



(51) International Patent Classification:
G06K 9/00 (2006.01)

(21) International Application Number:
PCT/EP2012/057887

(22) International Filing Date:
30 April 2012 (30.04.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11305717.8 9 June 2011 (09.06.2011) EP

(71) Applicant (for all designated States except US): **THOMSON LICENSING** [FR/FR]; 1-5 rue Jeanne d'Arc, F-92130 Issy-les-Moulineaux (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ALLEAUME, Vincent** [FR/FR]; Technicolor R&d France, 1 Avenue Belle Fontaine, F-35576 Cesson Sevigne Cedex (FR). **SINGH ATEENDRA, Kumar** [IN/FR]; Technicolor R&d France,

1 Avenue Belle Fontaine, F-35576 Cesson Sevigne Cedex (FR). **NARASIMHA, Ramya** [IN/FR]; Technicolor R&d France, 1 Avenue Belle Fontaine, F-35576 Cesson Sevigne Cedex (FR).

(74) Agents: **PERROT, Sebastien** et al.; 1-5 rue Jeanne d'Arc, F-92443 Issy-les-Moulineaux 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, QA, RO, RS, RU, RW, 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,

[Continued on next page]

(54) Title: METHOD AND DEVICE FOR DETECTING AN OBJECT IN AN IMAGE

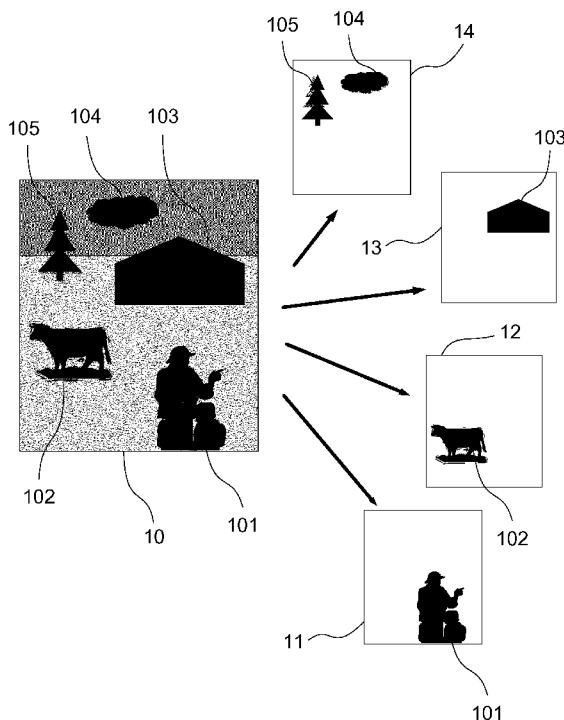


Fig 1

(57) Abstract: The invention relates to a method for detecting a first object comprised in a first image, the first image comprising a plurality of pixels, each pixel being assigned a first grey level information. As to speed up the detection of the first object, the method comprises the steps of segmenting the first object in the first image; assigning a second grey level information to pixels different from pixels forming the first object in the first image and detecting the first object by comparing the segmented first object of the first image with at least a second image representing a first model of the first object, the second grey level information being assigned to pixels of the at least a second image different from pixels forming the first model of the first object in the at least a second image. The invention also relates to a corresponding device.



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).

Declarations under Rule 4.17:

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

Published:

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

METHOD AND DEVICE FOR DETECTING AN OBJECT IN AN IMAGE

1. Scope of the invention.

The invention relates to the domain of object detection in images
5 and more specifically to the domain of object detection implementing a machine learning process.

2. Prior art.

Nowadays, many detection systems based upon computer vision
10 (targeting specific object detection, or more usually face detection) are using some kind of algorithm that usually needs to be very fast for the typical use cases they are involved in.

When based on machine learning approach, building (i.e. "training") such a detection algorithm is usually requiring a very long first
15 initial process, called learning process, that however needs to be done once to set up that detector. During such a typical learning process, the detector is built step by step by using some sets of a so-called "positive" images (i.e. images containing object to be detected, such as faces) on one hand, and some preferably huge sets of "negatives" images (containing all kind of object
20 and background but not the object to be detected) on the other hand. During the training step, the main encountered problem is to provide some relevant images sets. The efficiency of the built detector is so often linked to the number and type of learning images. The positive image set is usually built from gathering hundred or thousand of images including the object the
25 detector will later have to detect. Regarding the negatives images set however, a good set should ideally contain any possible other objects, each with any type of background. That later point is obviously not feasible, as usually the objects' background remains uncontrolled.

30 3. Summary of the invention.

The purpose of the invention is to overcome these disadvantages of the prior art.

More particularly, a particular purpose of the invention is to speed up the detection of an object in an image.

35 The invention relates to a method for detecting a first object comprised in a first image, the first image comprising a plurality of pixels,

each pixel being assigned a first grey level information. The method comprises the steps of:

- segmenting the first object in the first image;
- assigning a second grey level information to pixels different from pixels forming the first object in the first image;
- detecting the first object by comparing the segmented first object of the first image with at least a second image representing a first model of the first object, the second grey level information being assigned to pixels of the at least a second image different from pixels forming the first model of the first object in the at least a second image.

