



## (51) International Patent Classification:

G06T 13/40 (2011.01)

## (21) International Application Number:

PCT/CN2014/073695

## (22) International Filing Date:

19 March 2014 (19.03.2014)

## (25) Filing Language:

English

## (26) Publication Language:

English

(71) Applicant: INTEL CORPORATION [US/US]; 2200 Mission College Blvd., Santa Clara, California 95054 (US).

## (72) Inventors; and

(71) Applicants (for US only): DU, Yangzhou [CN/CN]; Room 502, Building 2-21-7, XinGangZhuangYuan, LiJi-aQiao, Shunyi District, Beijing 101304 (CN). KIM, Tae-Hoon [KR/KR]; 206-202 Raemian Greaten Apt., Yeoksam 2-Dong, Gangnam-Gu, Seoul, 135756 (KR). LI, Wenlong [CN/CN]; 4-502 Building 18, Tiantongyuan West 2 District, Beijing 102218 (CN). LI, Qiang [CN/CN]; Room 1112, Building 2, Jinxingyuan, Taiyangxingcheng, Taiyanggong, Chaoyang District, Beijing 100028 (CN). TONG, Xiaofeng [CN/CN]; R4-4-602, QinChunJiaYuan Community, XiSanQi Street, Haidian District, Beijing 100096 (CN). WANG, Tao [CN/CN]; ChunYin Park 8-6-5F, Center Town, Lan DianChang Middle Road, Beijing 100087 (CN). PARK, Minje [KR/KR]; 112-1302 Humansia Apt. Dochoon-dong Jungwon-gu, Seongnam 462-751 (KR). DUCHENNE, Olivier [FR/KR]; Prime Villate #501, 5 Seorae-ro, Seocho-gu, Seoul, 137-806 (KR). ZHANG, Yimin [CN/CN]; 11-1507 Taiyangyuan, Haidian District, Beijing 100086 (CN). CHEON, Yeongjae [KR/KR]; 705, SM-Ellui A.P.T., 791-40, Siheung-3dong, Geumcheon-gu, Seoul (KR). JUN, Bongjin [KR/KR]; 11F, 708-33, Venture Bldg, Yeoksam 2-dong, Gangnam-gu, Seoul, 135-919 (KR). RYU, Wooju [KR/KR];

301-403, 17, Yonggu-daero 2801beon-gil, Suji-gu, Yongin-si, Gyeonggi-do (KR). SACHSON, Thomas [US/US]; 13876 Alameda de las Pulgas, Menlo Park, California 94025 (US). SMILEY, Mary D. [US/US]; 3065 Bowers Avenue, Santa Clara, California 95054 (US).

(74) Agent: CHINA PATENT AGENT (H.K.) LTD.; 22/F., Great Eagle Center, 23 Harbour Road, Wanchai, Hong Kong (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, BG, 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).

## Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

## Published:

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

## (54) Title: FACIAL EXPRESSION AND/OR INTERACTION DRIVEN AVATAR APPARATUS AND METHOD

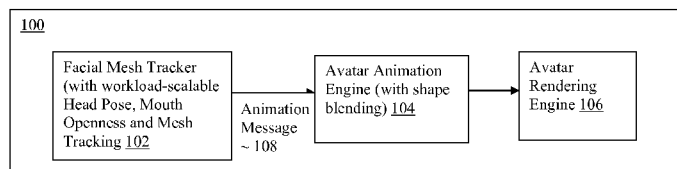


Figure 1

(57) Abstract: Apparatuses, methods and storage medium associated with animating and rendering an avatar are disclosed herein. In embodiments, an apparatus may include a facial mesh tracker to receive a plurality of image frames, detect facial action movements of a face and head pose gestures of a head within the plurality of image frames, and output a plurality of facial motion parameters and head pose parameters that depict facial action movements and head pose gestures detected, all in real time, for animation and rendering of an avatar. The facial action movements and head pose gestures may be detected through inter-frame differences for a mouth and an eye, or the head, based on pixel sampling of the image frames. The facial action movements may include opening or closing of a mouth, and blinking of an eye. The head pose gestures may include head rotation such as pitch, yaw, roll, and head movement along horizontal and vertical direction, and the head comes closer or goes farther from the camera. Other embodiments may be described and/or claimed.



## **FACIAL EXPRESSION AND/OR INTERACTION DRIVEN AVATAR APPARATUS AND METHOD**

### **Technical Field**

5           The present disclosure relates to the field of data processing. More particularly, the present disclosure relates to facial expression and/or interaction driven animation and rendering of avatar.

### **Background**

10           The background description provided herein is for the purpose of generally presenting the context of the disclosure. Unless otherwise indicated herein, the materials described in this section are not prior art to the claims in this application and are not admitted to be prior art by inclusion in this section.

          As user's graphic representation, avatar has been quite popular in virtual world. However, most existing avatar systems are static, and few of them are driven by text, script or voice. Some other avatar systems use graphics interchange format (GIF)  
15           animation, which is a set of predefined static avatar image playing in sequence. In recent years, with the advancement of computer vision, camera, image processing, etc., some avatar may be driven by facial performance. However, existing systems tend to be computation intensive, requiring high-performance general and graphics processor, and do  
20           not work well on mobile devices, such as smartphones or computing tablets.

### **Brief Description of the Drawings**

          Embodiments will be readily understood by the following detailed description in conjunction with the accompanying drawings. To facilitate this description, like reference numerals designate like structural elements. Embodiments are illustrated by way of  
25           example, and not by way of limitation, in the figures of the accompanying drawings.

          Figure 1 illustrates a block diagram of a pocket avatar system, according to the disclosed embodiments.

          Figure 2 illustrates a block diagram for the facial mesh tracker of Figure 1 in further detail, according to the disclosed embodiments.

30           Figures 3 and 4 illustrate interaction driven avatar, according to the disclosed embodiments.

          Figure 5 is a flow diagram illustrating a process for generating facial expression and interaction animation messages, according to the disclosed embodiments.

Figure 6 is a flow diagram illustrating a process for interleaving facial expression and interaction animations, according to the disclosed embodiments.

Figure 7 is a flow diagram illustrating a process for estimating head pose, according to the disclosed embodiments.

5        Figure 8 illustrates an example computer system suitable for use to practice various aspects of the present disclosure, according to the disclosed embodiments.

Figure 9 illustrates a storage medium having instructions for practicing methods described with references to Figures 2-7, according to disclosed embodiments.

### **Detailed Description**

10        Apparatuses, methods and storage medium associated with animating and rendering an avatar are disclosed herein. In embodiments, an apparatus may include a facial mesh tracker to receive a plurality of image frames, detect, through the plurality of image frames, facial action movements of a face of a user, and head pose gestures of a head of the user, and output a plurality of facial motion parameters that depict facial action  
15        movements detected, and a plurality of head pose gestures parameters that depict head pose gestures detected, all in real time, for animation and rendering of an avatar. The facial action movements and the head pose gestures may be detected through inter-frame differences for a mouth and an eye of the face, and the head, based on pixel sampling of the image frames.

20        In embodiments, the facial action movements may include opening or closing of a mouth, and blinking of an eye, and the plurality of facial motion parameters may include parameters that depict the opening or closing of the mouth and blinking of the eye. The head pose gestures may include pitch, yaw, roll of a head, horizontal and vertical movement of a head, and distance change of a head (becoming closer or farther to the  
25        camera capturing the image frames), and the plurality of head pose parameters may include parameters that depict the pitch, yaw, roll, horizontal /vertical movement, and distance change of the head.

In embodiments, the apparatus may further include an avatar animation engine coupled with the facial mesh tracker to receive the plurality of facial motion parameters  
30        outputted by the facial mesh tracker, and drive an avatar model to animate the avatar, replicating a facial expression of the user on the avatar, through blending of a plurality of pre-defined shapes. Further, the apparatus may include an avatar rendering engine, coupled with the avatar animation engine, to draw the avatar as animated by avatar animation engine.

In the following detailed description, reference is made to the accompanying drawings which form a part hereof wherein like numerals designate like parts throughout, and in which is shown by way of illustration embodiments that may be practiced. It is to be understood that other embodiments may be utilized and structural or logical changes may be made without departing from the scope of the present disclosure. Therefore, the following detailed description is not to be taken in a limiting sense, and the scope of embodiments is defined by the appended claims and their equivalents.

Aspects of the disclosure are disclosed in the accompanying description. Alternate embodiments of the present disclosure and their equivalents may be devised without parting from the spirit or scope of the present disclosure. It should be noted that like elements disclosed below are indicated by like reference numbers in the drawings.

Various operations may be described as multiple discrete actions or operations in turn, in a manner that is most helpful in understanding the claimed subject matter. However, the order of description should not be construed as to imply that these operations are necessarily order dependent. In particular, these operations may not be performed in the order of presentation. Operations described may be performed in a different order than the described embodiment. Various additional operations may be performed and/or described operations may be omitted in additional embodiments.

For the purposes of the present disclosure, the phrase “A and/or B” means (A), (B), or (A and B). For the purposes of the present disclosure, the phrase “A, B, and/or C” means (A), (B), (C), (A and B), (A and C), (B and C), or (A, B and C).

The description may use the phrases “in an embodiment,” or “in embodiments,” which may each refer to one or more of the same or different embodiments. Furthermore, the terms “comprising,” “including,” “having,” and the like, as used with respect to embodiments of the present disclosure, are synonymous.

As used herein, the term “module” may refer to, be part of, or include an Application Specific Integrated Circuit (ASIC), an electronic circuit, a processor (shared, dedicated, or group) and/or memory (shared, dedicated, or group) that execute one or more software or firmware programs, a combinational logic circuit, and/or other suitable components that provide the described functionality.

Referring now to Figure 1, wherein a pocket avatar system, according to the disclosed embodiments, is shown. As illustrated, pocket avatar system 100 may include facial mesh tracker 102, avatar animation engine 104, and avatar rendering engine 106, coupled with each other as shown. Facial mesh tracker 102 may be configured to receive

a plurality of image frames, e.g., from an image source, such as a camera (not shown), detect facial action movements of a face of a user and/or head pose gestures of a head of the user, within the plurality of image frames, and output a plurality of facial motion parameters that depict facial action movements detected, e.g., eye and/or mouth movements, and head pose gestures parameters that depict head pose gestures detected, such as head rotation, movement, and/or coming closer or farther from the camera, all in real time. Avatar animation engine 104 may be configured to receive the plurality of facial motion parameters outputted by the facial mesh tracker 102, and drive an avatar model to animate the avatar, replicating a facial expression and/or head movement of the user on the avatar. Avatar rendering engine 106 may be configured to draw the avatar as animated by avatar animation engine 104.

In embodiments, facial mesh tracker 102 may include at least head pose, mouth openness, and mesh tracking function blocks that are sufficiently accurate, yet scalable in their processing power required, making pocket avatar system 100 suitable to be hosted by a wide range of mobile computing devices, such as smartphones and/or computing tablets. Additionally, in embodiments, avatar animation engine 104 may replicate a facial expression of the user on the avatar, through blending of a plurality of pre-defined shapes, further making pocket avatar system 100 suitable to be hosted by a wide range of mobile computing devices.

