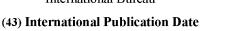
#### (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

# (19) World Intellectual Property Organization

International Bureau

19 September 2019 (19.09.2019)







(10) International Publication Number WO 2019/178114 A1

(51) International Patent Classification: *G06F 3/01* (2006.01) *G06K 9/00* (2006.01)

(21) International Application Number:

PCT/US2019/021872

(22) International Filing Date:

12 March 2019 (12.03.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/642,533

13 March 2018 (13.03.2018) US

(71) Applicant: MAGIC LEAP, INC. [US/US]; 7500 W. Sunrise Blvd., Plantation, FL 33322 (US).

- (72) Inventors: DASH, Denver, Halbert; 7500 W. Sunrise Blvd., Plantation, FL 33322 (US). TANIRGAN, Emre; 7500 W. Sunrise Blvd., Plantation, FL 33322 (US).
- (74) Agent: LEUNG, Kevin, K. et al.; Vista IP Law Group LLP, 2160 Lundy Ave., Suite 230, San Jose, California 95131 (US).
- (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

#### (54) Title: GESTURE RECOGNITION SYSTEM AND METHOD OF USING SAME

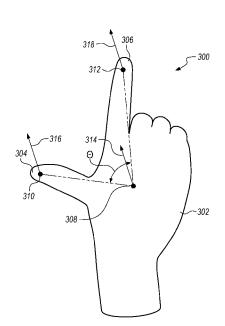


FIG. 16

(57) **Abstract:** A method of executing a gesture command includes identifying a hand centroid of a hand. The method also includes identifying a first finger tip of a first finger on the hand. The method further includes identifying a thumb tip of a thumb on the hand. Moreover, the method includes determining a surface normal relationship between the hand centroid, the first finger tip, and the thumb tip.



## **GESTURE RECOGNITION SYSTEM AND METHOD OF USING SAME**

#### **Cross-Reference to Related Applications**

**[0001]** The present disclosure claims priority to U.S. Provisional Application Nos. 62/642,533, filed on March 13, 2018, entitled "Gesture Recognition System and Method of Using Same." The contents of the aforementioned patent application are hereby expressly and fully incorporated by reference in their entirety, as though set forth in full.

5

15

20

25

30

# Field of the Invention

[0002] The present disclosure relates to mixed reality systems including gesture recognition systems, and methods for using same.

#### **Background**

**[0003]** Modern computing and display technologies have facilitated the development of "mixed reality" or "MR" systems for so called "virtual reality" or "augmented reality" experiences, wherein digitally reproduced images or portions thereof are presented to a user in a manner wherein they seem to be, or may be perceived as, real. A virtual reality, or "VR", scenario typically involves presentation of digital or virtual image information without transparency to actual real-world visual input. An augmented reality, or "AR", scenario typically involves presentation of digital or virtual image information as an augmentation to visualization of the actual world around the user (i.e., transparency to other actual real-world visual input). Accordingly, AR scenarios involve presentation of digital or virtual image information with transparency to other actual real-world visual input.

**[0004]** Various optical systems generate images at various depths for displaying mixed reality (VR and AR) scenarios. Some such optical systems are described in U.S. U.S. Utility Patent Application Serial No. 14/555,585 filed on November 27, 2014 (attorney docket number ML.20011.00), the contents of which are hereby expressly and fully incorporated by reference in their entirety, as though set forth in full.

**[0005]** Mixed reality systems typically employ wearable display devices (e.g., head-worn displays, helmet-mounted displays, or smart glasses) that are at least loosely coupled to a user's head, and thus move when the user's head moves. If the

user's head motions are detected by the display system, the data being displayed can be updated to take the change in head pose (i.e., the orientation and/or location of user's head) into account.

5

10

15

20

25

30

**[0006]** As an example, if a user wearing a head-worn display views a virtual representation of a three-dimensional (3-D) object on the display and walks around the area where the 3-D object appears, that 3-D object can be re-rendered for each viewpoint, giving the user the perception that he or she is walking around an object that occupies real space. If the head-worn display is used to present multiple objects within a virtual space (for instance, a rich virtual world), measurements of head pose can be used to re-render the scene to match the user's dynamically changing head location and orientation and provide an increased sense of immersion in the virtual space. However, there is an inevitable lag between rendering a scene and displaying/projecting the rendered scene.

**[0007]** Head-worn displays that enable AR provide concurrent viewing of both real and virtual objects. With an "optical see-through" display, the user can see through transparent (or semi-transparent) elements in the display system to view directly the light from real objects in the environment. The transparent element, often referred to as a "combiner," superimposes light from the display over the user's view of the real world, where light produced by the display projects an image of the virtual content over the see-through view of the real world. A camera (e.g., a field of view or FOV camera) may be mounted onto the wearable display to capture images or videos of the scene being viewed by the user.

**[0008]** Current 3-D optical systems, such as those in mixed reality systems, optically render virtual objects. Objects are "virtual" in that they are not real physical objects located in respective positions in 3-D space. Instead, virtual objects only exist in the brains (e.g., the optical centers) of viewers and/or listeners when stimulated by light beams directed to the eyes of audience members.

**[0009]** MR systems allow virtual objects to interact with real world physical objects. Some of these interactions make use of user input. Mixed reality systems can receive user input through gesture recognition sub-systems, thereby eliminating separate input devices (e.g., keyboards, mice, joysticks, totems, etc.) from the mixed reality systems. Examples of such gesture recognition system are described in U.S. U.S. Utility Patent Application Serial No. 14/331,218 filed on July 14, 2014 (attorney docket number ML-30020.00), the contents of which are hereby expressly and fully

incorporated by reference in their entirety, as though set forth in full. In such systems, a user's hands function as the input devices. However, the complexity of the human hand and it's wide range of articulation complicates gesture recognition. There exists a need for more accurate and precise/repeatable gesture recognition and systems for accomplishing same.

# **Summary**

5

10

15

20

25

30

**[0010]** In one embodiment, a method of executing a gesture command includes identifying a hand centroid of a hand. The method also includes identifying a first finger tip of a first finger on the hand. The method further includes identifying a thumb tip of a thumb on the hand. Moreover, the method includes determining a surface normal relationship between the hand centroid, the first finger tip, and the thumb tip.

**[0011]** In one or more embodiments, the surface normal relationship includes respective surface normals at the hand centroid on a caudal surface of the hand, at the first finger tip, and at the thumb tip are substantially parallel. A virtual cursor may be displayed in a field of view at a location corresponding to the first finger tip. The method may also include detecting a movement of the thumb tip to a mapped planar position between the hand centroid and the first finger tip, and executing a command at the cursor location in response to detecting the movement.

[0012] In one or more embodiments, the surface normal relationship includes a surface normal at the hand centroid on a caudal surface of the hand is substantially orthogonal to the respective surface normals at the first finger tip and at the thumb tip. A virtual cursor may be displayed in a field of view at a location between the first finger tip and the thumb tip. The method of may also include detecting that a first finger tip position of the first finger tip and a thumb tip position of the thumb tip are substantially similar, and executing a command at the cursor location in response to detecting of the substantially similar first finger tip and thumb tip positions.

**[0013]** In one or more embodiments, the hand centroid, the first finger tip, and the thumb tip are outside of a field of view of a user. In other embodiments, the hand centroid, the first finger tip, and the thumb tip may be inside of a field of view of a camera worn by the user.

**[0014]** In another embodiment, a method of executing a gesture command includes identifying a hand centroid of a hand. The method also includes identifying

a first finger tip of a first finger on the hand. The method further includes identifying a thumb tip of a thumb on the hand. Moreover, the method includes calculating a first line between the first finger tip and the hand centroid. In addition, the method includes calculating a second line between the thumb tip and the hand centroid. The method also includes determining an angular relationship between the first and second lines.

5

10

15

20

25

30

[0015] In one or more embodiments, the angular relationship is that the first and second lines are substantially orthogonal. A virtual cursor may displayed in a field of view at a location corresponding to the first finger tip. The method may also include detecting a change in the angular relationship between the first and second lines, and executing a command at the cursor location in response to detecting the change in the angular relationship. The change in the angular relationship may be a threshold amount of decrease in an angle between first and second lines. The threshold amount of decrease may be about 30 degrees..

**[0016]** Additional and other objects, features, and advantages of the invention are described in the detail description, figures and claims.

# **Brief Description of the Drawings**

[0017] The drawings illustrate the design and utility of various embodiments of the present invention. It should be noted that the figures are not drawn to scale and that elements of similar structures or functions are represented by like reference numerals throughout the figures. In order to better appreciate how to obtain the above-recited and other advantages and objects of various embodiments of the invention, a more detailed description of the present inventions briefly described above will be rendered by reference to specific embodiments thereof, which are illustrated in the accompanying drawings. Understanding that these drawings depict only typical embodiments of the invention and are not therefore to be considered limiting of its scope, the invention will be described and explained with additional specificity and detail through the use of the accompanying drawings in which:

**[0018]** Figure 1 depicts a user's view of augmented reality (AR) through a wearable AR user device according to one embodiment.

**[0019]** Figure 2 depicts an AR system including a gesture recognition system according to one embodiment.

**[0020]** Figures 3-11 are schematic diagrams illustrating gesture recognition according to various embodiments.

**[0021]** Figure 12 is a process flow diagram of an exemplary method of performing an action based on a recognized gesture according to one embodiment.

5 **[0022]** Figure 13 is a plan view illustrating various finger gestures according to various embodiments.

**[0023]** Figures 14-20 are a series of plan views illustrating steps in a method of recognizing a gesture and performing an action based on the recognized gesture according to one embodiment.

10 **[0024]** Figures 21-26 are a series of plan views illustrating steps in a method of recognizing a gesture and performing an action based on the recognized gesture according to another embodiment.

15

20

25

30

**[0025]** Figure 27 is a block diagram schematically depicting an illustrative computing system suitable for implementing an embodiment of the present disclosure.

#### **Detailed Description**

**[0026]** Various embodiments of the invention are directed to systems, methods, and articles of manufacture for recognizing gestures in a single embodiment or in multiple embodiments. Other objects, features, and advantages of the invention are described in the detailed description, figures, and claims.

[0027] Various embodiments will now be described in detail with reference to the drawings, which are provided as illustrative examples of the invention so as to enable those skilled in the art to practice the invention. Notably, the figures and the examples below are not meant to limit the scope of the present invention. Where certain elements of the present invention may be partially or fully implemented using known components (or methods or processes), only those portions of such known components (or methods or processes) that are necessary for an understanding of the present invention will be described, and the detailed descriptions of other portions of such known components (or methods or processes) will be omitted so as not to obscure the invention. Further, various embodiments encompass present and future known equivalents to the components referred to herein by way of illustration.