Advantageously, the segmented first object is further compared with at least a third image representing a second model of a second object different from the first object, the second grey level information being assigned to pixels of the at least a third image different from pixels forming the second model in the at least a third image.

According to a particular characteristic, the first object is segmented according to a first depth information associated with pixels of the first image.

In an advantageous manner, depth values associated with pixels forming the first object belong to a first interval of depth values.

According to another characteristic, the segmenting step comprises a step of slicing the first image into a plurality of slices according to depth information, pixels forming the first object belonging to one single slice among the slices.

Advantageously, the method further comprises the steps of:

- segmenting the first model of the first object in the at least a second image according to second depth information associated with pixels of said at least a second image;
- assigning the second grey level information to pixels of the at least a second image different from pixels forming the first model of the first object in the at least a second image;
- storing the at least a second image.

According to another characteristic, the method further comprises the steps of:

- segmenting the second model of the second object in the at least a third image according to third depth information associated with pixels of said at least a third image;

- assigning the second grey level information to pixels different from pixels forming the second model of the second object in the at least a third image;
- storing the at least a third image.

5 The invention also relates to a device configured for detecting a first object comprised in a first image, the first image comprising a plurality of pixels, each pixel being assigned a first grey level information, the device comprising:

- means for segmenting the first object in the first image;
- 10 - means for assigning a second grey level information to pixels different from pixels forming the first object in the first image;
- means for detecting the first object by comparing the segmented first object of the first image with at least a second image representing a first model of the first object, the second grey level information being
- 15 assigned to pixels of the at least a second image different from pixels forming the first model of the first object in the at least a second image.

Advantageously, the segmented first object is further compared with at least a third image representing a second model of a second object different from the first object, the second grey level information being

20 assigned to pixels of the at least a third image different from pixels forming the second model in the at least a third image.

According to a particular characteristic, the device further comprises:

- 25 - means for segmenting the first model of the first object in the at least a second image according to second depth information associated with pixels of said at least a second image;
- means for assigning the second grey level information to pixels of the at least a second image different from pixels forming the first model of
- 30 the first object in the at least a second image;
- means for storing the at least a second image.

According to another characteristic, the device further comprises:

- means for segmenting the second model of the second object in the at least a third image according to third depth information associated with
- 35 pixels of said at least a third image;

- means for assigning the second grey level information to pixels different from pixels forming the second model of the second object in the at least a third image;
- means for storing the at least a third image.

5 The invention also relates to a computer program product comprising instructions of program code for executing steps of the method for detecting the first object, when the program is executed on a computer.

4. List of figures.

10 The invention will be better understood, and other specific features and advantages will emerge upon reading the following description, the description making reference to the annexed drawings wherein:

- figure 1 illustrates a first image segmented into several slices, according to a particular embodiment of the invention,
- 15 - figure 2 illustrates a device implementing a method for detecting a first object in the first image of figure 1, according to a particular embodiment of the invention,
- figure 3 illustrates a method for detecting a first object in the first image of figure 1, according to a particular embodiment of the invention.

20

5. Detailed description of embodiments of the invention.

Invention will be described by reference to a particular and non-limitative embodiment of a method for detecting a first object, which is comprised in a first image. This method provides an efficient solution for speeding up the detection process by removing the background effect during the detection process. According to this embodiment, the first object is segmented in the first image, for example by using a depth information associated to the pixels of the first image or by using the color information associated to the pixels of the image (represented by grey levels) or by using the detection of edges in the first image. A second grey level information is then assigned to pixels of the image which are different from the pixels belonging or forming the first object. The first object is then detected by using its segmented representation with a controlled background, i.e. a background for which the grey level information is controlled and known. To that aim, the segmented first object is compared with second images stored in a data base, which each comprise a representation of a first model of the first object with a controlled background, i.e. the grey level information assigned to

25

30

35

pixels different from the pixels of the second images forming the first model being equal to the second grey level used for the representation of the segmented first object. The assignment of a predetermined grey level information to pixels different from the pixels forming the segmented first object and the first model of the first object enables to speed up the comparison process between the representation of the segmented first object and the second images comprising a model of the first object, the comparison process being focused on the object to be detected and on the model of the object.

According to another aspect of the invention, the purpose of the invention is to provide a specific training and recognition system that remove the objects' background effect during the detection process and/or during the learning process as well.

Figure 1 illustrates a first image 10 comprising several objects, among which some people 101, a cow 102, a house 103, a cloud 104 and a tree 105. At least a first grey level information is assigned to each pixel of the first image. In the case where the first image corresponds to a grayscale image, one grey level information is assigned to each pixel of the first image. In the case where the first image corresponds to a color image, for example a RGB image ("Red, Green and Blue" image), three grey level information are assigned to each pixel, i.e. one grey level information for each color channel R, G, B. Naturally, the number of grey level information assigned to the pixels of the first image may be different from one or three, depending on the representation of the image (for example 4 grey level information for a CMYK ("Cyan, Magenta, Yellow and Black") representation, i.e. a 4 color channels representation). The grey level information is for example coded on 8, 10 or 12 bits. The first image 10 is split into several layers or slices 11, 12, 13 and 14, each comprising one or several of the objects comprised in the first image 10. The first slice 11 comprises the people 101, the second slice 12 comprises the cow 102, the third slice 13 comprises the house 103 and the fourth slice 14 comprises the cloud 104 and the tree 105. The splitting of the first image 10 is advantageously obtained by segmenting the objects 101 to 105 comprised in the first image 10.

