patches, the computation of visual descriptors on each patch and their assignment to visual words of a reference codebook.

 Detecting interest points means computing a set of points from the input image
15 I . The points are usually characterized by their position but also by their scale. The aim of the interest point detection is to detect points that are stable under global or local transformation. Given the image I , an interest point detector is applied on it and the resulting interest points are stored. There are many different interest point detector algorithms in the state of the art : DoG detector (David G. Lowe, "Distinctive image
20 features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2 (2004), pp. 91-110), Hessian Affine... (T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," Foundations and Trends® in Computer Graphics and Vision, vol. 3, 2008, pp. 177–280.)

 Patch extraction consists in extracting small parts of the image around each
25 interest point. Patch extraction takes into account the positions of the points in the image I (as illustrated in Fig. 1 by dotted lines) but also its scale and eventually other parameters obtained from the interest point detection. The extracted patches $\mathbf{pm}$ are then resized to a fixed size.

 A visual descriptor V_m is then computed by gathering a set of features that
30 characterize an extracted patch $\mathbf{pm}$. Visual descriptors are usually vector of floats. They are meant to be compared together to determine how close the neighborhood they represent are close to each other. The comparison is often made with the Euclidean distance or the cosine similarity. A visual descriptor V_m should be both robust to transformations and discriminative. Many different visual descriptors are

available in the state of the art : SIFT, GLOH, CSLBP (K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, 2005, pp. 1615-1630.)

5 The assignment consists in finding the visual words VW_j of the reference codebook that are the most representative of a visual descriptor V_m . Classical algorithms choose the visual words VW_j which visual descriptors V_j are the closest to V_m . We say that the visual descriptor V_m is assigned to the visual word VW_j .

10 Once two images I and J have been indexed using the BoF indexation, a similarity score between them, based on histogram comparison, is computed. For this purpose, a histogram H_I (respectively H_J) of visual words present in the image I (respectively J) is computed and normalized to unit length. H_I (and H_J) has size $|C|$, where $|C|$ designates the cardinality of the reference codebook. The similarity score between the two images I and J is then usually defined as the scalar product of H_I and H_J . Sometimes, the similarity score is refined by using a geometric consistency check
15 which removes the geometrically incoherent pairs of corresponding image parts.

The BoF indexation is problematic in some specific image categories such as faces of scanned documents where the documents contain essentially texts and few and/or small images because BoF indexing such documents leads to assign very few visual words relative to images. Mainly visual words are relative to text parts and
20 represent black corner on a white background, or black edge on a white background, while visual words assigned to image parts such as eye, nose, upper lip are very common in face images. Thus, during the similarity score computation, a lot of possible corresponding pairs of image parts and text parts are generated and most of these pairs are irrelevant to determine the category of the image included in this document. This is problematic since the cosine similarity is no more relevant to sort the images (false correspondences generate much noise). Given a scanned document I as a query, a set of scanned documents R as references and J a near duplicate of I , $J \in R$, many images in R will have approximately the same score as $H_I \cdot H_J$. This is also
25 problematic when the similarity score takes into account the spatial position of visual words because it is often possible to find a spatially coherent set of correspondences among the large set of possible correspondences.
30

This phenomenon reduces both the recall of the algorithm (we are likely to miss the images to find) and its precision (irrelevant images are more likely to be returned).

M. Marszalek ("Past the limits of bag-of-features," rapport de these, 2008, Institut Polytechnique de Grenoble) proposes a method for detecting visual words which are representative of a specific image category. The author considers an image database which is split into training sets which gather each image belonging to a specific category. A codebook of visual words for each of this training set is learnt and the codebooks are concatenated to get a general reference codebook of the image database. The method allows thus to cluster an image database into multiple sets of specific image categories and to get visual words which are representative of each of these image categories, but this method does not allow to detect any new specific image category from the image database without a full re-indexation of the whole database.

The problem solved by the present invention is to remedy the above-cited drawbacks.

