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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 January 2016 (28.01.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/012655 A1

(51) International Patent Classification:
G06K 9/00 (2006.01) *G06T 7/00* (2006.01)
G06K 9/62 (2006.01)

(21) International Application Number:
PCT/FI2015/050506

(22) International Filing Date:
21 July 2015 (21.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201410372776.6 25 July 2014 (25.07.2014) CN

(71) Applicant: **NOKIA CORPORATION** [FI/FI]; Karakaari 7, FI-02610 Espoo (FI).

(72) Inventor: **JIANG, Xiaoheng**; Room 1812, Times Building, No. 265, Anshan Road, Nankai District, Tianjin (CN).

(74) Agents: **NOKIA TECHNOLOGIES OY** et al.; Ari Aarnio, IPR Department, Karakaari 7, FI-02610 Espoo (FI).

(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, 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, ST, 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, 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))*

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

(57) Abstract: Embodiments of the present invention disclose a method, apparatus and electronic device of detecting an object in an image. The method comprises: positioning a first sliding window to a start sampling point in the image; repetitively executing following steps until completing a traversal of the image so as to obtain a first group of sampling points including the sampling point positioned by the first sliding window: calculating a response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature of the object; determining a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value; and positioning the first sliding window to the next sampling point; and detecting a position of the object according to the first group of sampling points and their respective response values. The method can effectively shorten time for image search and improve the efficiency of the computer vision.

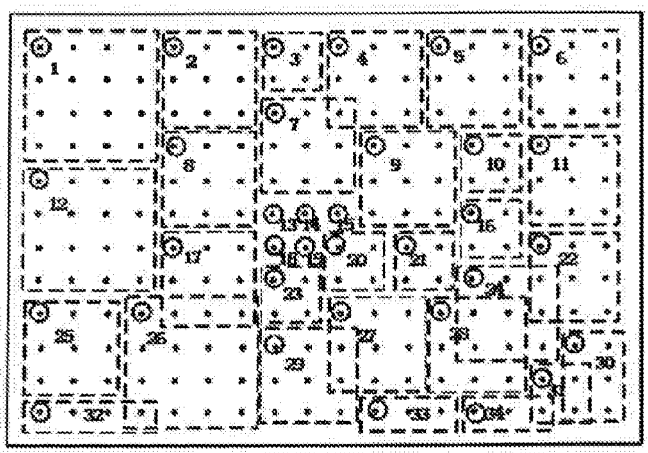

Fig. 3

WO 2016/012655 A1

METHOD, APPARATUS AND ELECTRONIC DEVICE FOR DETECTING OBJECT IN AN IMAGE

FIELD OF THE INVENTION

Embodiments of the present invention relate to the field of computer vision, and specifically to a method, apparatus and electronic device for detecting an object in an image.

BACKGROUND OF THE INVENTION

With the popularity of mobile devices, mobile interactions and services are promoting the broader usage of computer vision in mobile user interface and mobile solutions. For example, gesture tracking and recognition is becoming an important interaction enabler for wearable devices.

However, comparing to human vision, computer vision's efficiency is still so low that it can not yet meet the needs of many practical mobile use cases. The main reasons lie in the following two aspects:
1) Object detection is very timing costly. Unlike human vision which can always searches objects in the vision field in a parallel way, most of prior art computer vision methods are searching objects in the traversal way. The traversal schemes have to adopt exhaustive search steps as objects to be detected are unknown in either scales or orientations.
2) Object identification is not stable and recognition errors happen often. As an object to be recognized are usually unknown in either scale or orientation, the object can not be recognized correctly if the model fitting for the object happens to be missed.

It can be seen that on the one hand, a lot of computation in massive exhaustive search is redundant, and use of the exhaustive search is very time consuming. On the other hand, a lot of models with different scales and orientations are trained, but there are always some scales and orientations which cannot be covered, which makes some objects be missed in the exhaustive search.

SUMMARY OF THE INVENTION

To overcome the above problems in the prior art, embodiments of the present invention are intended to provide a method, apparatus and electronic device for detecting an object in an image.

According to an aspect of the present invention, there is provided a method of detecting an object in an image. The method comprises: positioning a first sliding window to a start sampling point in the image; repetitively executing following steps until completing a traversal of the image so as to obtain a first group of sampling points including sampling points positioned by the first sliding window: calculating a response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature of the object; determining a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value; and positioning the first sliding window to the next sampling point; and detecting a position of the object according to the first group of sampling points and their respective response values. Embodiments of this aspect of the present invention further comprise a respective computer program product and electronic device.

According to another aspect of the present invention, there is provided an apparatus for detecting an object in an image. The apparatus comprises: a start sampling point positioning unit configured to position a first sliding window to a start sampling point in the image; a repetitive execution unit configured to repetitively execute the following units until completing a traversal of the image so as to obtain a first group of sampling points including the sampling point positioned by the first sliding window: a response value calculating unit configured to calculate a response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature of the object; a next sampling point position determining unit configured to determine a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value; and a next sampling point positioning unit configured to position the first sliding window to the next sampling point; and an object position determining unit configured to detect a position of the object according to the first group of sampling points and their respective response values.

It will be appreciated through the following description that according to embodiments of the present invention, adaptive search is introduced, i.e., determine the position of a next sampling point in the image depending on the response value, so that the present invention focuses on potential object regions, avoiding redundant computation on non-object regions, so as to effectively shorten time for image search and improve the efficiency of the computer vision.

BRIEF DESCRIPTION OF THE ACCOMPANYING DRAWINGS

The above and other objects, features and advantages of embodiments of the present invention will be made apparent by reading through the following detailed depictions with reference to figures. In the figures, several embodiments of the present invention are shown in an exemplary and non-restrictive manner, wherein:

Fig. 1 is a schematic view of an electronic device in which an embodiment of the present invention may be implemented;

Fig. 2 is a flow chart of a method of detecting an object in an image according to an embedment of the present invention;

Fig. 3 is a schematic diagram of determining a position of a next sampling point in the image depending on a response value according to an embodiment of the present invention;

Fig. 4 is a first schematic diagram of covering a predetermined scale range of the object through a different number of sliding windows according to an embodiment of the present invention;

Fig. 5 is a second schematic diagram of covering a predetermined scale range of the object through a different number of sliding windows according to an embodiment of the present invention;

Fig. 6 is a third schematic diagram of covering a predetermined scale range of the object through a different number of sliding windows according to an embodiment of the present invention;

Fig. 7 is a schematic diagram of covering a predetermined scale range of the object through a minimum number of sliding windows with a given response value V_0 according to an embodiment of the present invention;

Fig. 8 is a schematic diagram of a specific example having nine sliding windows with a given response value $V_0=0.5$ according to an embodiment of the present invention;

Fig. 9 is a schematic diagram of position distribution with a response value equal to 0.5 when a 80×80 pixel point sliding window approaches an object with the same size according to an embodiment of the present invention;

Fig. 10 is a schematic diagram of inserting an intermediate sliding window according to an embodiment of the present invention;

Fig. 11 is a schematic diagram of an apparatus for detecting an object in an image according to an embedment of the present invention.

In the figures, identical or corresponding reference signs denote identical or corresponding parts.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Hereinafter, several exemplary embodiments shown in the drawings will be referenced to describe the principle and spirit of the present invention. It should be understood that these embodiments are described only for enabling those skilled in the art to better understand and then further implement the present invention, not intended to limit the scope of the present invention in any manner.

As summarized above and discussed in detail hereunder, embodiments of the present invention provide a method of detecting an object in an image, wherein the term “image” comprises a still image such as a picture as well as a dynamic image as in a finder frame of a camera, and wherein the term “object” used herein may comprise an object having a specific color and a specific shape such as a vehicle as well as an object with a non-specific color and a non-specific shape such as a human face, and wherein the object (if any) may be located at any position in the image.

Reference is first made to Fig. 1, in which a block diagram of an electronic device 100 in which exemplary embodiments of the present invention may be implemented therein is presented. According to embodiments of the present invention, the electronic device 100 may be a portable electronic device such as a mobile phone. However, it should be understood that this is only exemplary and non-limitative. Other types of user device may also easily employ embodiments of the present invention, such as a personal digital assistant (PDA), a pager, a mobile computer, a mobile TC, a game device, a laptop computer, a camera, a video camera, a GPS, or other type of voice and text communication system.

The electronic device 100 may have a communication function. To this end, as shown in the figure, the electronic device 100 may comprise one or more antennas 112 operable to communicate with a transmitter 114 and a receiver 116. The electronic device 100 further comprises at least one processor controller 120. It should be understood that the controller 120 comprises circuits needed for implementing all functions of the electronic device 100. For example, the controller 120 may comprise a digital signal processor device, a microprocessor device, an A/D converter, a D/A converter and other support circuits. The control and signal processing functions of the electronic device 100 are distributed based on these devices’ respective capabilities. The electronic device 100 may further comprise a user interface, e.g., a ringer 122, a speaker 124, a loudspeaker 126, a display, or a viewfinder 128 and a keypad 130, all of which are coupled to the controller 120.

In particular, the electronic device 100 comprises a camera 136, a video camera or similar image acquiring devices for capturing static and/or dynamic images. The electronic device may also only have another camera, video camera or similar image acquiring device.

