



- (51) **International Patent Classification:**
G06F 3/033 (2006.01)
- (21) **International Application Number:**
PCT/IL2012/050191
- (22) **International Filing Date:**
31 May 2012 (31.05.2012)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/491,334 31 May 2011 (31.05.2011) US
- (71) **Applicant (for all designated States except US):** POINT-GRAB LTD. [IL/IL]; HaCharash Street 4, P.O. Box 6442, 45251 Hod Hasharon (IL).
- (72) **Inventor; and**
- (75) **Inventor/Applicant (for US only):** EILAT, Eran [IL/IL]; Gansin St. 16, 53419 Givatayim (IL).
- (74) **Agents:** PEARL COHEN ZEDEK LATZER et al.; P.O. Box 12704, 46733 Herzlia (IL).

(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, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

(54) **Title:** COMPUTER VISION BASED CONTROL OF A DEVICE USING MACHINE LEARNING

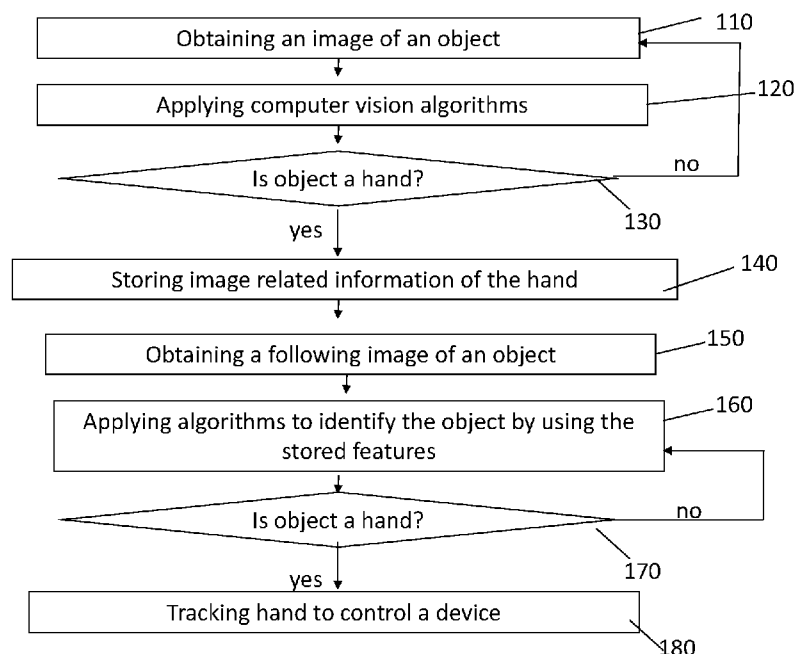


FIGURE 1A

(57) **Abstract:** A method for computer vision based control of a device, the method comprising: obtaining a first frame comprising an image of an object within a field of view; identifying the object as a hand by applying computer vision algorithms; storing image related information of the identified hand; obtaining a second frame comprising an image of an object within a field of view and identifying the object in the second frame as a hand by using the stored information of the identified hand; and controlling the device based on the hand identified in the first and second frames.

COMPUTER VISION BASED CONTROL OF A DEVICE USING MACHINE LEARNING

FIELD OF THE INVENTION

5 [0001] The present invention relates to the field of computer vision based control of electronic devices. Specifically, the invention relates to computer vision based hand identification using machine learning techniques.

BACKGROUND OF THE INVENTION

10 [0002] The need for more convenient, intuitive and portable input devices increases, as computers and other electronic devices become more prevalent in our everyday life.

[0003] Recently, human gesturing, such as hand gesturing, has been suggested as a user interface input tool in which a hand gesture is detected by a camera and is translated into a specific command. Gesture recognition enables humans to interface with machines naturally without
15 any mechanical appliances. The development of alternative computer interfaces (forgoing the traditional keyboard and mouse), video games and remote controlling are only some of the fields that may implement human gesturing techniques.

[0004] Recognition of a hand gesture usually requires identification of an object as a hand and tracking the identified hand to detect a posture or gesture that is being performed.

20 [0005] Known gesture recognizing systems detect a user hand by using color, shape and/or contour detectors.

[0006] Machine learning techniques can be used to train a machine to discriminate between features and thus to identify objects, typically different faces or facial expressions. Machines can be trained to identify objects belonging to a specific group (such as human faces) by
25 providing the machine with many training examples of objects belonging to the specific group. Thus, during manufacture a machine is supplied with abroad pre-made database with which to compare any new object that is later presented to the machine during use, after the machine has left the manufacturing facility.

[0007] However, identifying a human hand in the process of gesturing may prove to be a
30 challenge for these methods of detection because many environments include designs that may be similar enough to a human hand to cause too many cases of false identification and the

SUMMARY OF THE INVENTION

5 [0008] The method for computer vision based control of a device, according to embodiments of the invention, provides an efficient process for accurate hand identification, regardless of the background environment and of other complications such as the hand's posture or angle at which it is being viewed.

[0009] The method according to embodiments of the invention facilitates hand identification
10 so that in the process of tracking the hand, even if sight of the hand is lost (hand changes orientation or position, hand moves by confusing background, etc.), re-identifying the hand is quick, thereby enabling better tracking of the hand.

[0010] According to embodiments of the invention image related information is stored on-line, during use, rather than using pre-made databases. This enables each machine to learn its
15 specific environment and user enabling more accurate and quick identification of the user's hand.

[0011] According to one embodiment of the invention there is provided a method for computer vision based control of a device, the method including the steps of obtaining a first frame comprising an image of an object within a field of view; identifying the object as a hand
20 by applying computer vision algorithms; storing image related information of the identified hand; obtaining a second frame comprising an image of an object within a field of view and identifying the object in the second frame as a hand by using the stored information of the identified hand; and controlling the device based on the hand identified in the first and second frames.

25 [0012] This process may continue by storing image related information of the hand identified in the second frame. According to some embodiments an on-line database may thus be constructed.

[0013] Image related information may include Local Binary Pattern (LBP) features, statistical parameters of grey level or Speeded Up Robust Features (SURF) or other appropriate features.

30 [0014] The method may include tracking the hand identified in the first frame and continuing the tracking only if the hand is also identified in the second image. The device may be controlled according to the tracking of the hand.

[0015] The method may further include identifying a non-hand object and storing image related information of the non-hand object. According to some embodiments the image related information of the object identified as a hand and the image related information of the non-hand object are stored only if the information is different than any image related information already stored.

[0016] According to some embodiments the image related information of an object identified as a hand and/or the image related information of the non-hand object is stored for a pre-defined period. The pre-defined period may be based on use or on absolute time.

[0017] A non-hand object may be a portion of a frame, said portion not including a hand. The portion may be located at a pre-determined distance or further from the position of the hand within the frame. According to some embodiments the portion includes an area in which no movement was detected.

[0018] According to some embodiments identifying the object in the second frame as a hand by using the information of the identified hand includes detecting in the identified hand a set of features; assigning a value to each feature; and comparing the values of the features to a hand identification threshold, said hand identification threshold constructed by using values of features of formerly identified hands. A new hand identification threshold may be constructed every pre-defined period.

[0019] According to some embodiments the object in the first image is identified as a hand only if the object is moving in a pre-defined movement, such as a wave like movement.

[0020] The object identified as a hand may be a hand in any posture or post-posture. Thus, the method may include storing image related shape information of the hand in a predefined posture; and obtaining a second frame comprising an image of an object within a field of view and identifying the object in the second frame as a hand in the predefined posture by using the stored shape information.

[0021] A posture may be, for example, a hand with all fingers extended or a hand with all fingers brought together such that their tips are touching or almost touching. Post-posture may be, for example, a hand during the act of extending fingers after having held them in a fist or closed fingers posture.

[0022] The device may be controlled according to a posture or gesture of the hand.

[0023] According to another embodiment of the invention there is provided a system for computer vision based control of a device, the system comprising: an adaptive detector, said

WO 2012/164562
detector configured to identify an object in a first image as a hand; **PCT/IL2012/050191**
store image related information of the identified hand; and identify an object in a second image as a hand by using the stored image related information; a processor to track the identified hand; and a controller to control the device based on the identified hand.

5 [0024] The system may further include an image sensor to obtain the first and second images, said image sensor in communication with the adaptive detector. The sensor may be a 2D camera.