To this aim, the present invention concerns a method which comprises the following steps:

- forming a general training set which gathers images belonging to multiple image categories,
- forming a specific category training set which gathers images belonging to a specific image category to detect,
- obtaining a reference codebook from visual local descriptors extracted from images of said general training sets,
- computing a distribution of visual words of the reference codebook over the specific category training set and a distribution of visual words of the reference codebook over the general training set, and
- determining which visual words of the reference codebook are representative of the specific image category by comparing the two distributions of visual words.

The method is advantageous compared to the state of the art because visual words which are specific to an image category are detected from a reference codebook without requiring the re-computation of this reference codebook. Thus, when a new specific training set has been formed, visual words representative of a new image category are detected from this reference codebook without computing again neither the distribution of visual words of the reference codebook over the general training set nor the visual words which are representative of the already identified image categories.

Moreover, the computation of visual words for a specific image category, which is usually done off-line because the computing time is long, shall be done only once.

Another advantage is that the computation of a new set of visual words to characterize an image category requires only a collection of images which belong to this category.

Moreover, the method allows specific processing of visual words of the reference codebook during the detection of visual words such as removal (occurrences of some visual words of the reference codebook are rejected systematically) or weighting (visual word of the reference codebook are associated to different weight to emphasize the meaning of some visual words).

According to another aspect, the present invention concerns an apparatus adapted to implement the aforementioned method.

The present invention also concerns, in at least one embodiment, a computer program that can be downloaded from a communication network and/or stored on a medium that can be read by a computer and run by a processor. This computer program comprises instructions for implementing the aforementioned methods in any one of their various embodiments, when said program is run by the processor.

The present invention also concerns an information storage means, storing a computer program comprising a set of instructions that can be run by a processor for implementing the aforementioned methods in any one of their various embodiments, when the stored information is read by a computer and run by a processor.

The characteristics of the invention mentioned above, as well as others, will emerge more clearly from a reading of the following description given in relation to the accompanying figures, amongst which:

Fig. 1 represents a diagram of the BoF indexing of an image,
Fig. 2 represents a diagram of the method according to the invention,
Fig. 3 represents a chronogram of an embodiment of the method,
Fig. 4 schematically represents architecture of an apparatus according to the present invention.

The present invention is about a method for detecting visual words which are representative of a specific image category but is also applicable for detecting words more generally which are representative of a specific document category, words can be seen for example as text words.

Generally speaking, the method according to the present invention is applicable any time document matching is needed or when a specific process shall be applied on specific documents or on specific parts of documents of a collection.

Fig. 2 represents a diagram of the method according to the invention.

5 The method, for detecting visual words which are representative of a specific image category from a reference codebook obtained from visual local descriptors extracted from training images, comprises the following steps:

- forming (1) a general training set S which gathers images belonging to multiple image categories,
- 10 - forming (2) a specific category training set D which gathers images belonging to the specific image category to detect,
- obtaining (3) a reference codebook from visual local descriptors extracted from images of said general training sets S,
- computing (4) a distribution of visual words of the reference codebook over
15 the specific category training set and a distribution of visual words of the reference codebook over the general training set S, and
- determining (5) which visual words of the reference codebook are representative of the specific image category by comparing the two distributions of visual words.

20 The general training set S and the specific category training set D are both relatively large (a typical size is 100000).

The general training set S contains images that belong to different categories and/or obtained from different devices without any preference.

25 The specific category training set D contains images that belong to a specific image category to detect.

According to an embodiment, the general training set S and the specific category training set D can be built without any loss of generality from an image index of a search engine, from data of any image specialist and/or from a web crawl.

30 According to a characteristic of the invention, a reference codebook VCB is obtained from images of the training set S using the BoF indexing described in the introductive approach. In the following, it is assumed that the cardinality of the reference codebook VCB equals J, i.e. the reference codebook VCB is formed by J visual words VW_j .

According to an embodiment, the visual words, which are representative of the specific image category to detect, are visual words which are much more frequent in the specific category training set D than in the general training set S.