Additionally, the electronic device 100 further comprises a battery 134, such as a vibrating battery pack for supplying power to various circuits needed for operating the electronic device 100 and alternatively providing a mechanical vibration as a detectable output. The electronic device 100 may further comprise a user identity module (UIM) 138. The UIM 138 is generally a memory device having an in-built processor. The UIM 138 may, for example, comprise a subscriber identity module (SIM), a universal integrated circuit card (UICC), a universal user identity module (USIM), or a removable user identity module (R-UIM), etc.

The UIM 138 may comprise a card connection detecting module according to embodiments of the present invention.

The electronic device 100 further comprises a storage device. For example, the electronic device 100 may comprise a non-volatile memory 140, for example, a non-volatile random access memory (RAM) in a cache area for temporarily storing data. The electronic device 100 may further comprise other non-volatile memory 142 that may be either in-built or removable. The non-volatile memory 142 may additionally or alternatively comprise an EEPROM or a flash memory, etc., as an example. The memory may store any item in a plurality of information segments and data used by the electronic device 100, so as to perform functions of the electronic device 100, e.g., functions exemplarily illustrated hereunder with reference to one or more of Fig. 2 through Fig. 11.

It should be understood that the structural diagram in Fig. 1 is only for the purpose of illustration, not intended to limit the scope of the present invention. In some cases, some components may be added or reduced according to the actual needs.

Fig. 2 is a flow chart of a method 200 of detecting an object in an image according to an embedment of the present invention. It may be appreciated that the method 200 is implemented by the electronic device described above with reference to Fig. 1, for example, by a controller 120. For the sake of discussion, the method 200 will be described with reference to the electronic device 100 shown in Fig. 1.

After the method 200 starts, at step S202, a first sliding window is positioned to a start sampling point in the image.

Then, the method 200 proceeds to step S204 of repeating the following step until completing a traversal of the image so as to obtain a first group of sampling points including the sampling point positioned by the first sliding window:

calculating a response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature of the object;

determining a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value; and

positioning the first sliding window to the next sampling point.

The method 200 proceeds to step S206 of detecting a position of the object according to the first group of sampling points and their respective response values.

As stated above in Background Art, use of equidistance exhaustive search causes an efficiency of the computer vision relatively lower. Those skilled in the art may appreciate that the so-called equidistance exhaustive search means that a plurality of points are arranged in an image at a fixed interval (which may be an integer times an interval of pixel points in the image), the sliding windows are respectively positioned to each of these points to calculate the response value, and the response value indicates a matching degree of the pixel point in the sliding window with respect to a predetermined feature of the object. Obviously, in equidistance exhaustive search, a lot of positioning and calculating steps are redundant so that use of the exhaustive search is very time consuming.

To solve the problem, embodiments of the present invention introduces adaptive search, i.e., determining a position of a next sampling point in the image depending on the response value. Specifically, in an embodiment in which the response value increases as the matching degree of the pixel point in the sliding window with respect to a predetermined feature of the object increases, if the response value as calculated at the position of the current sampling point is relatively small, this indicates that a probability of the object in or nearby the sliding window is relatively small. Hence, if it is determined that the position of a next sampling point makes a distance between next sampling point and the current sampling point is relatively longer, i.e., some points between next sampling point and the current sampling point are excluded to avoid the positioning and calculating steps on these points, image searching time can be effectively shortened so as to improve the efficiency of computer vision.

Fig. 3 is a schematic diagram of determining a position of a next sampling point in the image depending on a response value according to an embodiment of the present invention. In Fig. 3, there exist a lot of black dots arranged at a fixed interval, and these black dots may be understood as corresponding to many sampling points arranged at a fixed interval in the prior art. Assume that all sampling points (as shown by circled black dot) are selected from these black dots (although this assumption is not requisite), the above embodiment will be described in detail by taking sampling points 1, 2 and 3 as an example.

Assume that the sampling point 1 is very far away from the object, so the pixel point in the first sliding window matches undesirably with the predetermined feature of the object, and the calculated response value is very small. According to the embodiment of the present invention, since the response value is very small, it is determined that the position of a next sampling point 2 makes a distance between next sampling point 2 and the current sampling point 1 is relatively longer, i.e., a larger number of pixel points (including three black dots) between the next sampling point 2 and the current sampling point 1 are excluded.

Then, assume that the sampling point 2 is nearby the object, so the pixel point in the first sliding window matches better with the predetermined feature of the object, and the calculated response value increases. According to the embodiment of the present invention, since the response value increases, it is determined that the position of a next sampling point

3 makes a distance between next sampling point 3 and the current sampling point 2 is shortened i.e., a smaller number of pixel points (including two black dots) between the next sampling point 3 and the current sampling point 2 are excluded.

Alternatively or additionally, in some embodiments, the above steps are repeated until completing a traversal of the image, i.e., the above steps are repeated until all pixel points in the image all have already been traversed, that is to say, all pixel points have already been positioned by the first sliding window or excluded in the following excluding manner, wherein the excluding manner comprises excluding pixel points between the current sampling point and next sampling point in all pixel points.

Alternatively or additionally, in some embodiments, in order to further shorten time for image search, more pixel points may be further excluded, for example, all pixel points (including 15 black dots therein) in a dotted line frame where the sampling point 1 lies are excluded, wherein the dotted line frame is a maximum square frame in a square frame with a distance between the sampling point 1 and sampling point 2 as an side and including a row and column where the sampling point 1 lies. In this embodiment, as compared with the equidistance exhaustive search method in the prior art in which the sampling is performed 216 (namely 18x12) times in Fig. 3, the method according to the embodiment of the present invention only needs to perform 34 times in Fig. 3 so as to substantially shorten time for image search and improve the efficiency of the computer vision.

Noticeably, although Fig. 3 shows that a sampling order (namely, a predetermined sliding direction of the sliding window) of sampling points is from left to right and from up to down, embodiments of the present invention are not limited to this. Any other sliding directions such as front right to left and from down to up are possible and should all be introduced in the protection scope of the present invention.

Noticeably, the embodiments of the present invention are not limited to the case in which there is only one object applied to the figure. In other words, even though a plurality of objects exist in the figure, embodiments of the present invention can also determine efficiently and precisely. For example, in Fig. 3, besides one object possibly existing at a center, there might exist another object in a right lower corner.

Additionally, those skilled in the art can understand that embodiments for example in which the response value decreases accordingly as the matching degree of the pixel point in the sliding window with respect to the predetermined feature of the object increases. In this embodiment, if the response value calculated at the current sampling point is relatively larger, this indicates a probability of the object in or nearby the sliding window is relatively smaller. Hence, if it is determined that the position of a next sampling point makes a distance between next sampling point and the current sampling point is relatively longer, i.e., more positioning and calculating steps of some points between the next sampling point and the current sampling point are excluded, image searching time can be effectively shortened so as to improve the efficiency of computer vision.

Alternatively or additionally, in some embodiments, a plurality of response value sections may be pre-set, wherein each response value section corresponds to a predetermined distance, respectively, and a predetermined distance corresponding to the response value section with a larger response value is smaller than a predetermined distance corresponding to the response value section with a smaller response value. In this embodiment, the position of a next sampling point may be determined in the following manner depending on the response value. In this manner, first a response value section to which the response value belongs is determined from the plurality of preset response value sections, secondly, a predetermined distance corresponding to the response value section is determined, and finally a position of a next sampling point is determined according to the determined predetermined distance, i.e., enable the distance between next sampling point and the current sampling point to be equal to or is approximate to the determined predetermined distance. In this embodiment, during determination of a next sampling point, a correspondence relationship between the response value section and the predetermined distance is relied on, excessive computation needn't be done for the response value, and the distance may be determined only by for example looking up a table, to effectively shorten time for positioning next sampling point and thereby effectively shorten time for image search and improve the efficiency of the computer vision.

Alternatively or additionally, in some embodiments, detecting the position of the object according to the first group of sampling points and respective response values may achieve the following: determining a response value among these response values greater than a predetermined response value threshold; determining one or more sampling points from the first group of sampling points, wherein the one or more sampling points have a response value greater than the predetermined response value threshold; and determining that the object is located at the one or more sampling points and a scale of the object equals to or is approximate to a scale of the first sliding window. Those skilled in the art can appreciate that this embodiment uses the sliding window with a given scale, the scale of the sliding window may vary according to needs.

In order to suit object of different sizes, embodiments of the present invention may use the above sliding method and use sliding windows with different sizes to sequentially slide in the image. Specifically, these sliding windows with different sizes may be respectively set according to a predetermined proportion based on the size of the first sliding window, wherein the predetermined proportion is set to enable a predetermined scale range of the object can be covered with a minimum number of the plurality of sliding windows, wherein the predetermined scale range is usually a scale range of pre-estimating the object in the image before image processing, e.g., from 30 pixel points to 300 pixel points. How to set the predetermined proportion will be described in detail with reference to Fig. 4 to Fig. 6.

Fig. 4 through Fig. 6 are schematic diagrams about covering the predetermined scale range of the object through a different number of sliding windows, wherein horizontal coordinate x represents a scale of the object, and a longitudinal coordinate y represents a response value. As can be seen from the above, to cover the same predetermined scale range, Fig. 4 uses seven sliding windows (marked as M1 to M7 respectively), Fig. 5 uses three sliding windows (marked as M1, M4 and M7 respectively), and Fig. 6 uses two sliding windows (marked as M1 and M5 respectively).