[0025] The system may also include a processor to identify a hand gesture or posture and the controller generates a user command based on the identified hand gesture or posture.

10 [0026] The device may be a TV, DVD player, PC, mobile phone, camera, STB (Set Top Box) and a streamer.

BRIEF DESCRIPTION OF THE FIGURES

15 [0027] The invention will now be described in relation to certain examples and embodiments with reference to the following illustrative figures so that it may be more fully understood. In the drawings:

[0028] Figs. 1A - C schematically illustrate methods for computer vision based control of a device according to embodiments of the invention;

20 [0029] Fig. 2A schematically illustrates a method for computer vision based control of a device including re-setting a database of hand objects, according to an embodiment of the invention;

[0030] Fig. 2B schematically illustrates a method for machine learning identification of a hand including re-setting a hand identification threshold, according to an embodiment of the invention;

25 [0031] Figs. 3A – 3E schematically illustrate a method for training a hand identification system on-line, according to an embodiment of the invention;

[0032] Fig. 4 is a schematic illustration of a system operable according to embodiments of the invention.

30

[0033] Computer vision based identification of a hand during a process of user-machine interaction has to sometimes deal with diverse backgrounds, some of which may include designs similar to hands.

- 5 [0034] The method for computer vision based control of a device, according to embodiments of the invention, uses machine learning techniques in a unique way which enables accurate and quick identification of a user's hand.

[0035] According to one embodiment, which is schematically illustrated in **Fig. 1A**, the method includes obtaining a first frame, the frame including an image of an object within a
10 field of view (110). In the next step computer vision algorithms are applied to identify the object (120). If the object is identified, by the computer vision algorithms, as a hand (130) then image related information of the identified object (hand) is stored (140). If the object is not identified by the computer vision algorithms as a hand a following image is obtained (110) and checked.

- 15 [0036] After information of an object identified as a hand is stored (140), the next frame obtained which includes an image of an object within a field of view (150) will be checked for the presence of a hand by applying algorithms which use the stored information (160). If the object in this next frame is identified as a hand by using the stored information (170) then the object is confirmed as a hand and it is further tracked to control the device (180). If the object
20 has not been identified as a hand by using the stored information then a following image is obtained and checked for the presence of a hand by using the stored information (steps 150 and 160).

[0037] Tracking of the object may be done also based on the first identification of the object as a hand, in step 130, so that tracking of a hand, which may begin immediately with an initial
25 identification of the hand, may be improved as time goes by. According to some embodiments, if an object is identified as a hand by using computer vision algorithms (step 130) it is tracked but the tracking is terminated if in a following image, which is checked for the presence of a hand by applying algorithms which use the stored information (step 160), it is determined that the object is not a hand. Thus, tracking of the hand identified in the first frame may be
30 continued only if the hand is also identified in the following image.

[0038] Computer vision algorithms which are applied to identify an object as a hand in the first frame (in step 120) may include known computer vision algorithms such as appropriate image analysis algorithms. A feature detector or a combination of detectors may be used. For

example, a texture detector and edge detector may be used. If both specific texture and specific edges are detected in a set of images then an identification of a hand may be made. One example of an edge detection method includes the CannyTM algorithm available in computer vision libraries such as IntelTM OpenCV. Texture detectors may use known algorithms such as texture detection algorithms provided by MatlabTM.

[0039] In another example, an object detector is applied together with a contour detector. In some exemplary embodiments, an object detector may use an algorithm for calculating Haar features. Contour detection may be based on edge detection, typically, of edges that meet some criteria, such as minimal length or certain direction.

[0040] According to some embodiments an image of a field of view is translated into values. Each pixel of the image is assigned a value that is comprised of 8 bits. According to one embodiment some of the bits (e.g., 4 bits) are assigned values that relate to grey level parameters of the pixel and some of the bits (e.g., 4 bits) relate to the location of the pixel (e.g., on X and Y axes) relative to a reference point within the hand (e.g., the assigned values may represent a distance to a pixel in the center of the hand). The values of the pixels are used to construct vectors (or other representations of the values assigned to pixels) which are used to represent hand objects. A classifier may be used to process these vectors.

[0041] Using image related information, such as vectors as described above, provides a more accurate identification of a hand since each pixel is compared to a reference pixel in the hand itself (e.g., to a pixel in the center of the hand) rather than to a reference pixel external to the hand (for example, to a pixel at the edge of the frame).

[0042] Other methods of hand identification may include the use of shape detection algorithms together with another parameter such as movement so that an object may be identified as a hand only if it is moving and if it is determined by the shape detection algorithms that the object has a (typically pre-defined) hand shape.

[0043] According to one embodiment the object in the first image may be identified using known machine learning techniques, such as supervised learning techniques, in which a set of training examples is presented to the computer. Each example typically includes a pair consisting of an input object and a desired output value. A supervised learning algorithm analyzes the training data and produces an inferred function (classifier), if the output is discrete, or a regression function, if the output is continuous. According to some embodiments training examples may include vectors which are constructed as described above.

first image may be identified as a hand by using a pre-constructed database. In this case, a hand is identified in the first frame by using a semi automated process in which a user assists or directs machine construction of a database of hands and in the following frames the hand is identified by using a fully automated process in which the machine construction of a database of hand objects is automatic. An identified hand or information of an identified hand may be added to the first, semi automatically constructed database or a newly identified hand (or information of the hand) may be stored or added to a new fully automatic machine-constructed database.

[0045] It should be appreciated that the term "hand" may refer to a hand in any posture, such as a hand open with all fingers extended, a hand open with some fingers extended, a hand with all fingers brought together such that their tips are touching or almost touching, or other postures.

[0046] According to one embodiment the "first frame" may include a set of frames. An object in the first frame (set of frames) may be identified as a hand (step 130) by using computer vision algorithms (step 120) but only if it is also determined that the object is moving in a pre-defined pattern. If, for example, an object is identified as having a hand shape (by computer vision algorithms) in five consecutive frames it will still not be identified as a hand unless it is determined that the object is moving, for example, in a specific pattern, e.g., in a repeating back and forth waving motion. According to this embodiment, identification of a hand in a set of frames by using computer vision algorithms will only result in storing information of the object (e.g., adding image related information of the object to a database of hand objects) (step 140) if the object has been determined to be moving and in some embodiment, only if the object has been determined to be moving in a pre-defined, rather than random, movement.

[0047] Storing or adding image related information of an object identified as a hand to the database of hand objects (step 140) may be done by applying machine learning techniques, such as by using an adaptive boosting algorithm. Machine learning techniques (such as adaptive boosting) are also typically used in step 160 in which the stored information is used to identify objects in a next frame.

[0048] Once an object is identified as a hand according to embodiments of the invention it may be tracked using known tracking methods. Tracking the identified hand (and possibly identifying specific gestures or postures) is then translated into control of a device. For

WO 2012/164562
example, a cursor on a display of a computer may be moved on the computer screen and/or
icons may be clicked on by tracking a user's hand.

PCT/IL2012/050191

[0049] Devices that may be controlled according to embodiments of the invention may include any electronic device that can accept user commands, e.g., TV, DVD player, PC,
5 mobile phone, camera, STB (Set Top Box), streamer, etc.

[0050] The method, as schematically illustrated in **Fig. 1B**, may continue such that once an object is identified as a hand by using the stored information (step 160) information of that object is also stored or added to a database of hand objects. According to some embodiments, once a hand is identified as a hand (in step 130 or 160) information of this hand is compared
10 to information already stored. If the information of an identified hand is very similar to information of a hand already stored (e.g. in a database of hand objects), there may be a decision not to store this additional information so as not to burden the system with redundant information. Thus, storing information of a hand identified in the second frame may be done, in some embodiments, only if the information of the hand identified in the second frame is
15 different than any information already stored.

[0051] Image related information may include values or other representations of image features or parameters such as pixels or vectors. Some features, for example, may include Local Binary Pattern (LBP) features, statistical parameters of grey level and/or Speeded Up Robust Features (SURF). Alternatively, image related information may include portions of
20 images or full images.

[0052] **Fig. 1C** schematically exemplifies the use of image related information according to embodiments of the invention.