In embodiments, facial mesh tracker 102 may be configured to generate and output animation messages 108 having the facial motion parameters that depict facial action movements detected and head pose gesture parameters that depict head pose gestures, for avatar animation engine 104. In embodiments, facial mesh tracker 102 and avatar animation engine 104 may be further configured to cooperate to support user interaction driven avatar animation, where a canned expression, e.g., sticking a tongue out, corresponding to a user interaction, e.g., a swipe gesture, may be animated, in lieu of detected facial expression and/or head pose. Similarly, facial mesh tracker 102 may be configured to detect, generate and output animation messages 108 having information about the user interaction, e.g., a start period, a keep period, and an end period, and/or the corresponding canned expression.

In embodiments, facial mesh tracker 102 may be configured to generate a normalized head pose of the user by using a 3D facial action model and a 3D neutral facial shape of the user pre-constructed using a 3D facial shape model. Both, the 3D facial action model and the 3D facial shape model may be pre-constructed through machine

learning of a 3D facial database.

While pocket avatar system 100 is designed to be particularly suitable to be operated on a mobile device, such as a smartphone, a phablet, a computing tablet, a laptop computer, or an e-reader, the disclosure is not to be so limited. It is anticipated that pocket avatar system 100 may also be operated on computing devices with more computing power than the typical mobile devices, such as a desktop computer, a game console, a set-top box, or a computer server. The foregoing and other aspects of pocket avatar system 100 will be described in further detail in turn below.

Figure 2 illustrates a block diagram for the facial mesh tracker of Figure 1 in further detail, according to the disclosed embodiments. As illustrated, in embodiments, facial mesh tracker 102 may include face detection function block 202, landmark detection function block 204, initial face mesh fitting function block 206, facial expression estimation function block 208, head pose tracking function block 210, mouth openness estimation function block 212, facial mesh tracking function block 214, tracking validation function block 216, eye blink detection and mouth correction function block 218, facial mesh adaptation function block 220 and blend shape mapping function block 222, coupled with each other as shown. Function blocks 202-222 may be implemented in hardware, e.g., ASIC or programmable devices programmed with the appropriate logic, software to be executed by general and/or graphics processors, or a combination of both.

In embodiments, face detection function block 202 may be configured to detect the face through window scan of one or more of the plurality of image frames received. At each window position, modified census transform (MCT) features may be extracted, and a cascade classifier may be applied to look for the face. Landmark detection function block 204 may be configured to detect landmark points on the face, e.g., eye centers, nose-tip, mouth corners, and face contour points. Given a face rectangle, an initial landmark position may be given according to mean face shape. Thereafter, the exact landmark positions may be found iteratively through an explicit shape regression (ESR) method.

In embodiments, initial face mesh fitting function block 206 may be configured to initialize a 3D pose of a face mesh based at least in part on a plurality of landmark points detected on the face. A Candide3 wireframe head model may be used. The rotation angles, translation vector and scaling factor of the head model may be estimated using the POSIT algorithm. Resultantly, the projection of the 3D mesh on the image plane may match with the 2D landmarks. Facial expression estimation function block 208 may be configured to initialize a plurality of facial motion parameters based at least in part on a

plurality of landmark points detected on the face. The Candide3 head model may be controlled by facial action parameters (FAU), such as mouth width, mouth height, nose wrinkle, eye opening. These FAU parameters may be estimated through least square fitting.

5           Head pose tracking function block 210 may be configured to calculate rotation angles of the user's head, including pitch, yaw and/or roll, and translation distance along horizontal, vertical direction, and coming closer or going farther from the camera. The calculation may be based on a subset of sub-sampled pixels of the plurality of image frames, applying dynamic template matching and re-registration. Mouth openness  
10          estimation function block 212 may be configured to calculate opening distance of an upper lip and a lower lip of the mouth. The correlation of mouth geometry (opening/closing) and appearance may be trained using a sample database. Further, the mouth opening distance may be estimated based on a subset of sub-sampled pixels of a current image frame of the plurality of image frames, applying FERN regression.

15           Facial mesh tracking function block 214 may be configured to adjust position, orientation or deformation of a face mesh to maintain continuing coverage of the face and reflection of facial movement by the face mesh, based on a subset of sub-sampled pixels of the plurality of image frames. The adjustment may be performed through image alignment of successive image frames, subject to pre-defined FAU parameters in  
20          Candide3 model. The results of head pose tracking function block 210 and mouth openness may serve as soft-constraints to parameter optimization. Tracking validation function block 216 may be configured to monitor face mesh tracking status, to determine whether it is necessary to re-locate the face. Tracking validation function block 216 may apply one or more face region or eye region classifiers to make the determination. If the  
25          tracking is running smoothly, operation may continue with next frame tracking, otherwise, operation may return to face detection function block 202, to have the face re-located for the current frame.

            Eye blink detection and mouth correction function block 218 may be configured to detect eye blinking status and mouth shape. Eye blinking may be detected through optical  
30          flow analysis, whereas mouth shape/movement may be estimated through detection of inter-frame histogram differences for the mouth. As refinement of whole face mesh tracking, eye blink detection and mouth correction function block 216 may yield more accurate eye-blinking estimation, and enhance mouth movement sensitivity. Face mesh adaptation function block 220 may be configured to reconstruct a face mesh according to

derived facial action units, and re-sample of a current image frame under the face mesh to set up processing of a next image frame. Shape mapping function block 222 may be configured to convert facial action units into blend-shape coefficients for the animation of the avatar. Since face tracking may use different mesh geometry and animation structure with avatar rendering side, shape mapping function block 220 may also be configured to perform animation coefficient conversion and face model retargeting for avatar animation engine 104. In embodiments, shape mapping function block 220 may output a number of face tracking parameters as the blend shape weights, for avatar animation engine 104. These face tracking parameters may include, but are not limited to “lower lip down” (LLIPD), “both lips widen” (BLIPW), “both lips up” (BLIPU), “nose wrinkle” (NOSEW) and “eyebrow down” (BROWD).

Having head pose tracking function block 210 estimates head pose angle and mouth openness estimation function block 212 estimates mouth opening distance, their results may serve as a soft constraint to the numerical optimization performed by facial mesh tracking function block 214. The arrangement may provide more stable estimation of facial movement parameters, and potentially prevent drifting problems in visual tracking, resulting in less computation requirement, and more suitable for operation on mobile devices, which typically have less computing resources/power than desktop devices or servers.

Additionally, the employment of tracking validation function block 216 to validate the face patch covered the face mesh provides in-time failure recovery in visual tracking, again making pocket avatar system 100 particularly suitable for operation on a wide range of mobile devices. The employment of eye blink detection and mouth correction function block 218, operating on more granular re-sampled pixels around the eyes and mouth areas, after tracking validation, may improve the eye blink detection accuracy and enhances the mouth movement sensitivity.

Further, with head pose tracking function block 210, mouth openness estimation 212 and facial mesh tracking function block 214 operating on subsets of sub-sampled pixels, the workload of these function blocks may be more scalable, as the workload may be substantially proportional to the number of pixels sampled. Accordingly, the workload may be adjusted, in view of the available computing power, by adjusting the density of pixel sampling. In embodiments, similar strategy may be adopted for face detection function block 202, landmark detection 204, tracking validation function block 216, and eye blink detection and mouth 218. The region of interest may be first resized to a smaller



size before the corresponding image analysis is performed. As a result, the workload of these function blocks 202-204 and 216-218 may be made substantially independent of the image frame size, and may be more scalable in view of available computing resource/power, making portable avatar system 100 more suitable for mobile devices.

Referring back to Figure 1, as described earlier, avatar animation engine 104 may be configured to animate avatar employing shape blending, to speed up its operations. In embodiments, a model with neutral expression and some typical expressions, such as mouth open, mouth smile, brow-up, and brow-down, blink, etc., may be first pre-constructed, prior to facial tracking and animation. The blend shapes may be decided or selected for various tracker 102 capabilities and target mobile device system requirements. During operation, as described earlier facial mesh tracker 202 may output the blend shape weights for avatar animation engine 104.

Upon receiving the blend shape weights ( $\alpha_i$ ) for the various blend shapes, avatar animation engine 104 may generate the expressed facial results with the formula:

$$B^* = B_0 + \sum_i \alpha_i \cdot \Delta B_i$$

where  $B^*$  is the target expressed facial,

$B_0$  is the base model with neutral expression, and

$\Delta B_i$  is  $i^{th}$  blend shape that stores the vertex position offset based on base model for specific expression.

Compared with other facial animation techniques, such as motion transferring and mesh deformation, using blend shape for facial animation may have several advantages: 1) Expressions customization: expressions may be customized according to the concept and characteristics of the avatar, when the avatar models are created. The avatar models may be made more funny and attractive to users. 2) Low computation cost: the computation may be configured to be proportional to the model size, and made more suitable for parallel processing. 3) Good scalability: addition of more expressions into the framework may be made easier.

Still referring to Figure 1, as described earlier, in embodiments, facial mesh tracker 102 may be configured to generate and output animation messages 108 having the facial motion parameters that depict facial action movements detected, for avatar animation engine 104. In embodiments, facial mesh tracker 102 and avatar animation engine 104 may be further configured to cooperate to support user interaction driven avatar animation,

where a canned expression, e.g., sticking a tongue out, corresponding to a user interaction, e.g., a swipe gesture, may be animated, in lieu of detected facial expression. See Figure 3, wherein the example animation of the canned expression 300 of sticking a tongue out, corresponding to a user interaction, is illustrated. Similarly, facial mesh tracker 102 may be configured to detect, generate and output animation messages 108 having information about the user interaction, e.g., a start period 402, a keep period 404, and an end period 406, as illustrated in Figure 4, and/or the corresponding canned expression.

In embodiments, there may be two types of animation messages 108, facial expression animation messages, and interaction messages. The facial expression messages may be used to support facial expression driven avatar animation, whereas the interaction messages may be used to support interaction event driven avatar animation, e.g., touch events driven avatar animation, for devices with touch sensitive screens. In embodiments, a facial expression message may be 88 bytes in length. The first 12 bytes may be used to specify an avatar type, a version and a message size. The remaining 76 bytes may be used to specify various attributes or characteristics of the facial expressions. For the facial expression data, in embodiments, the first 12 bytes may specify the head pose, the next 36 bytes may specify various pre-defined blend shapes, with the remaining 28 bytes reserved. In embodiments, animation message 108 may be compressed, with the head pose and blend shape data quantized to 16-bit short and 8-bit byte respectively.

In embodiments, the interaction messages may specify the interaction type, and duration information. The interaction type may index to a corresponding canned expression to be animated, e.g., but not limited, tongue-out 300, wink (not shown), kiss (not shown), and so forth. The duration information may specify a start period 402, a keep period 404 and an end period 406. In embodiments, start period 402 may define the number of frames in the starting stage. For example, for the TongueOut example, the avatar will stick out the tongue in this stage. Keep period 404 may define the time to keep the current status, whereas end period 406 may define when the avatar should recover back to neutral expression. In other words, end period 406 may define the recovering time from the interaction expression to the neutral face.