Obviously, in Fig. 4, since two adjacent sliding windows (e.g., M1 and M2) substantially overlap therebetween, a lot of redundant positioning and calculating steps exist and thereby consume a lot of time. However, on the other hand, in Fig. 6, as an uncovered scale scope exists between the only two sliding windows M1 and M5, it is possible to cause missed detection of an object which is right located in the uncovered scale range. As can be seen from the above, when Fig. 5 uses three appropriate sliding windows M1, M4 and M7 to cover the predetermined scale range (i.e., a minimum number of sliding windows cover the predetermined scale range), massive overlapping will not occur and uncovered scale range does not exist, so that high efficiency of the computer vision and avoidance of missed detection of objects can be simultaneously considered from on the whole.

Then, the position of the object may be detected according to a plurality of groups of sampling points and their respective response values. Specifically, it may achieves the following: determining a response value among these response values greater than a predetermined response value threshold; determining a second group of sampling point from the plurality of groups of sampling points, wherein the second group of sampling points include one or more sampling points having a response value greater than the predetermined response value threshold; and determining that the object is located at the one or more sampling points and a scale of the object equals to or is approximate to a scale of the sliding window associated with the second group of sampling points.

Furthermore, to further avoid missed detection of objects, embodiments of the present invention further propose inserting an intermediate sliding window between the two adjacent sliding windows set according to the predetermined proportion. Meanwhile, to still ensure the high efficiency of the computer vision, this embodiment neither uses the intermediate sliding window for exhaustive search, nor uses this intermediate sliding window to perform adaptive search proposed by the embodiment of the present invention (if the adaptive search is performed, the above-mentioned minimum number will not be satisfied), and uses the intermediate sliding window to only preform positioning and response value calculation at the sampling point selected from the above already-determined plurality of groups of sampling points. As such, the missed detection of the objects may be further avoided almost without adding the time for image search. This embodiment will be describe below in detail through step (1) through step (4).

Step (1): selecting sampling points whose response values are greater than the predetermined response value threshold from the plurality of groups of sampling points, wherein the sampling points are the sampling points already positioned by a respective one of a plurality of sliding windows and have a respective response value, wherein the predetermined response value threshold is set to filter smaller values in the response values corresponding to the plurality of groups of sampling points to avoid using the above manner of inserting the intermediate sliding window at sampling points corresponding to these smaller values so as to avoid increasing too much time for image search.

Step (2): positioning an intermediate sliding window to the position of the sampling point, wherein a proportion of a scale of the intermediate sliding window to a scale of the above corresponding sliding window is closer to one as compared with the predetermined proportion. Those skilled in the art may understand considering that the scale of the plurality of sliding windows might be set to increase gradually or set to decrease gradually, i.e., the predetermined proportion might be greater than one or less than 1, in this embodiment, setting the proportion of a scale of the intermediate sliding window to a scale of the above corresponding sliding window to be closer to one as compared with the predetermined proportion can more clearly define the above intermediate sliding window to avoid unnecessary confusion.

Step (3): determining the response value calculated according to matching degree of the pixel point in the intermediate sliding window with respect to the predetermined feature and a maximum response value in the above corresponding response values.

Step (4): determining that the object is located at the sampling point and a scale of the object equals to or is approximate to a scale of the sliding window corresponding to the maximum response value.

Those skilled in the art may understand that the technical content described from step (3) to step (4) has already been described in the above embodiments of the present invention and will not be detailed here.

Now, return to the description about how to set the predetermined proportion, wherein the predetermined proportion may be further set to enable a response value corresponding to the overlapping portion when two adjacent sliding windows in the above plurality of sliding windows covers the respective scale range of the object to be greater than or equal to the predetermined response value threshold used in the above step (2) so that the two adjacent sliding widows in the plurality of sliding windows can completely cover a respective object scale (i.e., a portion of object scale will not be missed between two sliding windows), as shown in Fig. 7.

To make the above embodiments of the present invention transparent, the present invention further provides a specific embodiment for detecting a human face in an image, which will be described in detail below.

(1) Determine the number of the sliding windows

In this embodiment, assume that an estimated scale range of the human face in the image is between 20 pixel points to 240 pixel points, so, for the sliding window proportion 1.2 usually used by the prior art method, at least 14 sliding windows with different scales ($20\times1.2^L\leqslant 240$, $L\leqslant\log_{1.2}12$) are needed.

However, referring to Fig. 8, for the method proposed by the embodiment of the present invention, a sliding window scale 1.34 is used to obtain nine sliding windows with different scales, which has the same scale range as the prior art method using 14 sliding windows with different scales. The proportion 1.34 is pre-calculated based on the following minimum risk conditions, e.g., assume that a classifier is used to determine a predetermined response value threshold to obtain an object missed detection minimum risk, and therefore $V_0=0.5$ can be determined in Fig. 7 corresponding to the minimum risk and thereby the proportion 1.34 and the sliding windows using nine different scales can be determined based on $V_0=0.5$.

(2) Create a confidence graph corresponding to the sliding window through adaptive search

Regarding the sliding window with a different scale determined in step (1), the embodiment respectively applies the above adaptive search for each sliding window to create the confidence graph corresponding to the sliding window so as to obtain the plurality of groups of sampling points and their respective sampling values.

First, a horizontal step Δx and a vertical step Δy are determined. Specifically, these steps may be based on a scale S of the respective sliding window, and a scale P_1, P_2 of a maximum inscribed rectangle in an isogram as determined when the calculated response value equals to the predetermined response value threshold. For example:

$\Delta x=P_1\times S\times0.5$ (pixel points)

$\Delta y=P_2\times S\times0.5$ (pixel points)

Wherein, referring to Fig. 9, P_1 and P_2 are determined by the extending isogram obtained from response value distribution through a XY plane. In Fig. 9, the ellipse is the isogram corresponding to a scale (e.g., 80x80 pixel points) of the respective sliding window, and the isogram uses a value equal to the predetermined response value threshold, namely V_0=0.5. On this basis, P_1=0.09375 and P_2=0.0625 can be calculated by making statistics on a testing set. Then, Δx can be calculated and its integral part is 3, and Δy can be calculated and its integral part is 2.

Correspondence relationship between the plurality of response value sections and the predetermined distance as shown in the following Table 1 can be built based on the above determined steps. For example, as for the response values in the response value section 0.15 to 0.3 in x direction, 3 times the step may be used, namely nine pixel points.

Table 1: Correspondence relationship between response value sections and the predetermined distance

Response value（RV）	0<RV⩽0.15	0.15<RV⩽0.3	0.3<RV⩽0.5	RV>0.5
Distance in x direction	4×Δx	3×Δx	2×Δx	Δx
Distance in y direction	4×Δy	3×Δy	2×Δy	Δy

Then, as for the above sliding window, if the response value obtained by calculating at a position of the start sampling point is RV (i, j), wherein (i, j) represents coordinates of the point. The position of a next sampling point is decided based on this response value, i.e., the smaller the response value is, the longer next sampling point is; on the contrary, the larger the response value is, e.g., greater than the above V_0, the closer next sampling point will be or is approximate to the current sampling point. If the sliding window is used to implement the adaptive search proposed by the embodiment of the present invention across the whole image, the confidence graph for this sliding window may be built, and the confidence graph may be as shown in Fig. 3.

(3) Insert the intermediate sliding window

First, selecting sampling points whose response values are greater than the predetermined response value threshold from the plurality of groups of sampling points. As shown in Fig. 10, assume that the sampling point (i, j) is a black dot in a sliding window S, and the respective response value is RV (S, i, j), the predetermined response value threshold is less than V_0, and it is 0.4 in this embodiment.

Then, the intermediate sliding window S-1 is inserted. The intermediate sliding window S-1 has a larger scale, and a proportion of the scale to the scale of the sliding window S is closer to one, e.g., equal to 1.1, as compared with the above proportion 1.2. Furthermore, the intermediate sampling window S-1 is positioned to the sampling point (i, j) and calculation is carried out to obtain the respective response value RV (S-1, i, j).

Assume that RV（S-1，i，j）>RV（S，i，j）, according to the embodiment of the present invention, it will determined that an object exists at the position of the sampling point (i, j) and a scale of the object equals to or is approximate to a scale of the intermediate sliding window S-1.

Similarly, an intermediate sliding window S+1 with a smaller scale may be inserted, and its processing manner is similar to the intermediate sliding window S-1 and will not be detailed here.

Additionally, the sliding window S may be positioned to the pixel point (i, k) adjacent to the sampling point (i, j) and calculation is carried out to obtain the respective response value RV (S, i, k). Furthermore, assume that RV（S，i，k）>RV（S，i，j）, according to the embodiment of the present invention, it will determined that an object exists at the position of the sampling point (i, k) and a scale of the object equals to or is approximate to a scale of the intermediate sliding window S.

(4) Performance comparison

The following comparison results can be obtained by respectively applying the conventional equidistance exhaustive search and the adaptive search in the embodiment of the present invention to a MIT+CMU front face test set which includes 130 images.

Table 2: Comparison of results of exhaustive search and adaptive search

Dataset	Approach	times of sliding windows	Misreported missed detection rate	Speedup(times)
MIT+CMU	Exhaustive search	997860（average）	0.091	1.0
	Adaptive search	244610（average）	0.089	2.83

As can be seen from Table 2, in the case of similar detection performance, the adaptive search needs to use the times of using sliding window only about one quarter of the exhaustive search, and exhibits an operation time 2.83 times faster than the exhaustive search.