The segmentation of the objects is implemented by using a clustering method. According to the clustering method, the first image 10 is first partitioned into N clusters by picking N cluster centers, either randomly

or based on some heuristic. Then, each pixel of the first image 10 is assigned to the cluster that minimizes the distance between the pixel and the cluster center, the distance corresponding to the squared or absolute distance between the pixel and the cluster center, the distance being for example based on the grey level information associated to the pixel and the cluster center. According to a variant, the distance is based on a depth information associated to the pixel and the cluster center, in the case where a depth map or a disparity map is associated to the first image. According to this variant, the depth map or the disparity map is determined from source images (according to any method known by the person skilled in the art) or generated directly during the acquisition of the first image, for example via a depth sensor. Then, the cluster centers are re-computed by averaging all of the pixels of the clusters. The pixels of the first image 10 are then re-assigned to the clusters in order to minimize the distance between each pixel and a re-computed cluster center. The steps of re-computing the cluster centers and re-assignment of the pixels to the clusters are repeated until convergence is obtained, the convergence being obtained for example when no pixel change clusters.

According to a variant, the segmentation of the objects is implemented by using an edge detection method, the edges detected in the first image corresponding to the limits between objects and background for example. The detection of the edges is for example based on the detection of important variation of the grey level values associated to neighbor pixels in a given area of the first image 10. According to a variant, the detection of the edges is based on important variations (i.e. variations more than a threshold value) of depth values associated to neighbor pixels.

Once the objects have been segmented and the slices performed, a second grey level value is assigned to pixels of the slices which do not correspond to pixels of the objects 101 to 105. As an example, the value 0 is assigned to these pixels, which enables to obtain a white background, the pixels of the objects 101 to 105 keeping their original grey level value(s). Of course, another value different may be assigned to the pixels different from the pixels of the objects as to obtain another color for the background (the background corresponding to all the pixels of a slice except the pixels forming the object(s) comprised in the slice).

Figure 2 diagrammatically illustrates a hardware embodiment of a device 2 adapted and configured for the detection of at least an object comprised in the first image 10 and adapted to the creation of display signals of one or several images or layers/slices 11 to 14 of the first image 10. The device 2 corresponds for example to a personal computer PC, to a laptop, to a set top box or to a work station.

The device 2 comprises the following elements, connected to each other by an address and data bus 24, which also transports a clock signal:

- a microprocessor 21 (or CPU),
- a graphical card 22 comprising:
 - several graphic processor units GPUs 220
 - a volatile memory of the GRAM ("Graphical Random Access Memory") type 221,
- a non-volatile memory of the ROM ("Read Only Memory") type 26,
- a Random Access Memory (RAM) 27,
- one or several I/O ("Input/Output") devices 24, such as for example a keyboard, a mouse, a webcam, and so on,
- a power supply 28.

The device 2 also comprises a display device 23 of the type of display screen directly connected to the graphical card 22 for notably displaying the rendering of synthesis images lighted by an environment map which are computed and composed in the graphical card, for example in real time. The use of a dedicated bus for connecting the display device 23 to the graphical card 22 has the advantage of having more important throughput of data transmission, thus reducing latency time for displaying images composed by the graphical card. According to a variant, a display device is outside the device 2 and is connected to the device 2 with a cable transmitting display signals. The device 2, for example the graphical card 22, comprises transmission means or a connector (non illustrated on Figure 2) adapted for the transmission of display signals to external display means such as for example a LCD or plasma screen, a video projector.

It is noted that the word "register" used in the description of the memories 22, 26 and 27 designates, in each of the memories mentioned, a memory zone of low capacity (some binary data) as well as a memory zone of large capacity (enabling a whole program to be stored or all or part of the data representative of computed data or data to be displayed).

When powered up, microprocessor 21 loads and runs the instructions of the program stored in RAM 27.

The random access memory 27 notably comprises:

- 5 - in a register 270, the operating program of the microprocessor 21 loaded at power up of the device 2,
- parameters 271 representative of the first image (for example, grey level information for each pixel and for each color channel, depth information for each pixel),
- 10 - parameters 272 representative of second image(s) (for example, grey level information for each pixel and for each color channel),
- parameters 273 representative of third image(s) (for example, grey level information for each pixel and for each color channel).

15 The algorithms implementing the steps of the method specific to the invention and described below are stored in the GRAM 221 of the graphical card 22 associated with the device 2 implementing these steps. When powering up and once the parameters 271, 272 and 273 representative of the first, second and third images loaded in RAM 27, the
20 GPUs 220 of the graphical card 22 load these parameters in GRAM 221 and executes the instructions of these algorithms under the form of micro-programs such as "shader" using the HLSL ("High Level Shader Language") language, the GLSL ("OpenGL Shading language") language for example.

