

DESCRIPCIÓN

GENERACIÓN DE UNA REPRESENTACIÓN 3D DE UNA CABEZA DE UN PARTICIPANTE EN UNA SESIÓN DE COMUNICACIÓN POR VÍDEO

CAMPO TÉCNICO

5 La invención se refiere a un dispositivo informático para generar una representación tridimensional (3D) de una cabeza de un participante en una sesión de comunicación por vídeo, un procedimiento de generación de una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo, un correspondiente programa de ordenador, un
10 portador de datos legible por ordenador correspondiente, y una señal portadora de datos correspondiente.

ANTECEDENTES

Se conocen diferentes alternativas para generar y mostrar representaciones tridimensionales (3D) de cabezas humanas a un
15 espectador, por ejemplo, durante una sesión de comunicación por vídeo entre dos o más participantes.

Un primer tipo de soluciones se basa en avatares 3D generados por ordenador. Tales avatares se generan usando modelos de Aprendizaje Automático (ML) que se entrenan usando representaciones 3D capturadas de
20 cabezas humanas, y subsecuentemente se adaptan a una cabeza específica usando una imagen bidimensional (2D) que representa la cabeza. Durante el uso, la postura de cabeza de la cabeza del participante emisor se detecta

continuamente y se usa como entrada para el modelo de ML, que genera un avatar animado dinámicamente que refleja el movimiento real de la cabeza del participante emisor. En los últimos años se ha logrado una mejora considerable en los avatares 3D generados a través del uso de Redes
5 Generativas Adversarias (GANs), como se ha demostrado por H. Luo et al. ("Normalized Avatar Synthesis Using StyleGAN and Perceptual Refinement", arXiv:2106.11423, arXiv, 2021).

Un segundo tipo de soluciones se basa en capturar la cabeza del participante emisor usando sensores 3D, tales como cámaras estéreo, y
10 transmitir la representación 3D capturada en tiempo real para mostrar a un participante receptor, por ejemplo, como un flujo de nube de puntos o un flujo de malla. Las soluciones basadas en captura y transmisión en tiempo real son superiores en la representación de detalles de la cabeza capturada, en comparación con los avatares 3D animados. Estos detalles, en particular los
15 detalles del rostro capturado, son importantes para transmitir las emociones del participante emisor. Sin embargo, transmitir representaciones capturadas 3D de cabezas en tiempo real requiere anchos de banda considerablemente mayores de los enlaces de comunicación usados para transmitir los datos 3D capturados, por ejemplo, como una nube de puntos o flujo de malla. Además,
20 las soluciones actuales que se basan en capturar la cabeza del participante emisor en tiempo real típicamente requieren sensores 3D relativamente complejos, tales como varias cámaras, para garantizar que las partes

exteriores de la cabeza del participante emisor (fuera del rostro) se capturen por completo cuando se mueve la cabeza, lo cual aumenta la complejidad y costes.

SUMARIO

5 Un objeto de la invención es proporcionar una alternativa mejorada a las técnicas anteriores y a la técnica anterior.

Más específicamente, un objeto de la invención es proporcionar soluciones mejoradas para generar representaciones 3D de cabezas humanas basadas en capturar la cabeza usando un dispositivo sensor 3D.

10 Estos y otros objetos de la invención se logran por medio de diferentes aspectos de la invención, como se define mediante las reivindicaciones independientes. Realizaciones de la invención se caracterizan por las reivindicaciones dependientes.

De acuerdo con un primer aspecto de la invención, se proporciona un
15 dispositivo informático para generar una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo. El dispositivo informático comprende circuitería de procesamiento que hace que el dispositivo informático esté operativo para adquirir una representación 3D capturada de la cabeza, y para identificar posiciones de un conjunto de puntos
20 de referencia faciales en la representación 3D capturada. El conjunto de puntos de referencia faciales comprende puntos de referencia faciales que son indicativos de un límite del rostro humano. El dispositivo informático es

además operativo para determinar una pose de la cabeza, y para determinar un límite entre una parte interior y una parte exterior de la representación 3D capturada. El límite se determina en base a las posiciones identificadas del conjunto de puntos de referencia faciales. La parte interior de la
5 representación 3D capturada representa el rostro del participante. El dispositivo informático es además operativo para generar una representación de avatar que corresponde a la parte exterior de la representación 3D capturada. La representación de avatar se genera usando un modelo de ML que se entrena para cabezas humanas. La pose determinada de la cabeza se
10 usa como entrada para el modelo de ML.