To conclude, the present invention has the following advantages: first, first in scale domain, a minimum number of sliding windows are used to reduce the times of using the sliding window. Secondly, in space domain, adaptive search is used for each sliding window to reduce the number of the sampling points. Finally, better evaluation is executed in the space domain and the scale domain to ensure performance. Hence, embodiments of the present invention focus on potential object regions, avoiding redundant computation on non-object regions, so as to effectively shorten time for image search and improve the efficiency of the computer vision.

Now referring to Fig. 11, Fig. 11 is a schematic diagram of an apparatus 1100 of detecting an object in an image according to an embedment of the present invention. As shown in Fig. 11, according to the embodiment of the present invention, the apparatus 1100 comprises: a start sampling point positioning unit 1102 configured to position a first sliding window to a start sampling point in the image; a repetitive execution unit 1104 configured to repetitively execute the following units until completing a traversal of the image so as to obtain a first group of sampling points including the sampling point positioned by the first sliding window: a response value calculating unit 11042 configured to calculate a response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature of the object; a next sampling point position determining unit 11044 configured to determine a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value; and a next sampling point positioning unit 11046 configured to position the first sliding window to the next sampling point; and an object position determining unit 1106 configured to detect a position of the object according to the first group of sampling points and their respective response values.

In some embodiments, the next sampling point position determining unit 11044 is configured to determine the position of a next sampling point so that a distance between next sampling point and the current sampling point is longer in the case that the response value is smaller.

In some embodiments, the next sampling point position determining unit 11044 further comprises a response value section determining unit configured to determine a response value section to which the response value belongs from a plurality of response value sections, wherein each of the plurality of response value sections corresponds to a predetermined distance, wherein a predetermined distance corresponding to the response value section with a larger response value is smaller than a predetermined distance corresponding to the response value section with a smaller response value; and a distance determining unit configured to determine a position of the next sampling point so that the distance between the next sampling point and the current sampling point is the predetermined distance corresponding to the response value section to which the response value belongs.

In some embodiments, the predetermined distance corresponding to the response value section with the larger response value is a first integer times a step, the predetermined distance corresponding to the response value section with the smaller response value is a second integer times the step, and the first integer is smaller than the second integer.

In some embodiments, the step is determined according to at least one of: a scale of the first sliding window, and a scale of a maximum inscribed rectangle in an isogram as determined when the calculated response value equals to the predetermined response value threshold.

In some embodiments, the step is an integer times the distance between adjacent pixel points in the image.

In some embodiments, the predetermined sliding direction is from left to right and from up to down.

In some embodiments, completing a traversal of the image comprises: all pixel points in the image have already been positioned by the first sliding window or excluded in the following excluding manner, wherein the excluding manner comprises excluding pixel points located between the current sampling point and next sampling point in all pixel points.

In some embodiments, the excluding manner further comprises excluding, from within the all pixel points, pixel points that are located in a particular square, wherein the particular square includes the current sampling point, and two perpendicular sides of the particular square start from the current sampling point, and wherein a length of each side of the particular square is one or more pixel points less than a distance between the current sampling point and the next sampling point.

In some embodiments, the response value calculating unit 11042 is configured to calculate the response value to enable the response value larger in the case that there are more pixel points in the first sliding window matching the predetermined feature.

In some embodiments, the next sampling point determining unit 11044 is further configured to other groups of sampling points so as to obtain a plurality of groups of sampling points including the first group of sampling points, wherein each of the plurality of groups of sampling points comprises sampling points positioned by a respective one of a plurality of sliding windows, wherein a scale of each of the plurality of sliding windows is respectively set based on the scale of the first sliding window and according to a predetermined proportion; the object position determining unit 1106 is further configured to detect the position of the object according to the plurality of groups of sampling points and their respective response values.

In some embodiments, the predetermined proportion is set to enable a predetermined scale range of the object to be covered with a minimum number of the plurality of sliding windows.

In some embodiments, the predetermined proportion being set to enable a predetermined scale range of the object to be covered with a minimum number of the plurality of sliding windows comprises: the predetermined proportion is set to enable a response value corresponding to an overlapping portion when any two adjacent sliding windows of the plurality of sliding windows cover the respective scale range of the object to be greater than or equal to the predetermined response value threshold.

In some embodiments, the apparatus further comprises: a sampling point selecting unit configured to select sampling points whose response values are greater than the predetermined response value threshold from the plurality of groups of sampling points, wherein the sampling points are the sampling points already positioned by a respective one of a plurality of sliding windows and have a respective response value; a sampling point positioning unit configured to position an intermediate sliding window to the position of the sampling point, wherein a proportion of a scale of the intermediate sliding window to a scale of the respective sliding window is closer to one as compared with the predetermined proportion; and an response value comparing unit configured to compare the respective response value with the response value calculated according to matching degree of the pixel point in the intermediate sliding window with respect to the predetermined feature; the object position determining unit 1106 is further configured to determine that the object is located at the sampling point and a scale of the object equals to or is approximate to a scale of the intermediate sliding window.

It should be appreciated that features described with reference to Fig.1-Fig. 10 also apply to the apparatus 1100 and therefore will not detailed here.

Furthermore, it should be noted that for the sake of clarity, FIG. 11 does not show some optional units or sub-units included in the apparatus 1100. Moreover, the term “unit” here may be either a hardware module or a software unit module. Correspondingly, the apparatus 1100 may be implemented through various manners. For example, in some embodiments, the apparatus 1100 may be implemented partially or completely by software and/or firmware, e.g., implemented as a computer program product embodied on a computer readable medium. Alternatively or additionally, the apparatus 1100 may be implemented partially or completely based on hardware, e.g., implemented as an integrated circuit (IC), an application-specific integrated circuit (ASIC), a system on chip (SOC), a field programmable gate array (FPGA), and the like. The scope of the present invention is not limited in this aspect.

Only for the illustration purpose, several exemplary embodiments of the present invention have been described above. Embodiments of the present invention may be implemented by hardware, software or a combination of the software and combination. The hardware part may be implemented using a dedicated logic; the software part may be stored in the memory, executed by an appropriate instruction executing system, e.g., a microprocessor or a dedicatedly designed hardware. In particular, both of the above methods described with reference to Figs. 1 and 10 may be implemented as a computer program product of detecting an object in the image, the computer program product being tangibly stored on a non-transient computer readable medium and including machine executable instructions, the machine executable instructions, when being executed, causing the machine to perform steps of the method 200.

Those ordinary skilled in art may understand that the above apparatus and method may be implemented using a computer-executable instruction and/or included in processor control code. In implementation, such code is provided on a medium such as magnetic disk, CD or DVD/ROM, a programmable memory such as read-only memory (firmware), or a data carrier such as optical or electronic signal carrier. The system of the present invention can be implemented by a very large scale integrated circuit or gate array, semiconductor such as logic chips and transistors, or hardware circuitry of programmable hardware devices like field programmable gate arrays and programmable logic devices, or implemented by various kinds of processor-executable software, or implemented by a combination of the above hardware circuits and software, such as firmware.

It is to be understood that although several units or sub-units of the apparatus have been mentioned in the above detailed description, such division is merely exemplary and not mandatory. In fact, according to embodiments of the present invention, the features and functions of two or more modules described above may be embodied in one module. On the contrary, the features and functions of one module described above may be embodied by a plurality of modules. In addition, although in the accompanying drawings operations of the method of the present invention are described in specific order, it is not required or suggested these operations be necessarily executed in the specific order or the desired result should be achieved by executing all illustrated operations. On the contrary, the steps depicted in the flowcharts may change their execution order. Additionally or alternatively, some steps may be omitted, a plurality of steps may be combined into one step for execution, and/or one step may be decomposed into a plurality of steps for execution.

Although the present invention has been described with reference to several embodiments, it is to be understood the present invention is not limited to the embodiments disclosed herein. The present invention is intended to embrace various modifications and equivalent arrangements comprised in the spirit and scope of the appended claims. The scope of the appended claims accords with the broadest interpretation, thereby embracing all such modifications and equivalent structures and functions.

Claims:

1. A method of detecting an object in an image, comprising:

positioning a first sliding window to a start sampling point in the image;

repetitively executing following steps until completing a traversal of the image so as to obtain a first group of sampling points including sampling points positioned by the first sliding window:

calculating a response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature of the object;

determining a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value; and

positioning the first sliding window to the next sampling point; and

detecting a position of the object according to the first group of sampling points and their respective response values.

2. The method according to claim 1, wherein the determining a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value, comprises: determining the position of the next sampling point so that a distance between the next sampling point and a current sampling point is longer in the case that the response value is smaller.

3. The method according to claim 2, wherein the determining a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value, further comprises:

determining a response value section to which the response value belongs from a plurality of response value sections, wherein each of the plurality of response value sections corresponds to a predetermined distance, wherein a predetermined distance corresponding to the response value section with a larger response value is smaller than a predetermined distance corresponding to the response value section with a smaller response value; and

determining a position of the next sampling point so that the distance between the next sampling point and the current sampling point is the predetermined distance corresponding to the response value section to which the response value belongs.