The GRAM 221 notably comprises:

- 25 - in a register 2210, the parameters representative of the first image 10,
- in a register 2211, the parameters representative of the second image(s),
- in a register 2212, the parameters representative of the third
30 image(s),
- in a register 2213, parameters representative of at least a first object 101 to 105 segmented from the first image 10 (for examples parameters of the pixels of the layer/slice comprising the first object)
- 35 - value(s) 2214 representative of the second grey level information associated to pixels of the slice comprising the first

object different from the pixels forming the first object in the slice.

According to a variant, a part of the RAM 27 is allocated by the CPU 21 for storing the data 2210 to 2214 if memory space available in GRAM 221 is not sufficient. Nevertheless, this variant brings more important latency time in the detection of the first object in the first image composed form the micro-programs comprised in the GPUs as the data have to be transmitted from the graphical card to the RAM 27 through the bus 25 that has transmission capacities generally less than capacities available in the graphical card for transmitting the data from the GPUs to the GRAM and inversely.

According to a variant, the power supply 28 is outside the device 5.

According to a variant, the instructions of the algorithm implementing the steps for detecting the first object in the first image are all performed on CPU only.

Figure 3 illustrates a method for detecting a first object comprised in the first image 10, according to a particular and non limitative embodiment of the invention.

During an initialization step 30, the various parameters of the device 2 are updated. In particular, the parameters representative of the first image are initialized in any manner.

Next, during a step 31, the first object comprised in the first object is segmented, for example by using a clustering method or an edge detection method. The segmentation is advantageously based on depth information associated to the pixels of the first image. According to a variant, the segmentation is based on the grey level information, which is associated to the pixels of the first image. When based on depth information associated to the pixels of the first image, the first object is segmented by selecting the pixels having an associated depth information comprised in a first interval of depth values, i.e. comprised between a minimal depth value and a maximal depth value, as to select the object of the first image located at a given depth. According to a variant, the segmentation of the first image comprises a step of slicing the first image into a plurality of slices, each slice corresponding to a layer of the first image at a given depth. The slicing of the first image enables to classify the objects of the first image according to their depth, i.e.

by grouping foreground objects, background objects and middle-ground(s) objects. According to this variant, the pixels forming the segmented first object belong all to a specific single slice.

Then, during a step 32, a second grey level information is
5 assigned to pixels of the first image different from the pixels forming the first object which has been segmented in step 31. According to a variant, the second grey level is applied to the pixels of the slice comprising the first object, which are different from the pixels belonging to the first object. Such an assignment enables to obtain an image comprising only the segmented
10 first object with a controlled background, i.e. a background with known and controlled parameters.

Then, during a step 63, the segmented first object is compared is compared with one or several second images comprising a first model of this first object. The second images correspond advantageously to so-called
15 positive images used in a machine learning process as to detect an object corresponding to the model represented in the positive images. If a hand of a person is to be detected in an image, the segmented hand of the image is compared to a set of positive images representing different hands of people and forming models of a hand. If the segmented hand match with a majority
20 of the models of the hand comprised in the positive images or with a percentage of the model bigger than a threshold (for example bigger than 60%, 70% or 80%), it means that the segmented object of the image is really a hand. As to speed up the comparison process, the pixels of the second images different from the pixels forming the first model of the first objet to be
25 detected are assigned the second grey level information. The second image is for example obtained by incrusting the first model on an image, the background of which being filled with the second grey level information. It enables to focus the comparison process on the pixels forming the first model, the background of the second image(s) being fully controlled as for
30 the first image comprising the segmented first object.

According to a variant, the segmented first object is also compared to one or several third images comprising second models of second objects, which are all different from the first object. The set of third images form a set of so-called negative images used in a machine learning process. The
35 comparison between the segmented first objet and the second models enables to refine the comparison process. By comparing the segmented first object with an important set of second models of objects different from the

first object and if no second model matches to the first object, then the probability that the first object corresponds to the first model is more important. According to this variant, the pixels of the third images different from the pixels forming the second models are assigned the second grey level information. A third image is for example obtained by incrusting the second model on an image, the background of which being filled with the second grey level information. Thus, less negative images are required for training the detector as the comparison process is focused on the second models and gathering a wide range of second models images with different backgrounds is useless, the background being controlled according to this variant. Reducing the number of third images enables to reduce the number of comparisons and thus speed up the detection of the first image.

According to an advantageous variant of the invention, the segmentation step and the assignment step described above are implemented for the generation of the second images and the third images for supplying the learning machine with positive and negative images. According to this variant, the method further comprises the steps of segmenting the first model of the first object in the second image(s), for example based on depth information associated to the second image(s) or based on grey level values associated to the pixels of the second image(s), in a same way as in the step 31 described previously; of assigning the second grey level information to pixels of the second image(s) which are different from pixels forming the first model in the second image(s), in a same way as in the step 32 described previously; and of storing the second image(s) in registers of a data base. In a same manner, the method further comprises the steps of segmenting the second model(s) of second object(s) different from the first object in the third image(s), for example based on depth information associated to the second image(s) or based on grey level values associated to the pixels of the second image(s), in a same way as in the step 31 described previously; of assigning the second grey level information to pixels of the third image(s) which are different from pixels forming the second model(s) in the third image(s), in a same way as in the step 32 described previously; and of storing the second image(s) in registers of the data base. Applying the same process to the training as to the detection enables to speed up the overall process for detecting an object in an image by using a machine learning process.