[0053] The method illustrated in **Fig. 1C** shows one way of how stored information assists and facilitates hand identification in a following image. According to one embodiment, once
25 a hand is identified in a first frame (by computer vision algorithms possibly using known machine learning techniques), a set of features is detected in that hand (111). Features, which are typically image related features, may include, for example, Local Binary Pattern (LBP) features, statistical parameters of grey level and/or Speeded Up Robust Features (SURF). Each detected feature is assigned a value (112). A hand identification threshold is then
30 constructed based on the assigned values (113).

[0054] A second frame (which includes an object) is obtained (114). The object is checked for the set of features (115) and each detected feature is assigned a value (116). The values are then calculated and if the calculated values are above the hand identification threshold

WO 2012/164562 **PCT/IL2012/050191**
then the object is identified as a hand (117). If the calculated values do not exceed the hand identification threshold then a following frame is obtained (118) and further checked.

[0055] Thus, a hand identification threshold constructed by using values of features of formerly identified hands is used in identification of hands in subsequent images.

5 [0056] The method described in **Figs. 1A -C** may be applied, for example, during routine use of a gesture controlled device. A user may wave his hand in front of a gesture controlled system. An image sensor included in the system obtains images of the user's hand and a computer vision algorithm is employed by the system to identify the user's hand. Once the user's hand is identified by the computer vision algorithm, the image of that hand (or image
10 related information of that hand) is stored or added to a database, information which is then used to identify the user's hand in subsequent images. Thus, according to embodiments of the invention, a database of training examples of a hand which are used by learning algorithms is created on-line, while the user is using the system. The advantage of this method, as opposed to using pre-constructed databases of known machine learning techniques, is that the
15 examples in this on-line database are user specific, since it is information of the user's hand itself that is being added to the database each time. A database constructed according to embodiments of the invention includes examples of a user's specific hand and typical background environments of this specific user (machine learning of "background" will be discussed below) so that with each use identifying the hand of the user becomes easier and
20 quicker.

[0057] It may be advantageous in some cases to delete stored information or "reset" the database once in a while, for example, so that the database does not become too specific.

[0058] Reference is now made to **Fig. 2A**, which schematically illustrates a method for resetting a database of hand objects.

25 [0059] In one embodiment information of an object which has been identified as a hand (for example as described with reference to **Fig. 1A**) is stored (e.g., added to a database of hand objects) (240). Each information added is stored in the system for a pre-defined period. Once the pre-defined period has passed the information is deleted (244) and the process of machine learning and database construction (for example, as described with reference to **Fig. 1A**)
30 starts again.

[0060] According to some embodiments the pre-defined period is based on use. For example, the database of information of hand objects may be erased after a specific number of sessions. A session may include the time between activation of a program until the program is

WO 2012/164562
terminated. According to some embodiments a session includes the time between
identification of a hand until the hand is no longer identified (e.g., if the hand exits the frame
or field of view). According to one embodiment stored information of hand objects is deleted
each time a user ends a session. Thus, according to some embodiments new information is
5 used in each use.

[0061] According to other embodiments the pre-defined period is based on absolute time. For
example, information may be deleted every day (24 hours) or every week, regardless of its
use during that day or week. In some embodiments information may be deleted at a specific
time after a session has begun.

10 [0062] According to one embodiment, information may be deleted manually by the user.
According to another embodiment information is automatically deleted, for example, after
each use (e.g., session).

[0063] Similarly, the hand identification threshold (described in **Fig. 1C**) may be “re-set”
once in a while. As schematically illustrated in **Fig. 2B**, if an object is detected as a hand, a
15 hand identification threshold is constructed (211). After a predetermined period (which may
be based on absolute time or on use, such as described with reference to **Fig. 2A**) the hand
identification threshold is erased (212) and in a subsequently obtained frame which includes
an object (213) the set of features will be detected in the object and a new hand identification
threshold may be constructed (214).

20 [0064] Training a hand identification system according to embodiments of the invention may
include presenting to the machine learning algorithm training data which includes both
examples of a hand (in different postures) and examples of a “non-hand” object. As opposed
to standard machine learning methods, the method according to embodiments of the
invention can train an algorithm in a way that is tailored to a user and/or to a specific
25 environment (e.g., specific backgrounds). Thus, according to one embodiment, when
applying machine learning techniques to add information of an object identified as a hand to
a database of hand objects, information of a non-hand object may at the same time also be
stored or added to a non-hand object database.

[0065] Methods for training a hand identification system according to embodiments of the
30 invention are schematically illustrated in **Figs. 3A – E**.

[0066] In **Fig. 3A** a frame or image is divided to portions (31) and each portion is checked
for the presence of a hand (33). If the portion does not include a hand then that portion or
information of that portion is presented to the machine learning algorithm as a non-hand

portion or information of that portion of the image is presented to the machine learning algorithm as a hand object (37). Alternatively, only information of the image of the hand (or part of the hand) itself, rather than information of the portion which includes the hand (or part of hand) may be presented to the machine learning algorithm as a "hand information".

[0067] The frame or image that is divided to portions may be the "first frame" (in which an object is identified as a hand by applying computer vision algorithms) and/or the "following frame" (in which an object is identified as a hand by using the information stored on-line).

[0068] The frame may be divided to portions based on a pre-determined grid, for example, the frame may be divided into 16 equal portions. Alternatively the frame may be divided to areas having certain characteristics (e.g., areas which include dark or colored features or a specific shape, and areas that do not).

[0069] In one embodiment, which is schematically described in **Fig. 3B**, the frame is divided to portions (31) and the portions are checked for the presence of a hand (33). If a checked portion does not include a hand then the distance of that portion to the portion that does include a hand is determined. If the determined distance is equal to or above a predetermined value (32) then that portion is presented to the machine learning algorithm as a non-hand object (34). According to this embodiment, only portions of an image which are far from the portion including the hand are defined as "non-hand".

[0070] According to another embodiment a set of frames is checked for the presence of a hand in each of the frames. The set of frames is also checked for movement. Movement may indicate the presence of a hand, for example, in cases where a user is expected to move his hand as a means for activating and/or controlling a program.

[0071] According to one embodiment a portion (or information of that portion) is presented as a non-hand object only if it is at a distance that is equal to or above the predetermined value and if no movement was detected in that portion.

[0072] According to one embodiment, which is schematically described in **Fig. 3C**, a set of frames is checked. Each of the frames in the set of frames is divided to portions (31') and each portion is checked to see if movement was detected in that portion (38). If no movement was detected in the area of the checked portion then that portion (or information of that portion) is presented to the machine learning algorithm as a non hand object (39). In some embodiments, a determination must be made that no hand and no movement were detected in

WO 2012/164562
a portion in order for that portion (or information of that portion) to be presented to the machine learning algorithm as a non-hand object. **PCT/IL2012/050191**

[0073] These embodiments may raise the accuracy of identification of non-hand objects, thus lowering the false positive reading rate of the system.

5 [0074] According to one embodiment, which is schematically described in **Fig. 3D**, a set of frames is obtained (301) and each frame is divided to portions (303). Movement is searched for in the set of frames. If movement is detected in a certain portion then that portion is searched for the presence of a hand (304). If a hand is detected then information of the identified hand (or the portion which includes the hand) is presented to the machine learning
10 algorithm as a hand object (306) and may be stored or added to the database of hand objects.

[0075] If movement is not detected in the set of frames then each frame in the set of frames is searched for portions that do not include a hand (305). Portions detected which do not include a hand may then be presented to the machine learning algorithm as a non-hand object (307).

[0076] This embodiment may lower the rate of false positive identifications of the system
15 and may reduce computation time by applying algorithms to identify a hand only in cases where movement was detected (thus indicating possible presence of a hand).

[0077] In general, the method of hand identification using on-line machine learning, according to embodiments of the invention, takes up less computing time than known (“off-line”) machine learning techniques because only limited data (user specific scenes) needs to
20 be learnt on-line, compared with the many examples presented to a machine learning algorithm off-line.

[0078] According to one embodiment a hand searched in the methods described above may be a hand in a specific posture, for example, a posture in which a hand has all fingers brought together such that their tips are touching or almost touching. If such a posture of a hand is
25 detected in an image, by computer vision methods, information of this image or of a portion of this image is stored, for example, in a first posture hand database. If a second, different posture is detected, in a second image, by computer vision methods, information of the second image, or of a portion of the second image is stored, for example, in a second posture hand database. Thus, several databases may be concurrently constructed on-line, according to
30 embodiments of the invention.