According to a preferred embodiment, the distributions of visual words of the reference codebook VCB over the training sets are computed by:

- computing a first histogram $H_s(j)$ for the general training set S. The value of a bin j of $H_s(j)$ is the frequency of a visual word VW_j of the reference codebook VCB in this set;
 - computing a second histogram $H_d(j)$ for the specific category training set D.
- The value of a bin j of $H_d(j)$ is the frequency of a visual word VW_j of the reference codebook VCB in this set.

Then, the visual words of the reference codebook which are representative of the specific image category are determined by computing for each bin j of the histograms, a score $r(j)$ which indicates the likelihood for each of the visual words of the reference codebook VCB to be representative of the specific image category to detect.

According to an embodiment, the two histograms H_s and H_d are compared bin to bin. For each visual word VW_j of the reference codebook VCB, a score $r(j)$ is computed. For instance, $r(j)$ is defined by: $r(j) = \frac{H_d(j)}{H_s(j)}$. This quantity induces a ranking of the visual words of the reference codebook VCB according to their “affinity” with the specific image category to detect. The visual word with the highest score r is the visual word that is the most characteristic of the specific image category to detect. The visual words representative of this category are then built thanks to this score r .

According to an embodiment, the visual words which are representative of the specific image category to detect are visual words with a score $r(j)$ greater than a given threshold TH or, alternatively, the N (integer) visual words with the highest scores.

Fig. 3 illustrates the former alternative. The cardinality of the reference codebook VCB equals 10 in this example. The black rectangles represent the values of the bin j of H_s and white rectangles represent the values of the bin j of H_d . The threshold TH equals 2. According to this example, the score for the bins 2 and 6 is greater than TH. Consequently, the visual words VW_2 and VW_6 are detected as being representative to the specific image category to detect.

According to another aspect, the present invention concerns an apparatus 30 which comprises means for implementing the aforementioned method for detecting visual words which are representative of a specific image category from a reference codebook VCB.

5 **Fig. 4** schematically represents an architecture of the apparatus 40. According to the shown architecture, the apparatus 40 comprises the following components interconnected by a communications bus 410: a processor, microprocessor, microcontroller or CPU (*Central Processing Unit*) 400; a RAM (*Random-Access Memory*) 401; a ROM (*Read-Only Memory*) 402; a HDD (*Hard-Disk Drive*) 403, or
10 any other device adapted to read information stored on storage means.

CPU 400 is capable of executing instructions loaded into RAM 401 from ROM 402 or from an external memory, such as HDD 403. After the apparatus 400 has been powered on, CPU 400 is capable of reading instructions from RAM 401 and executing these instructions. The instructions form one computer program that causes CPU 400
15 to perform some or all of the steps of the methods disclosed in relation with Fig. 2. Any and all steps of these methods may be implemented in software by execution of a set of instructions or program by a programmable computing machine, such as a PC (*Personal Computer*), a DSP (*Digital Signal Processor*) or a microcontroller; or else implemented in hardware by a machine or a dedicated component, such as an FPGA
20 (*Field-Programmable Gate Array*) or an ASIC (*Application-Specific Integrated Circuit*).

According to an embodiment, the apparatus 40 also comprises at least one communication interface, for example a first wireless communication interface 404 and a wired communication interface 405.

25 The images which form the general training set S and/or the specific category training set D may be stored in the ROM but they may be also, at least partially, obtained from one communication interface by querying, for example, an external database or by using a search engine.

The reference codebook VCB and/or the distributions of visual words and/or the
30 visual words may be stored in the ROM or in external storing means. In the latter case, a communication interface is used to read or store them.

CLAIMS

1) A method for detecting visual words which are representative of a specific image category from a reference codebook obtained from visual local descriptors
5 extracted from training images, said method comprising the following steps:

- forming (1) a general training set (S) which gathers images belonging to multiple image categories,
- forming (2) a specific category training set (D) which gathers images belonging to the specific image category to detect,
- 10 - obtaining (3) a reference codebook from visual local descriptors extracted from images of said general training sets (S),
- computing (4) a distribution of visual words of the reference codebook over the specific category training set and a distribution of visual words of the reference codebook over the general training set (S), and
- 15 - determining (5) which visual words of the reference codebook are representative of the specific image category by comparing the two distributions of visual words.