[0028] The gesture recognition systems may be implemented independently of mixed reality systems, but some embodiments below are described in relation to

MR/AR systems for illustrative purposes only. Further, the gesture recognition systems described herein may also be used in an identical manner with VR systems.

## Illustrative Mixed Reality Scenario and System

**[0029]** The description that follows pertains to an illustrative augmented reality system with which the gesture recognition system may be practiced. However, it is to be understood that the embodiments also lends themselves to applications in other types of display systems (including other types of mixed reality systems), and therefore the embodiments are not to be limited to only the illustrative system disclosed herein.

5

10

15

25

30

[0030] Mixed reality (e.g., AR) scenarios often include presentation of virtual content (e.g., images and sound) corresponding to virtual objects in relationship to real-world objects. For example, referring to Figure 1, an augmented reality scene 100 is depicted wherein a user of an AR technology sees a real-world, physical, park-like setting 102 featuring people, trees, buildings in the background, and a real-world, physical concrete platform 104. In addition to these items, the user of the AR technology also perceives that he "sees" a virtual robot statue 106 standing upon the real-world, physical platform 104, and a virtual cartoon-like avatar character 108 flying by which seems to be a personification of a bumblebee, even though these virtual objects 106, 108 do not exist in the real world.

20 **[0031]** Like AR scenarios, VR scenarios can also receive user input from gestures to allow interaction with virtual/digital objects in the scenarios. Accurately recognizing user gestures can improve the AR/VR scenarios.

**[0032]** The description that follows pertains to an illustrative augmented reality system with which the invention may be practiced. However, it is to be understood that the invention also lends itself to applications in other types of augmented reality and virtual reality systems, and therefore the invention is not to be limited to only the illustrative system disclosed herein.

[0033] Referring to Figure 2, one embodiment of an augmented reality system 200 constructed in accordance with present embodiments will now be described. The augmented reality system 200 may include one or more at least partially transparent surfaces through which the ambient environment can be seen and through which the augmented reality system 200 produces images of virtual objects (e.g., using one or more light projectors).

**[0034]** The augmented reality system 200 includes a frame structure 202 worn by a user 250, a display system 204 carried by the frame structure 202, such that the display system 204 is positioned in front of the eyes of the user 250, and a speaker 206 and a FOV camera 208 incorporated into or connected to the frame structure 202. In the illustrated embodiment, the speaker 206 is carried by the frame structure 202, such that the speaker 206 is positioned adjacent (in or around) the ear canal of the user 250, e.g., an earbud or headphone. In the illustrated embodiment, the FOV camera 208 is carried by the frame structure 202, such that the FOV camera 208 is forward facing and adjacent the user's eyes. The augmented reality system 200 may be operated in conjunction with a gesture recognition subsystem 220, and operatively coupled by a wired or wireless connection 210 thereto.

5

10

15

20

25

30

**[0035]** The gesture recognition subsystem 220 includes an object recognizer 222, a geometric analyzer 224, and a memory ("RAM") 226. The object recognizer 222 is configured to process image data and identify objects therein. The geometric analyzer 224 is configured to process object data and identify geometric characteristics thereof. The identified objects and corresponding geometric characteristics can be stored in the RAM 226.

**[0036]** The forward facing FOV camera 208 may be employed for any number of purposes, such as recording of images/video from the forward direction of the system 200. In addition, the FOV cameras 208 may be used to capture information about the environment in which the user 250 is located, such as information indicative of distance, orientation, and/or angular position of the user 250 with respect to that environment and specific objects in that environment. The FOV cameras 208 may also be used to capture information about the user's hands. That information may be analyzed by the gesture recognition subsystem 220 to attempt to recognize a gesture for user input.

[0037] The augmented reality system 200 may further include one or more sensors mounted to the frame structure 202 for detecting the position (including orientation) and movement of the head of the user 250 and/or the eye position and inter-ocular distance of the user 250. Such sensor(s) may include image capture devices, microphones, inertial measurement units (IMUs), accelerometers, compasses, GPS units, radio devices, and/or gyros. For example, in one embodiment, the augmented reality system 200 comprises a head worn transducer subsystem that includes one or more inertial transducers to capture inertial

measures indicative of movement of the head of the user 250. Such devices may be used to sense, measure, or collect information about the head movements of the user 250. For instance, these devices may be used to detect/measure movements, speeds, acceleration and/or positions of the head of the user 250. The position (including orientation) of the head of the user 250 is also known as the user 250 "head pose."

5

10

15

20

25

30

**[0038]** The augmented reality system 200 may further include rearward facing cameras to track angular position (the direction in which the eye or eyes are pointing), blinking, and depth of focus (by detecting eye convergence) of the eyes of the user 250. Such eye tracking information may, for example, be discerned by projecting light at the end user's eyes, and detecting the return or reflection of at least some of that projected light.

[0039] The augmented reality system 200 further comprises a control subsystem (not shown) that may take any of a large variety of forms. The control subsystem includes a number of controllers, for instance one or more microcontrollers, microprocessors or central processing units (CPUs), digital signal processors, graphics processing units (GPUs), other integrated circuit controllers, such as application specific integrated circuits (ASICs), programmable gate arrays (PGAs), for instance field PGAs (FPGAs), and/or programmable logic controllers (PLUs). The control subsystem may be separate from the gesture recognition subsystem 220 or may include the gesture recognition subsystem 220.

may be physically contained in a distributed subsystem. For example, the augmented reality system 200 comprises a local processing and data module operatively coupled, such as by a wired lead or wireless connectivity 210, to a portion of the display subsystem 204. The local processing and data module may be mounted in a variety of configurations, such as fixedly attached to the frame structure 202, fixedly attached to a helmet or hat, embedded in headphones, removably attached to the torso of the user 250, or removably attached to the hip of the user 250 in a belt-coupling style configuration. The augmented reality system 200 may further include a remote processing module and remote data repository operatively coupled, such as by a wired lead or wireless connectivity to the local processing and data module, such that these remote modules are operatively coupled to each other and available as resources to the local processing and data

module. The local processing and data module may comprise a power-efficient processor or controller, as well as digital memory, such as flash memory, both of which may be utilized to assist in the processing, caching, and storage of data captured from the sensors and/or acquired and/or processed using the remote processing module and/or remote data repository, possibly for passage to the display subsystem 204 after such processing or retrieval. The remote processing module may comprise one or more relatively powerful processors or controllers configured to analyze and process data and/or image information. The remote data repository may comprise a relatively large-scale digital data storage facility, which may be available through the internet or other networking configuration in a "cloud" resource configuration. In one embodiment, all data is stored and all computation is performed in the local processing and data module, allowing fully autonomous use from any remote modules. The couplings between the various components described above may include one or more wired interfaces or ports for providing wires or optical communications, or one or more wireless interfaces or ports, such as via RF, microwave, and IR for providing wireless communications. In some implementations, all communications may be wired, while in other implementations all communications may be wireless, with the exception of the optical fiber(s).

5

10

15

#### Summary of Problems and Solutions

20 [0041] The inaccuracy and lack of repeatability in current gesture recognition systems can result in user input errors that can frustrate users. The embodiments described herein identify the centroid and various surface normals of a user's hand. By comparing these geometric characteristics of a user's hand, the embodiments described herein more accurately recognize user gestures, therein improving user input.

#### Gesture Recognition Systems and Methods

**[0042]** In some implementations, the AR system may detect and be responsive to one or more finger/hand gestures. These gestures can take a variety of forms and may, for example, be based on inter-finger interaction, pointing, tapping, rubbing, etc. Other gestures may, for example, include 2D or 3D representations of characters (e.g., letters, digits, punctuation). To enter such, a user swipes their finger in the defined character pattern. Other gestures may include thumb/wheel selection type gestures, which may, for example be used with a "popup" circular radial menu which may be rendered in a field of view of a user, according to one illustrated embodiment.

5

10

15

20

25

30

**[0043]** Embodiments of the AR system can therefore recognize various commands using gestures, and in response perform certain functions mapped to the commands. The mapping of gestures to commands may be universally defined, across many users, facilitating development of various applications which employ at least some commonality in user interface. Alternatively or additionally, users or developers may define a mapping between at least some of the gestures and corresponding commands to be executed by the AR system in response to detection of the commands.

**[0044]** For example, a pointed index finger may indicate a command to focus, for example to focus on a particular portion of a scene or virtual content at which the index finger is pointed. A pinch gesture can be made with the tip of the index finger touching a tip of the thumb to form a closed circle, e.g., to indicate a grab and/or copy command. Another example pinch gesture can be made with the tip of the ring finger touching a tip of the thumb to form a closed circle, e.g., to indicate a select command.

**[0045]** Yet another example pinch gesture can be made with the tip of the pinkie finger touching a tip of the thumb to form a closed circle, e.g., to indicate a back and/or cancel command. A gesture in which the ring and middle fingers are curled with the tip of the ring finger touching a tip of the thumb may indicate, for example, a click and/or menu command. Touching the tip of the index finger to a location on the head worn component or frame may indicate a return to home command.

**[0046]** Embodiments of the invention provide an advanced system and method for performing gesture tracking and identification. In one embodiment, a rejection cascade approach is performed, where multiple stages of gesture analysis is

performed upon image data to identify gestures. As shown in the cascade 3600 of Figure 3, incoming images (e.g., an RGB image at a depth D) is processed using a series of permissive analysis nodes. Each analysis node performs a distinct step of determining whether the image is identifiable as a gesture. Each stage in this process (i.e., analysis node) performs a targeted computation so that the sequence of stages/analysis nodes can be used to efficiently perform the gesture processing. This means, for example, that the amount of processing power at each stage of the process, along with the sequence/order of the nodes, can be used to optimize the ability to remove non-gestures while doing so with minimal computational expenses. For example, computationally less-expensive algorithms may be applied in the earlier stages to remove large numbers of "easier" candidates, thereby leaving smaller numbers of "harder" data to be analyzed in later stages using more computationally expensive algorithms.

5

10

15

20

25

30

[0047] The general approach to perform this type of processing in one embodiment is shown in the flowchart 3601 of Figure 4. The first step is to generate candidates for the gesture processing (step 3602). These include, for example, images captured from sensor measurements of the wearable device, e.g., from camera(s) mounted on the wearable device. Next, analysis is performed on the candidates to generate analysis data (step 3604). For example, one type of analysis may be to check on whether the contour of the shapes (e.g., fingers) in the image is sharp enough. Sorting is then performed on the analyzed candidates (step 3606). Finally, any candidate that corresponds to a scoring/analysis value that is lower than a minimum threshold is removed from consideration (step 3608).