4. The method according to claim 3, wherein the predetermined distance corresponding to the response value section with the larger response value is a first integer times a step, the predetermined distance corresponding to the response value section with the smaller response value is a second integer times the step, and the first integer is smaller than the second integer.

5. The method according to claim 4, wherein the step is determined according to at least

one of: a scale of the first sliding window, and a scale of a maximum inscribed rectangle in an isogram as determined when the calculated response value equals to a predetermined response value threshold.

6. The method according to claim 4, wherein the step is an integer times the distance between adjacent pixel points in the image.

7. The method according to claim 1, wherein the predetermined sliding direction is from left to right and from up to down.

8. The method according to claim 1, wherein completing a traversal of the image comprises: all pixel points in the image have already been positioned by the first sliding window or excluded in the following excluding manner, wherein the excluding manner comprises excluding pixel points located between the current sampling point and the next sampling point in all pixel points.

9. The method according to claim 8, wherein the excluding manner further comprises excluding, from within the all pixel points, pixel points that are located in a particular square, wherein the particular square includes the current sampling point, and two perpendicular sides of the particular square start from the current sampling point, and wherein a length of each side of the particular square is one or more pixel points less than a distance between the current sampling point and the next sampling point.

10. The method according to claim 1, wherein the calculating the response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature comprises: calculating the response value to enable the response value larger in the case that there are more pixel points in the first sliding window matching the predetermined feature.

11. The method according to any of claims 1-10, wherein after determining the first group of sampling points, the method further comprises:

determining other groups of sampling points so as to obtain a plurality of groups of sampling points including the first group of sampling points, wherein each of the plurality of groups of sampling points comprises sampling points positioned by a respective one of a plurality of sliding windows, wherein a scale of each of the plurality of sliding windows is respectively set based on the scale of the first sliding window and according to a predetermined proportion;

detecting the position of the object according to the plurality of groups of sampling points and their respective response values.

12. The method according to claim 11, wherein the predetermined proportion is set to enable a predetermined scale range of the object to be covered with a minimum number of the plurality of sliding windows.

13. The method according to claim 12, wherein the predetermined proportion being set to enable a predetermined scale range of the object to be covered with a minimum number of the plurality of sliding windows comprises: the predetermined proportion is set to enable a response value corresponding to an overlapping portion when any two adjacent sliding windows of the plurality of sliding windows cover the respective scale range of the object to be greater than or equal to a predetermined response value threshold.

14. The method according to claim 11, wherein the detecting the position of the object according to the plurality of groups of sampling points and their respective response values comprises:

selecting sampling points whose response values are greater than a predetermined response value threshold from the plurality of groups of sampling points, wherein the sampling points are the sampling points already positioned by a respective one of the plurality of sliding windows and have corresponding response values;

positioning an intermediate sliding window to the position of the sampling point, wherein a proportion of a scale of the intermediate sliding window to a scale of the respective sliding window is closer to one as compared with the predetermined proportion;

comparing the respective response value with the response value calculated according to matching degree of the pixel point in the intermediate sliding window with respect to the predetermined feature; and

determining that the object is located at the sampling point and a scale of the object equals to or is approximate to a scale of the intermediate sliding window.

15. An apparatus for detecting an object in an image, comprising:

a start sampling point positioning unit configured to position a first sliding window to a start sampling point in the image;

a repetitive execution unit configured to repetitively execute following units until completing a traversal of the image so as to obtain a first group of sampling points including the sampling point positioned by the first sliding window:

a response value calculating unit configured to calculate a response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature of the object;

a next sampling point position determining unit configured to determine a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value; and

a next sampling point positioning unit configured to position the first sliding window to the next sampling point; and

an object position determining unit configured to detect a position of the object according to the first group of sampling points and their respective response values.

16. The apparatus according to claim 15, wherein the next sampling point position determining unit is configured to determine the position of the next sampling point so that a distance between the next sampling point and a current sampling point is longer in the case that the response value is smaller.

17. The apparatus according to claim 16, wherein the next sampling point position determining unit further comprises:

a response value section determining unit configured to determine a response value section to which the response value belongs from a plurality of response value sections, wherein each of the plurality of response value sections corresponds to a predetermined distance, wherein a predetermined distance corresponding to the response value section with a larger response value is smaller than a predetermined distance corresponding to the response value section with a smaller response value; and

a distance determining unit configured to determine a position of the next sampling point so that the distance between the next sampling point and the current sampling point is the predetermined distance corresponding to the response value section to which the response value belongs.

18. The apparatus according to claim 17, wherein the predetermined distance corresponding to the response value section with the larger response value is a first integer times a step, the predetermined distance corresponding to the response value section with the smaller response value is a second integer times the step, and the first integer is smaller than the second integer.

19. The apparatus according to claim 18, wherein the step is determined according to at least one of: a scale of the first sliding window, and a scale of a maximum inscribed rectangle in an isogram as determined when the calculated response value equals to a predetermined response value threshold.

20. The apparatus according to claim 18, wherein the step is an integer times the distance between adjacent pixel points in the image.

21. The apparatus according to claim 15, wherein the predetermined sliding direction is from left to right and from up to down.

22. The apparatus according to claim 15, wherein completing a traversal of the image

comprises: all pixel points in the image have already been positioned by the first sliding window or excluded in the following excluding manner, wherein the excluding manner comprises excluding pixel points located between the current sampling point and the next sampling point in all pixel points.

23. The apparatus according to claim 22, wherein the excluding manner further comprises excluding, from within the all pixel points, pixel points that are located in a particular square, wherein the particular square includes the current sampling point, and two perpendicular sides of the particular square start from the current sampling point, and wherein a length of each side of the particular square is one or more pixel points less than a distance between the current sampling point and the next sampling point.

24. The apparatus according to claim 15, wherein the response value calculating unit is configured to calculate the response value to enable the response value larger in the case that there are more pixel points in the first sliding window matching the predetermined feature.

25. The apparatus according to any of claims 15-24, wherein

the next sampling point determining unit is further configured to determine other groups of sampling points so as to obtain a plurality of groups of sampling points including the first group of sampling points, wherein each of the plurality of groups of sampling points comprises sampling points positioned by a respective one of a plurality of sliding windows, wherein a scale of each of the plurality of sliding windows is respectively set based on the scale of the first sliding window and according to a predetermined proportion;

the object position determining unit is further configured to detect the position of the object according to the plurality of groups of sampling points and their respective response values.

26. The apparatus according to claim 25, wherein the predetermined proportion is set to enable a predetermined scale range of the object to be covered with a minimum number of the plurality of sliding windows.

27. The apparatus according to claim 26, wherein the predetermined proportion being set to enable a predetermined scale range of the object to be covered with a minimum number of the plurality of sliding windows comprises: the predetermined proportion is set to enable a response value corresponding to an overlapping portion when any two adjacent sliding windows of the plurality of sliding windows cover the respective scale range of the object to be greater than or equal to a predetermined response value threshold.

28. The apparatus according to claim 25, further comprising:

a sampling point selecting unit configured to select sampling points whose response values are greater than a predetermined response value threshold from the plurality of groups of sampling points, wherein the sampling points are the sampling points already positioned by a respective one of the plurality of sliding windows and have a respective response value;

a sampling point positioning unit configured to position an intermediate sliding window to the position of the sampling point, wherein a proportion of a scale of the intermediate sliding window to a scale of the respective sliding window is closer to one as compared with the predetermined proportion; and

an response value comparing unit configured to compare the respective response value with the response value calculated according to matching degree of the pixel point in the intermediate sliding window with respect to the predetermined feature;

the object position determining unit is further configured to determine that the object is located at the sampling point and a scale of the object equals to or is approximate to a scale of the intermediate sliding window.

29. A computer program product of detecting an object in an image, the computer program product being tangibly stored on a non-transient computer readable medium and including machine executable instructions, the machine executable instructions, when being executed, causing the machine to perform steps of the method according to any of claims 1 to 14.

30. An electronic device comprising a memory and a processor, wherein the processor is configured to perform the method according to any of claims 1 to 14.