2) A method according to claim 1, wherein the visual words which are
20 representative of the specific image category to detect are visual words which are much more frequent in the specific category training set (D) than in the general training set (S).

3) A method according to claim 2, in which computing the two distributions
25 comprises:

- computing a first histogram (Hs) for the general training set, the value of a bin of which is the frequency of a visual word of the reference codebook in this set,
 - computing a second histogram (Hd) for the specific category training set, the value of a bin of which is the frequency of a visual word of the reference
30 codebook in this set, and
- comparing comprises
- computing for each bin of the histograms, a score which indicates the likelihood for each of the visual words of the reference codebook to be representative of the specific image category to detect.

4) A method according to the claim 3, in which the visual words which are representative of the specific image category to detect are visual words with a score higher than a given threshold or a predefined number (N) of visual words with the highest scores.

5) Method according to any one of previous claims, wherein the specific training set (D) gathers images obtained from different devices and/or from an image index of a search engine, and/or from a web crawl.

6) An apparatus for detecting visual words which are representative of a specific image category from a reference codebook obtained from visual local descriptors extracted from training images, said apparatus comprising means for:

- forming (1) a general training set (S) which gathers images belonging to multiple image categories,
- forming (2) a specific category training set (D) which gathers images belonging to the specific image category to detect,
- obtaining (3) a reference codebook from visual local descriptors extracted from images of said general training sets (S),
- computing (4) a distribution of visual words of the reference codebook over the specific category training set and a distribution of visual words of the reference codebook over the general training set (S), and
- determining (5) which visual words of the reference codebook are representative of the specific image category by comparing the two distributions of visual words.

7) A computer program characterized in that it comprises program code instructions which can be loaded in a programmable apparatus for implementing the method according to any one of claims 1 to 6, when the program code instructions are run by the programmable apparatus.

8) Information storage means, characterized in that they store a computer program comprising program code instructions which can be loaded in a

programmable apparatus for implementing the method according to any one of claims 1 to 6, when the program code instructions are run by the programmable apparatus.