A specific and non limitative embodiment of the invention mainly consist in adding and using a depth camera to the vision system used for object acquiring. That depth camera is calibrated and registered to the other color (or grey image) sensor. This set up provides colored (or grey-level) images plus depth information of each image, used for training or detection. Based upon the different depth area detected on combined date images, each "object" (regarding depth range) gets a background-free image from the process described below:

- Group formed from similar depth items are gathered as objects,
- Each object from above step is used to segment its counterpart in the colored (or grey-level) related image, providing a sub-set of the original image,
- The remaining color (or grey-level) of the sub-set image area that do not belongs to the object is colored with a specific color (or grey) value, being defined as a uniform background color.

The resulting image is a segmented object with uniform and controlled background.

Using that background removal knowledge, the training process is now only focusing on differentiating positive objects (faces for instance) from any negative ones (hands, items, ...), with no extra process due to the background specific appearance.

In turn, applying the same segmentation process in the recognition acquiring and process system, then the detection algorithm efficiency is not affected by any background condition being observed during acquiring.

A particular and non limitative embodiment of the invention is a face detector, that uses both color, or grey-level, image and its related depth information (i.e. each pixel of that image has a related depth information) provided by some appropriate mean or determined by using at least two views of a same scene (for estimating for example the disparity between a first image, i.e. for example a left image, and a second image, i.e. a right image, of a stereoscopic image). A efficient mean could be taking as source a device that combines both the depth and color image acquiring (such as a Kinect® device for example).

According to this specific embodiment, the face detector first needs to be build, meaning that it is going to be trained to acquire some

accurate detection rules regarding any object images it has later to recognize as a face, or to discard as a non-face.

As usual in machine learning process, the training process will use a set of “positives” (object being faces) images (second images) and “negatives” ones (objects being anything but faces) (third images). However, the particularity of these sets is the following:

- Each object of the training image as a related depth that is known, or easy to find (for example each object image may be centered to put that object in the center of both the depth and color images),
- 10 - Each training object image (face or not) is computed from the original color image having a well-known (predefined) color applied to any pixel that does not match the centered object regarding its depth area, if any. Typically the background of the image will be “paint” with that well-defined specific color (let’s call it “out-of-object pixel color”).
- 15 - The detector will then follow the training process (usually through iterative steps) using these object images having a perfectly controlled background.

A color (or grey-level) image with related depth information will be provided as input to the detector, that in turn will provided a list of any detected faces coordinates and size, if found.

First, to enable use of the specific detector built as above, the candidate input image (plus related depth information) to be analyzed by the detector is segmented in sub-plane images, depending of the depth area being detected through analysis of the depth information:

- 25 - Each pixel with a close depth are gathered as candidate “object”, in a dedicated plane image, with “out-of-object pixel color” being applied to other pixels of that image. That image could be seen as a “slice” of the original image, containing a depth sliced part of it, with any other object being removed (or “paint” with the specific non-object color).
- 30 - For each depth slice having an object being detected (i.e. non empty slice image), then that image is pass to the detector, that in turn can detect if some face(s) are present in that objects plane.
- As the background (and foreground) of any candidate object image is controlled in the same way than during the learning process, the detector is expected to retrieve faces with the same detection accuracy than during that learning & testing step.
- 35

According to this particular embodiment, a very accurate and background invariant object detector is provided, which is also faster to train than with classical approach as its is requiring less training images.

5 Naturally, the invention is not limited to the aforementioned embodiments.

In particular, the invention is not limited to a method for detecting an object in an image but also extends to any device implementing this method and notably all devices comprising at least a GPU, to the computer
10 program product, which comprises instructions of program code for executing the steps of the method, when said program is executed on a computer and to storage device for storing the instructions of the program code. Implementation of the calculations needed for detecting the first object in the first object is not limited to an implementation in micro-programs of the
15 shader type but also extends to an implementation in every type of program, for example some programs to be executed by a microprocessor of CPU type.

The invention also extends to a method for training a detector used for detecting an object in an image and for supplying the detector with
20 positive and negative images.

CLAIMS

1. Method for detecting a first object comprised in a first image, said first
5 image comprising a plurality of pixels, each pixel being assigned a first grey
level information, characterized in that the method comprises the steps of:
- segmenting the first object in the first image;
 - assigning a second grey level information to pixels different from
pixels forming the first object in the first image;
 - 10 - detecting the first object by comparing the segmented first object of
the first image with at least a second image representing a first model
of said first object, said second grey level information being assigned
to pixels of the at least a second image different from pixels forming
said first model of the first object in the at least a second image.
- 15
2. Method according to claim 1, characterized in that the segmented first
object is further compared with at least a third image representing a second
model of a second object different from the first object, the second grey level
information being assigned to pixels of the at least a third image different
20 from pixels forming said second model in the at least a third image.
3. Method according to one of claims 1 to 2, characterized in that the first
object is segmented according to a first depth information associated with
pixels of the first image.
- 25
4. Method according to claim 3, characterized in that depth values
associated with pixels forming the first object belong to a first interval of
depth values.
- 30
5. Method according to claim 3, characterized in that the segmenting
step comprises a step of slicing the first image into a plurality of slices
according to depth information, pixels forming the first object belonging to
one single slice among said slices.
- 35
6. Method according to one of claims 1 to 5, characterized in that it
further comprises the steps of:

- segmenting the first model of the first object in the at least a second image according to second depth information associated with pixels of said at least a second image;
- assigning the second grey level information to pixels of the at least a second image different from pixels forming the first model of the first object in the at least a second image;
- storing the at least a second image.