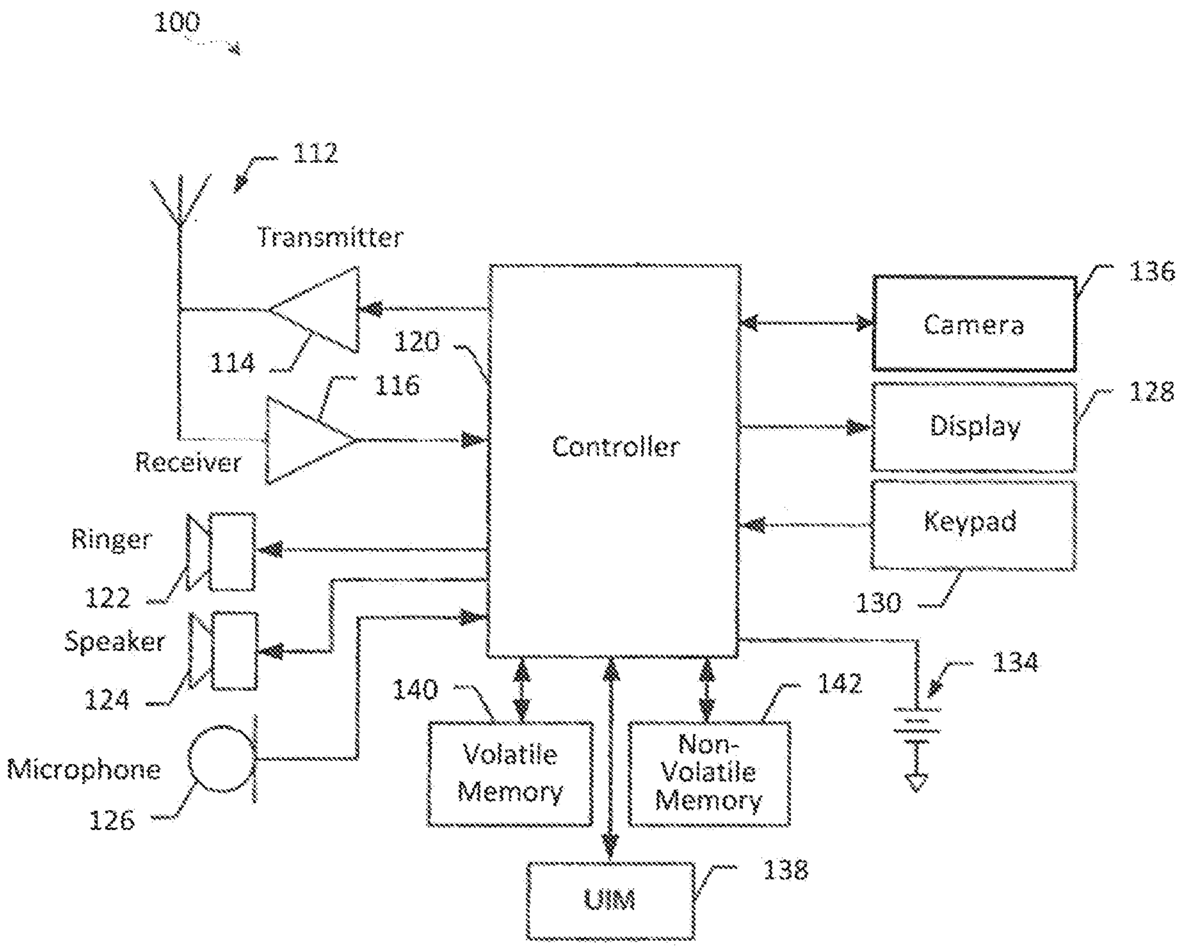
100
112
Transmitter
136
Camera
114
120
116
128
Display
Controller
Receiver
Keypad
Ringer
122
130
Speaker
134
140
142
124
Volatile Memory
Non-Volatile Memory
Microphone
126
138
UIM

Fig. 1

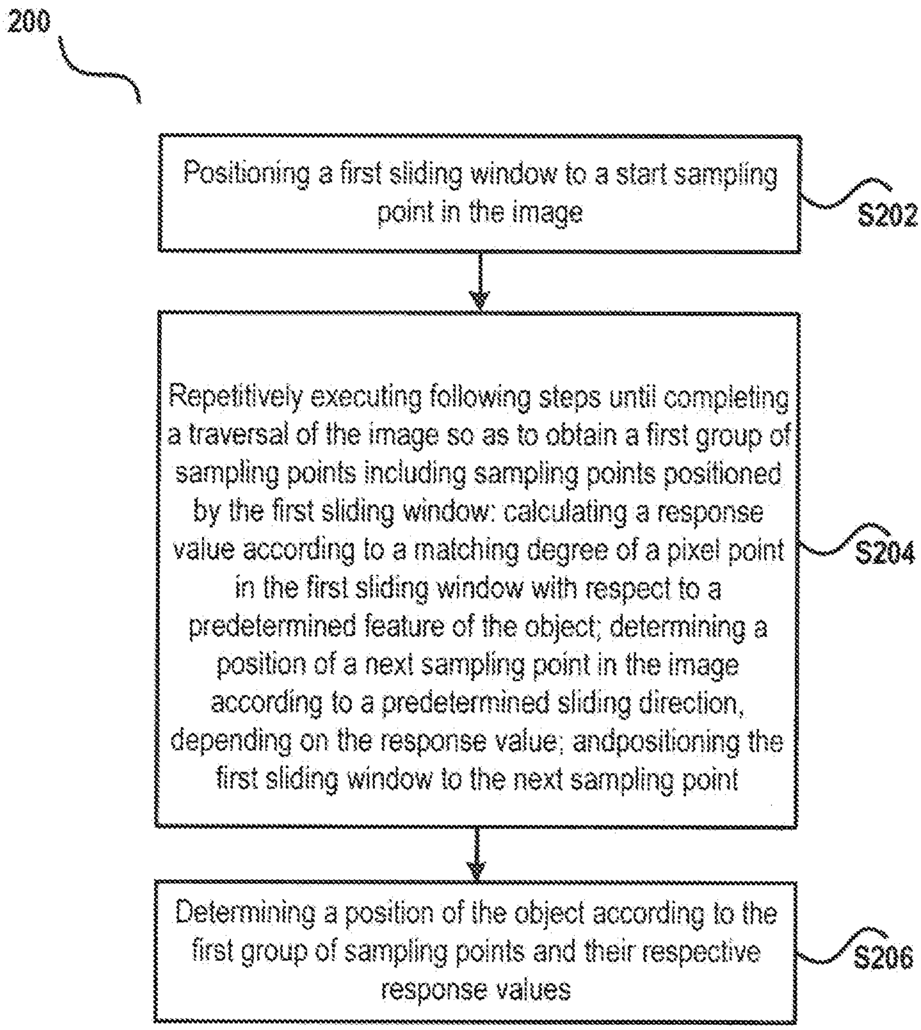
200
Positioning a first sliding window to a start sampling point in the image
S202
Repetitively executing following steps until completing a traversal of the image so as to obtain a first group of sampling points including sampling points positioned by the first sliding window: calculating a response value according to a matching degree of a pixel point in the first sliding window with respect to a predetermined feature of the object; determining a position of a next sampling point in the image according to a predetermined sliding direction, depending on the response value; andpositioning the first sliding window to the next sampling point
S204
Determining a position of the object according to the first group of sampling points and their respective response values
S206

Fig. 2

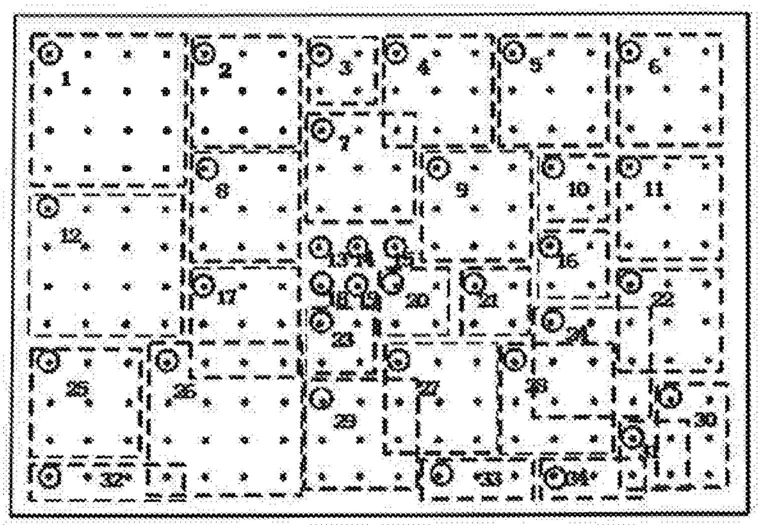

Fig. 3

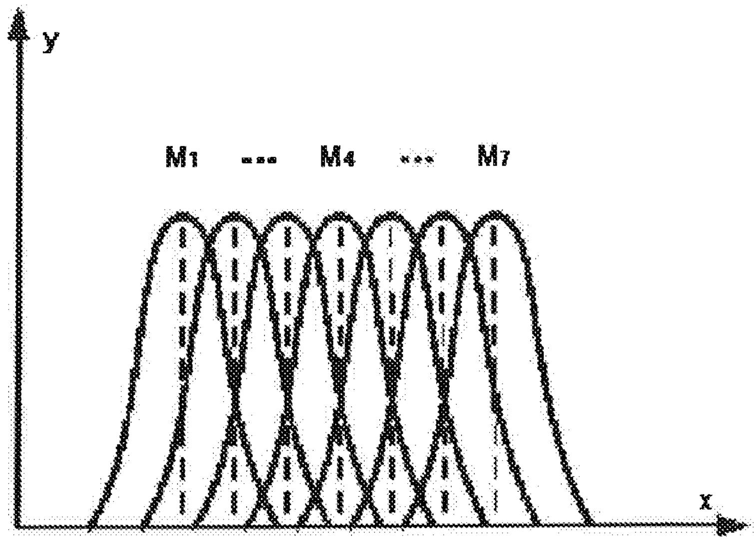

Fig. 4

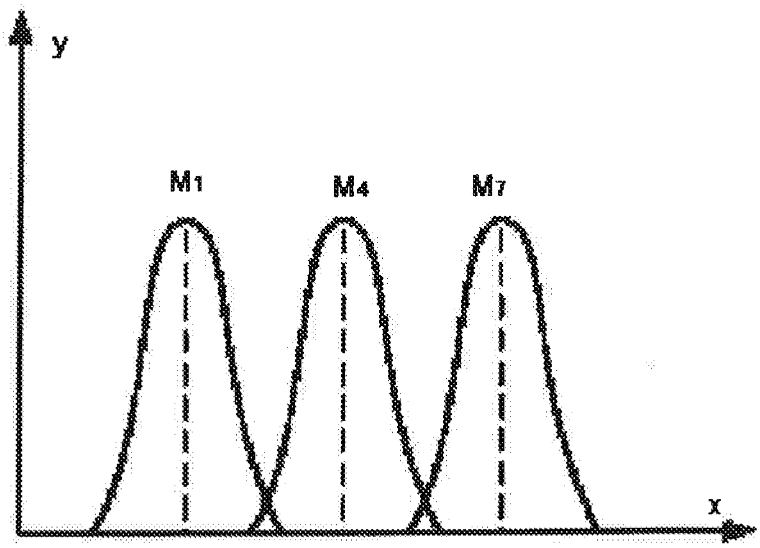

Fig. 5

SUBSTITUTE SHEET (RULE 26)