[0079] According to one embodiment a database may include a post-posturing hand. For example, one database may include hand objects (or information of hand objects) in which the hand is closed in a fist or a hand that has all fingers brought together such that their tips

are touching or almost touching. Another database may include hands which are opening; extending fingers after having held them in a fist or closed fingers posture. The present inventor has found that "post posture" hands are specific to users (namely, each user moves his hand between hand postures in a unique way). Thus, using a "post-posture" database may add to the specificity and thus to the efficiency of methods according to the invention.

[0080] A method according to one embodiment, which is schematically illustrated in **Fig. 3E**, includes obtaining an image of an object within a field of view (332). The object is compared to a plurality of databases (334) and a grade is assigned (336) according to the similarity of the object to the database in each case. A decision is made regarding the object (e.g., whether it is a hand in a specific posture, whether it is a hand in "post-posture", whether it is a "non-hand" object, etc.) based on the highest grade (338).

[0081] According to one embodiment a "wild card" database can be created and used in a case where two grades are too similar to enable a decision. The wild card database is typically made up of information of the previous frame, the frame before the one being checked at present.

[0082] Reference is now made to **Fig. 4** which schematically illustrates system 400 according to an embodiment of the invention.

[0083] System 400 includes an image sensor 403 for obtaining a sequence of images of a field of view (FOV) 414, which may include an object (such as a hand 415). The image sensor 403 is typically associated with processor 402, and storage device 407 for storing image data. The storage device 407 may be integrated within the image sensor 403 or may be external to the image sensor 403. According to some embodiments image data may be stored in processor 402, for example in a cache memory.

[0084] The processor 402 is in communication with a controller 404 which is in communication with a device 401. Image data of the field of view is sent to processor 402 for analysis. A user command is generated by processor 402, based on the image analysis, and is sent to a controller 404 for controlling device 401. Alternatively, a user command may be generated by controller 404 based on data from processor 402.

[0085] The device 401 may be any electronic device that can accept user commands from controller 404, e.g., TV, DVD player, PC, mobile phone, camera, STB (Set Top Box), streamer, etc. According to one embodiment, device 401 is an electronic device available with an integrated standard 2D camera. According to other embodiments a camera is an external accessory to the device. According to some embodiments more than one 2D camera are

WO 2012/164562
provided to enable obtaining 3D information. According to some embodiments the system includes a 3D camera.

PCT/IL2012/050191

[0086] The processor 402 may be integrated within the device 401. According to other embodiments a first processor may be integrated within the image sensor 403 and a second processor may be integrated within the device 401.

[0087] The communication between the image sensor 403 and processor 402 and/or between the processor 402 and controller 404 and/or device 401 may be through a wired or wireless link, such as through IR communication, radio transmission, Bluetooth technology and/or other suitable communication routes.

[0088] According to one embodiment image sensor 403 is a forward facing camera. Image sensor 403 may be a standard 2D camera such as a webcam or other standard video capture device, typically installed on PCs or other electronic devices. According to some embodiments, image sensor 403 can be IR sensitive.

[0089] The processor 402 can apply computer vision algorithms, such as motion detection and shape recognition algorithms to identify and further track an object, typically, the user's hand. The processor 402 or another associated processor may comprise an adaptive detector which can identify an object in a first image as a hand and can add the identified hand to a database of hand objects. The detector can then identify an object in a second image as a hand by using the database of hand objects (for example, by implementing methods described above).

[0090] Once the object is identified as a hand it is tracked by processor 402 or by a different dedicated processor. The controller 404 may generate a user command based on identification of a movement of the user's hand in a specific pattern based on the tracking of the hand. A specific pattern of movement may be for example, a repetitive movement of the hand (e.g., wave like movement).

[0091] Optionally, system 400 may include an electronic display 406. According to embodiments of the invention, mouse emulation and/or control of a cursor on a display, are based on computer visual identification and tracking of a user's hand, for example, as detailed above. Additionally, display 406 may be used to indicate to the user the position of the user's hand within the field of view.

[0092] System 400 may be operable according to methods, some embodiments of which were described above.

WO 2012/164562
[0095] According to some embodiments systems distributed to users may be later used to
PCT/IL2012/050191
construct a new, more accurate database of hand objects by obtaining data from the users and
combining the databases of all the different users' systems to create a new database of hand
(and/or non-hand) objects.

1. A method for computer vision based control of a device, the method comprising:
 obtaining a first frame comprising an image of an object within a field of view;
 5 identifying the object as a hand by applying computer vision algorithms;
 storing image related information of the identified hand ;
 obtaining a second frame comprising an image of an object within a field of view and
 identifying the object in the second frame as a hand by using the stored information of
 the identified hand; and
 10 controlling the device based on the hand identified in the first and second frames.
2. The method according to claim 1 comprising tracking the hand identified in the first
 frame and continuing the tracking only if the hand is also identified in the second
 image.
3. The method of claim 2 comprising controlling the device according to the tracking of
 15 the hand.
4. The method according to claim 1 comprising storing image related information of the
 hand identified in the second frame.
5. The method according to claim 1 comprising identifying a non-hand object and
 storing image related information of the non-hand object.
- 20 6. The method according to claim 5 comprising storing the image related information of
 the object identified as a hand and the image related information of the non-hand
 object, only if the information is different than any image related information already
 stored.
7. The method according to claim 1 comprising storing image related information of an
 25 object identified as a hand for a pre-defined period.
8. The method according to claim 7 wherein the pre-defined period is based on use.
9. The method according to claim 7 wherein the pre-defined period is based on absolute
 time.
10. The method according to claim 5 comprising storing image related information of the
 30 non-hand object for a pre-defined period.

11. The method according to claim 10 wherein the pre-defined period is based on use.

12. The method according to claim 10 wherein the pre-defined period is based on absolute time.

13. The method according to claim 5 wherein the non-hand object comprises a portion of a frame, said portion not including a hand.

14. The method according to claim 13 wherein the portion is located at a pre-determined distance or further from the position of the hand within the frame.

15. The method according to claim 13 wherein the portion includes an area in which no movement was detected.

16. The method according to claim 1 wherein the image related information comprises features selected from the group consisting of Local Binary Pattern (LBP) features, statistical parameters of grey level and Speeded Up Robust Features (SURF).

17. The method according to claim 1 wherein identifying the object in the second frame as a hand by using the information of the identified hand comprises:

detecting in the identified hand a set of features;

assigning a value to each feature; and

comparing the values of the features to a hand identification threshold, said hand identification threshold constructed by using values of features of formerly identified hands.

18. The method according to claim 17 comprising constructing a new hand identification threshold every pre-defined period.

19. The method according to claim 1 comprising identifying the object in the first image as a hand only if the object is moving in a pre-defined movement.

20. The method according to claim 19 wherein the pre-defined movement is a wave like movement.

21. The method according to claim 1 wherein an object identified as a hand comprises a hand in any posture or post-posture.

22. The method according to claim 1 wherein an object identified as a hand is selected from the group consisting of a hand with all fingers extended, a hand with all fingers brought together such that their tips are touching or almost touching and a hand during the act of extending fingers after having held them in a fist or closed fingers posture.

23. The method according to claim 21 comprising controlling the device according to a posture of the hand.
24. The method according to claim 21 wherein the object identified as a hand comprises a hand in a predefined posture, the method comprising
- 5 storing image related shape information of the hand in the predefined posture ;
- obtaining a second frame comprising an image of an object within a field of view and identifying the object in the second frame as a hand in the predefined posture by using the stored shape information; and
- controlling the device based on the identified predefined posture.
- 10 25. A system for computer vision based control of a device, the system comprising:
- an adaptive detector, said detector configured to
- identify an object in a first image as a hand;
- store image related information of the identified hand; and
- identify an object in a second image as a hand by using the stored
- 15 image related information;
- a processor to track the identified hand; and
- a controller to control the device based on the identified hand.
26. The system according to claim 25 comprising an image sensor to obtain the first and second images, said image sensor in communication with the adaptive detector.
- 20 27. The system according to claim 25 comprising a processor to identify a hand gesture and wherein the controller generates a user command based on the identified hand gesture.
28. The system according to claim 25 comprising a processor to identify a hand posture and wherein the controller generates a user command based on the identified hand
- 25 posture.
29. The system according to claim 25 wherein the device is selected from the group consisting of a TV, DVD player, PC, mobile phone, camera, STB (Set Top Box) and a streamer.