7. Method according to claim 2, characterized in that it further comprises the steps of:

- segmenting the second model of the second object in the at least a third image according to third depth information associated with pixels of said at least a third image;
- assigning the second grey level information to pixels different from pixels forming the second model of the second object in the at least a third image;
- storing the at least a third image.

8. Device configured for detecting a first object comprised in a first image, said first image comprising a plurality of pixels, each pixel being assigned a first grey level information, characterized in that the device comprises:

- means for segmenting the first object in the first image;
- means for assigning a second grey level information to pixels different from pixels forming the first object in the first image;
- means for detecting the first object by comparing the segmented first object of the first image with at least a second image representing a first model of said first object, said second grey level information being assigned to pixels of the at least a second image different from pixels forming said first model of the first object in the at least a second image.

9. Device according to claim 8, characterized in that characterized in that the segmented first object is further compared with at least a third image representing a second model of a second object different from the first object, the second grey level information being assigned to pixels of the at least a

third image different from pixels forming said second model in the at least a third image.

10. Device according to one of claims 8 to 9, characterized in that it further
5 comprises:
- means for segmenting the first model of the first object in the at least a second image according to second depth information associated with pixels of said at least a second image;
 - means for assigning the second grey level information to pixels of the
10 at least a second image different from pixels forming the first model of the first object in the at least a second image;
 - means for storing the at least a second image.
11. Device according to claim 9, characterized in that it further comprises:
- 15 - means for segmenting the second model of the second object in the at least a third image according to third depth information associated with pixels of said at least a third image;
 - means for assigning the second grey level information to pixels different from pixels forming the second model of the second object in
20 the at least a third image;
 - means for storing the at least a third image.
12. Computer program product, characterized in that it comprises instructions of program code for executing steps of the method according to
25 one of claims 1 to 7, when said program is executed on a computer.