**[0048]** Figure 5 depicts a more detailed approach 3650 for gesture analysis according to one embodiment of the invention. The first action is to perform depth segmentation upon the input data. For example, typically the camera providing the data inputs (e.g., the camera producing RGB+depth data) will be mounted on the user's head, where the camera FOV (field of view) will cover the range in which the human could reasonably perform gestures. As shown in illustration 3660 of Figure 6, a line search can be performed through the data (e.g., from the bottom of the FOV).

**[0049]** If there are identifiable points along that line, then potentially a gesture has been identified. Performing this analysis over a series of lines can be used to generate the depth data. In some embodiment, this type of processing can be quite sparse –perhaps where 50 points are acquired relatively really quickly. Of course,

different kinds of line series can be employed, e.g., in addition to or instead of flat lines across the bottom, smaller diagonal lines are employed in the area where there might be a hand/arm.

**[0050]** Any suitable pattern may be employed, selecting ones that are most effective at detecting gestures. In some embodiments, a confidence-enhanced depth map is obtained, where the data is flood filled from cascade processing where a "flood fill" is performed to check for and filter whether the identified object is really a hand/arm. The confidence enhancement can be performed, for example, by getting a clear map of the hand and then checking for the amount of light that is reflected off the hand in the images to the sensor, where the greater amount of light corresponds to a higher confidence level to enhance the map.

5

10

15

20

25

30

**[0051]** From the depth data, one can cascade to perform immediate/fast processing, e.g., where the image data is amenable to very fast recognition of a gesture. This works best for very simple gestures and/or hand/finger positions.

**[0052]** In many cases, deeper processing is to be performed to augment the depth map. For example, one type of depth augmentation is to perform depth transforms upon the data. Another type of augmentation is to check for geodesic distances from specified point sets, such as boundaries, centroids, etc. For example, from a surface location, a determination is made of the distance to various points on the map. This attempts to find, for example, the farthest point to the tip of the fingers (by finding the end of the fingers). The point sets may be from the boundaries (e.g., outline of hand) or centroid (e.g., statistical central mass location).

[0053] Surface normalization may also be calculated. In addition, curvatures may also be estimated, which identifies how fast a contour turns — and to perform a filtering process to go over the points and removing concave points from fingers. In some embodiments, orientation normalization may be performed on the data. To explain, consider that a given image of the hand may be captured with the hand in different positions. However, the analysis may be expecting a canonical position of the image data of the hand. In this situation, as shown in illustration 3670 of Figure 7, the mapped data may be re-oriented to change to a normalized/canonical hand position.

**[0054]** One advantageous approach in some embodiments of invention is to perform background subtraction on the data. In many cases, a known background exists in a scene, e.g., the pattern of a background wall. In this situation, the map of

the object to be analyzed can be enhanced by removing the background image data. An example of this process is shown in illustration 3680 of Figure 8, where the left portion of the figure shows an image of a hand over some background image data. The right-hand portion of the figure shows the results of removing the background from the image, leaving the augmented hand data with increased clarity and focus.

5

10

15

20

25

30

**[0055]** Depth comparisons may also be performed upon points in the image to identify the specific points that pertain to the hand (as opposed to the background non-hand data). For example, as shown in illustration 3690 of Figure 9, it can be seen a first point A is located at a first depth and a second point B is located at a significantly different second depth. In this situation, the difference in the depths of these two points makes it very evident that they likely belong to different objects. Therefore, if one knows that the depth of the hand is at the same depth value as point A, then one can conclude that point A is part of the hand. On the other hand, since the depth value for point B is not the same as the depth of the hand, then one can readily conclude that point B is not part of the hand.

**[0056]** At this point a series of analysis stages is performed upon the depth map. Any number of analysis stages can be applied to the data. The present embodiment shows three stages, but one of ordinary skill in the art would readily understand that any other number of stages (either smaller or larger) may be used as appropriate for the application to which the invention is applied.

**[0057]** In the current embodiment, stage 1 analysis is performed using a classifier mechanism upon the data. For example, a classification/decision forest can be used to apply a series of yes/no decisions in the analysis to identify the different parts of the hand for the different points in the mapping.

**[0058]** This identifies, for example, whether a particular point belongs to the palm portion, back of hand, non-thumb finger, thumb, fingertip, and/or finger joint. Any suitable classifier can be used for this analysis stage. For example, a deep learning module or a neural network mechanism can be used instead of or in addition to the classification forest. In addition, a regression forest (e.g., using a Hough transformation) can be used in addition to the classification forest.

**[0059]** The next stage of analysis (stage 2) can be used to further analysis the mapping data. For example, analysis can be performed to identify joint locations, articular, or to perform skeletonization on the data. Illustration 3695 of Figure 10 provides an illustration of skeletonization, where an original map of the hand data is

used to identify the locations of bones/joints within the hand, resulting in a type of "stick" figure model of the hand/hand skeleton. This type of model provides with clarity a very distinct view of the location of the fingers and the specific orientation and/or configuration of the hand components. Labelling may also be applied at this stage to the different parts of the hand.

5

10

15

20

25

30

**[0060]** At this point, it is possible that the data is now directly consumable by a downstream application without requiring any further analysis. This may occur, for example, if the downstream application itself includes logic to perform additional analysis/computations upon the model data. In addition, the system can also optionally cascade to perform immediate/fast processing, e.g., where the data is amenable to very fast recognition of a gesture, such as the (1) fist gesture; (2) open palm gesture; (3) finger gun gesture; (4) pinch; (5) etc. This processing may include using a neural network with buckets with confidence values. For example, as shown in illustration 3698 of Figure 11, various points on the hand mapping (e.g., point on extended thumb and point on extended first finger) can be used to immediately identify a pointing gesture. The outputs will then proceed to a world engine, e.g., to take action upon a recognized gesture.

**[0061]** In addition, deeper processing can be performed in the stage 3 analysis. This may involve, for example, using a decision forest/tree to classify the gesture. This additional processing can be used to identify the gesture, determine a hand pose, identify context dependencies, and/or any other information as needed.

**[0062]** Prior/control information can be applied in any of the described steps to optimize processing. This permits some biasing for the analysis actions taken in that stage of processing. For example, for game processing, previous action taken in the game can be used to bias the analysis based upon earlier hand positions/poses. In addition, a confusion matrix can be used to more accurately perform the analysis.

**[0063]** Using the principles of gesture recognition discussed above, the AR system may use visual input gathered from the user's FOV cameras and recognize various gestures that may be associated with a predetermined command or action. Referring now to flowchart 3700 of Figure 12, in step 3102, the AR system may detect a gesture as discussed in detail above. In some embodiments, the AR system may detect one or more gestures, for example, a series of gestures.

[0064] As described above, the movement of the fingers (e.g., one or more gestures) may be compared to a database to detect a predetermined command, in

step 3104. The AR system recognizes a command in step 3106. If the one or more gestures do correspond to any known command, the AR system determines the desired action and/or desired virtual content based on the one or more gestures, in step 3108. If the one or more gestures do not correspond to any known command, the AR system simply goes back to detecting other gestures or movements to step 3102.

5

10

15

20

25

30

**[0065]** In step 3108, the AR system determines the type of action necessary in order to satisfy the command. For example, the user may want to switch an application, or may want to turn a page, may want to generate a user interface, may want to connect to a friend located at another physical location, etc. Based on the desired action/virtual content, the AR system determines whether to retrieve information from the cloud servers, or whether the action can be performed using local resources on the user device, in step 3110.

**[0066]** For example, if the user simply wants to turn a page of a virtual book, the required data may already have been downloaded or may reside entirely on the local device, in which case, the AR system simply retrieves data associated with the next page and may display the next page to the user. Similarly, if the user wants to create a user interface such that the user can draw a picture in the middle of space, the AR system may simply generate a virtual drawing surface in the desired location without needing data from the cloud.

**[0067]** Data associated with many applications and capabilities may be stored on the local device such that the user device does not need to unnecessarily connect to the cloud or access the passable world model. Thus, if the desired action can be performed locally, local data may be used to display virtual content corresponding to the detected gesture (step 3112).

**[0068]** Alternatively, in step 3114, if the system needs to retrieve data from the cloud or the passable world model, the system may send a request to the cloud network, retrieve the appropriate data and send it back to the local device such that the action or virtual content may be appropriately displayed to the user. For example, if the user wants to connect to a friend at another physical location, the AR system may need to access the passable world model to retrieve the necessary data associated with the physical form of the friend in order to render it accordingly at the local user device.

**[0069]** Thus, based on the user's interaction with the AR system, the AR system may create many types of user interfaces as desired by the user. The following represent some exemplary embodiments of user interfaces that may be created in a similar fashion to the exemplary process described above.

**[0070]** It should be appreciated that the above process is simplified for illustrative purposes, and other embodiments may include additional steps based on the desired user interface. The following discussion goes through various types of finger gestures, that may all be recognized and used such that the AR system automatically performs an action and/or presents virtual content to the user that is either derived from the cloud or retrieved locally.

5

10

15

20

25

30

#### Finger Gestures

**[0071]** Finger gestures can take a variety of forms and may, for example, be based on inter-finger interaction, pointing, tapping, rubbing, etc.

[0072] Other gestures may, for example, include 2D or 3D representations of characters (e.g., letters, digits, punctuation). To enter such, a user swipes their finger in the defined character pattern. In one implementation of a user interface, the AR system renders three circles, each circle with specifically chosen characters (e.g., letters, digits, punctuation) arranged circumferentially around the periphery. The user can swipe through the circles and letters to designate a character selection or input. In another implementation, the AR system renders a keyboard (e.g., QWERTY keyboard) low in the user's field of view, proximate a position of the user's dominate hand in a bent-arm position. The user can than perform a swipe-like motion through desired keys, and then indicate that the swipe gesture selection is complete by performing another gesture (e.g., thumb-to-ring finger gesture) or other proprioceptive interaction.

**[0073]** Other gestures may include thumb/wheel selection type gestures, which may, for example be used with a "popup" circular radial menu—which may be rendered in a field of view of a user, according to one illustrated embodiment.

**[0074]** Gestures 3800 of Figure 13 shows a number of additional gestures. The AR system recognizes various commands, and in response performs certain functions mapped to the commands. The mapping of gestures to commands may be universally defined, across many users, facilitating development of various applications which employ at least some commonality in user interface. Alternatively or additionally, users or developers may define a mapping between at least some of

the gestures and corresponding commands to be executed by the AR system in response to detection of the commands.