1/2

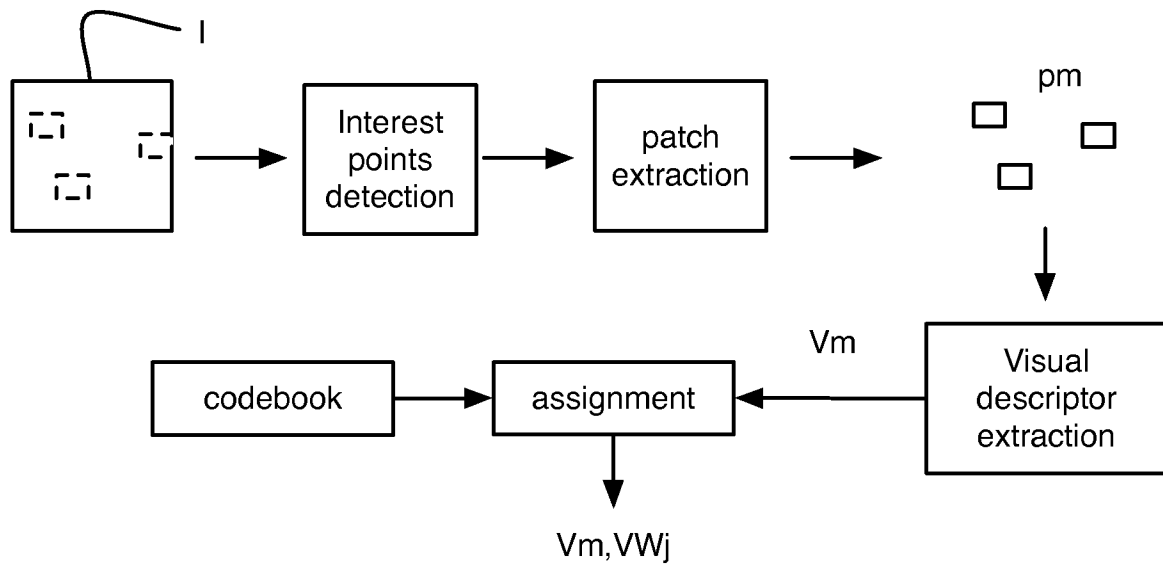


Fig. 1

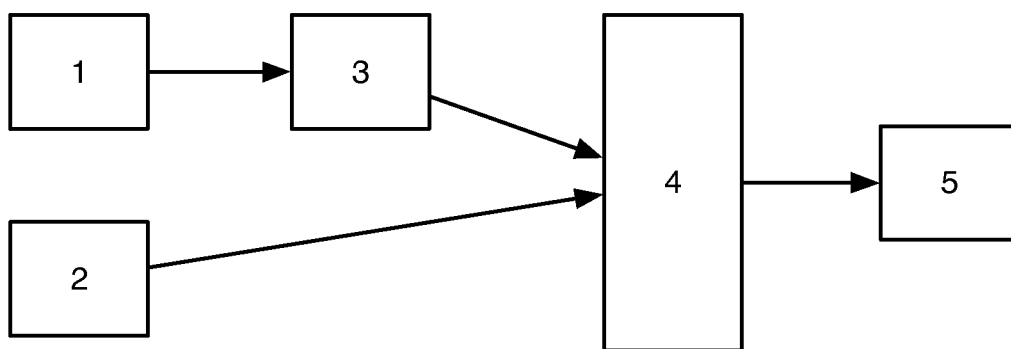


Fig. 2

2/2

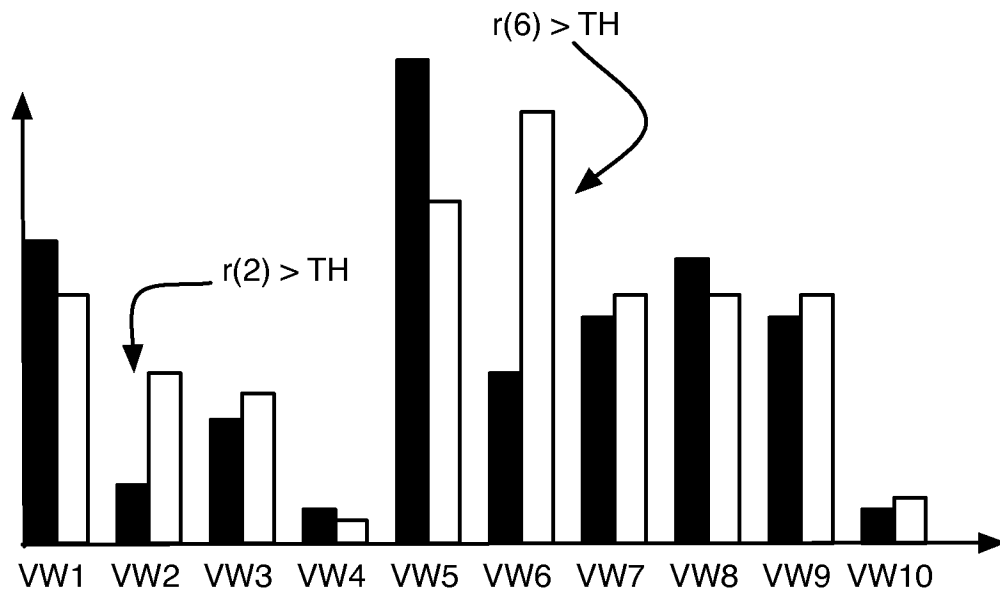


Fig. 3

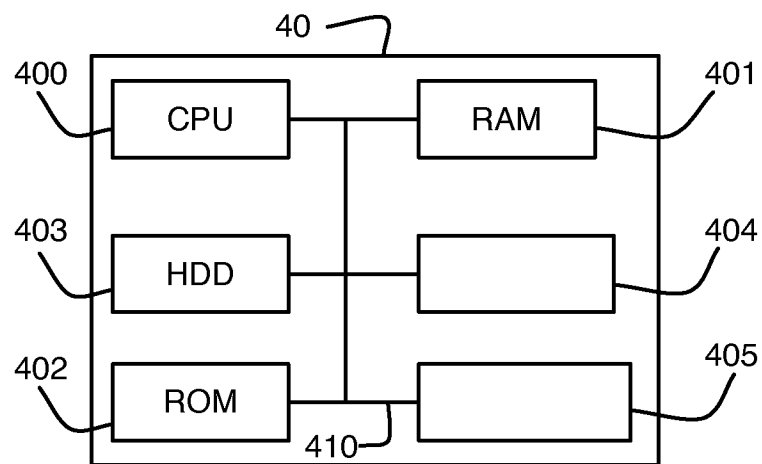


Fig. 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2011/001459

A. CLASSIFICATION OF SUBJECT MATTER INV. G06K9/46 G06K9/62 ADD.		
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) G06K		
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 practical, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SHILIANG ZHANG ET AL: "Descriptive visual words and visual phrases for image applications", PROCEEDINGS OF THE SEVENTEEN ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA, MM '09, 1 January 2009 (2009-01-01), page 75, XP55007017, New York, New York, USA DOI: 10.1145/1631272.1631285 ISBN: 978-1-60-558608-3 Section 3.1 ----- -/--	1-9
☒ 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 document 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	
15 September 2011	23/09/2011	
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Alecú, Teodor Iulian	

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2011/001459

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JIANG HAO ET AL: "Improved bags-of-words algorithm for scene recognition", SIGNAL PROCESSING SYSTEMS (ICSPS), 2010 2ND INTERNATIONAL CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, 5 July 2010 (2010-07-05), pages V2-279, XP031737959, ISBN: 978-1-4244-6892-8 Section II -----	1-9
A	FLORENT PERRONNIN ET AL: "Adapted Vocabularies for Generic Visual Categorization", 1 January 2006 (2006-01-01), COMPUTER VISION - ECCV 2006 LECTURE NOTES IN COMPUTER SCIENCE;;LNCS, SPRINGER, BERLIN, DE, PAGE(S) 464 - 475, XP019036558, ISBN: 978-3-540-33838-3 Sections 2 and 3 -----	1-9