1/3

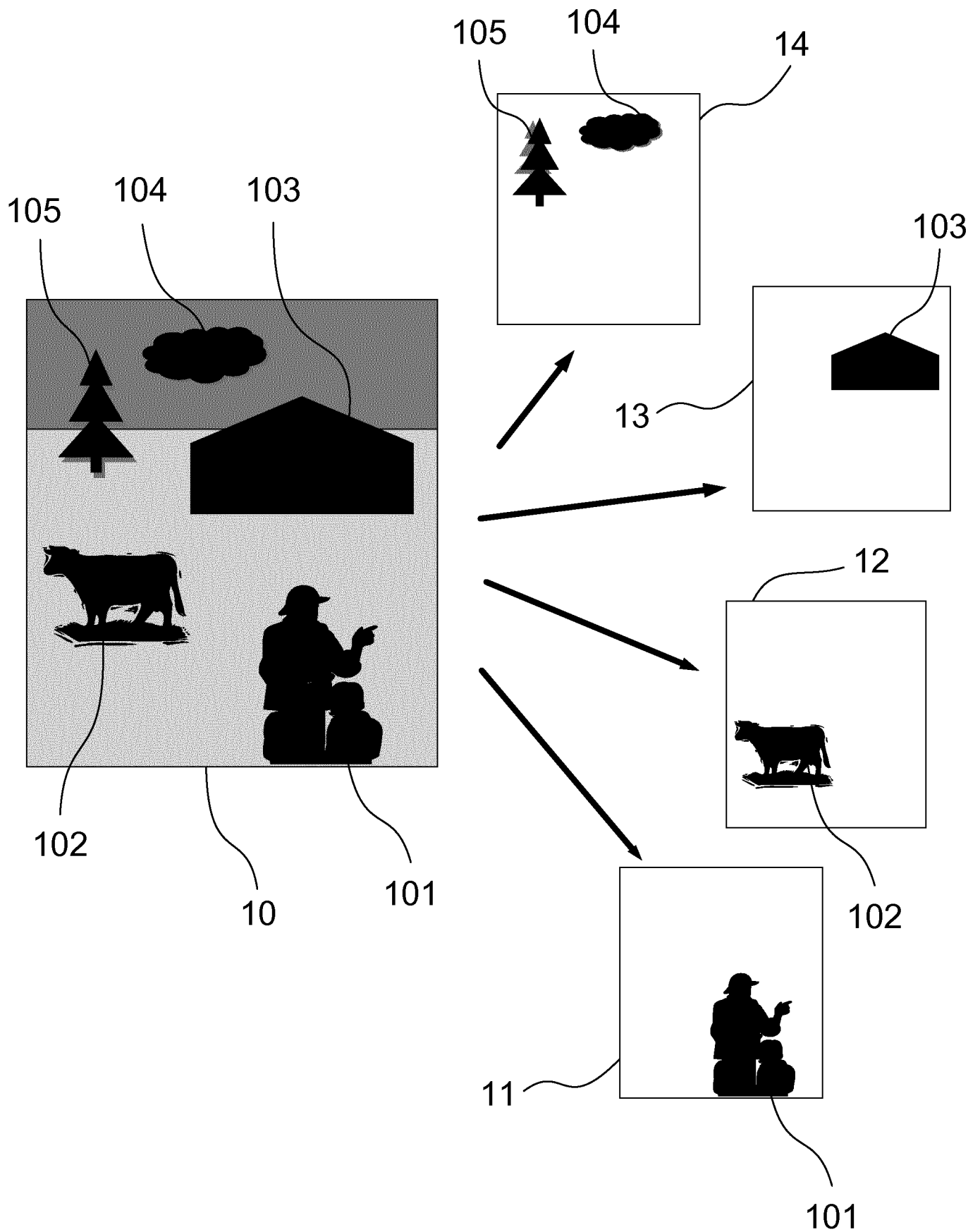


Fig 1

2/3

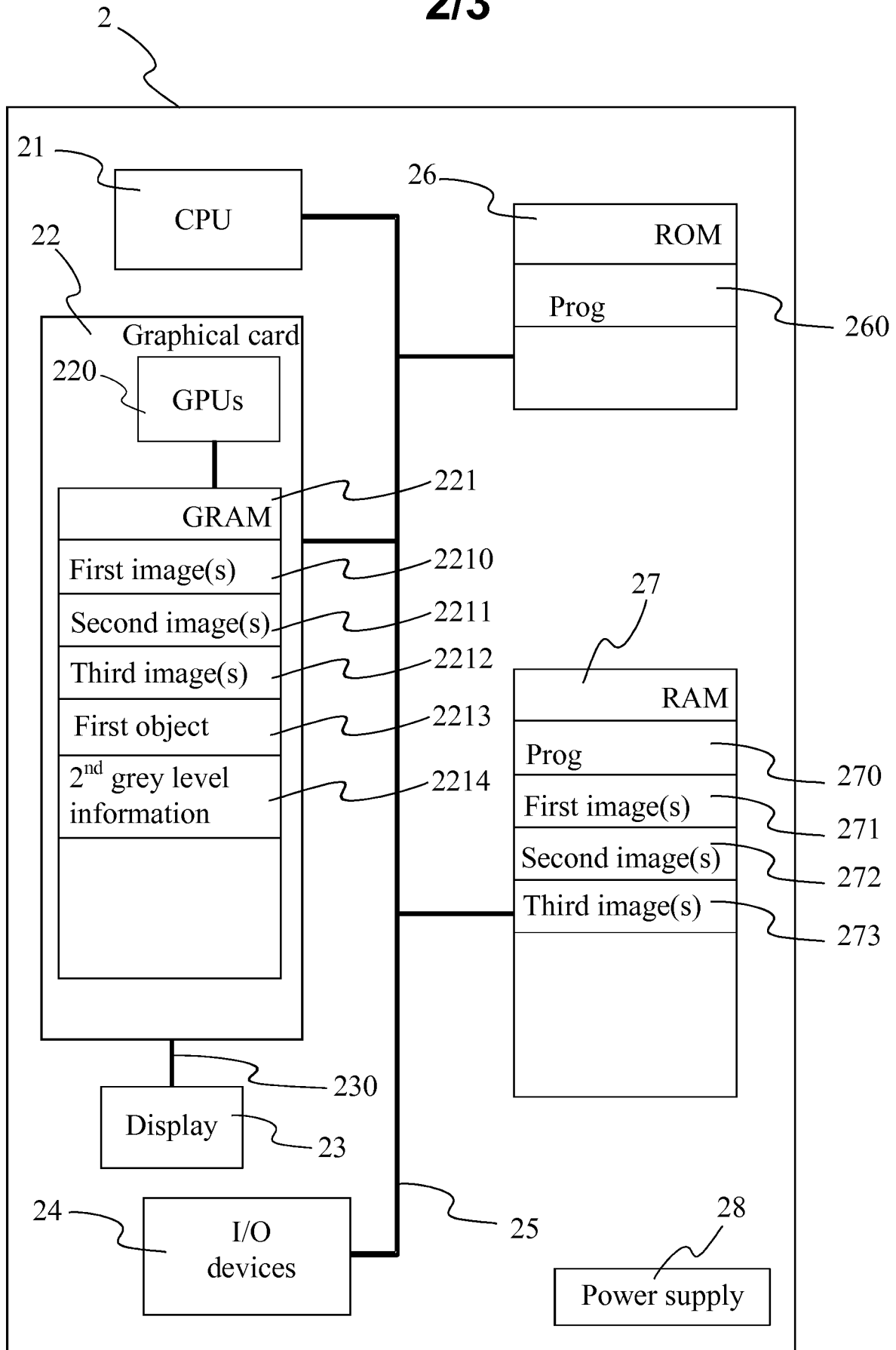


Fig 2

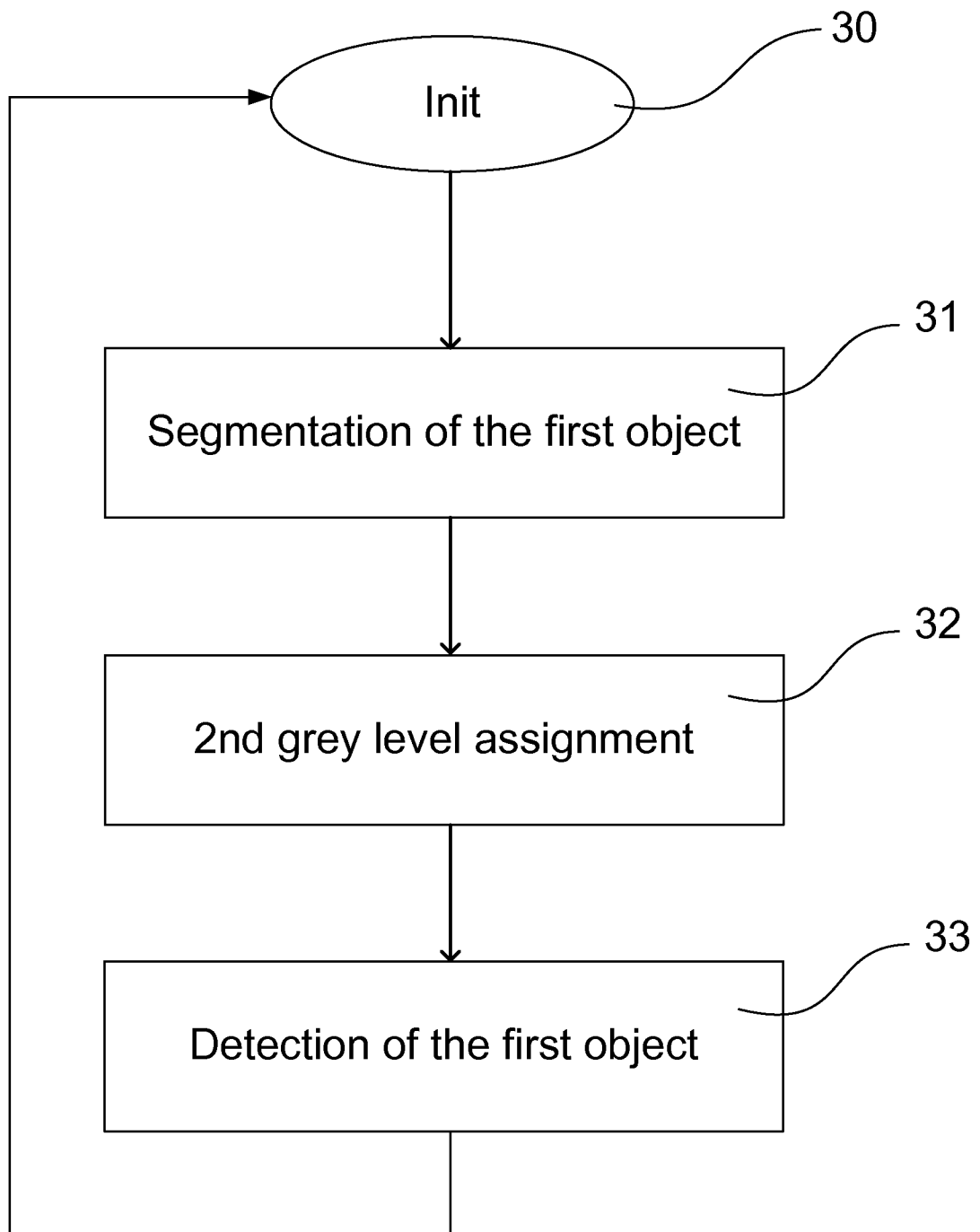
3/3

Fig 3

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/057887

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06K9/00
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 practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/096096 A1 (HUBER ERIC [US]) 20 May 2004 (2004-05-20) paragraphs [0005] - [21;32]; figures 1-7 -----	1-12
X	US 7 542 624 B1 (KOCH MARK W [US]) 2 June 2009 (2009-06-02) columns 1,2,11,12 figures 1-3 -----	1-12
X	David A Forsyth; Jean Ponce: "Computer Vision: A Modern Approach", 1 January 2003 (2003-01-01), Prentice Hall, XP002678726, pages viii-xi; p.240 Fig.11.8; p.314-317; p.467-468; p.481 fig.21.10; p.495-497; p.500-501 ----- -/-	1-12



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

27 June 2012

Date of mailing of the international search report

05/07/2012

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

Darolti, Cristina

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/057887

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011/026764 A1 (WANG SEN [US]) 3 February 2011 (2011-02-03) paragraph [0006] - paragraph [0032]; figures 1-4 -----	1-12

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2012/057887

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2004096096	A1	20-05-2004	NONE
US 7542624	B1	02-06-2009	NONE
US 2011026764	A1	03-02-2011	NONE