**[0075]** In the top row left-most position, a pointed index finger may indicate a command to focus, for example to focus on a particular portion of a scene or virtual content at which the index finger is pointed. In the top row middle position, a first pinch gesture with the tip of the index finger touching a tip of the thumb to form a closed circle may indicate a grab and/or copy command. In the top row right-most position, a second pinch gesture with the tip of the ring finger touching a tip of the thumb to form a closed circle may indicate a select command.

5

10

15

20

25

30

[0076] In the bottom row left-most position, a third pinch gesture with the tip of the pinkie finger touching a tip of the thumb to form a closed circle may indicate a back and/or cancel command. In the bottom row middle position, a gesture in which the ring and middle fingers are curled with the tip of the ring finger touching a tip of the thumb may indicate a click (e.g., right click) and/or menu command. In the bottom row right-most position, touching the tip of the index finger to a location on the head worn component or frame may indicate a return to home command. Such may cause the AR system to return to a home or default configuration, for example displaying a home or default menu.

[0077] It should be appreciated that there may be many more types of user input not limited to the ones discussed above. For example, the system may measure neurological signals and use that as an input for the system. The system may have a sensor that tracks brain signals and map it against a table of commands. In other words, the user input is simply the user's thoughts, that may be measured by the user's brain signals. This may also be referred to as subvocalization sensing. Such a system may also include apparatus for sensing EEG data to translate the user's "thoughts" into brain signals that may be decipherable by the system.

# <u>Hand Centroid Gesture Recognition Systems and Methods</u>

#### "L"-Shaped Gesture

**[0078]** Figures 14-20 depict a hand centroid gesture recognition system according to one embodiment. Figure 14 depicts an image 300 of a user's right hand 302. This image may have been acquired using a front facing FOV camera 208. The user's hand 302 is making an "L"-shaped gesture by positioning a thumb 304 and a first finger 306 at an almost right angle to each other.

**[0079]** In one embodiment, the object recognizer 222 in the gesture recognition subsystem 220 identifies the hand 302 in the image. The object recognizer 222 receives the image data from the front facing FOV camera 208 and/or the RAM 226. The object recognizer 222 may identify the hand 302 as a hand using the methods described above. Data relating to the identification of the hand 302 is stored in the RAM 226.

5

10

15

20

25

30

**[0080]** Figure 15 depicts map points 308, 310, 312 identified in the image 300 of the hand 302. These map points correspond to a hand centroid 308, a thumb tip 310, and a first finger tip 312. In some embodiments, fewer or more map points may be identified. For example, one, two, eight, seventeen, or more points may be identified. In some embodiments, one or more images of the hand may be used to determine a 3D representation of the hand 302.

[0081] In one embodiment, the object recognizer 222 and/or the geometric analyzer 224 in the gesture recognition subsystem 220 identifies the map points correspond to the hand centroid 308, the thumb tip 310, and the first finger tip 312 in the image. The object recognizer 222 and/or the geometric analyzer 224 receives the image data from the front facing FOV camera 208 and/or the RAM 226. The object recognizer 222 and/or the geometric analyzer 224 may identify the map points 308, 310, 312 using the methods described above and/or geometric analysis methods. Data relating to the map points 308, 310, 312 in the image 300 of the hand 302 is stored in the RAM 226.

**[0082]** Figure 16 depicts surface normals 314, 316, 318 at the map points 308, 310, 312 identified in the image 300 of the hand 302. The surface normal 314 at the hand centroid map point 308 is on a caudal surface of the hand 302. The surface normals 316, 318 at the thumb tip and first finger tip map points 310, 312 are on a surface defined by the extended thumb 304 and finger 306.

**[0083]** In one embodiment, the geometric analyzer 224 in the gesture recognition subsystem 220 identifies the surface normals 314, 316, 318 at the map points 308, 310, 312 in the image. The geometric analyzer 224 receives the image and map point data from the front facing FOV camera 208 and/or the RAM 226. The geometric analyzer 224 may identify the surface normals 314, 316, 318 using geometric analysis methods. Data relating to the surface normals 314, 316, 318 in the image 300 of the hand 302 is stored in the RAM 226.

**[0084]** Because these three surface normals 314, 316, 318 are substantially parallel (e.g., within 10 degrees of each other), the gesture recognition subsystem 220 recognizes the gesture as an "L"-shape. In some embodiments, the gesture recognition subsystem 220, for "L"-shape gestures, determines an angle  $\theta$  between map points 312, 308, and 310. In some embodiments, determining that  $\theta$  is approximately 90 degrees is part of the gesture recognition subsystem 220 recognizing the gesture as an "L"-shape. In response to recognizing the "L"-shape gesture, the augmented reality system 200 displays a cursor 320 at a location adjacent the first finger tip map point 312, as shown in Figure 17. The augmented reality system 200 may "stick" the cursor 320 to the first finger tip map point 312 to maintain the spatial relationship there between even when the first finger 306 moves. In some embodiments, the AR system 200 may project the cursor 320 into the environment based on a relationship of the map points 320, 308, and/or 310, for example, using ray casting.

[0085] In one embodiment, the geometric analyzer 224 in the gesture recognition subsystem 220 analyzes the surface normals 314, 316, 318 at the map points 308, 310, 312 in the image to determine whether they are substantially parallel. The geometric analyzer 224 receives the image and map point data from the front facing FOV camera 208 and/or the RAM 226. The geometric analyzer 224 may determine whether the surface normals 314, 316, 318 are substantially parallel using geometric analysis methods. Data relating geometric relationship between the surface normals 314, 316, 318 in the image 300 of the hand 302 and the identified "L"-gesture is stored in the RAM 226.

**[0086]** Figures 18-20 depict a click command executed by a hand 302 displaying the "L"-gesture. Figure 18 shows that the tip of the user's thumb 304 has moved behind the user's hand 302. Figure 19 depicts in shadow a position of the tip of the user's thumb 304, including the map point 310 at the thumb tip, as projected by the gesture recognition subsystem 220. Figure 20 depicts a line 322 calculated by the gesture recognition subsystem 220 between the map points 312 and 308, from the first finger tip to the hand centroid. As the thumb tip 304 approaches line 322 to become occluded by the hand at map point 310,  $\theta$  decreases in value. At a threshold limit of a change in the value of  $\theta$  (from  $\theta$  in Figure 16 to  $\theta$ ' in Figure 20), in some embodiments less than about 30 degrees, the gesture recognition subsystem 220 identifies a click command executed by the hand 302 displaying the "L"-gesture.

In identifying a threshold change in the value of  $\theta$ , the gesture recognition subsystem 220 determines that the thumb tip 304 approaches line 322. In response to identification of the click command, the augmented reality system 200 executes a command (e.g., select, activate, close, etc.) at the location of the cursor 320.

5

10

15

20

25

30

[0087] In one embodiment, the object recognizer 222 and/or the geometric analyzer 224 in the gesture recognition subsystem 220 identifies the movement of the user's thumb 304 towards the position of map point 310 in Figures 19 and 20, projects the tip of the thumb 304, and analyzes the various map point 308, 310, 312 positions, all of which correspond to identification of the click command. The object recognizer 222 and/or the geometric analyzer 224 receives the image data from the front facing FOV camera 208 and/or the RAM 226. The object recognizer 222 and/or the geometric analyzer 224 may identify, project, and analyze using the methods described above and/or geometric analysis methods. Data relating to identification of the click command is stored in the RAM 226.

#### "C"-Shaped Gesture

**[0088]** Figures 21-26 depict a hand centroid gesture recognition system according to another embodiment. Figure 21 depicts an image 300 of a user's right hand 302. This image may have been acquired using a front facing FOV camera 208. The user's hand 302 is making a "C"-shaped gesture by selective positioning of a thumb 304 and a first finger 306 of the hand 302.

**[0089]** In one embodiment, the object recognizer 222 in the gesture recognition subsystem 220 identifies the hand 302 in the image. The object recognizer 222 receives the image data from the front facing FOV camera 208 and/or the RAM 226. The object recognizer 222 may identify the hand 302 as a hand using the methods described above. Data relating to the identification of the hand 302 is stored in the RAM 226.

**[0090]** Figure 22 depicts map points 308, 310, 312 identified in the image 300 of the hand 302. These map points correspond to a hand centroid 308, a thumb tip 310, and a first finger tip 312.

**[0091]** In one embodiment, the object recognizer 222 and/or the geometric analyzer 224 in the gesture recognition subsystem 220 identifies the map points correspond to the hand centroid 308, the thumb tip 310, and the first finger tip 312 in the image. The object recognizer 222 and/or the geometric analyzer 224 receives the image data from the front facing FOV camera 208 and/or the RAM 226. The

object recognizer 222 and/or the geometric analyzer 224 may identify the map points 308, 310, 312 using the methods described above and/or geometric analysis methods. Data relating to the map points 308, 310, 312 in the image 300 of the hand 302 is stored in the RAM 226. In some embodiments, fewer or more map points may be identified. For example, one, two, eight, seventeen, or more points may be identified. In some embodiments, one or more images of the hand may be used to determine a 3D representation of the hand 302.

5

10

15

20

25

30

**[0092]** Figure 23 depicts surface normals 314, 316, 318 at the map points 308, 310, 312 identified in the image 300 of the hand 302. The surface normal 314 at the hand centroid map point 308 is on a caudal surface of the hand 302. The surface normals 316, 318 at the thumb tip and first finger tip map points 310, 312 are on a surface defined by the thumb 304 and finger 306.

[0093] In one embodiment, the geometric analyzer 224 in the gesture recognition subsystem 220 identifies the surface normals 314, 316, 318 at the map points 308, 310, 312 in the image. The geometric analyzer 224 receives the image and map point data from the front facing FOV camera 208 and/or the RAM 226. The geometric analyzer 224 may identify the surface normals 314, 316, 318 using geometric analysis methods. Data relating to the surface normals 314, 316, 318 in the image 300 of the hand 302 is stored in the RAM 226.

**[0094]** Because the surface normal at the hand centroid map point 308 is substantially orthogonal (e.g., within 10 degrees of perpendicular) to the respective surface normals 316, 318 at the thumb tip and first finger tip map points 310, 312, the gesture recognition subsystem 220 recognizes the gesture as an "C"-shape. In response to recognizing the "C"-shape gesture, the augmented reality system 200 displays a cursor 320 at a location (e.g., midway) adjacent a line 324 between the thumb tip and the first finger tip map points 310, 312, as shown in Figure 24. The augmented reality system 200 may "stick" the cursor 320 to the location (e.g., midway) between the thumb tip and the first finger tip map points 310, 312 to maintain the spatial relationship there between even when the hand 302 moves.