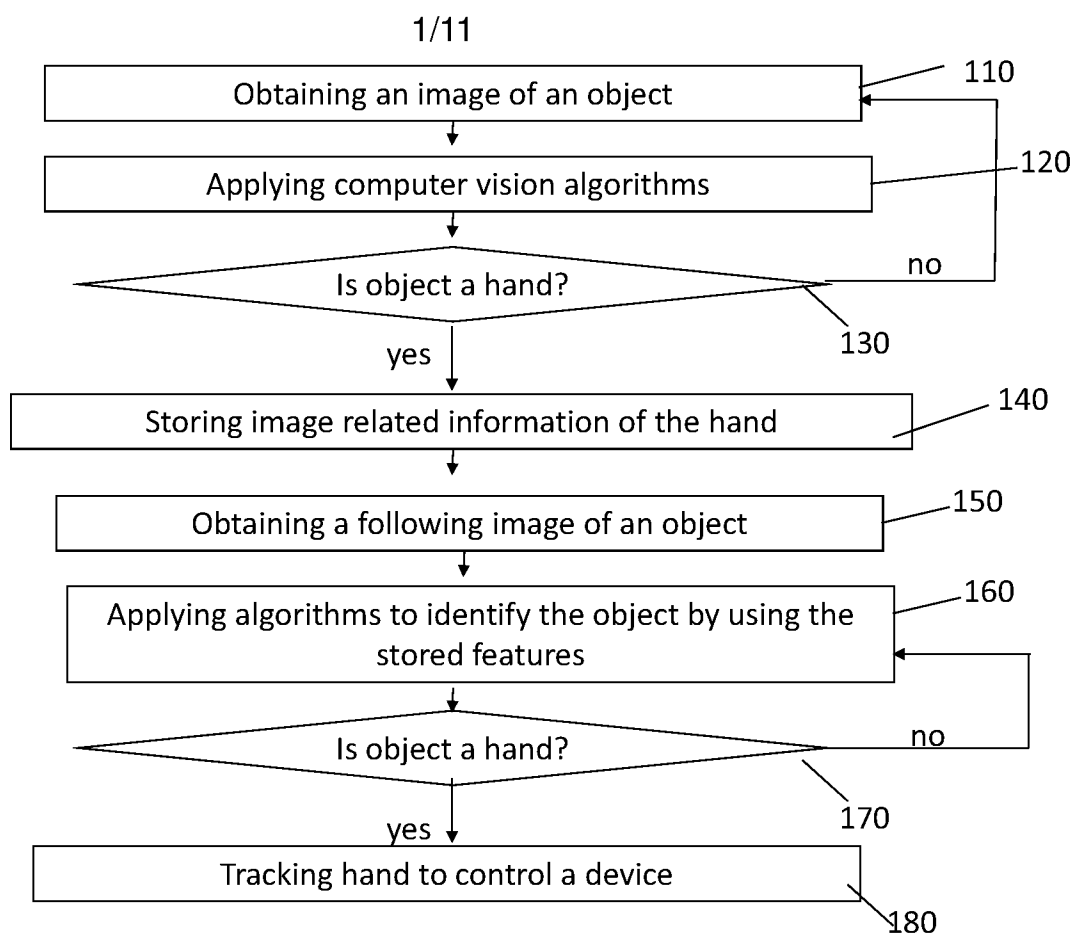


FIGURE 1A

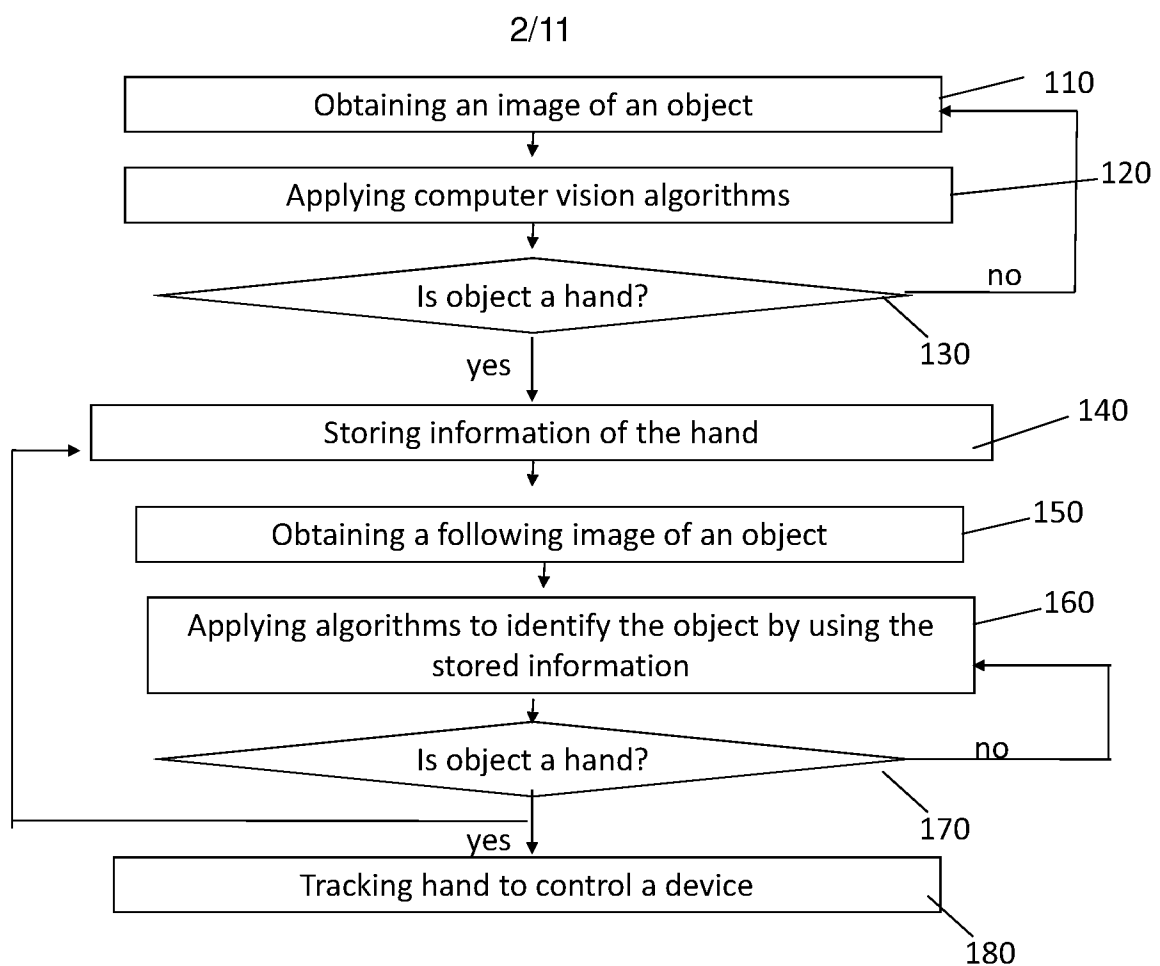


FIGURE 1B

3/11

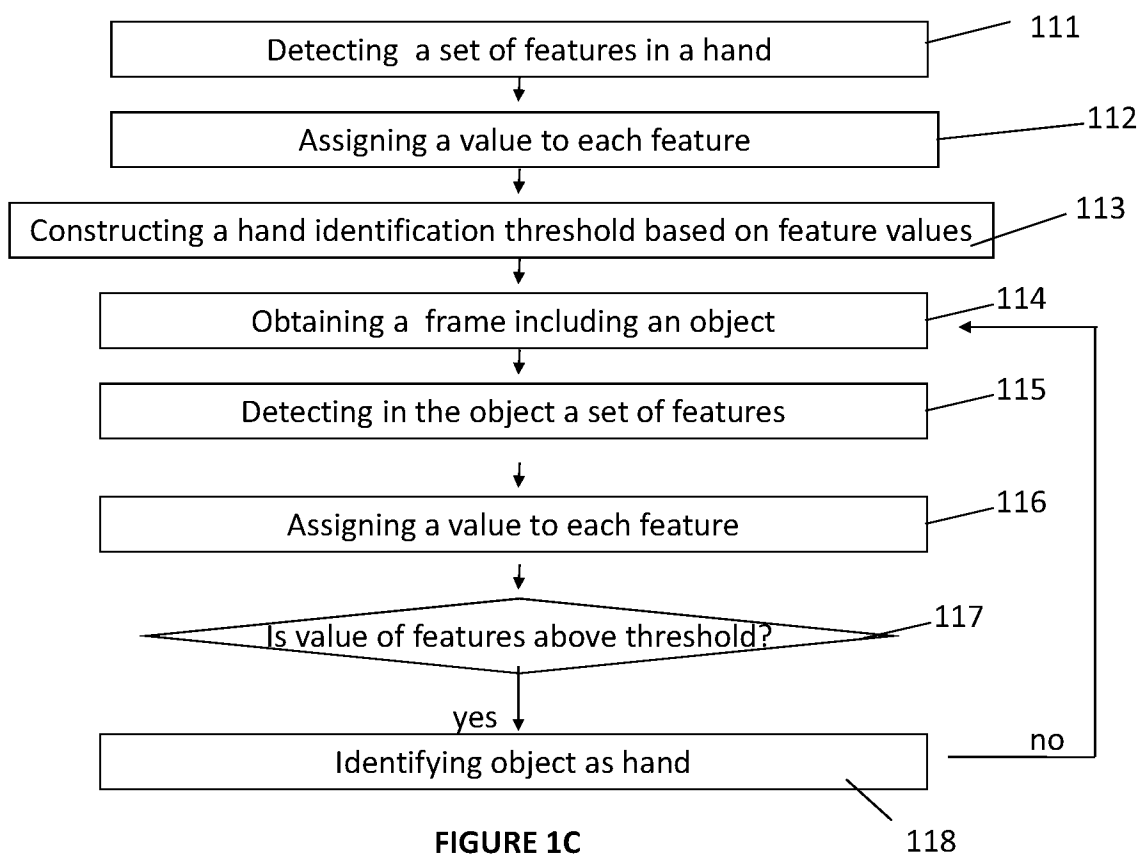


FIGURE 1C

4/11

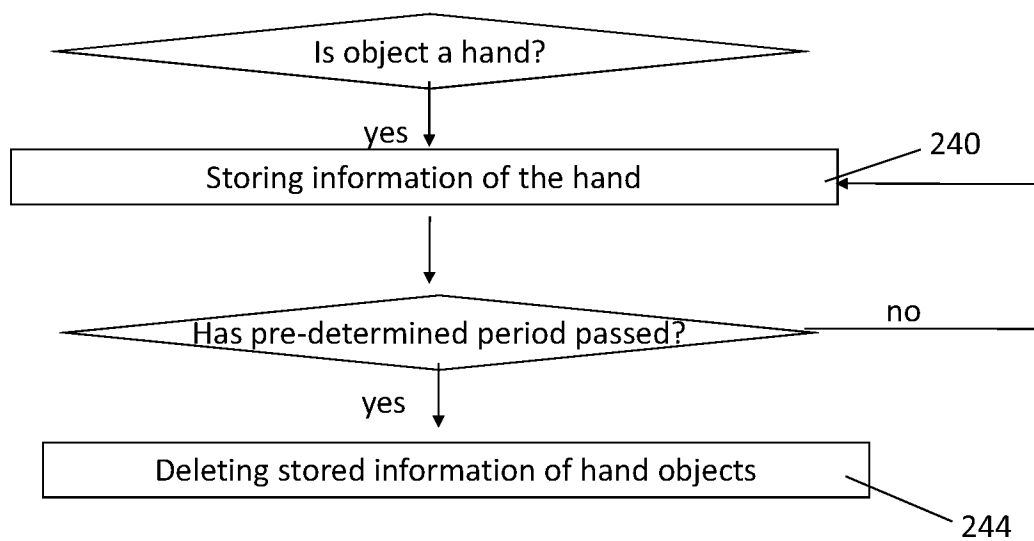


FIGURE 2A

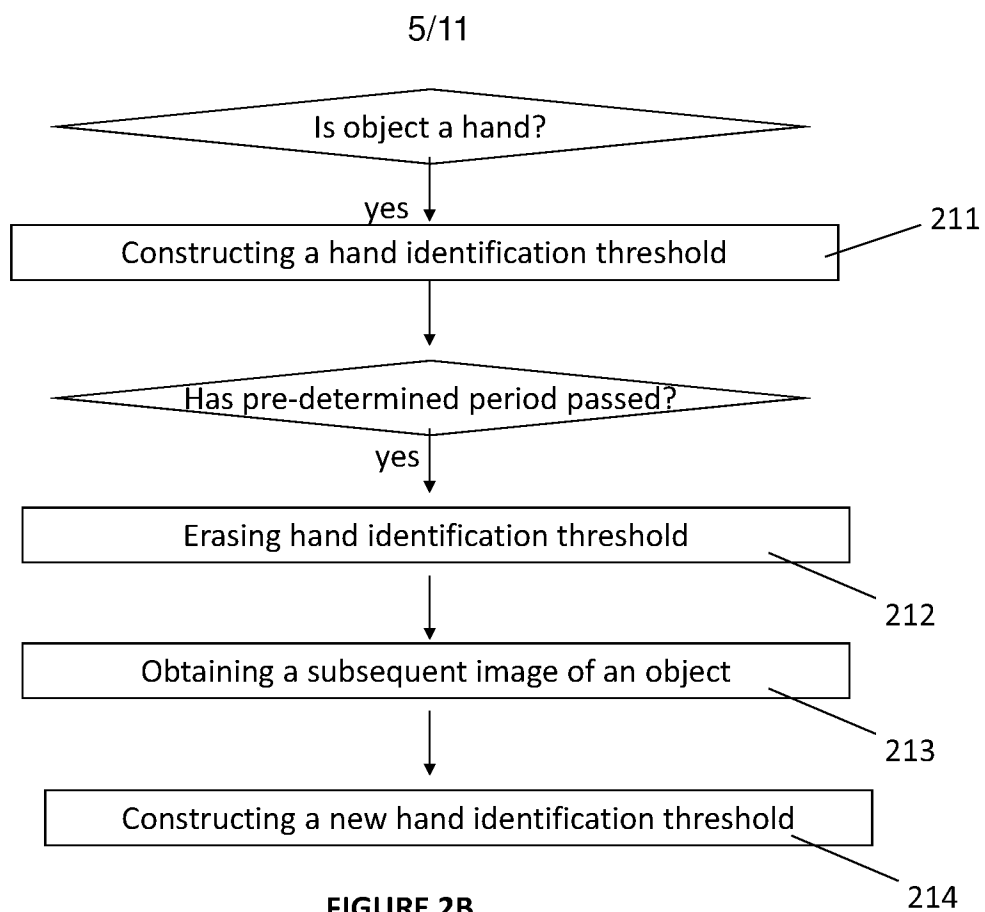


FIGURE 2B

6/11

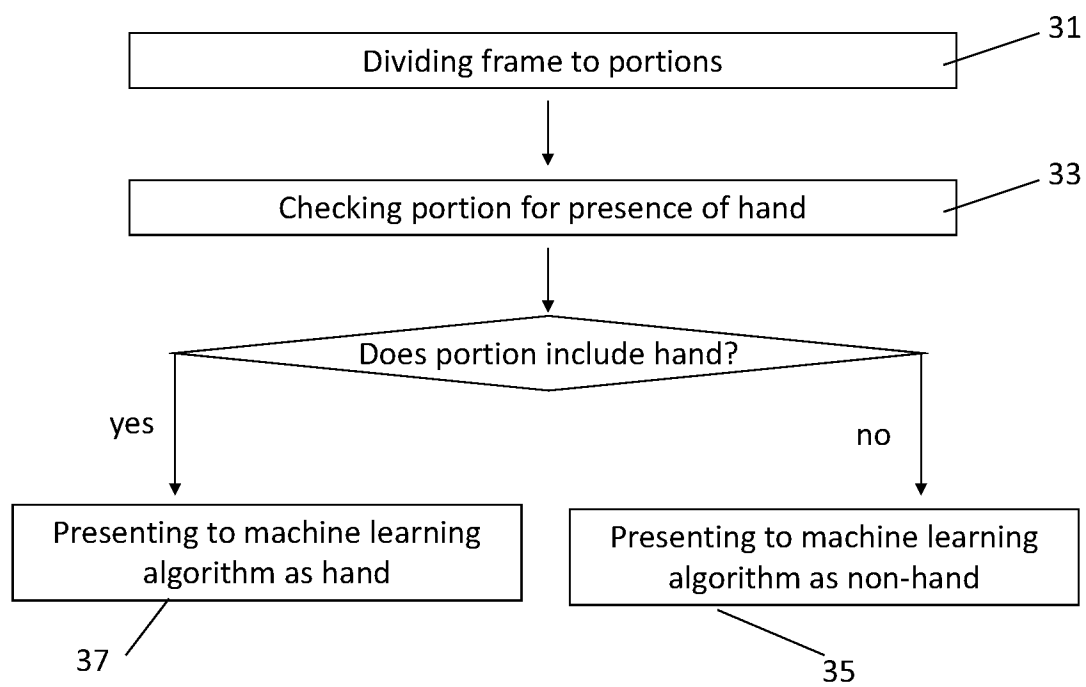


FIGURE 3A

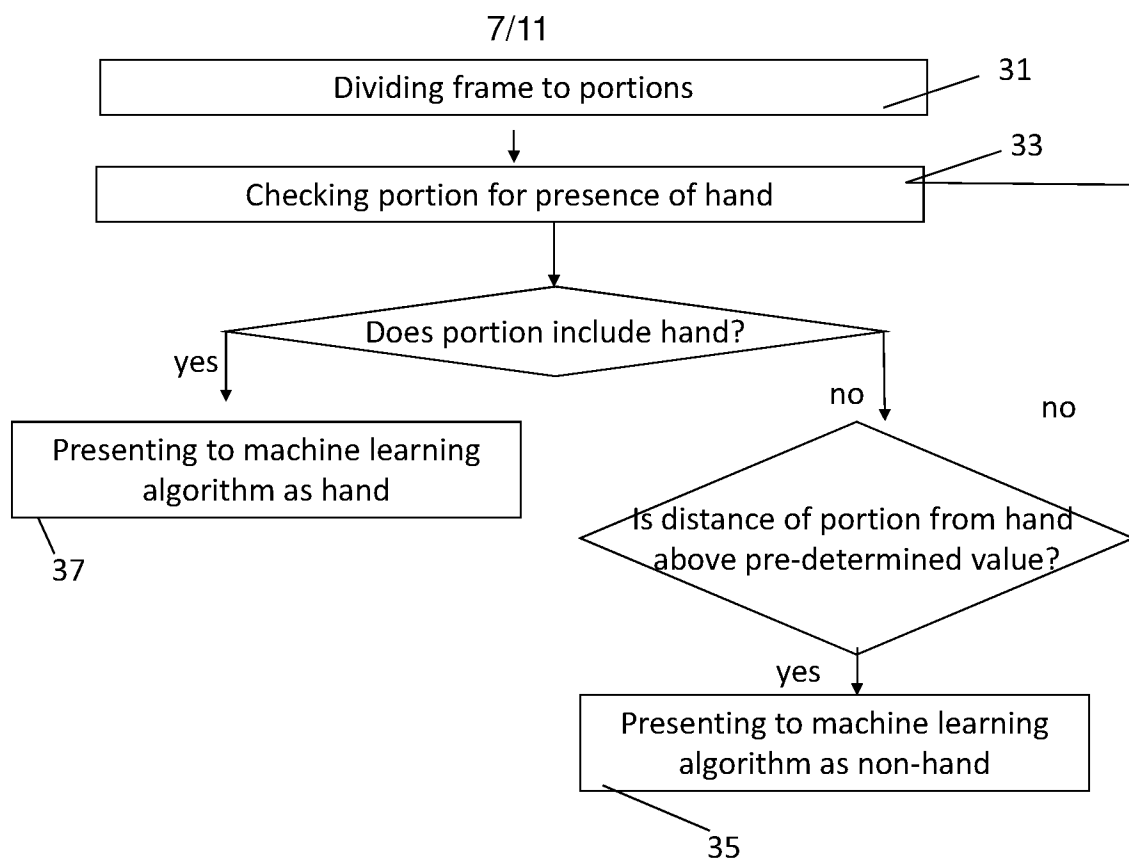


FIGURE 3B

8/11

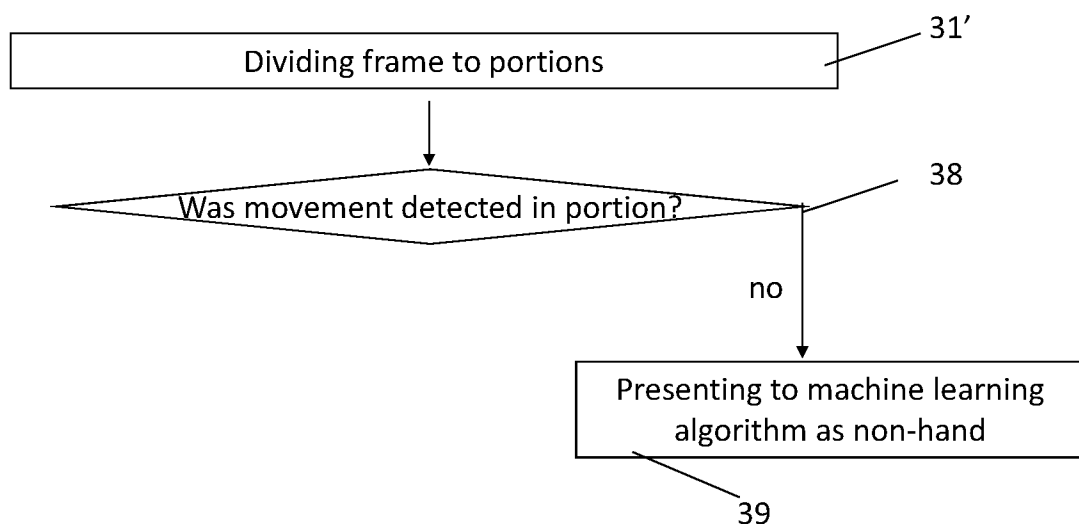


FIGURE 3C

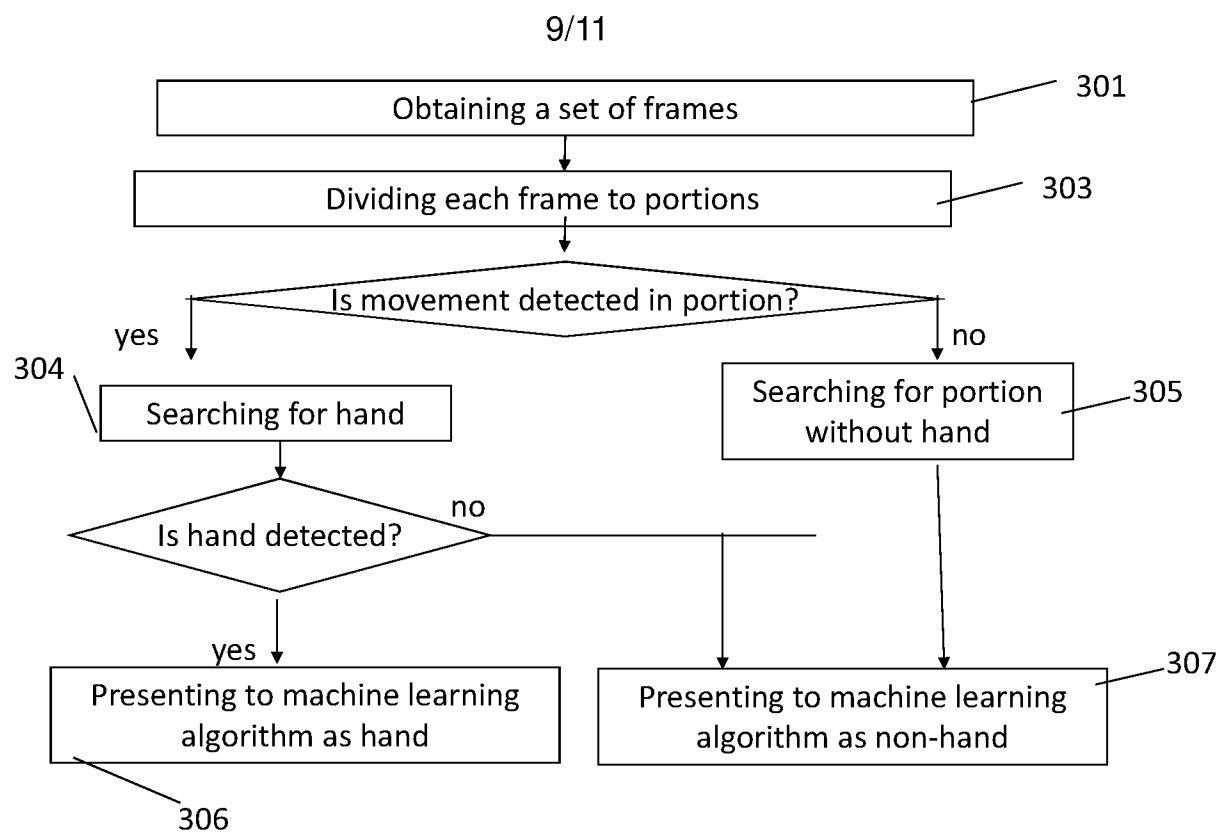
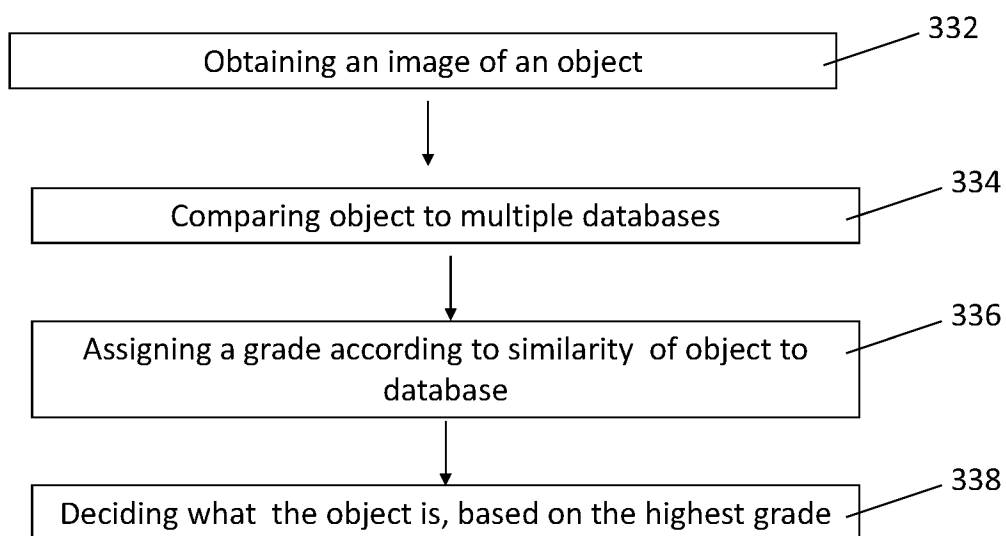


FIGURE 3D

10/11

**FIGURE 3E**

11/11

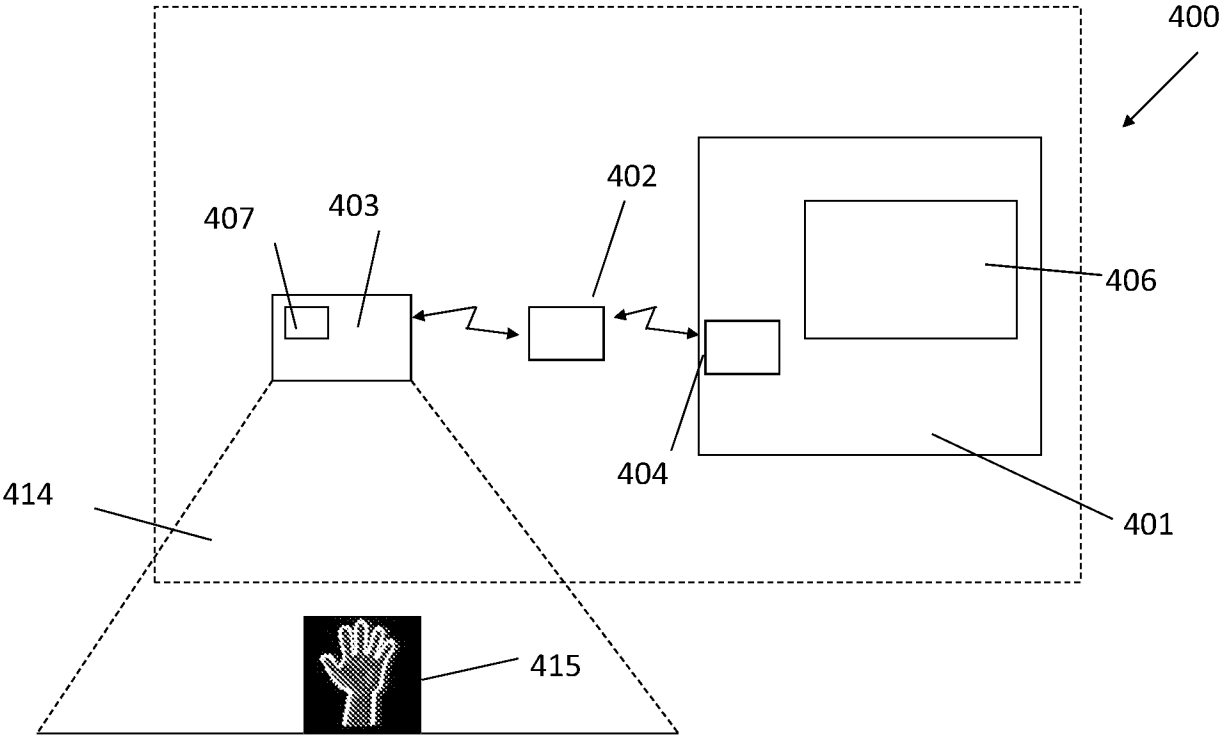


FIGURE 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL 12/50191