In embodiments, all interaction events have the same priority, and all facial expression events have the same priority, while interaction events have higher priority than the facial expression events. That means: 1) an interaction event cannot interrupt other interaction events. It will take effect only after the end of the current interaction event. During an interaction event, the event queue will not accept another interaction

event; 2) an interaction event can interrupt facial expression events anytime. When a new interaction event is detected, the facial mesh tracker 102 will replace the facial expression event with the interaction event at that time frame. After that, the facial expression event will resume to take effect.

- 5 As described earlier, in embodiments, avatar animation engine 104 may employ blend shapes. For these embodiments, the expression may be animated for the start, keep and end periods 402-404 as follows:

$$B_t = \begin{cases} B_0 + \frac{t}{N_s} \Delta B, & \text{if } 0 \leq t < N_s \\ B_0 + \Delta B, & \text{if } N_s \leq t < N_s + N_k \\ B_0 + \frac{t - N_s - N_k}{N_e} \Delta B, & \text{if } N_s + N_k \leq t < N_s + N_k + N_e \end{cases}$$

where  $B_t$  is the expression at a point in time,

- 10  $B_0$  and  $\Delta B$  are as earlier defined,

$t$  is time, and

$N_s$ ,  $N_k$ , and  $N_e$  are the number of frames for the start, keep and end periods.

- Referring now to Figure 5, wherein a process for generating facial expression and interaction animation messages, is illustrated, according to the disclosed embodiments.
- 15 Process 500 for generating facial expression and interaction animation messages may be performed e.g., by the earlier described facial mesh tracker 102 of Figure 1. As shown, the process may start at block 502 where recording of animation messages may begin. Message recording may begin in response to e.g., a user providing a start recording
- 20 instruction, such as a click on a start recording button in a user interface provided by pocket avatar system 100. At block 504, an image frame may be read. At block 506, a face and facial movements within the image frame may be detected.

- At block 508, a determination may be made as to whether a new interaction has been detected, or a prior interaction event remains not completed. If no new interaction
- 25 has been detected, nor any prior interaction event remains in progress, at block 510, a facial expression message with facial movement data may be generated, for facial expression animation. From block 510, process 500 may continue at block 504 as earlier described.

- At block 512, if a new interaction has been detected, a new interaction message
- 30 with the interaction and duration information may be generated, to facilitate animation of the corresponding canned expression. However, if a prior interaction event remains in

progress, neither facial expression nor interaction message will be generated, allowing interaction animation of the corresponding canned expression of the prior interaction to continue. From block 512, process 500 may continue at block 504 as earlier described, if neither a stop recording instruction has been received, nor a recording length limit threshold has been reached. On the other hand, if either a stop recording instruction has been received, or a recording length limit threshold has been reached, process 500 may proceed to block 514 and terminates.

Figure 6 is a flow diagram illustrating a process for interleaving facial expression and interaction driven animation, according to the disclosed embodiments. Process 600 for interleaving facial expression and interaction driven animation may be performed e.g., by the earlier described avatar animation engine 104 of Figure 1. As shown, the process may start at block 602 where playing of animation messages may begin. Message playing may begin contemporaneously with recording, in response to e.g., a user providing a start recording/playing instruction, such as a click on a start recording/playing button in a user interface provided by pocket avatar system 100. At block 604, an animation message corresponding to an image frame may be read, and its data extracted.

At block 606, if the extracted data has interaction event inside, animation of the index canned expression is performed. Further, a marking of the beginning of a new interaction event may be made. However, if the extracted data has no interaction event inside, and currently there is no incomplete animation of any canned expression for a prior interaction event, animation of facial expression, in accordance with the facial expression data in the animation message is performed. On the other hand, if the extracted data has no interaction event inside, but currently there is incomplete animation of a canned expression for a prior interaction event, then animation of the canned expression corresponding to the prior interaction event continues.

From block 606, process 600 may continue at block 604 as earlier described, if neither a stop recording/playing instruction has been received, nor end of messages has been reached. On the other hand, if either a stop recording/playing instruction has been received, or end of messages has been reached, process 600 may proceed to block 608 and terminates.

Referring now to Figure 7, wherein a flow diagram illustrating a process for estimating head pose, is illustrated, according to the disclosed embodiments. As shown, process 700 for estimating head pose may include model training operations 702, 3D shape reconstruction for neutral face operations 704, frontal view prediction operations

706, and visual tracking operations 708. Model training operations 702 may be performed offline, prior to operation of tracking, animation and rendering by portable avatar system 100, whereas 3D shape reconstruction for neutral face operations 704, frontal view prediction operations 706, and visual tracking operations 708 may be performed by the earlier described facial mesh tracker 102.

As shown, model training operations 702 may include using a learner 714 to learn a 3D Facial Shape Units Model (FSU) 716 and a 3D Facial Action Units (FAU) 718 from a 3D face database having a substantial collection of different facial expressions, e.g., hundreds of identities, each having several typical expressions, and the key landmark points are provided. The 3D FSU model may describe a space with variant face shapes, whereas the 3D FAU model may describe local motion of facial components (facial expression). More specifically, in embodiments, a principal component analysis (PCA) may be first performed on all 3D shapes with neutral expression. After that, mean shapes for each expression may be computed. The difference between the means shapes with expression, and the mean shape of neutral may be taken as the FAU model. In embodiments, each FAU may be designed for just one component's motion in one dimension. Examples of components may include eye, eyebrow, nose, mouth, and so forth. Thus, the FAUs are independent, and can be composed together to obtain a complex facial expression, e.g., a surprise expression may include mouth-open and brow-up FAUs.

In embodiments, 3D shape reconstruction for neutral face operations 704 may be performed during registration of a user, wherein a number of neutral faces may be collected, and employed to construct a 3D neutral face. More specifically, in embodiments, 3D FSUs describing face shape variance may be used to reconstruct the 3D face shape through minimization of the difference between 2D projection and a neutral face registered  $B_0$ , by solving the following optimization problem:

$$\min_{\alpha} \|B_0 - T_{2d}(P_0 + \alpha P)\|$$

where  $P_0$  is the mean shape of 3D FSU,

$P$  is an eigen vector of 3D FSU,

$\alpha$  is a linear combination coefficient, and

$T_{2d}$  is a projection from 3D space to 2D image space.

A 3D shape thus may be constructed by computing:

$$B_{2d} = F_0 + \alpha F$$

In embodiments, frontal view prediction operations 706 may be performed to reconstruct a 3D shape  $S_{3d}$ , using the 3D face shape of the user constructed during registration and the 3D FAU model, by minimizing the difference between 2D projection from 3D shape and the 2D image landmarks  $S_0$  provided by visual tracking operations 708, as follows:

$$S_{2d} = bR(B_{3d} + \gamma Q) + t$$

where  $b$ ,  $R$  and  $t$  are rigid transformation parameters (scale, rotation, and translation,

$Q$  is the 3D FAU model, and

$\gamma$  is the 3D FAU model's coefficients.

Similar to 3D shape reconstruction for neutral face operations 704, the solution may be obtained by solving the optimization problem of:

$$\min_{\gamma, b, R, t} \|S_0 - T_{2d}[bR(B_{3d} + \gamma Q) + t]\|$$

In embodiments, the optimization problem may be solved by updating the value of the rigid transformation parameters and 3D FAU coefficients separately, and iteratively. In other words, dividing the optimization problem into two sub-problems:

$$\begin{aligned} & \min_{\gamma} \|S_0 - T_{2d}[bR(B_{3d} + \gamma Q) + t]\|, \text{ given } b, R \text{ and } t. \\ & \min_{R, b, t} \|S_0 - T_{2d}[bR(B_{3d} + \gamma Q) + t]\|, \text{ given } \gamma. \end{aligned}$$

Thereafter, the landmark in the front view without 3D rigid transformation may be obtained with the optimization of:

$$S_{2d} = T_{2d}(B_{3d} + \gamma Q)$$

where  $S_{2d}$  is the 2D projection of 3D shape with FAUs for the user with a specific face shape.

The head pose tracking may complement the facial mesh tracking. In combination, the two tracking may validate each other, and improve overall tracking robustness.

Experiments have shown the disclosed portable avatar system 100 to be very efficient for mobile devices, capable of processing 70 frames per second on a Samsung Galaxy S3 phone and 110 frames per second on Apple iPhone5.

Figure 8 illustrates an example computer system that may be suitable for use as a

client device or a server to practice selected aspects of the present disclosure. As shown, computer 800 may include one or more processors or processor cores 802, and system memory 804. For the purpose of this application, including the claims, the terms “processor” and “processor cores” may be considered synonymous, unless the context clearly requires otherwise. Additionally, computer 800 may include mass storage devices 806 (such as diskette, hard drive, compact disc read only memory (CD-ROM) and so forth), input/output devices 808 (such as display, keyboard, cursor control and so forth) and communication interfaces 810 (such as network interface cards, modems and so forth). The elements may be coupled to each other via system bus 812, which may represent one or more buses. In the case of multiple buses, they may be bridged by one or more bus bridges (not shown).

Each of these elements may perform its conventional functions known in the art. In particular, system memory 804 and mass storage devices 806 may be employed to store a working copy and a permanent copy of the programming instructions implementing the operations associated with facial mesh tracker 102, avatar animation engine 104 and avatar rendering engine 106, earlier described, collectively referred to as computational logic 822. The various elements may be implemented by assembler instructions supported by processor(s) 802 or high-level languages, such as, for example, C, that can be compiled into such instructions.

The number, capability and/or capacity of these elements 810 - 812 may vary, depending on whether computer 800 is used as a client device or a server. When use as client device, the capability and/or capacity of these elements 810 - 812 may vary, depending on whether the client device is a stationary or mobile device, like a smartphone, computing tablet, ultrabook or laptop. Otherwise, the constitutions of elements 810-812 are known, and accordingly will not be further described.

As will be appreciated by one skilled in the art, the present disclosure may be embodied as methods or computer program products. Accordingly, the present disclosure, in addition to being embodied in hardware as earlier described, may take the form of an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware aspects that may all generally be referred to as a “circuit,” “module” or “system.” Furthermore, the present disclosure may take the form of a computer program product embodied in any tangible or non-transitory medium of expression having computer-usable program code embodied in the medium. Figure 9 illustrates an example computer-readable non-transitory storage medium that may

be suitable for use to store instructions that cause an apparatus, in response to execution of the instructions by the apparatus, to practice selected aspects of the present disclosure. As shown, non-transitory computer-readable storage medium 902 may include a number of programming instructions 904. Programming instructions 904 may be configured to  
5 enable a device, e.g., computer 800, in response to execution of the programming instructions, to perform, e.g., various operations associated with facial mesh tracker 102, avatar animation engine 104 and avatar rendering engine 106. In alternate embodiments, programming instructions 904 may be disposed on multiple computer-readable non-transitory storage media 902 instead. In alternate embodiments, programming instructions  
10 904 may be disposed on computer-readable transitory storage media 902, such as, signals.