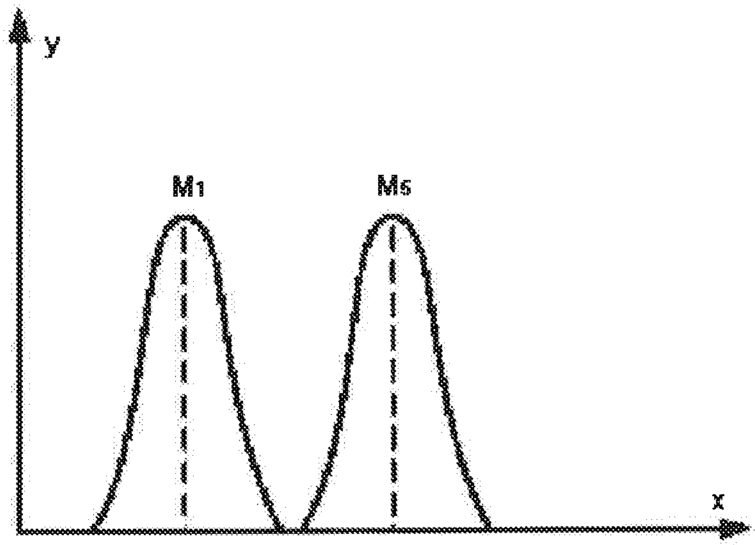

Fig. 6

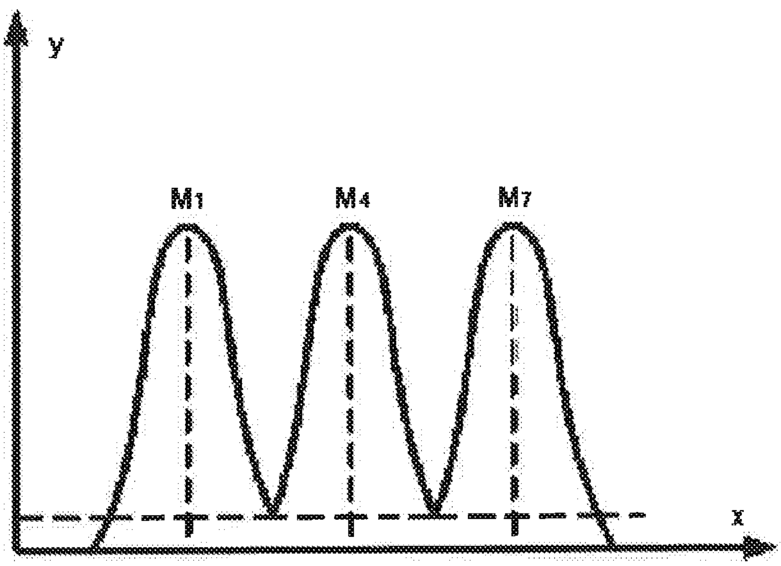

Fig. 7

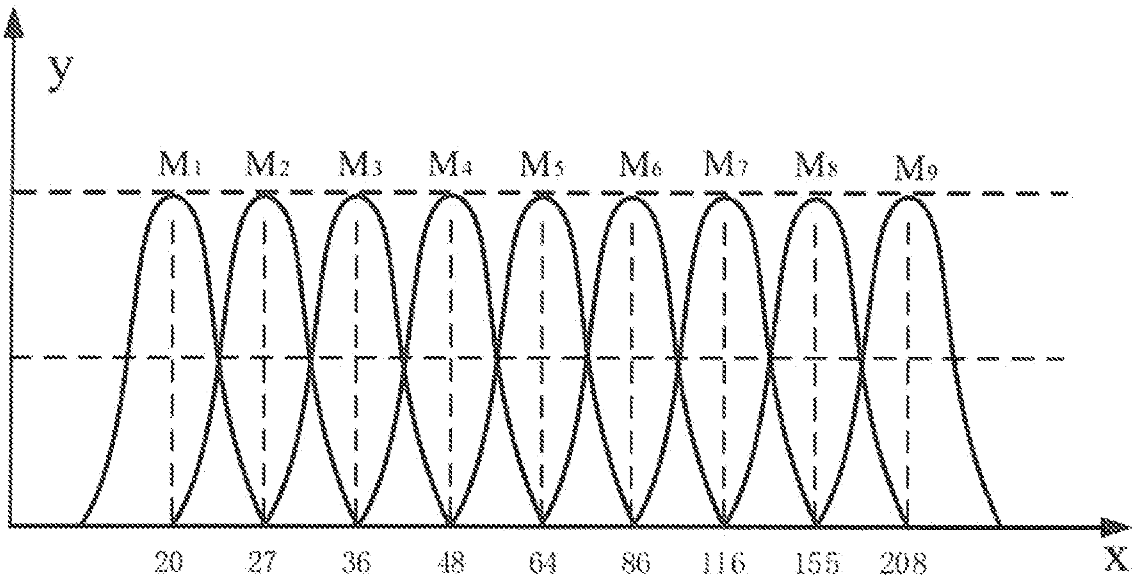

Fig. 8

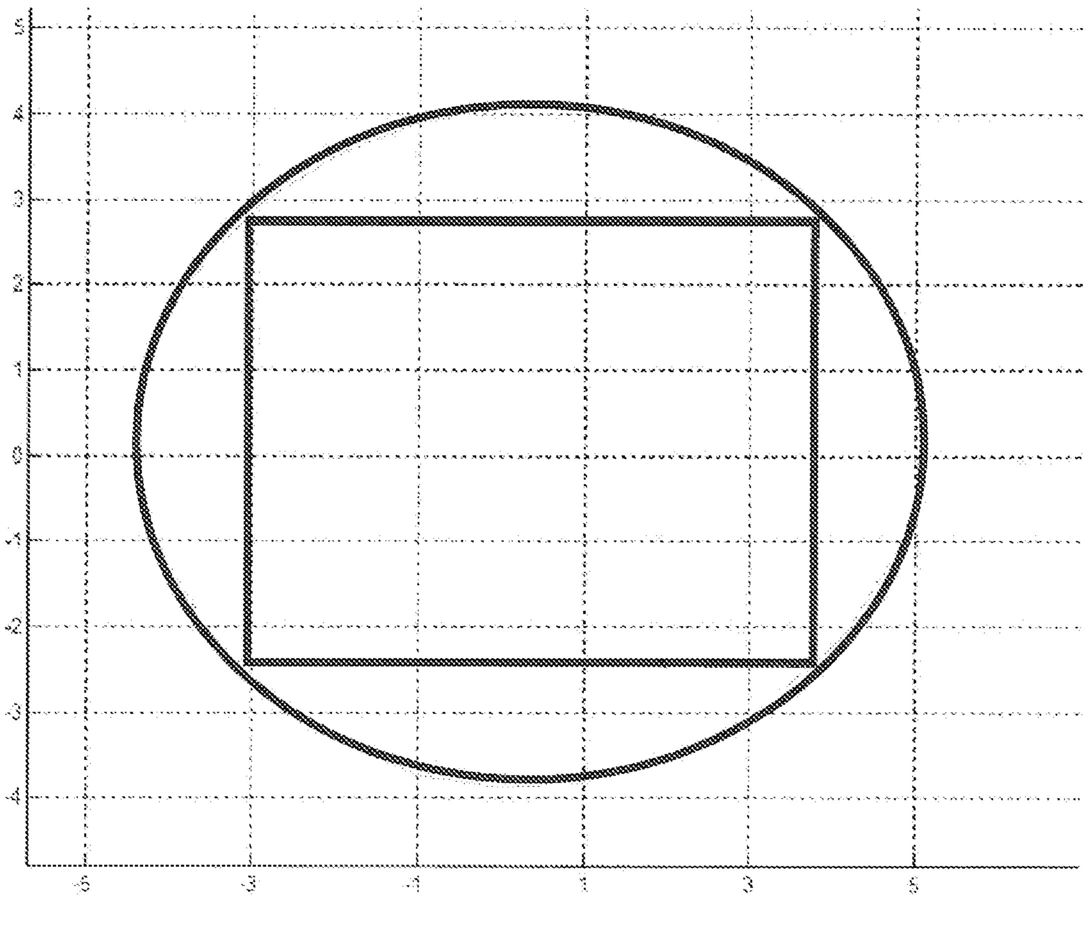

Fig. 9

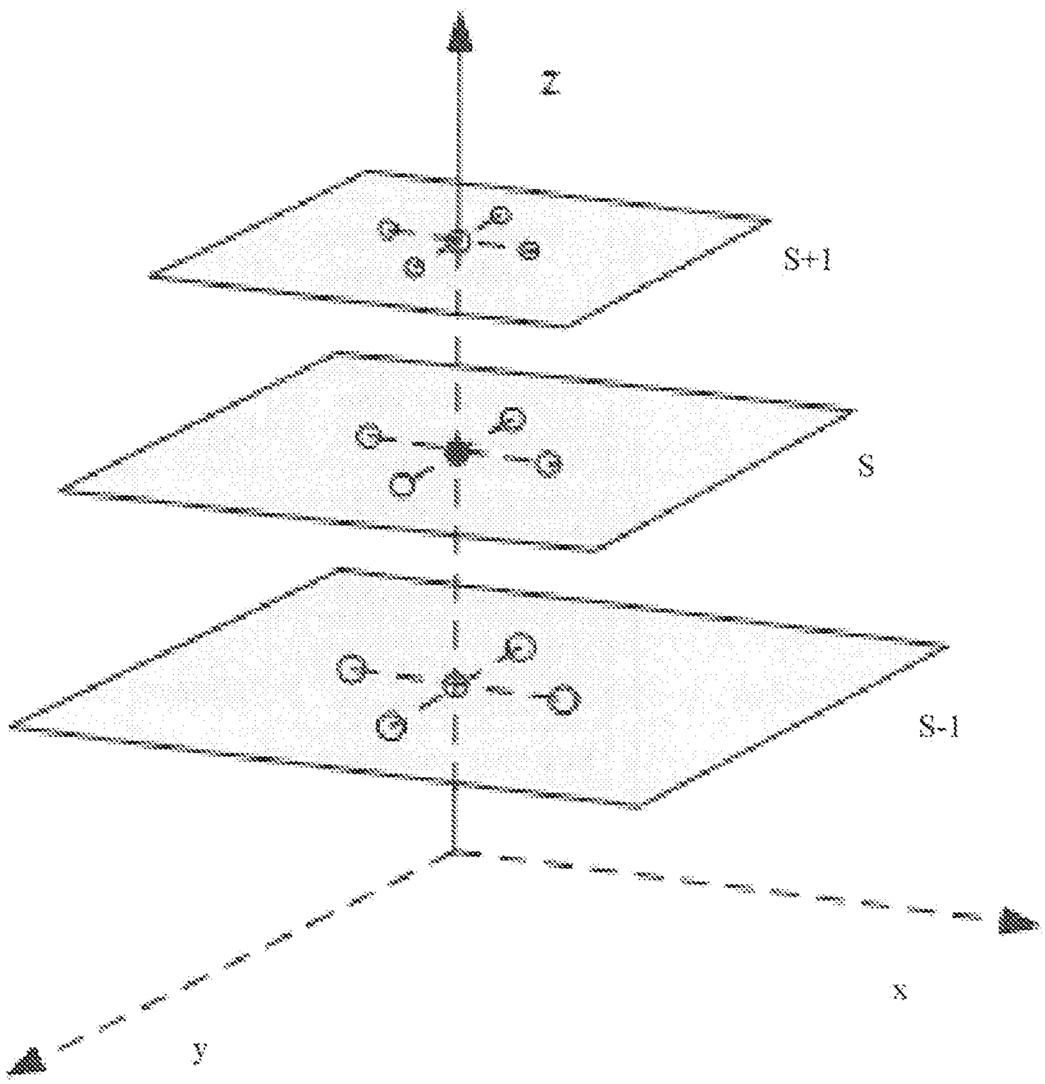

Fig. 10

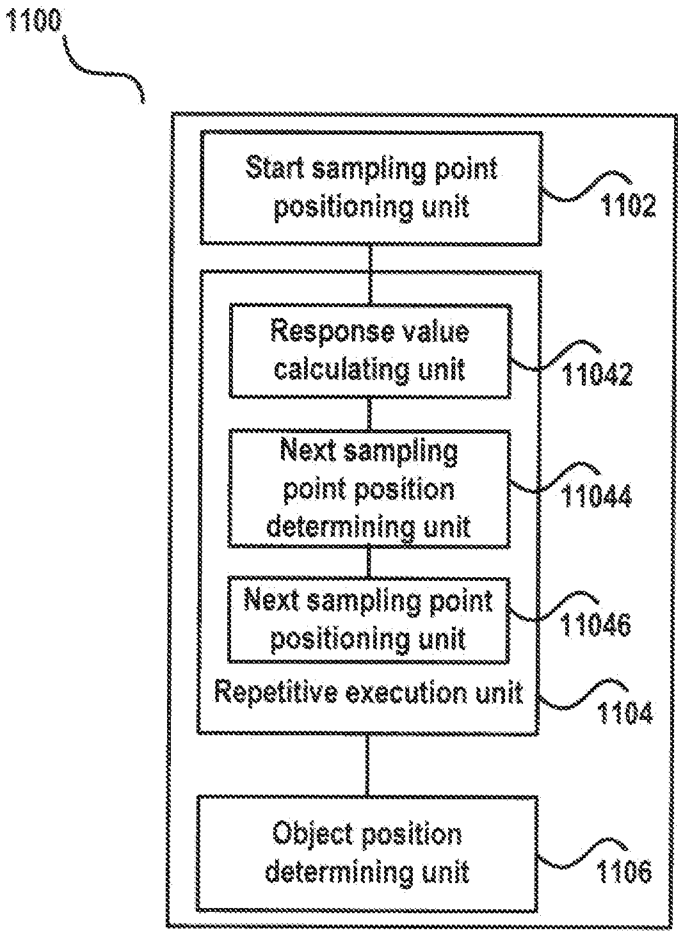
1100
Start sampling point positioning unit
1102
Response value calculating unit
11042
Next sampling point position determining unit
11044
Next sampling point positioning unit
11046
Repetitive execution unit
1104
Object position determining unit
1106

Fig. 11

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI2015/050506

A. **CLASSIFICATION OF SUBJECT MATTER**

See extra sheet

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)

IPC: G06K, G06T

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

FI, SE, NO, DK

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

EPO-Internal, WPI, INSPEC, XPI3E, XPESP, Internet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	COMASCHI, F et al. 'RASW: a Run-time Adaptive Sliding Window to Improve Viola-Jones Object Detection'. Seventh International Conference on Distributed Smart Cameras (ICDSC '13) [online], October 29, 2013 [retrieved on 2015-10-28]. Retrieved from the Internet: <URL: ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6778224> <DOI: 10.1109/ICDSC.2013.6778224> the whole document, especially abstract; sections III-V; Figures 1-2	1-30
X	VERSCHAE, R et al. 'A unified learning framework for object detection and classification using nested cascades of boosted classifiers'. Machine Vision and Applications [online], Volume 19, Issue 2, pages 85-103, March 2008 [available online 2007-10-09], [retrieved on 2015-10-28]. Retrieved from the Internet: <URL: http://link.springer.com/article/10.1007%2Fs00138-007-0084-0> <DOI: 10.1007/s00138-007-0084-0> the whole document, especially sections 3.1.2 and 3.3	1-30

☒ 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