De acuerdo con un segundo aspecto de la invención, se proporciona un procedimiento de generación de una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo. El procedimiento se realiza mediante un dispositivo informático y comprende adquirir una
15 representación 3D capturada de la cabeza, e identificar posiciones de un conjunto de puntos de referencia faciales en la representación 3D capturada. El conjunto de puntos de referencia faciales comprende puntos de referencia faciales que son indicativos de un límite del rostro humano. El procedimiento comprende además determinar una pose de la cabeza, y determinar un límite
20 entre una parte interior y una parte exterior de la representación 3D capturada. El límite se determina en base a las posiciones identificadas del conjunto de puntos de referencia faciales. La parte interior de la representación 3D

capturada representa el rostro del participante. El procedimiento comprende además generar una representación de avatar que corresponde a la parte exterior de la representación 3D capturada. La representación de avatar se genera usando un modelo de ML que se entrena para cabezas humanas. La

5 pose determinada de la cabeza se usa como entrada para el modelo de ML.

De acuerdo con un tercer aspecto de la invención, se proporciona un programa de ordenador. El programa de ordenador comprende instrucciones que, cuando el programa de ordenador se ejecuta mediante el dispositivo informático, hacen que el dispositivo informático lleve a cabo el procedimiento

10 de acuerdo con una realización del segundo aspecto de la invención.

De acuerdo con un cuarto aspecto de la invención, se proporciona un portador de datos legible por ordenador. El portador de datos legible por ordenador tiene almacenado en el mismo el programa de ordenador de acuerdo con el tercer aspecto de la invención.

15 De acuerdo con un quinto aspecto de la invención, se proporciona una señal portadora de datos. La señal portadora de datos porta el programa de ordenador de acuerdo con el tercer aspecto de la invención.

La invención hace uso de un entendimiento de que se puede generar una representación 3D de una cabeza de un participante en una sesión de

20 comunicación por vídeo en base a extraer una parte interior de la representación 3D capturada de la cabeza y haciendo que esa parte interior esté disponible para mostrar en un usuario receptor, por ejemplo, como un

flujo en tiempo real. La parte interior extraída corresponde a la región facial, o rostro de la cabeza, que generalmente incluye los ojos, nariz, oídos, y boca. El resto de la representación 3D capturada, denominada en este documento como la parte exterior, representa partes de la cabeza fuera del rostro. Esta
5 parte exterior se reemplaza por un avatar que se genera usando un modelo de ML entrenado para cabezas humanas, usando la pose de la cabeza capturada como entrada para el modelo de ML. Esto da como resultado un avatar animado que refleja la pose y movimiento reales de la cabeza capturada.

10 Realizaciones de la invención son ventajosas ya que un usuario receptor que está viendo la representación 3D generada de la cabeza capturada no sufre una representación 3D capturada de manera incompleta de la cabeza, lo cual puede producirse debido a limitaciones en los sensores 3D usados para capturar representaciones 3D. Al mismo tiempo, se
15 conservan la estructura detallada y movimientos del rostro de la cabeza capturada y sus partes, que son importantes para transmitir emociones en las comunicaciones interhumanas. De esa manera, se puede mejorar la experiencia de usuario. Realizaciones de la invención son además ventajosas ya que se reduce la cantidad de datos, que se capturan como la
20 representación 3D de la cabeza, que necesita transferirse al usuario receptor en tiempo real, por ejemplo, mediante generación de flujo. Este es el caso dado que solo es necesario que se transfiera un subconjunto de los datos 3D

capturados, la parte interior de la representación 3D capturada que representa el rostro de la cabeza capturada, desde el sensor 3D al dispositivo de visualización. De esa manera, se reduce el ancho de banda necesario para realizar una sesión de comunicación por vídeo 3D.

- 5 Incluso aunque en algunos casos se han descrito ventajas de la invención con referencia a realizaciones del primer aspecto de la invención, el razonamiento correspondiente se aplica a realizaciones de otros aspectos de la invención.

 Objetivos de, características de, y ventajas adicionales con, la
10 invención serán evidentes al estudiar la siguiente divulgación detallada, los dibujos y las reivindicaciones anexas. Los expertos en la técnica se dan cuenta de que se pueden combinar diferentes características de la invención para crear realizaciones distintas a las que se describen a continuación.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

- 15 Lo anterior, así como objetos, características y ventajas adicionales de la invención, se entenderán mejor a través de la siguiente descripción detallada ilustrativa y no limitante de realizaciones de la invención, con referencia a los dibujos anexas, en los cuales:

 La figura 1 ilustra una sesión de comunicación por vídeo entre dos
20 participantes, de acuerdo con realizaciones de la invención.

 La figura 2 ilustra una representación 3D capturada de una cabeza humana, de acuerdo con realizaciones de la invención.

La figura 3 ilustra la determinación de un límite entre una parte interior y una parte exterior de una representación 3D capturada de una cabeza, usando puntos de referencia faciales, de acuerdo con realizaciones de la invención.

5 La figura 4 ilustra esquemáticamente la generación de una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo, de acuerdo con realizaciones de la invención.