Any combination of one or more computer usable or computer readable medium(s) may be utilized. The computer-usable or computer-readable medium may be, for example but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. More specific examples  
15 (a non-exhaustive list) of the computer-readable medium would include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a transmission media such as  
20 those supporting the Internet or an intranet, or a magnetic storage device. Note that the computer-usable or computer-readable medium could even be paper or another suitable medium upon which the program is printed, as the program can be electronically captured, via, for instance, optical scanning of the paper or other medium, then compiled, interpreted, or otherwise processed in a suitable manner, if necessary, and then stored in a computer  
25 memory. In the context of this document, a computer-usable or computer-readable medium may be any medium that can contain, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device. The computer-usable medium may include a propagated data signal with the computer-usable program code embodied therewith, either in baseband or as part of a  
30 carrier wave. The computer usable program code may be transmitted using any appropriate medium, including but not limited to wireless, wireline, optical fiber cable, RF, etc.

Computer program code for carrying out operations of the present disclosure may be written in any combination of one or more programming languages, including an object



oriented programming language such as Java, Smalltalk, C++ or the like and conventional procedural programming languages, such as the “C” programming language or similar programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's  
5 computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider).

10 The present disclosure is described with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the disclosure. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart  
illustrations and/or block diagrams, can be implemented by computer program instructions.

15 These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for  
20 implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

These computer program instructions may also be stored in a computer-readable medium that can direct a computer or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer-readable medium produce an article of manufacture including instruction means which implement  
25 the function/act specified in the flowchart and/or block diagram block or blocks.

The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer  
30 implemented process such that the instructions which execute on the computer or other programmable apparatus provide processes for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

The flowchart and block diagrams in the figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods and computer program products according to various embodiments of the present disclosure. In

this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of code, which comprises one or more executable instructions for implementing the specified logical function(s). It should also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the disclosure. As used herein, the singular forms “a,” “an” and “the” are intended to include plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specific the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operation, elements, components, and/or groups thereof.

Embodiments may be implemented as a computer process, a computing system or as an article of manufacture such as a computer program product of computer readable media. The computer program product may be a computer storage medium readable by a computer system and encoding a computer program instructions for executing a computer process.

The corresponding structures, material, acts, and equivalents of all means or steps plus function elements in the claims below are intended to include any structure, material or act for performing the function in combination with other claimed elements are specifically claimed. The description of the present disclosure has been presented for purposes of illustration and description, but is not intended to be exhaustive or limited to the disclosure in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill without departing from the scope and spirit of the disclosure. The embodiment was chosen and described in order to best explain the principles of the disclosure and the practical application, and to enable others of ordinary skill in the art to understand the disclosure for embodiments with various modifications as are suited to the

particular use contemplated.

Referring back to Figure 8, for one embodiment, at least one of processors 802 may be packaged together with memory having computational logic 822 (in lieu of storing on memory 804 and storage 806). For one embodiment, at least one of processors 802 may be packaged together with memory having computational logic 822 to form a System in Package (SiP). For one embodiment, at least one of processors 802 may be integrated on the same die with memory having computational logic 822. For one embodiment, at least one of processors 802 may be packaged together with memory having computational logic 822 to form a System on Chip (SoC). For at least one embodiment, the SoC may be utilized in, e.g., but not limited to, a smartphone or computing tablet.

Thus various example embodiments of the present disclosure have been described including, but are not limited to:

Example 1 may be an apparatus for rendering avatar. The apparatus may comprise one or more processors; and a facial mesh tracker. The facial mesh tracker may be operated by the one or more processors, to receive a plurality of image frames, detect, through the plurality of image frame, facial action movements of a face of a user and head pose gestures of a head of the user, and output a plurality of facial motion parameters that depict facial action movements detected, a plurality of head gesture parameters that depict head pose gestures detected, all in real time, for animation and rendering of an avatar.

Further, detection of facial action movements, and head pose gestures may include detection of inter-frame differences for a mouth and an eye on the face, and the head, based on pixel sampling of the image frames.

Example 2 may be example 1, wherein the facial action movements may include opening or closing of the mouth, and blinking of the eye, and the plurality of facial motion parameters may include first one or more facial motion parameters that depict the opening or closing of the mouth and second one or more facial motion parameters that depict blinking of the eye.

Example 3 may be example 1 or 2, wherein the plurality of image frames may be captured by a camera, and the head pose gestures may include head rotation, movement along horizontal and vertical directions, and the head comes closer or goes farther from the camera; and wherein the plurality of head pose gesture parameters may include head pose gesture parameters that depict head rotation, head movement along horizontal and vertical directions, and head comes closer or goes farther from the camera.

Example 4 may be any one of examples 1-3, wherein the facial mesh tracker may include a face detection function block to detect the face through window scan of one or more of the plurality of image frames; wherein window scan may comprise extraction of modified census transform features and application of a cascade classifier at each window position.

Example 5 may be any one of examples 1-4, wherein the facial mesh tracker may include a landmark detection function block to detect landmark points on the face; wherein detection of landmark points may comprise assignment of an initial landmark position in a face rectangle according to mean face shape, and iteratively assign exact landmark positions through explicit shape regression.

Example 6 may be any one of examples 1-5, wherein the facial mesh tracker may include an initial face mesh fitting function block to initialize a 3D pose of a face mesh based at least in part on a plurality of landmark points detected on the face, employing a Candide3 wireframe head model.

Example 7 may be any one of examples 1-6, wherein the facial mesh tracker may include a facial expression estimation function block to initialize a plurality of facial motion parameters based at least in part on a plurality of landmark points detected on the face, through least square fitting.

Example 8 may be any one of examples 1-7, wherein the facial mesh tracker may include a head pose tracking function block to calculate rotation angles of the user's head, based on a subset of sub-sampled pixels of the plurality of image frames, applying dynamic template matching and re-registration.

Example 9 may be any one of examples 1-8, wherein the facial mesh tracker may include a mouth openness estimation function block to calculate opening distance of an upper lip and a lower lip of the mouth, based on a subset of sub-sampled pixels of the plurality of image frames, applying FERN regression.

Example 10 may be any one of examples 1-9, wherein the facial mesh tracking function block may adjust position, orientation or deformation of a face mesh to maintain continuing coverage of the face and reflection of facial movement by the face mesh, based on a subset of sub-sampled pixels of the plurality of image frames, and image alignment of successive image frames.

Example 11 may be any one of examples 1-10, wherein the facial mesh tracker may include a tracking validation function block to monitor face mesh tracking status,

applying one or more face region or eye region classifiers, to determine whether it is necessary to relocate the face.

Example 12 may be any one of examples 1-11, wherein the facial mesh tracker may include a mouth shape correction function block to correct mouth shape, through  
5 detection of inter-frame histogram differences for the mouth.

Example 13 may be any one of examples 1-12, wherein the facial mesh tracker may include an eye blinking detection function block to estimate eye blinking, through optical flow analysis.

Example 14 may be any one of examples 1-13, wherein the facial mesh tracker  
10 may include a face mesh adaptation function block to reconstruct a face mesh according to derived facial action units, and re-sample a current image frame under the face mesh to set up processing of a next image frame.

Example 15 may be any one of examples 1-14, wherein the facial mesh tracker may include blend-shape mapping function block to convert facial action units into blend-  
15 shape coefficients for the animation of the avatar.

Example 16 may be any one of examples 1-15, further comprising an avatar animation engine coupled with the facial mesh tracker to receive the plurality of facial motion parameters outputted by the facial mesh tracker, and drive an avatar model to animate the avatar, replicating a facial expression of the user on the avatar, through  
20 blending of a plurality of pre-defined shapes.

Example 17 may be any one of examples 1-16, further comprising an avatar rendering engine coupled with the avatar animation engine to draw the avatar as animated by avatar animation engine.

Example 18 may be any one of examples 1-17, wherein the apparatus is a selected  
25 one of a smartphone, a phablet, a computing tablet, a laptop computer, a e-reader, a desktop computer, a game console, a set-top box, or a computer server.

Example 19 may be a method for rendering an avatar. The method may comprise receiving, by a facial mesh tracker operating on a computing device, a plurality of image frames; detecting, by the facial mesh tracker, through the plurality of image frames, facial  
30 action movements of a face of a user, and head pose gestures of a head of the user; and outputting, by the facial mesh tracker, a plurality of facial motion parameters that depict facial action movements detected, and a plurality of head pose gesture parameters that depict head pose gestures detected. Additionally, receiving, detecting and outputting may all be performed in real time, for animation and rendering of an avatar. Further, detecting

facial action movements and head pose gestures may include detecting inter-frame differences for a mouth and an eye of the face, and the head, based on pixel sampling of the image frames.

Example 20 may be example 19, wherein the facial action movements may include opening or closing of the mouth, and blinking of the eye, and the plurality of facial motion parameters include first one or more facial motion parameters that depict the opening or closing of the mouth and second one or more facial motion parameters that depict blinking of the eye.

Example 21 may be example 19 or 20, wherein the plurality of image frames may be captured by a camera, and the head pose gestures may include head rotation, movement along horizontal and vertical directions, and the head comes closer or goes farther from the camera; and wherein the plurality of head pose gesture parameters may include head pose gesture parameters that depict head rotation, head movement along horizontal and vertical directions, and head comes closer or goes farther from the camera.

Example 22 may be any one of examples 19-21, wherein detecting may comprise detecting the face through window scanning of one or more of the plurality of image frames; wherein window scanning comprises extracting modified census transform features and applying a cascade classifier at each window position.

Example 23 may be any one of examples 19-22, wherein detecting may comprise detecting landmark points on the face; wherein detecting landmark points may comprise assigning an initial landmark position in a face rectangle according to mean face shape, and iteratively assigning exact landmark positions through explicit shape regression.

Example 24 may be any one of examples 19-23, , wherein detecting may comprise initializing a 3D pose of a face mesh based at least in part on a plurality of landmark points detected on the face, employing a Candide3 wireframe head model.

Example 25 may be any one of examples 19-24, , wherein detecting may include initializing a plurality of facial motion parameters based at least in part on a plurality of landmark points detected on the face, through least square fitting.

Example 26 may be any one of examples 19-25, , wherein detecting may comprise calculating rotation angles of the user's head, based on a subset of sub-sampled pixels of the plurality of image frames, applying dynamic template matching and re-registration.

Example 27 may be any one of examples 19-26, wherein detecting may include calculating opening distance of an upper lip and a lower lip of the mouth, based on a subset of sub-sampled pixels of the plurality of image frames, applying FERN regression.

Example 28 may be any one of examples 19-27, , wherein detecting may comprise adjusting position, orientation or deformation of a face mesh to maintain continuing coverage of the face and reflection of facial movement by the face mesh, based on a subset of sub-sampled pixels of the plurality of image frames, and aligning successive image

5 frames.

Example 29 may be any one of examples 19-28, wherein detecting may comprise monitoring face mesh tracking status, applying one or more face region or eye region classifiers, to determine whether it is necessary to re-locate the face.

Example 30 may be any one of examples 19-29, , wherein detecting may include  
10 correcting mouth shape, through detection of inter-frame histogram differences for the mouth.

Example 31 may be any one of examples 19-30,, wherein detecting may comprise estimating eye blinking, through optical flow analysis.

Example 32 may be any one of examples 19-31, , wherein detecting may comprise  
15 reconstructing a face mesh according to derived facial action units, and re-sampling a current image frame under the face mesh to set up processing of a next image frame.

Example 33 may be any one of examples 19-32, , wherein detecting may comprise converting facial action units into blend-shape coefficients for the animation of the avatar.