A	EP 1 739 593 A1 (XEROX CORP [US]) 3 January 2007 (2007-01-03) paragraph [0026] - paragraph [0040] abstract; figures 1-6 -----	1-9
A	LEI WANG ED - PHILIP TUDDENHAM ET AL: "Toward A Discriminative Codebook: Codeword Selection across Multi-resolution", CVPR '07. IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION; 18-23 JUNE 2007; MINNEAPOLIS, MN, USA, IEEE, PISCATAWAY, NJ, USA, 1 June 2007 (2007-06-01), pages 1-8, XP031114604, ISBN: 978-1-4244-1179-5 sections 4 and 5 -----	1-9
A	WINN J ET AL: "Object Categorization by Learned Universal Visual Dictionary", COMPUTER VISION, 2005. TENTH IEEE INTERNATIONAL CONFERENCE ON BEIJING, CHINA 17-20 OCT. 2005, PISCATAWAY, NJ, USA, IEEE, vol. 2, 17 October 2005 (2005-10-17), pages 1800-1807, XP010857031, DOI: 10.1109/ICCV.2005.171 ISBN: 978-0-7695-2658-4 Section 4 ----- -/--	1-9

INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2011/001459

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BRIAN FULKERSON ET AL: "Localizing Objects with Smart Dictionaries", 12 October 2008 (2008-10-12), COMPUTER VISION & ECCV 2008; [LECTURE NOTES IN COMPUTER SCIENCE], SPRINGER BERLIN HEIDELBERG, BERLIN, HEIDELBERG, PAGE(S) 179 - 192, XP019109136, ISBN: 978-3-540-88681-5 Section 3.1 -----	1-9
A	Wojcikiewicz ET AL: "Optimizing Visual Vocabularies for Image Classification", 17 November 2009 (2009-11-17), XP55007022, [retrieved on 2011-09-13] cited in the application Sections 4 and 5 -----	1-9

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2011/001459

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
EP 1739593	A1	03-01-2007	US 2007005356 A1	04-01-2007