30 October 2015 (30.10.2015)	06 November 2015 (06.11.2015)

Name and mailing address of the ISA/FI	Authorized officer
Finnish Patent and Registration Office P.O. Box 1160, FI-00101 HELSINKI, Finland	Mika Inki
Facsimile No. +358 9 6939 5328	Telephone No. +358 9 6939 500

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GUALDI, G et al. 'Multistage Particle Windows for Fast and Accurate Object Detection'. IEEE Transactions on Pattern Analysis & Machine Intelligence [online], Volume 34, Issue 08, August 2012 [retrieved on 2015-10-28]. Retrieved from the Internet: <URL: ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6109271> <DOI: 10.1109/TPAMI.2011.247> the whole document, especially sections 1, 3, 4.1.2, 4.4; Figure 1; Table 1	1-30
A	VIOLA, P et al. 'Rapid Object Detection using a Boosted Cascade of Simple Features'. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) [online], Volume I, pages 511-518, December 8, 2001 [retrieved on 2015-11-03]. Retrieved from the Internet: <URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=990517> <DOI: 10.1109/CVPR.2001.990517> the whole document, especially sections 2, 5; Figure 1	1-30
A	ZHANG, J et al. 'The Research of Face Detection Based on Double Skin Model and Adaboost Algorithm'. IEEE Conference Anthology [online], January 1, 2013 [retrieved on 2015-10-29]. Retrieved from the Internet: <URL: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6784860> <DOI: 10.1109/ANTHOLOGY.2013.6784860> the whole document, especially section III-B	14, 28
P, X	JIANG, X et al. 'Flexible sliding windows with adaptive pixel strides'. Signal Processing, Volume 110, pages 37-45, May 2015 [available online 2014-08-23], [retrieved on 2015-10-28]. Retrieved from the Internet: <URL: http://www.sciencedirect.com/science/article/pii/S0165168414003673> <DOI: 10.1016/j.sigpro.2014.08.004> the whole document	1-30

CLASSIFICATION OF SUBJECT MATTER

IPC
G06K 9/00 (2006.01)
G06K 9/62 (2006.01)
G06T 7/00 (2006.01)