**[0095]** In one embodiment, the geometric analyzer 224 in the gesture recognition subsystem 220 analyzes the surface normals 314, 316, 318 at the map points 308, 310, 312 in the image to determine whether they have the substantially orthogonal relationship described above. The geometric analyzer 224 receives the image and map point data from the front facing FOV camera 208 and/or the RAM 226. The

geometric analyzer 224 may determine whether the surface normals 314, 316, 318 have the substantially orthogonal relationship using geometric analysis methods. Data relating geometric relationship between the surface normals 314, 316, 318 in the image 300 of the hand 302 and the identified "C"-gesture is stored in the RAM 226.

5

10

15

20

25

30

[0096] Figures 25-26 depict a click command executed by a hand 302 displaying the "C"-gesture. Figure 25 shows that the tip of the user's thumb 304 and the tip of the user's first finger 306 have moved so that their positions are substantially similar. In one embodiment, the gesture recognition subsystem 220 may recognize that the positions are substantially similar when the first finger tip map point 312 is within a certain radius 326 of the thumb tip map point 310, as shown in Figure 26. Because the respective positions of the tip of the user's thumb 304 and the tip of the user's first finger 306 are substantially similar, the gesture recognition subsystem 220 identifies a click command executed by the hand 302 displaying the "C"-gesture. In response to identification of the click command, the augmented reality system 200 executes a command (e.g., select, activate, close, etc.) at the location of the cursor 320.

[0097] In one embodiment, the object recognizer 222 and/or the geometric analyzer 224 in the gesture recognition subsystem 220 identifies the movement of the user's thumb 304 and/or first finger 306 and analyzes the map point 312, 314 positions corresponding to their respective tips, all of which correspond to identification of the click command. The object recognizer 222 and/or the geometric analyzer 224 receives the image data from the front facing FOV camera 208 and/or the RAM 226. The object recognizer 222 and/or the geometric analyzer 224 may identify and analyze using the methods described above and/or geometric analysis methods. Data relating to identification of the click command is stored in the RAM 226.

**[0098]** While the gesture recognition embodiments described herein include analyzing images from front facing FOV cameras 208, the FOV of the cameras 208 may be larger than the FOV of the user (i.e., of the user's eyes). Accordingly, the "L"-shaped and/or "C"-shaped gestures described above may be used tactilely outside of the user's FOV, but within the FOV of the front facing FOV cameras 208. In one such embodiment, the "L"-shaped and/or "C"-shaped gestures may be used on a surface (e.g., desktop, lap, etc.) to perform the functions of a virtual mouse.

#### System Architecture Overview

**[0099]** Figure 27 is a block diagram of an illustrative computing system 1200 suitable for implementing an embodiment of the present invention. Computer system 1200 includes a bus 1206 or other communication mechanism for communicating information, which interconnects subsystems and devices, such as processor 1207, system memory 1208 (e.g., RAM), static storage device 1209 (e.g., ROM), disk drive 1210 (e.g., magnetic or optical), communication interface 1214 (e.g., modem or Ethernet card), display 1211 (e.g., CRT or LCD), input device 1212 (e.g., keyboard), and cursor control.

5

20

25

30

[00100] According to one embodiment of the invention, computer system 1200 performs specific operations by processor 1207 executing one or more sequences of one or more instructions contained in system memory 1208. Such instructions may be read into system memory 1208 from another computer readable/usable medium, such as static storage device 1209 or disk drive 1210. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions to implement the invention. Thus, embodiments of the invention are not limited to any specific combination of hardware circuitry and/or software. In one embodiment, the term "logic" shall mean any combination of software or hardware that is used to implement all or part of the invention.

**[00101]** The term "computer readable medium" or "computer usable medium" as used herein refers to any medium that participates in providing instructions to processor 1207 for execution. Such a medium may take many forms, including but not limited to, non- volatile media and volatile media. Non-volatile media includes, for example, optical or magnetic disks, such as disk drive 1210. Volatile media includes dynamic memory, such as system memory 1208.

**[00102]** Common forms of computer readable media includes, for example, floppy disk, flexible disk, hard disk, magnetic tape, any other magnetic medium, CD-ROM, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, RAM, PROM, EPROM, FLASH-EPROM (e.g., NAND flash, NOR flash), any other memory chip or cartridge, or any other medium from which a computer can read.

**[00103]** In an embodiment of the invention, execution of the sequences of instructions to practice the invention is performed by a single computer system 1200. According to other embodiments of the invention, two or more computer systems

1200 coupled by communication link 1215 (e.g., LAN, PTSN, or wireless network) may perform the sequence of instructions required to practice the invention in coordination with one another.

**[00104]** Computer system 1200 may transmit and receive messages, data, and instructions, including program, i.e., application code, through communication link 1215 and communication interface 1214. Received program code may be executed by processor 1207 as it is received, and/or stored in disk drive 1210, or other non-volatile storage for later execution. Database 1232 in storage medium 1231 may be used to store data accessible by system 1200 via data interface 1233.

5

10

15

20

25

30

**[00105]** The invention includes methods that may be performed using the subject devices. The methods may comprise the act of providing such a suitable device. Such provision may be performed by the user. In other words, the "providing" act merely requires the user obtain, access, approach, position, set-up, activate, power-up or otherwise act to provide the requisite device in the subject method. Methods recited herein may be carried out in any order of the recited events which is logically possible, as well as in the recited order of events.

**[00106]** Exemplary aspects of the invention, together with details regarding material selection and manufacture have been set forth above. As for other details of the present invention, these may be appreciated in connection with the above-referenced patents and publications as well as generally known or appreciated by those with skill in the art. The same may hold true with respect to method-based aspects of the invention in terms of additional acts as commonly or logically employed.

**[00107]** In addition, though the invention has been described in reference to several examples optionally incorporating various features, the invention is not to be limited to that which is described or indicated as contemplated with respect to each variation of the invention. Various changes may be made to the invention described and equivalents (whether recited herein or not included for the sake of some brevity) may be substituted without departing from the true spirit and scope of the invention. In addition, where a range of values is provided, it is understood that every intervening value, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. **[00108]** Also, it is contemplated that any optional feature of the inventive variations described may be set forth and claimed independently, or in combination with any

one or more of the features described herein. Reference to a singular item, includes the possibility that there are plural of the same items present. More specifically, as used herein and in claims associated hereto, the singular forms "a," "an," "said," and "the" include plural referents unless the specifically stated otherwise. In other words, use of the articles allow for "at least one" of the subject item in the description above as well as claims associated with this disclosure. It is further noted that such claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

5

10

15

20

25

**[00109]** Without the use of such exclusive terminology, the term "comprising" in claims associated with this disclosure shall allow for the inclusion of any additional element--irrespective of whether a given number of elements are enumerated in such claims, or the addition of a feature could be regarded as transforming the nature of an element set forth in such claims. Except as specifically defined herein, all technical and scientific terms used herein are to be given as broad a commonly understood meaning as possible while maintaining claim validity.

**[00110]** The breadth of the present invention is not to be limited to the examples provided and/or the subject specification, but rather only by the scope of claim language associated with this disclosure.

**[00111]** In the foregoing specification, the invention has been described with reference to specific embodiments thereof. It will, however, be evident that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention. For example, the above-described process flows are described with reference to a particular ordering of process actions. However, the ordering of many of the described process actions may be changed without affecting the scope or operation of the invention. The specification and drawings are, accordingly, to be regarded in an illustrative rather than restrictive sense.

#### Claims

A method of executing a gesture command, comprising:
 identifying a hand centroid of a hand;
 identifying a first finger tip of a first finger on the hand;
 identifying a thumb tip of a thumb on the hand; and
 determining a surface normal relationship between the hand centroid, the first

5

15

25

finger tip, and the thumb tip.

- 2. The method of claim 1, wherein the surface normal relationship comprises respective surface normals at the hand centroid on a caudal surface of the hand, at the first finger tip, and at the thumb tip are substantially parallel.
  - 3. The method of claim 2, wherein a virtual cursor is displayed in a field of view at a location corresponding to the first finger tip.
    - 4. The method of claim 3, further comprising,

detecting a movement of the thumb tip to a mapped planar position between the hand centroid and the first finger tip, and

executing a command at the cursor location in response to detecting the 20 movement.

- 5. The method of claim 1, wherein the surface normal relationship comprises a surface normal at the hand centroid on a caudal surface of the hand is substantially orthogonal to the respective surface normals at the first finger tip and at the thumb tip.
- 6. The method of claim 5, wherein a virtual cursor is displayed in a field of view at a location between the first finger tip and the thumb tip.
- The method of claim 6, further comprising, detecting that a first finger tip position of the first finger tip and a thumb tip position of the thumb tip are substantially similar, and

executing a command at the cursor location in response to detecting of the substantially similar first finger tip and thumb tip positions.

- 8. The method of claim 1, wherein the hand centroid, the first finger tip, and the thumb tip are outside of a field of view of a user.
  - 9. The method of claim 8, wherein the hand centroid, the first finger tip, and the thumb tip are inside of a field of view of a camera worn by the user.
- 10. A method of executing a gesture command, comprising: identifying a hand centroid of a hand; identifying a first finger tip of a first finger on the hand; identifying a thumb tip of a thumb on the hand; calculating a first line between the first finger tip and the hand centroid; calculating a second line between the thumb tip and the hand centroid; and determining an angular relationship between the first and second lines.
  - 11. The method of claim 10, wherein the angular relationship is that the first and second lines are substantially orthogonal.
  - 12. The method of claim 10, wherein a virtual cursor is displayed in a field of view at a location corresponding to the first finger tip.
    - 13. The method of claim 12, further comprising,

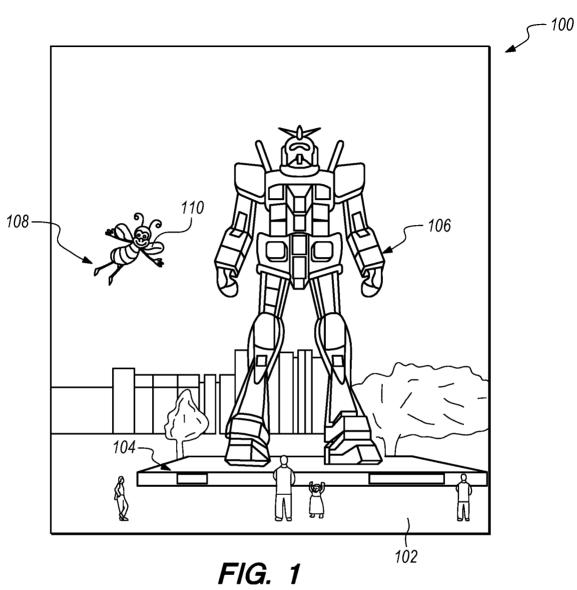
20

30

detecting a change in the angular relationship between the first and second lines, and

executing a command at the cursor location in response to detecting the change in the angular relationship.