Example 34 may be any one of examples 19-33, further comprising:  
20 receiving, by an avatar animation engine operating on the computing device, the plurality of facial motion parameters outputted; and

driving, by the avatar animation engine, an avatar model to animate the avatar, replicating a facial expression of the user on the avatar, through shape blending.

Example 35 may be any one of examples 19-34, further comprising drawing, by an  
25 avatar rendering engine operating on the computing device, the avatar as animated by avatar animation engine.

Example 36 may be one or more computer-readable storage medium comprising a plurality of instructions to cause a computing device, in response to execution of the instructions by the computing device, to perform any one of the method examples of 19-35.

30 Example 37 may be an apparatus for rendering avatar. The apparatus may comprise: facial mesh tracking means for receiving a plurality of image frames, detecting, through the plurality of image frames, facial action movements of a face of a user, and head pose gestures of the user, and outputting a plurality of facial motion parameters that depict facial action movements detected, and a plurality of head pose gestures parameters,

all in real time, for animation and rendering of an avatar. Further, detecting facial action movements and head pose gestures may include detecting inter-frame differences for a mouth and an eye of the face, and the head, based on pixel sampling of the image frames.

Example 38 may be example 37 further comprising avatar animation means for  
5 receiving the plurality of facial motion parameters, and driving an avatar model to animate the avatar, replicating a facial expression of the user on the avatar, through shape blending.

Example 39 may be example 38 further comprising avatar rendering means for drawing the avatar as animated by avatar animation engine.

Example 40 may be an apparatus for rendering an avatar. The apparatus may  
10 comprise: one or more processors, a facial mesh tracker and an animation engine. The facial mesh tracker may be operated by the one or more processors, to receive a plurality of image frames, first detect facial action movements of a face within the plurality of image frames, first generate first one or more animation messages recording the facial action movements, second detect one or more user interactions with the apparatus during  
15 receipt of the plurality of image frames and first detection of facial action movements of a face within the plurality of image frames, and second generate second one or more animation messages recording the one or more user interactions detected, all in real time. Further, the animation engine may be coupled with the facial mesh tracker, to drive an avatar model to animate an avatar, interleaving replication of the recorded facial action  
20 movements on the avatar based on the first one or more animation messages, with animation of one or more canned facial expressions corresponding to the one or more recorded user interactions based on the second one or more animation messages.

Example 41 may be example 40, wherein each of the first one or more animation messages may comprise a first plurality of data bytes to specify an avatar type, a second  
25 plurality of data bytes to specify head pose parameters, and a third plurality of data bytes to specify a plurality of pre-defined shapes to be blended to animate the facial expression.

Example 42 may be example 40 or 41, wherein each of the second one or more animation messages may comprise a first plurality of data bits to specify a user interaction, and a second plurality of data bits to specify a duration for animating the canned facial  
30 expression corresponding to the user interaction specified.

Example 43 may be any one of examples 40-42, wherein the duration may comprise a start period, a keep period and an end period for the animation.



Example 44 may be example 43, wherein the avatar animation engine may animate the corresponding canned facial expression blending one or more pre-defined shapes into a neutral face based at least in part on the start, keep and end periods.

Example 45 may be any one of examples 40-42, wherein second detect may  
5 comprise second detect of whether a new user interaction occurred and whether a prior detected user interaction has completed, during first detection of facial action movements of a face within an image frame.

Example 46 may be any one of examples 40-42, wherein the facial mesh tracker to  
10 start performance of the receipt, the first detect, the first generate, the second detect and the second generate, in response to a start instruction, and to stop performance of the receipt, the first detect, the first generate, the second detect and the second generate, in response to a stop instruction, or the number or a total size of the first and second animation messages reach a threshold.

Example 47 may be any one of examples 40-42, wherein the avatar animation  
15 engine is to determine whether data within an animation message comprises recording of occurrence of a new user interaction or incompleteness of a prior detected user interaction, during recovery of facial action movement data from an animation message for an image frame.

Example 48 may be any one of examples 40-42, wherein the avatar animation  
20 engine is to start performance of the animation, in response to a start instruction, and to stop performance of the animation, in response to a stop instruction, or completion of processing of all first and second animation messages.

Example 49 may be a method for rendering an avatar. The method may comprise:  
receiving, by a facial mesh tracker operating on a computing device, a plurality of image  
25 frames; first detecting, by the facial mesh tracker, facial action movements of a face within the plurality of image frames; first generating, by the facial mesh tracker, first one or more animation messages recording the facial action movements; second detecting, by the facial mesh tracker, one or more user interactions with the computing device during receipt of the plurality of image frames and first detecting of facial action movements of a face  
30 within the plurality of image frames; and second generating second one or more animation messages recording the one or more user interactions detected. Further, the may include driving, by an avatar animation engine, an avatar model to animate an avatar, interleaving replication of the recorded facial action movements on the avatar based on the first one or more animation messages, with animation of one or more canned facial expressions

corresponding to the one or more recorded user interactions based on the second one or more animation messages. Additionally, the receiving, the first detecting, the first generating, the second detecting, the second generating, and the driving, may all be performed in real time.

5           Example 50 may be example 49, wherein each of the first one or more animation messages may comprise a first plurality of data bytes to specify an avatar type, a second plurality of data bytes to specify head pose parameters, and a third plurality of data bytes to specify a plurality of pre-defined shapes to be blended to animate the facial expression.

10           Example 51 may be example 49 or 50, wherein each of the second one or more animation messages comprises a first plurality of data bits to specify a user interaction, and a second plurality of data bits to specify a duration for animating the canned facial expression corresponding to the user interaction specified.

          Example 52 may be example 51, wherein the duration may comprise a start period, a keep period and an end period for the animation.

15           Example 53 may be example 52, wherein animating the corresponding canned facial expression comprises blending one or more pre-defined shapes into a neutral face based at least in part on the start, keep and end periods.

          Example 54 may be any one of examples 49-53, wherein second detecting may comprise second detecting whether a new user interaction occurred and whether a prior detected user interaction has completed, during first detecting of facial action movements of a face within an image frame.

20           Example 55 may be any one of examples 49-54, wherein performance of receiving, first detecting, first generating, second detecting and second generating, is in response to a start instruction, and performance to stop, in response to a stop instruction, or the number or a total size of the first and second animation messages reaching a threshold.

          Example 56 may be any one of examples 49-55, wherein driving may comprise determining whether data within an animation message comprises recording of occurrence of a new user interaction or incompleteness of a prior detected user interaction, during recovery of facial action movement data from an animation message for an image frame.

30           Example 57 may be any one of methods of examples 49-56, wherein performance of driving, is in response to a start instruction, and performance to stop, in response to a stop instruction, or completion of processing of all first and second animation messages.

Example 58 may be one or more computer-readable storage medium comprising a plurality of instructions to cause a computing device, in response to execution of the instructions by the computing device, to perform any one of the example methods of 49-57.

Example 59 may be an apparatus for rendering an avatar. The apparatus may  
5 comprise: facial mesh tracking means for receiving a plurality of image frames, first detecting facial action movements of a face within the plurality of image frames, first generating first one or more animation messages recording the facial action movements, second detecting one or more user interactions with the apparatus during receiving of the plurality of image frames and first detecting of facial action movements of a face within  
10 the plurality of image frames, and second generating second one or more animation messages recording the one or more user interactions detected, all in real time; and avatar animation means for driving an avatar model to animate an avatar, interleaving replication of the recorded facial action movements on the avatar based on the first one or more animation messages, with animation of one or more canned facial expressions  
15 corresponding to the one or more recorded user interactions based on the second one or more animation messages.

Example 60 may be example 59, wherein each of the first one or more animation messages may comprise a first plurality of data bytes to specify an avatar type, a second plurality of data bytes to specify head pose parameters, and a third plurality of data bytes  
20 to specify a plurality of pre-defined shapes to be blended to animate the facial expression.

Example 60 may be example 59 or 60, wherein each of the second one or more animation messages may comprise a first plurality of data bits to specify a user interaction, and a second plurality of data bits to specify a duration for animating the canned facial expression corresponding to the user interaction specified.

25 Example 62 may be example 61, wherein the duration may comprise a start period, a keep period and an end period for the animation.

Example 63 may be example 62, wherein the avatar animation means may  
comprise means for animating the corresponding canned facial expression, by blending one or more pre-defined shapes into a neutral face based at least in part on the start, keep  
30 and end periods.

Example 64 may be an apparatus for rendering avatar. The apparatus may comprise: one or more processors; and a facial mesh tracker. The facial mesh tracker may be operated by the one or more processors, to receive a plurality of image frames, detect facial action movements of a face within the plurality of image frames, and output a

plurality of facial motion parameters that depict facial action movements detected, all in real time, for animation and rendering of an avatar. Additionally, the face may be a face of a user, and the facial mesh tracker may detect facial action movements of the face through a normalized head pose of the user. Further, the facial mesh tracker may generate the normalized head pose of the user by using a 3D facial action model and a 3D neutral facial shape of the user pre-constructed using a 3D facial shape model.

Example 65 may be example 64, wherein the facial mesh tracker may generate the normalized head pose of the user through minimization of differences between 2D projection of the 3D neutral facial shape and detected 2D image landmarks.

Example 66 may be example 64 or 65, wherein the 3D facial action model is pre-developed offline through machine learning of a 3D facial database.

Example 67 may be any one of examples 64 - 66, wherein 3D neutral facial shape of the user may be pre-constructed using the 3D facial shape model, during registration of the user.

Example 68 may be any one of examples 64 - 67, wherein the 3D facial shape model may be pre-developed offline through machine learning of a 3D facial database.

Example 69 may be a method for rendering avatar. The method may comprise: receiving, by a facial mesh tracker operating on a computing device, a plurality of image frames; detecting, by the facial mesh tracker, facial action movements of a face within the plurality of image frames; and outputting, by the facial mesh tracker, a plurality of facial motion parameters that depict facial action movements detected, for animation and rendering of an avatar. Further, the face may be a face of a user, and detecting facial action movements of the face may be through a normalized head pose of the user, and may comprise generating the normalized head pose of the user by using a 3D facial action model and a 3D neutral facial shape of the user pre-constructed using a 3D facial shape model.

Example 70 may be example 69, wherein generating the normalized head pose of the user may comprise minimizing differences between 2D projection of the 3D neutral facial shape and detected 2D image landmarks.

Example 71 may be example 69 or 70, further comprising pre-developing offline the 3D facial action model through machine learning of a 3D facial database.

Example 72 may be example 69 or 71, further comprising pre-constructing the 3D neutral facial shape of the user using the 3D facial shape model, during registration of the user.

Example 73 may be example 69 or 72, further comprising pre-developing the 3D facial shape model offline through machine learning of a 3D facial database.

Example 74 may be one or more computer-readable storage medium comprising a plurality of instructions to cause a computing device, in response to execution of the  
5 instructions by the computing device, to perform any one of the methods of examples 69-73.