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 3/033 (2012.01)

USPC - 715/863

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)

USPC: 715/863; IPC(8): G06F 3/033 (2012.01)

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

USPC: 345/156; 382/103, 154, 155, 159

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

PubWEST: (PGPB, USPT, EPAB, JPAB); Google Scholar

Search Terms: control device, computer vision, gesture detect, motion detect, hand detect, finger detect, obtain frame, obtain image, field view, FOV, identify hand, algorithm, code, software, program, store image, second frame, track hand, non-hand, stationary, motionless

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2011/0026765 A1 (IVANICH et al.) 03 February 2011 (03.02.2011), abstract and para [0010]-[0014], [0018]-[0019], [0022]-[0025], [0030], [0033], [0036]-[0040], [0045]-[0047], [0052], [0055], [0059]-[0063], [0066].	1-6, 17, 19-29 ----- 7-16, 18
Y	US 2011/0102570 A1 (WILF et al.) 05 May 2011 (05.05.2011), abstract and para [0167], [0177], [0192], [0195], [0197], [0199], [0203], [0213]-[0219].	7-15
Y	US 2010/0199232 A1 (MISTRY et al.) 05 August 2010 (05.08.2010), abstract and para [0125].	16
Y	US 2010/0050134 A1 (CLARKSON) 25 February 2010 (25.02.2010), abstract and para [0032], [0050], [0071], [0077], [0106].	18
A	US 2011/0025601 A1 (WILSON et al.) 03 February 2011 (03.02.2011), entire document.	1-29
A	US 2010/0281440 A1 (UNDERKOFFLER et al.) 04 November 2010 (04.11.2010), entire document	1-29
A	US 7,274,803 B1 (SHARMA et al.) 25 September 2007 (25.09.2007), entire document	1-29
A	US 6,996,460 B1 (KRAHNSTOEVEER et al.) 07 February 2006 (07.02.2006), entire document	1-29

☐ Further documents are listed in the continuation of Box C.

* 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

06 September 2012 (06.09.2012)

Date of mailing of the international search report

18 SEP 2012

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774