- 14. The method of claim 13, wherein the change in the angular relationship is a threshold amount of decrease in an angle between first and second lines.
  - 15. The method of claim 14, wherein the threshold amount of decrease is about 30 degrees.



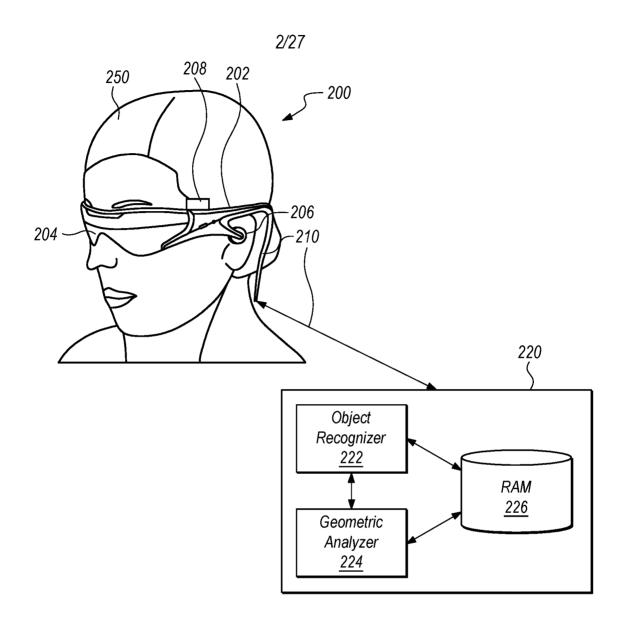


FIG. 2

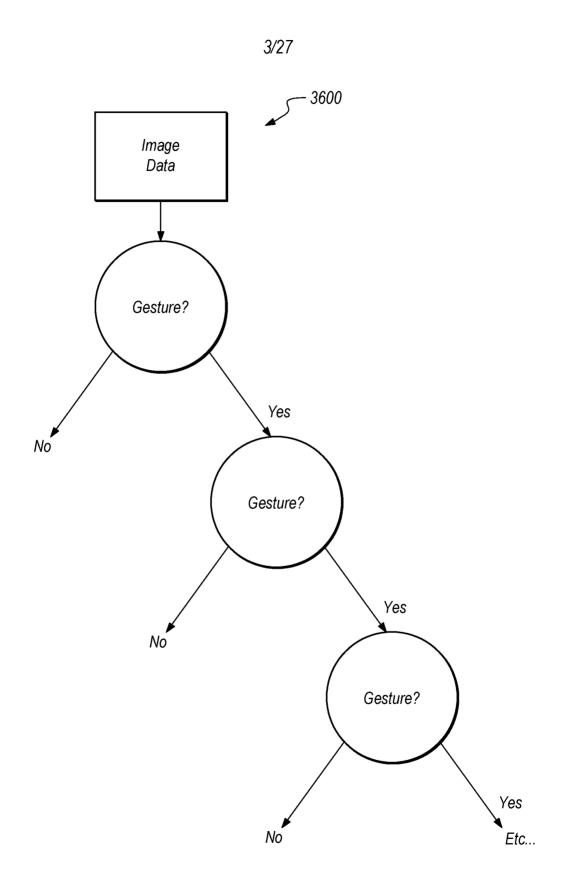


FIG. 3

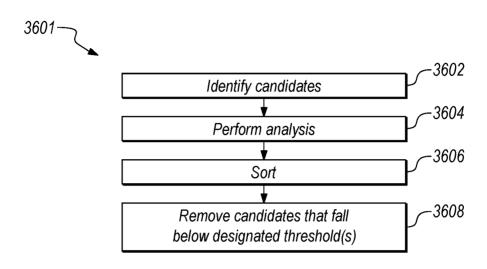


FIG. 4

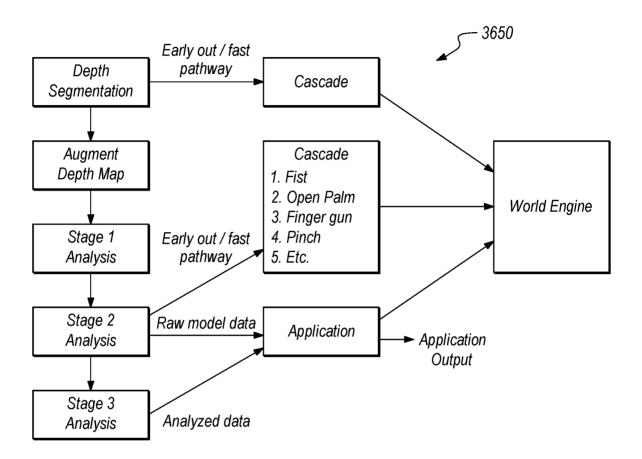


FIG. 5

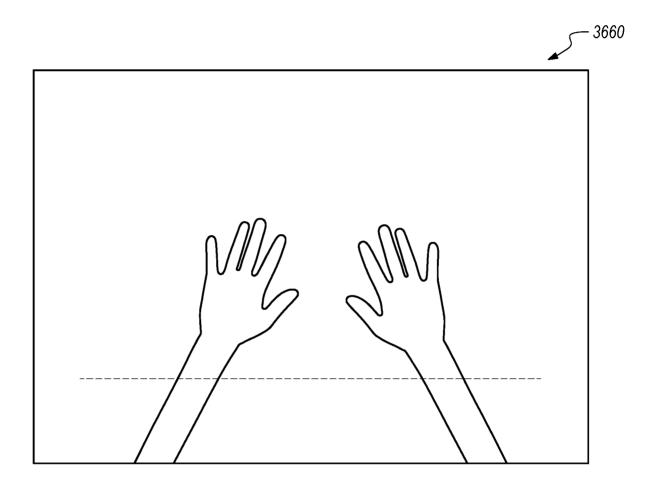


FIG. 6

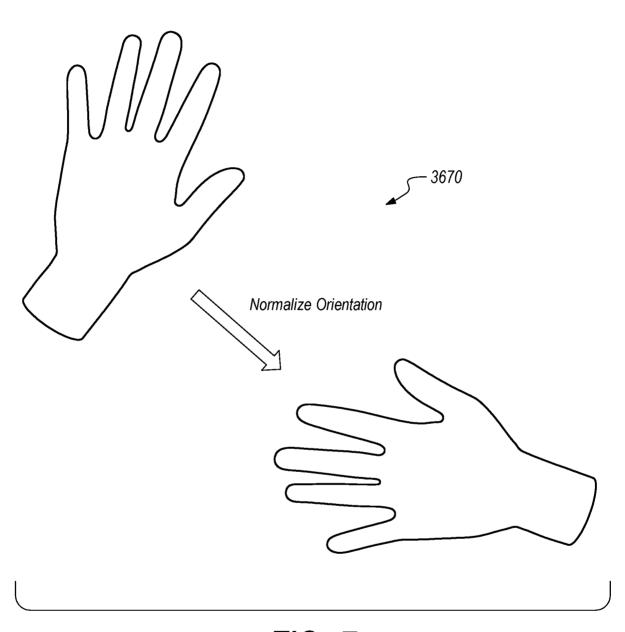


FIG. 7

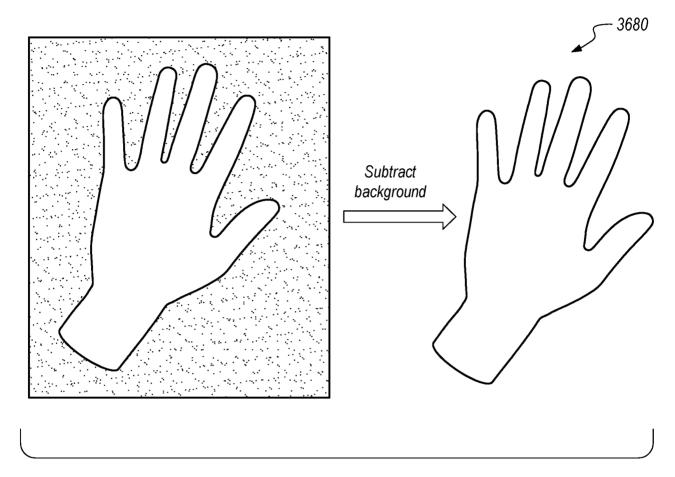


FIG. 8

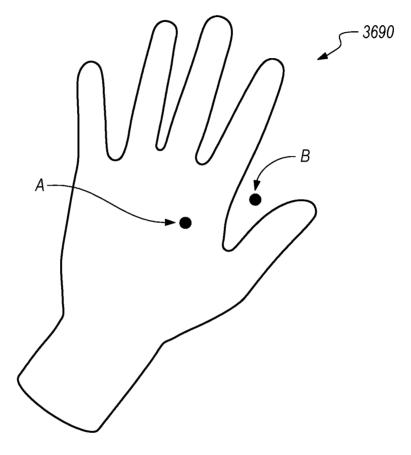
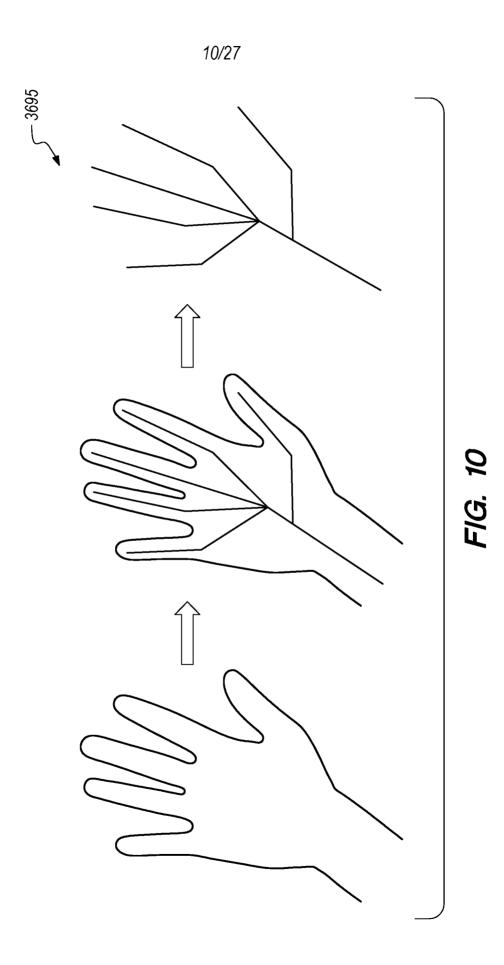