Example 75 may be an apparatus for rendering avatar. The apparatus may comprise: facial mesh tracking means for receiving a plurality of image frames, detecting facial action movements of a face within the plurality of image frames, and outputting a  
10 plurality of facial motion parameters that depict facial action movements detected, all in real time, for animation and rendering of an avatar. Further, the face may be a face of a user, and facial mesh tracker means may comprise means for detecting facial action movements of the face through a normalized head pose of the user, and means for generating the normalized head pose of the user by using a 3D facial action model and a  
15 3D neutral facial shape of the user pre-constructed using a 3D facial shape model.

Example 76 may example 75, wherein means for generating the normalized head pose of the user may comprise means for minimizing differences between 2D projection of the 3D neutral facial shape and detected 2D image landmarks.

It will be apparent to those skilled in the art that various modifications and  
20 variations can be made in the disclosed embodiments of the disclosed device and associated methods without departing from the spirit or scope of the disclosure. Thus, it is intended that the present disclosure covers the modifications and variations of the embodiments disclosed above provided that the modifications and variations come within the scope of any claims and their equivalents.

25

## Claims

What is claimed is:

1. An apparatus for rendering avatar, comprising:  
one or more processors; and  
5 a facial mesh tracker, to be operated by the one or more processors, to receive a plurality of image frames, detect, through the plurality of image frame, facial action movements of a face of a user and head pose gestures of a head of the user, and output a plurality of facial motion parameters that depict facial action movements detected, a plurality of head gesture parameters that depict head pose gestures detected, all in real  
10 time, for animation and rendering of an avatar;  
wherein detection of facial action movements, and head pose gestures includes detection of inter-frame differences for a mouth and an eye on the face, and the head, based on pixel sampling of the image frames.
2. The apparatus of claim 1, wherein the facial action movements include  
15 opening or closing of the mouth, and blinking of the eye, and the plurality of facial motion parameters include first one or more facial motion parameters that depict the opening or closing of the mouth and second one or more facial motion parameters that depict blinking of the eye.
3. The apparatus of claim 1, wherein the plurality of image frames are  
20 captured by a camera, and the head pose gestures include head rotation, movement along horizontal and vertical directions, and the head comes closer or goes farther from the camera; and wherein the plurality of head pose gesture parameters include head pose gesture parameters that depict head rotation, head movement along horizontal and vertical directions, and head comes closer or goes farther from the camera.
- 25 4. The apparatus of claim 1, wherein the facial mesh tracker includes a face detection function block to detect the face through window scan of one or more of the plurality of image frames; wherein window scan comprises extraction of modified census transform features and application of a cascade classifier at each window position.
- 30 5. The apparatus of claim 1, wherein the facial mesh tracker includes a landmark detection function block to detect landmark points on the face; wherein detection of landmark points comprises assignment of an initial landmark position in a

face rectangle according to mean face shape, and iteratively assign exact landmark positions through explicit shape regression.

6. The apparatus of claim 1, wherein the facial mesh tracker includes an initial face mesh fitting function block to initialize a 3D pose of a face mesh based at least in part on a plurality of landmark points detected on the face, employing a Candide3 wireframe head model.

7. The apparatus of claim 1, wherein the facial mesh tracker includes a facial expression estimation function block to initialize a plurality of facial motion parameters based at least in part on a plurality of landmark points detected on the face, through least square fitting.

8. The apparatus of claim 1, wherein the facial mesh tracker includes a head pose tracking function block to calculate rotation angles of the user's head, based on a subset of sub-sampled pixels of the plurality of image frames, applying dynamic template matching and re-registration.

9. The apparatus of claim 1, wherein the facial mesh tracker includes a mouth openness estimation function block to calculate opening distance of an upper lip and a lower lip of the mouth, based on a subset of sub-sampled pixels of the plurality of image frames, applying FERN regression.

10. The apparatus of claim 1, wherein the facial mesh tracking function block is to adjust position, orientation or deformation of a face mesh to maintain continuing coverage of the face and reflection of facial movement by the face mesh, based on a subset of sub-sampled pixels of the plurality of image frames, and image alignment of successive image frames.

11. The apparatus of claim 1, wherein the facial mesh tracker includes a tracking validation function block to monitor face mesh tracking status, applying one or more face region or eye region classifiers, to determine whether it is necessary to relocate the face.

12. The apparatus of claim 1, wherein the facial mesh tracker includes a mouth shape correction function block to correct mouth shape, through detection of inter-frame histogram differences for the mouth.

13. The apparatus of claim 1, wherein the facial mesh tracker includes an eye  
5 blinking detection function block to estimate eye blinking, through optical flow analysis.

14. The apparatus of claim 1, wherein the facial mesh tracker includes a face mesh adaptation function block to reconstruct a face mesh according to derived facial action units, and re-sample a current image frame under the face mesh to set up processing of a next image frame.

10 15. The apparatus of claim 1, wherein the facial mesh tracker includes blend-shape mapping function block to convert facial action units into blend-shape coefficients for the animation of the avatar.

16. The apparatus of claim 1 further comprising:  
an avatar animation engine coupled with the facial mesh tracker to receive the  
15 plurality of facial motion parameters outputted by the facial mesh tracker, and drive an avatar model to animate the avatar, replicating a facial expression of the user on the avatar, through blending of a plurality of pre-defined shapes; and  
an avatar rendering engine coupled with the avatar animation engine to draw the avatar as animated by avatar animation engine.

20 17. An apparatus for rendering an avatar, comprising:  
one or more processors; and  
a facial mesh tracker, to be operated by the one or more processors, to receive a plurality of image frames, first detect facial action movements of a face within the plurality of image frames, first generate first one or more animation messages recording  
25 the facial action movements, second detect one or more user interactions with the apparatus during receipt of the plurality of image frames and first detection of facial action movements of a face within the plurality of image frames, and second generate second one or more animation messages recording the one or more user interactions detected, all in real time; and  
30 an animation engine, coupled with the facial mesh tracker, to drive an avatar model to animate an avatar, interleaving replication of the recorded facial action movements on



the avatar based on the first one or more animation messages, with animation of one or more canned facial expressions corresponding to the one or more recorded user interactions based on the second one or more animation messages.

18. The apparatus of claim 17, wherein each of the first one or more animation  
5 messages comprises a first plurality of data bytes to specify an avatar type, a second plurality of data bytes to specify head pose parameters, and a third plurality of data bytes to specify a plurality of pre-defined shapes to be blended to animate the facial expression.

19. The apparatus of claim 17, wherein each of the second one or more  
10 animation messages comprises a first plurality of data bits to specify a user interaction, and a second plurality of data bits to specify a duration for animating the canned facial expression corresponding to the user interaction specified.

20. The apparatus of claim 19, wherein the duration comprises a start period, a keep period and an end period for the animation; and wherein the avatar animation engine to animate the corresponding canned facial expression blending one or more pre-defined  
15 shapes into a neutral face based at least in part on the start, keep and end periods.

21. The apparatus of any one of claims 17 - 20, wherein second detect  
comprises second detect of whether a new user interaction occurred and whether a prior detected user interaction has completed, during first detection of facial action movements of a face within an image frame; and wherein the avatar animation engine to determine  
20 whether data within an animation message comprises recording of occurrence of a new user interaction or incompleteness of a prior detected user interaction, during recovery of facial action movement data from an animation message for an image frame.

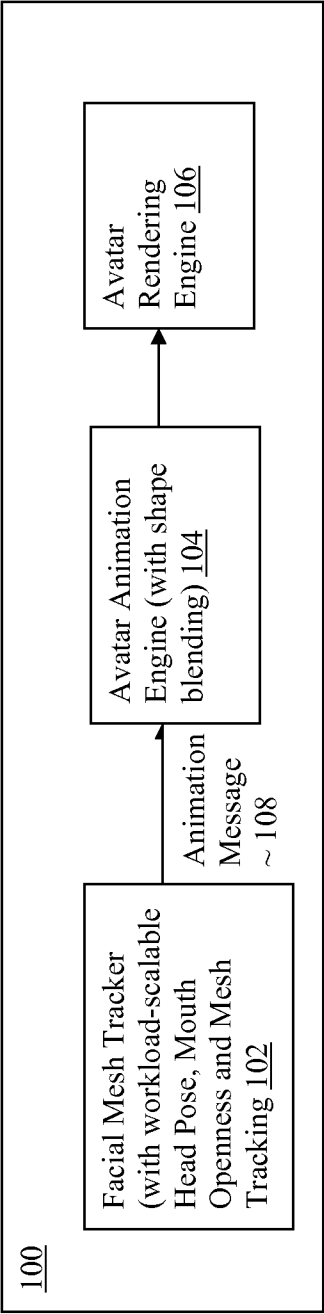
22. A method for rendering avatar, comprising:  
receiving, by a facial mesh tracker operating on a computing device, a plurality of  
25 image frames;  
detecting, by the facial mesh tracker, facial action movements of a face within the plurality of image frames; and  
outputting, by the facial mesh tracker, a plurality of facial motion parameters that depict facial action movements detected, for animation and rendering of an avatar;

wherein the face is a face of a user, and detecting facial action movements of the face is through a normalized head pose of the user, comprises generating the normalized head pose of the user by using a 3D facial action model and a 3D neutral facial shape of the user pre-constructed using a 3D facial shape model.

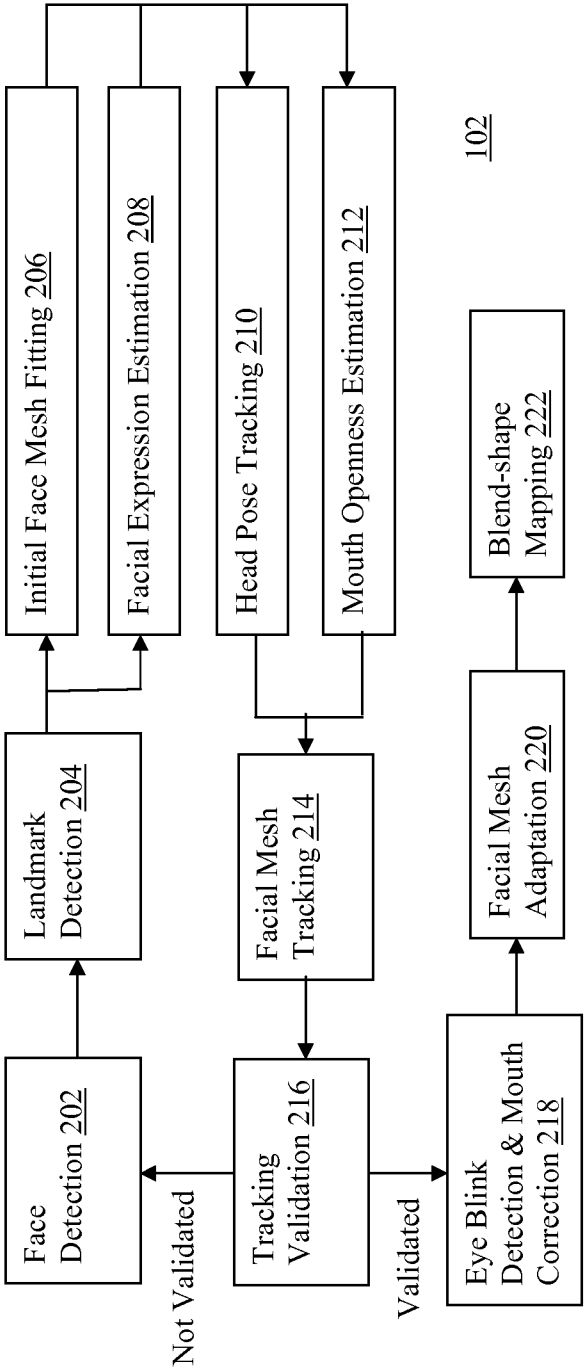
5           23.     The method of claim 22, wherein generating the normalized head pose of the user comprises minimizing differences between 2D projection of the 3D neutral facial shape and detected 2D image landmarks.