FIG. 9



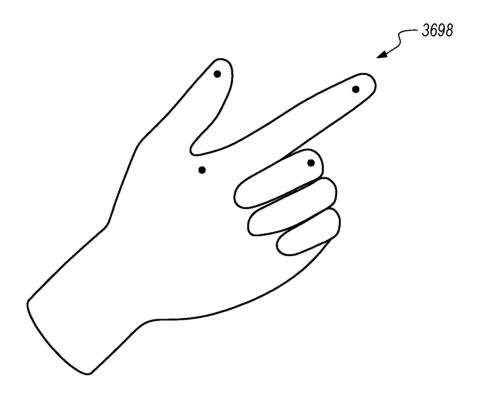


FIG. 11

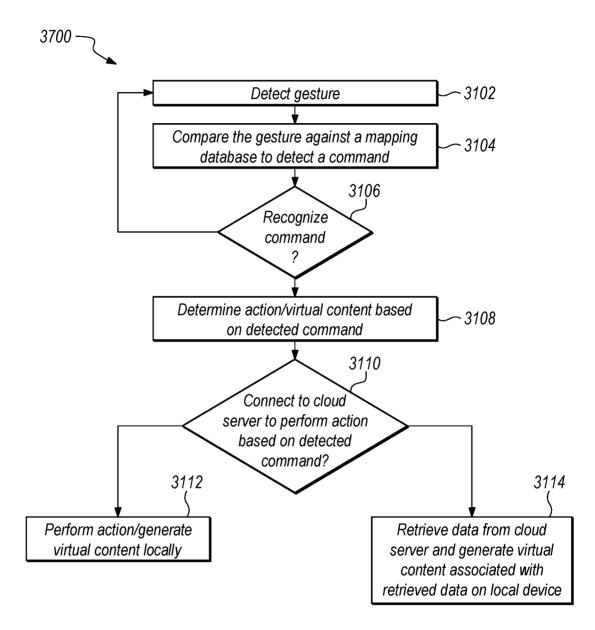
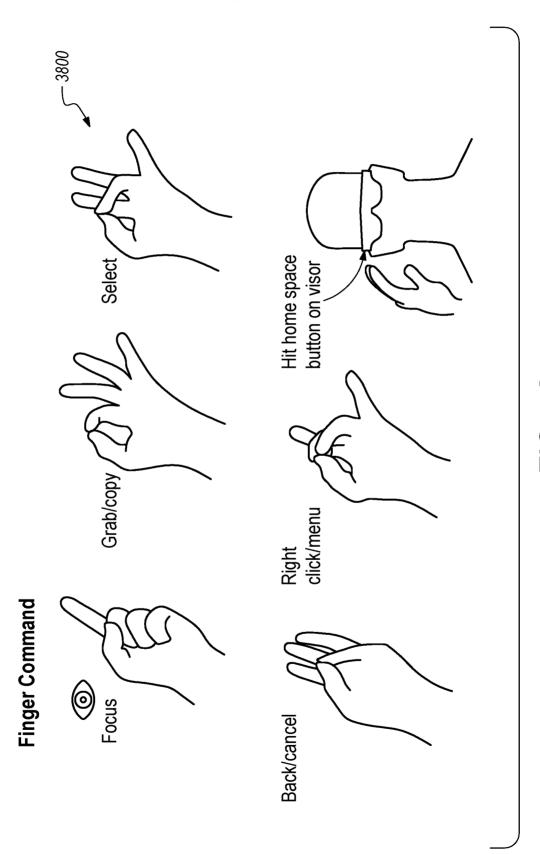