          24.     The method of claim 22, further comprises pre-developing offline the 3D facial action model and the 3D facial shape model, through machine learning of a 3D  
10   facial database.

          25.     The method of claim 22, further comprising pre-constructing the 3D neutral facial shape of the user using the 3D facial shape model, during registration of the user.



**Figure 1**



**Figure 2**



Figure 3

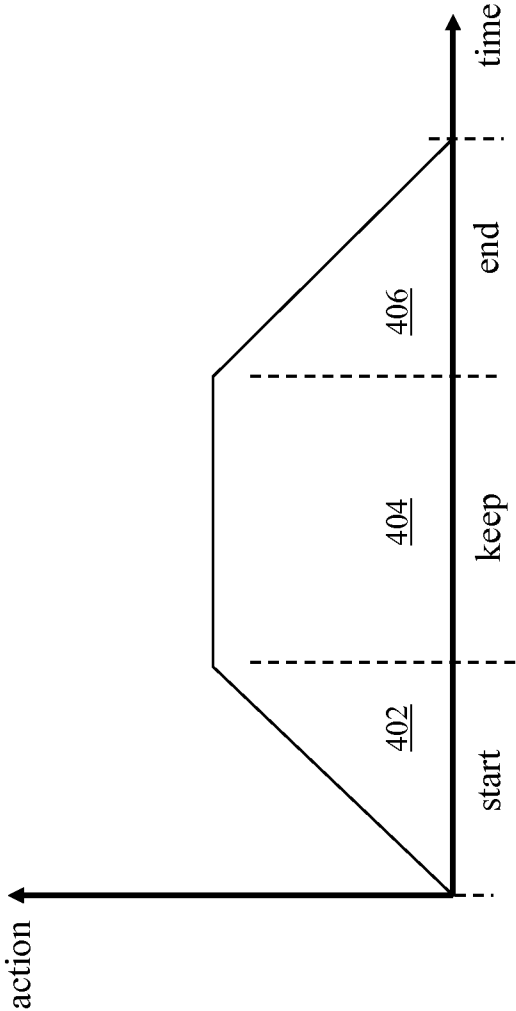


Figure 4

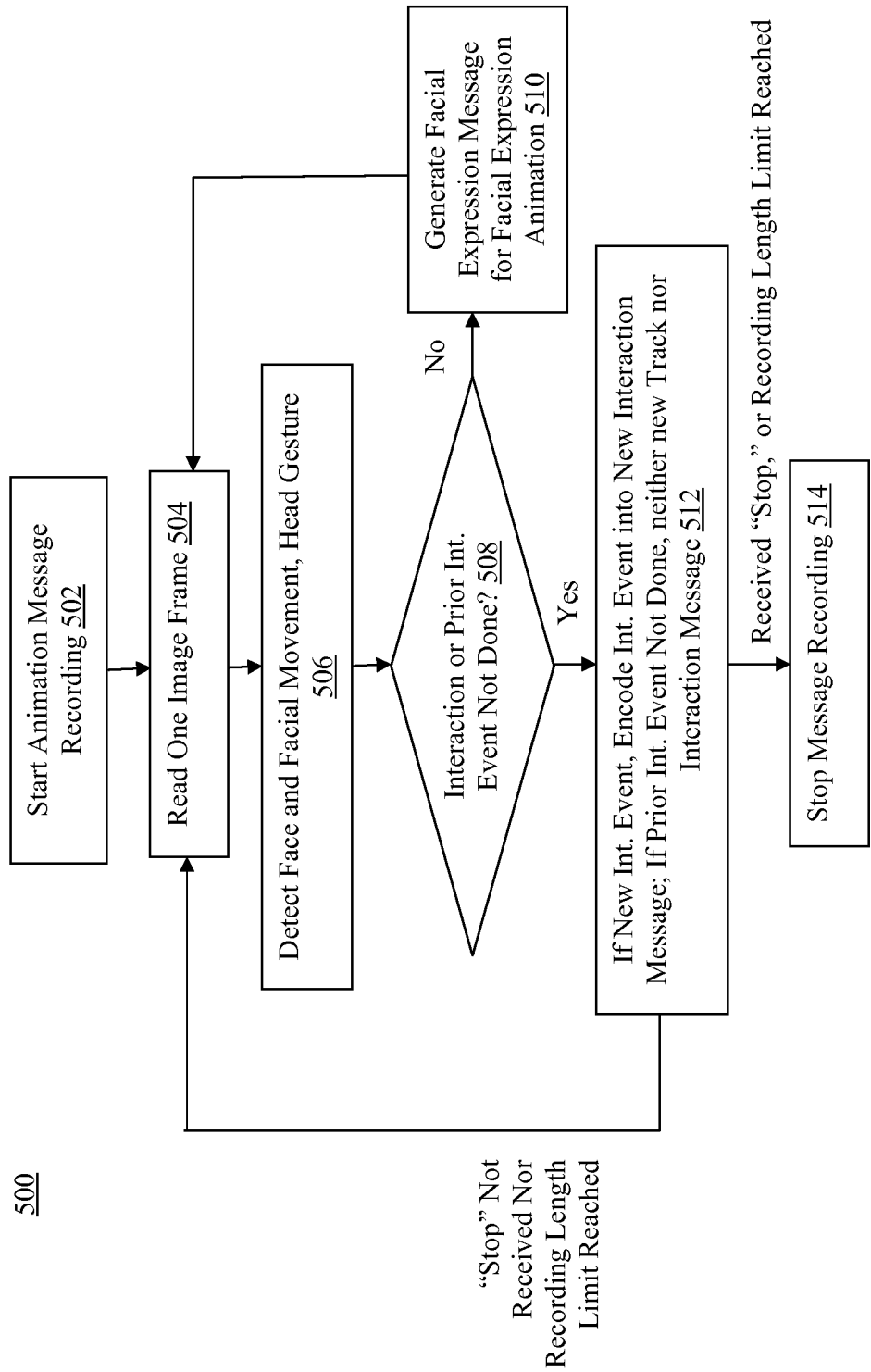


Figure 5

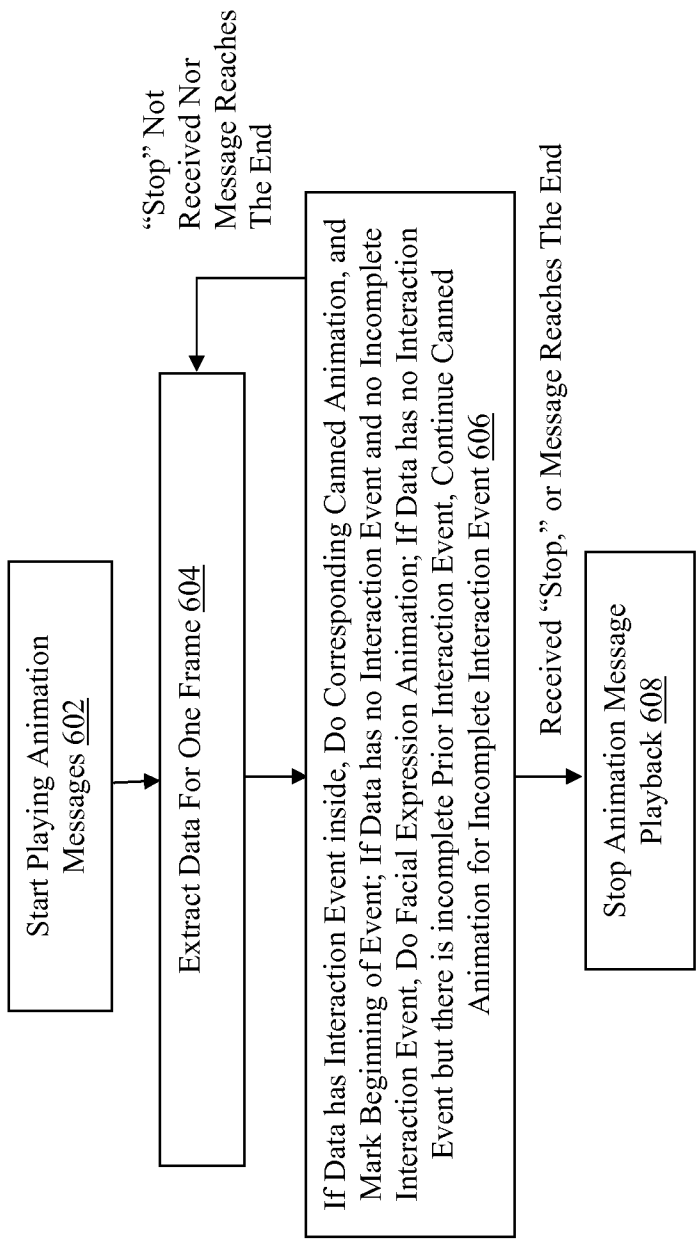


Figure 6

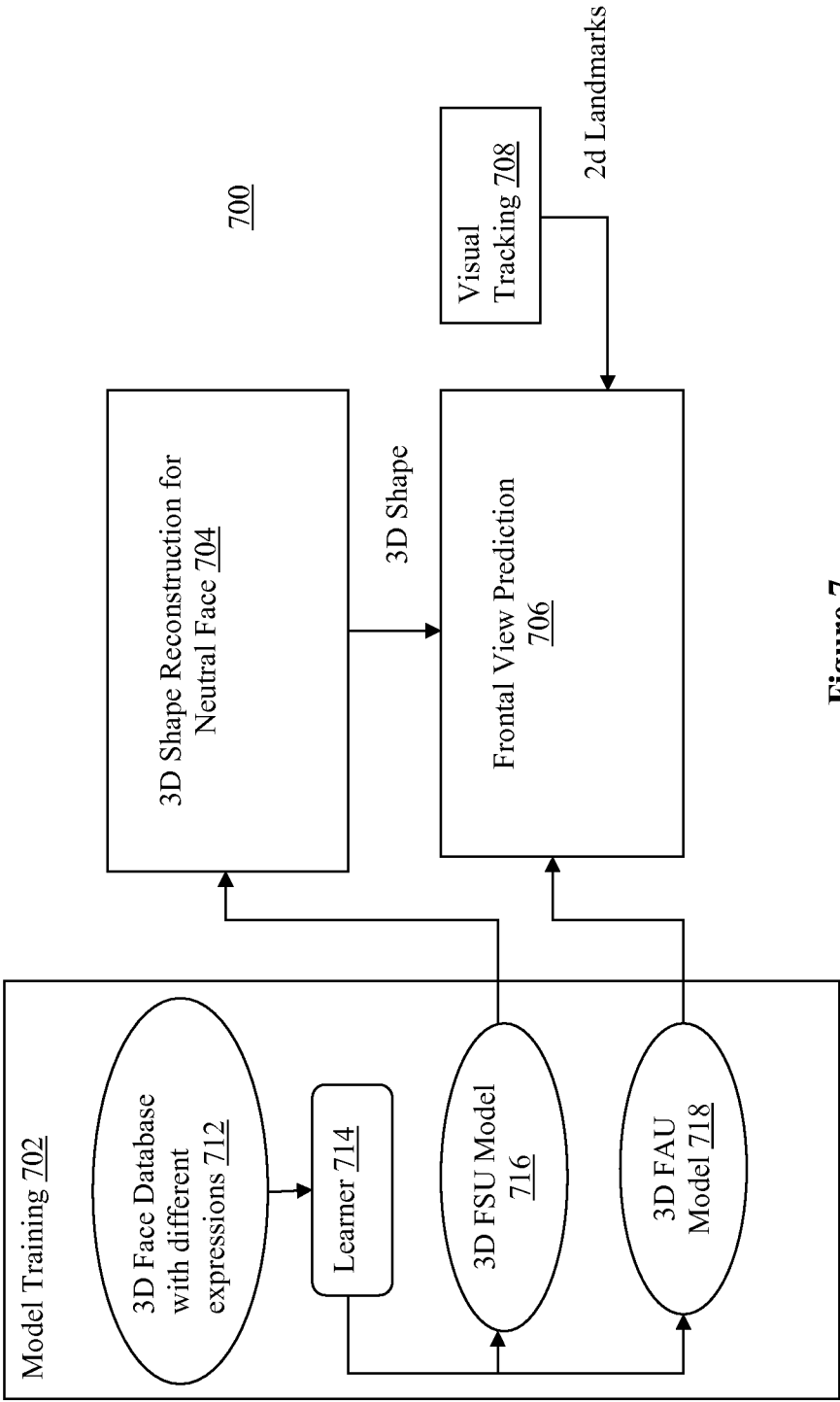


Figure 7

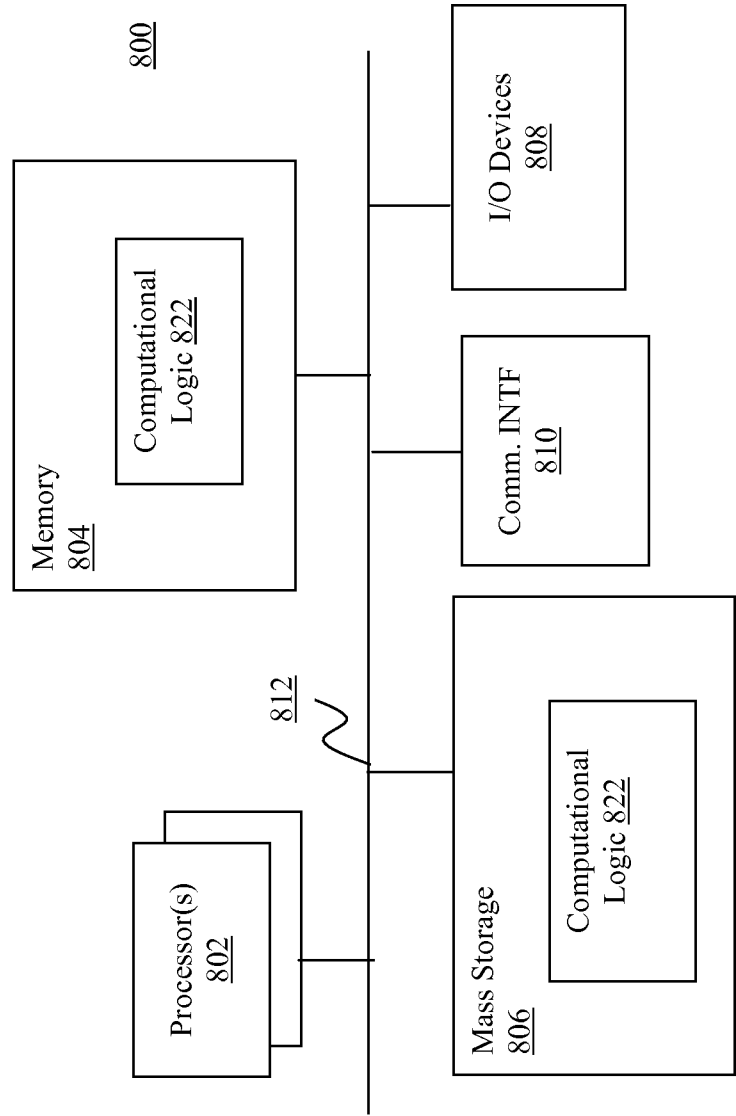
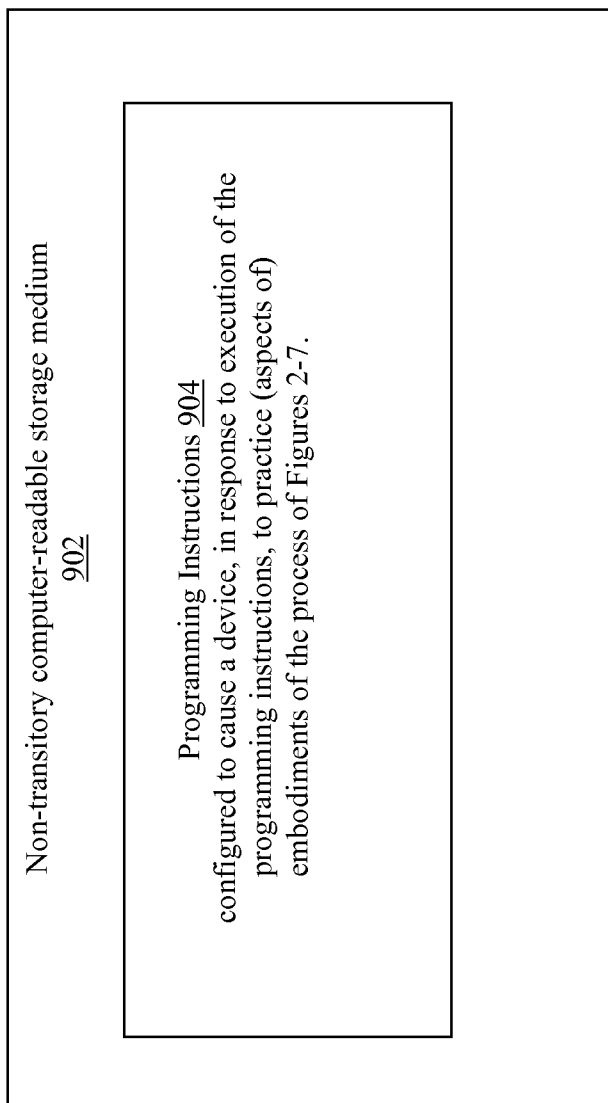


Figure 8



**Figure 9**

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/073695

**A. CLASSIFICATION OF SUBJECT MATTER**

G06T13/40 (2011.01)i

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)

G06T; G06F

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

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

WPI;EPODOC;PAJ;CNPAT;CNKI: avatar, pose, gesture, expression, face, mouth, eyes, rotation, movement, 3D, image, frame

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                    | Relevant to claim No. |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y         | CN 103093490 A (ZHEJIANG UNIVERSITY) 08 May 2013 (2013-05-08)<br>the abstract, claims 1-3 and 5, description, paragraphs [0016]-[0093], and figures 1-5               | 1-16, 22-25           |
| Y         | CN 1573660 A (MICROSOFT CORPORATION) 02 February 2005 (2005-02-02)<br>the abstract, description, pages 2-3                                                            | 1-16                  |
| Y         | CN 103473801 A (CHINA ACADEMY OF SCIENCES INSTITUTE OF AUTOMATION)<br>25 December 2013 (2013-12-25)<br>the abstract, description, paragraphs [0025]-[0026] and [0081] | 9、22-25               |
| X         | CN 102934144 A (MICROSOFT CORPORATION) 13 February 2013 (2013-02-13)<br>the abstract, description, paragraphs [0025]-[0034], and figures 1 and 6-7                    | 17-21                 |
| A         | CN 102934144 A (MICROSOFT CORPORATION) 13 February 2013 (2013-02-13)<br>the abstract, description, paragraphs [0025]-[0034], and figures 1 and 6-7                    | 1-16, 22-25           |
| A         | CN 103093490 A (ZHEJIANG UNIVERSITY) 08 May 2013 (2013-05-08)<br>the abstract, claims 1-3 and 5, description, paragraphs [0016]-[0093], and figures 1-5               | 17-21                 |

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

01 December 2014

Date of mailing of the international search report

22 December 2014

Name and mailing address of the ISA/CN

STATE INTELLECTUAL PROPERTY OFFICE OF THE  
P.R.CHINA(ISA/CN)  
6,Xitucheng Rd., Jimen Bridge, Haidian District, Beijing  
100088 China

Authorized officer

WANG,Jing

Facsimile No. (86-10)62019451

Telephone No. (86-10)62413685

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2014/073695**

| Patent document<br>cited in search report |           |   | Publication date<br>(day/month/year) | Patent family member(s) |             |    | Publication date<br>(day/month/year) |
|-------------------------------------------|-----------|---|--------------------------------------|-------------------------|-------------|----|--------------------------------------|