FIG. 12



F/G. 13

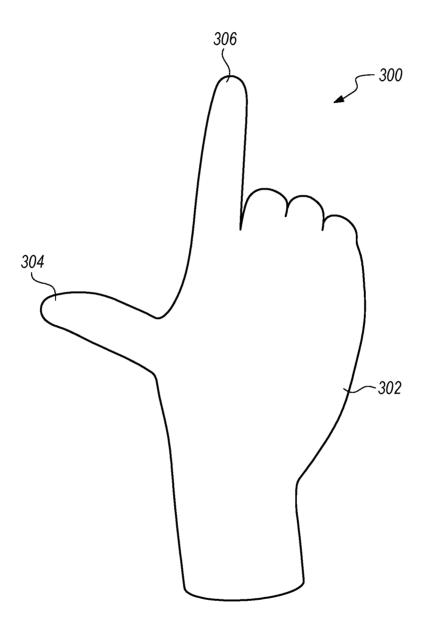


FIG. 14

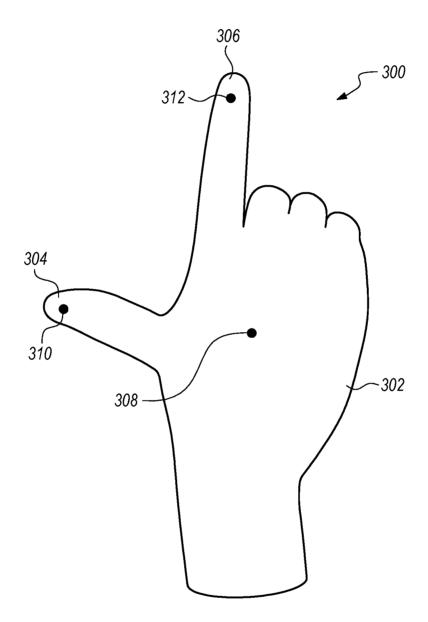


FIG. 15

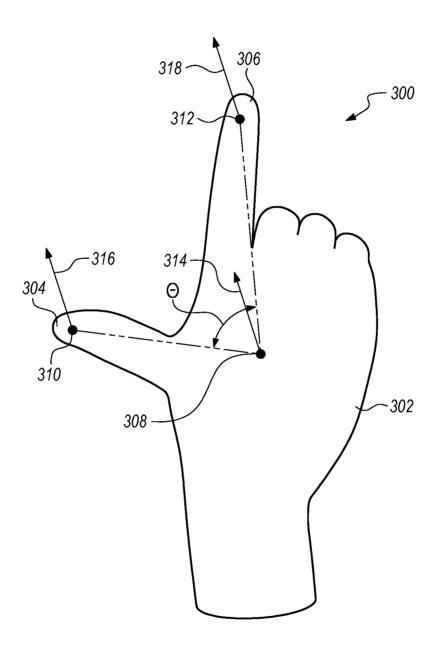


FIG. 16

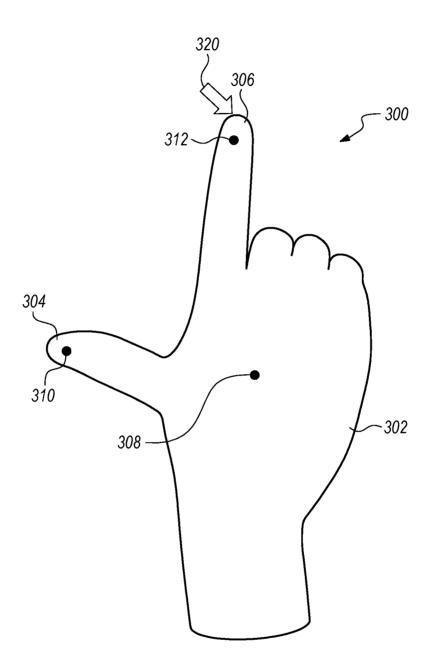


FIG. 17

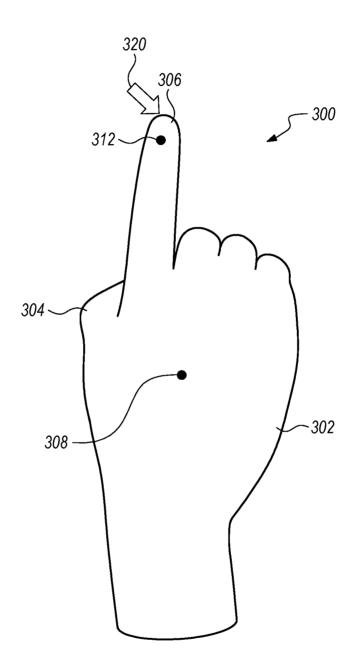


FIG. 18

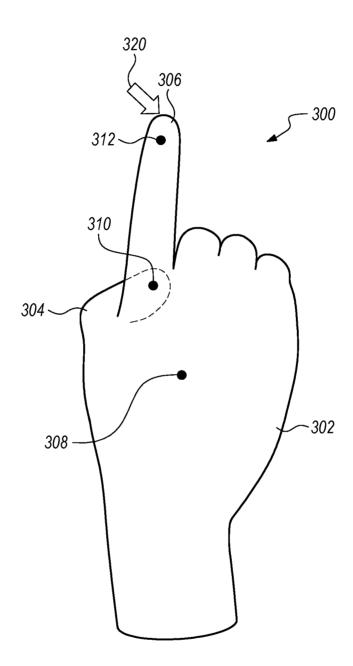


FIG. 19

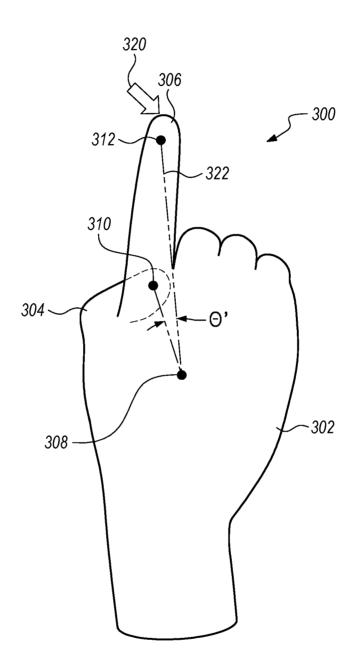


FIG. 20

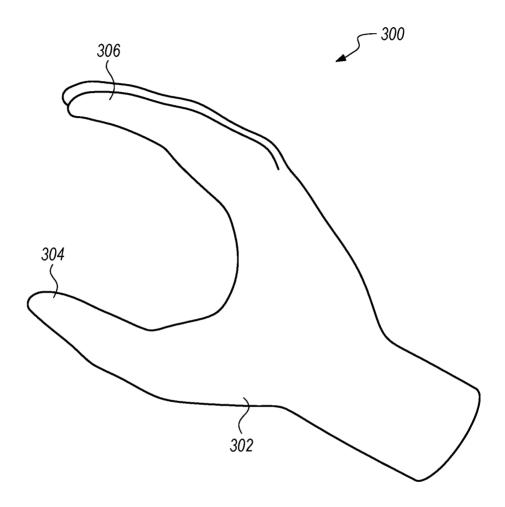


FIG. 21

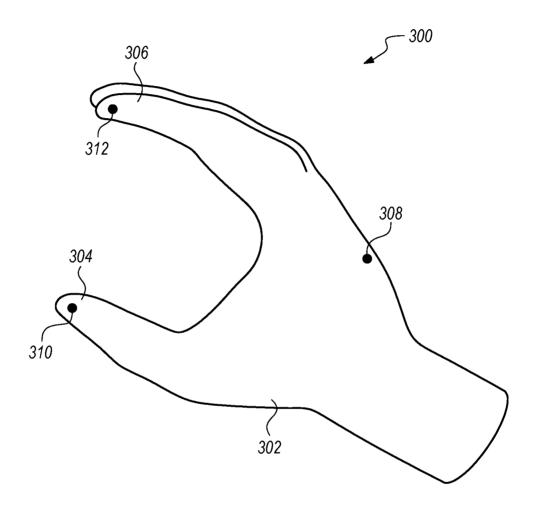


FIG. 22

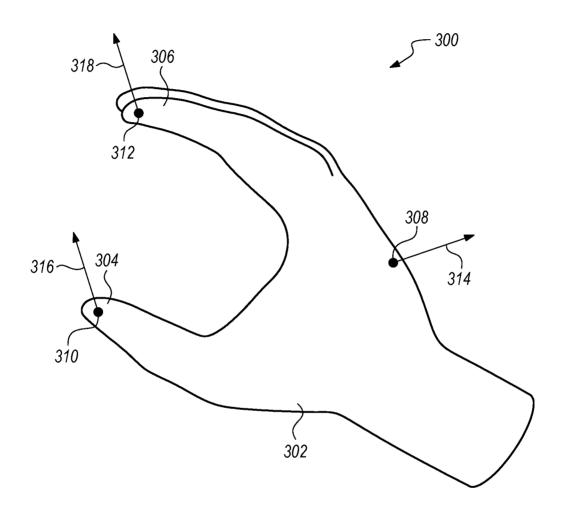


FIG. 23

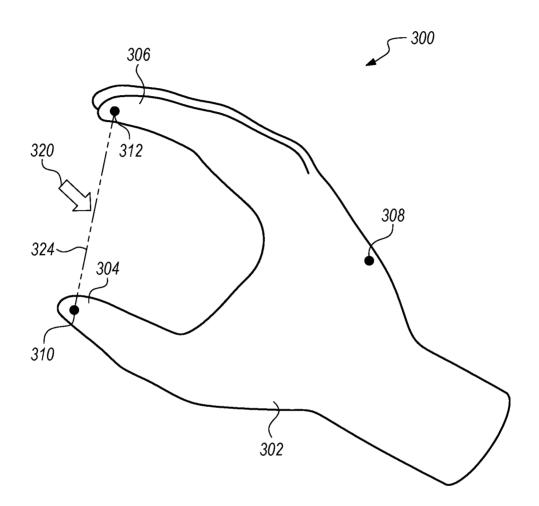


FIG. 24

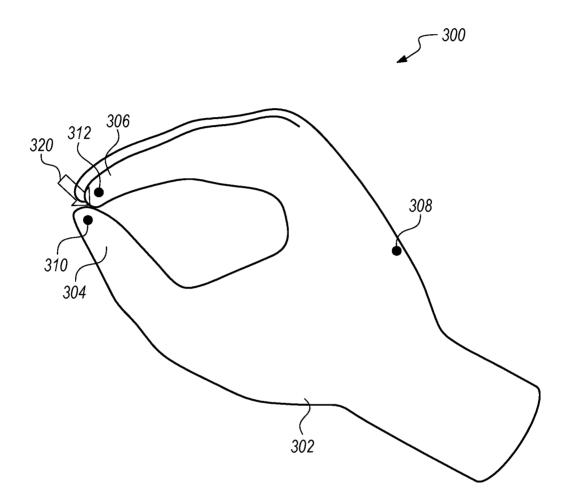


FIG. 25

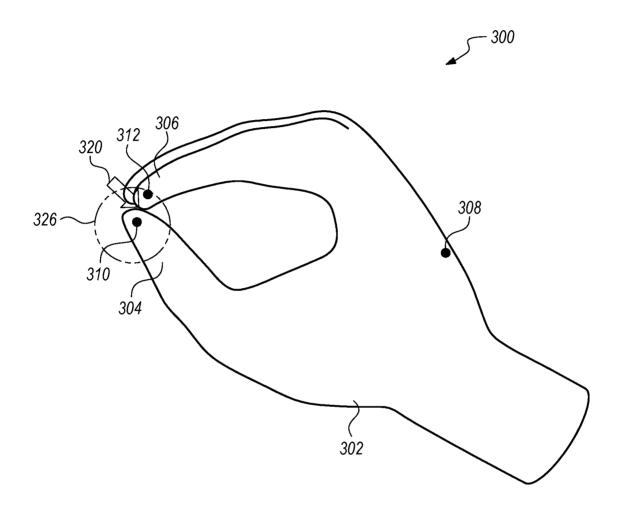


FIG. 26

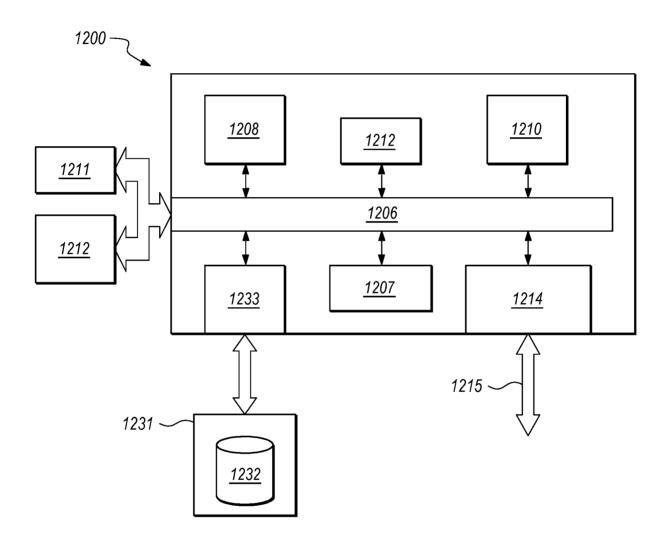


FIG. 27

## INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 19/21872

A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - G06F 3/01, G06K 9/00 (2019.01)  CPC - G06F 3/011, G06F 3/0304, G06F 3/042, G06F 3/04104, G06F 2203/04108	06K 9/00214, G06K 9/00389, G06K 9/4671, G06F	
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)		
See Search History Document		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched See Search History Document		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) See Search History Document		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category* Citation of document, with indication, where a	appropriate, of the relevant passages Relevant to claim No.	
X US 2015/0234477 A1 (MAGIC LEAP, INC.) 20 Augus especially abstract and para [0449]-[0461], [0466], [0466]	st 2015 (20.08.2015), entire document, 4-9 468], [0482]-[0484].	
A US 9,207,773 B1 (AQUIFI, INC.) 08 December 2015	(09.12.2015) entire degument	
7 03 9,207,773 B1 (AQOIFI, INC.) 08 December 2013	(08.12.2015), entire document.	
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.	"T" later document published after the international filing date or priority	
to be of particular relevance	the principle or theory underlying the invention	
filing date	considered novel or cannot be considered to involve an inventive	
"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)	"Y" document of particular relevance; the claimed invention cannot be	
"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	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	
08 July 2019 (08.07.2019)	17 JUL 2019	
Name and mailing address of the ISA/US	Authorized officer:	
Aail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450	Lee W. Young	
Facsimile No. 571-273-8300	PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774	

## INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 19/21872

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)
This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2. Claims Nos.:  because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1.
Group I: Claims 1-9, directed to executing a gesture command, comprising determining a surface normal relationship between a hand centroid, a first finger tip, and a thumb tip.
Group II: Claims 10-15, directed to executing a gesture command, comprising calculating a first line between a first finger tip and a hand centroid, calculating a second line between the thumb tip and the hand centroid, and determining an angular relationship between the first and second lines.
(See extra sheet)
As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  1-9
Remark on Protest  The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.  The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.  No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No. PCT/US 19/21872

In continuation of Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet):

The groups of inventions listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of the Group I invention is determining a surface normal relationship between the hand centroid, the first finger tip, and the thumb tip, not required by Group II.

The special technical feature of the Group II invention is calculating a first line between the first finger tip and the hand centroid; calculating a second line between the thumb tip and the hand centroid; and determining an angular relationship between the first and second lines, not required by Group I.

Groups I-II share the technical features of, in various combination, identifying a hand centroid of a hand; identifying a first finger tip of a first finger on the hand; identifying a thumb tip of a thumb on the hand.

However, these shared technical features fail to represent a contribution over the prior art of US 9,207,773 B1 to Aquifi, Inc. (hereinafter 'Aquifi'), which discloses identifying a hand centroid of a hand (col. 5, ln 57-60, Figs. 1E-1I - "The perhaps one hundred potential landmarks, or more typically the dozen or so exemplary landmark points defined by the present invention can be the fingertips, centroids of the user's hand(s)"); identifying a first finger tip of a first finger on the hand (col. 5, ln 57-60, col. 16, ln 54-67, Figs. 1E-1I); identifying a thumb tip of a thumb on the hand (col. 16, ln 54 - col. 17, ln 13, Figs. 1E-1I - "Dashed lines are depicted in FIG. 1D-1 between the centroid O of user object 60, here a right hand, and the landmarks of the tip of the thumb, N, and of the fingertips"; "user-desired icon may be caused to enlarge in size or to be highlighted, or the like, to indicate selection, as a function of system 10 recognizing the user gesture. In other applications, the user's hand gestures may command movement of cursor 40' so as to select an icon on menu ribbon").
Thus, the inventions listed as Groups I-II lack unity of invention because they do not share a same or corresponding special technical feature providing a contribution over the prior art.