| CN                                        | 103093490 | A | 08 May 2013                          | WO                      | 2014117446  | A1 | 07 August 2014                       |
| CN                                        | 1573660   | A | 02 February 2005                     | CN                      | 100405261   | C  | 23 July 2008                         |
|                                           |           |   |                                      | BR                      | PI0401782   | A  | 18 January 2005                      |
|                                           |           |   |                                      | EP                      | 1484665     | A3 | 20 June 2012                         |
|                                           |           |   |                                      | US                      | 2011050568  | A1 | 03 March 2011                        |
|                                           |           |   |                                      | EP                      | 1484665     | A2 | 08 December 2004                     |
|                                           |           |   |                                      | JP                      | 4758073     | B2 | 24 August 2011                       |
|                                           |           |   |                                      | US                      | 2004240708  | A1 | 02 December 2004                     |
|                                           |           |   |                                      | US                      | 2008298637  | A1 | 04 December 2008                     |
|                                           |           |   |                                      | US                      | 2012139832  | A1 | 07 June 2012                         |
|                                           |           |   |                                      | JP                      | 2004362569  | A  | 24 December 2004                     |
|                                           |           |   |                                      | KR                      | 20040103403 | A  | 08 December 2004                     |
| CN                                        | 103473801 | A | 25 December 2013                     | Non                     | e           |    |                                      |
| CN                                        | 102934144 | A | 13 February 2013                     | KR                      | 20130080442 | A  | 12 July 2013                         |
|                                           |           |   |                                      | JP                      | 2013535051  | A  | 09 September 2013                    |
|                                           |           |   |                                      | EP                      | 2580741     | A2 | 17 April 2013                        |
|                                           |           |   |                                      | WO                      | 2011156115  | A2 | 15 December 2011                     |
|                                           |           |   |                                      | US                      | 2011304629  | A1 | 15 December 2011                     |
|                                           |           |   |                                      | HK                      | 1179736     | A0 | 04 October 2013                      |