Las figuras 5A-5C muestran diagramas de secuencia que ilustran la generación de una representación 3D de una cabeza de un participante en
10 una sesión de comunicación por vídeo, de acuerdo con realizaciones de la invención.

La figura 6 ilustra esquemáticamente el dispositivo informático para generar una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo, de acuerdo con realizaciones de la
15 invención.

La figura 7 muestra un procedimiento de generación de una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo, de acuerdo con realizaciones de la invención.

Todas las figuras son esquemáticas, no necesariamente a escala, y
20 generalmente solo muestran partes que son necesarias con el fin de dilucidar la invención, en las que pueden omitirse o simplemente sugerirse otras partes.

DESCRIPCIÓN DETALLADA

La invención se describirá ahora más completamente de aquí en adelante con referencia a los dibujos acompañantes, en los cuales se muestran ciertas realizaciones de la invención. Sin embargo, esta invención puede realizarse de muchas formas diferentes y no debe considerarse como

5 limitada a las realizaciones establecidas en este documento. Más bien, estas realizaciones se proporcionan a modo de ejemplo de tal manera que esta divulgación será exhaustiva y completa, y transmitirá completamente el alcance de la invención a los expertos en la técnica.

La figura 1 ilustra una sesión de comunicación por vídeo entre dos

10 participantes 101 y 103, ejemplificada como una sesión de comunicación por vídeo unidireccional entre un dispositivo informático emisor 110 y un dispositivo informático receptor 130. Durante la sesión de comunicación por vídeo, que también puede denominarse como una sesión de comunicación por vídeo 3D, se captura una representación 3D de la cabeza 102 del

15 participante emisor 101 usando un sensor 3D 111, tal como una cámara estéreo, y se transmite (por ejemplo, genera en flujo) a través de una red de comunicaciones 140 al dispositivo informático receptor 130 para visualización al participante receptor 103, usando un dispositivo de visualización 131 tal como una pantalla de ordenador o una Pantalla Montada en la Cabeza (HMD).

20 Hay que anotar que realizaciones de la invención no se limitan a sesiones de comunicación por vídeo unidireccional entre dos participantes, como se ilustra en la figura 1. Más bien, se pueden prever realizaciones de la invención que

permitan sesiones de comunicación por vídeo unidireccional (por ejemplo, una presentación generada en flujo desde un presentador a muchos espectadores) o bidireccional (por ejemplo, una videollamada durante una reunión virtual) entre dos o más participantes. Más específicamente, en el

5 caso de una sesión de comunicación por vídeo bidireccional, realizaciones de la invención también soportan generar una representación 3D de una cabeza de otro participante, en la figura 1 la cabeza del participante 103, en la dirección inversa. Por tanto, un dispositivo informático que soporta sesiones de comunicación por vídeo bidireccional de acuerdo con realizaciones de la

10 invención comprende, o está conectado operativamente a, tanto un sensor 3D 111 como un dispositivo de visualización 131.

El dispositivo informático para generar una representación 3D de una cabeza 102 de un participante 101 en una sesión de comunicación por vídeo puede realizarse de diferentes formas, por ejemplo, como el dispositivo

15 informático emisor 110, como el dispositivo informático receptor 130, como un dispositivo informático de borde 120 que se proporciona en el borde de la red de comunicaciones 140 a través de la cual pasa el tráfico entre el dispositivo informático emisor 110 y el dispositivo informático receptor 130, o como una combinación de los mismos. El dispositivo informático de borde 120 puede,

20 por ejemplo, proporcionarse cerca de una Red de Acceso por Radio (RAN) que es parte de la red de comunicaciones 140, y a través de la cual el

dispositivo informático emisor 110 y/o el dispositivo informático receptor 130 se comunican entre sí y/o con el dispositivo informático de borde 120.

El dispositivo informático para generar una representación 3D de una cabeza 102 de un participante 101 en una sesión de comunicación por vídeo, en particular si se realiza como el dispositivo informático emisor 110 o como el dispositivo informático receptor 130, puede ser uno cualquiera de un teléfono inteligente, una tableta, un ordenador portátil, un dispositivo de Realidad Aumentada (AR), un dispositivo de Realidad Virtual (VR), un dispositivo de Realidad Mixta (MR), un dispositivo de Realidad eXtendida (XR), y una HMD. Alternativamente, el dispositivo informático para generar una representación 3D de una cabeza 102 de un participante 101 en una sesión de comunicación por vídeo, en particular si se realiza como el dispositivo informático de borde 120, puede ser uno cualquiera de un servidor de borde, un servidor de aplicaciones, y un ordenador en la nube. También se apreciará que el dispositivo informático para generar una representación 3D de una cabeza 102 puede realizarse de una manera distribuida. Es decir, las diferentes operaciones involucradas en la generación de una representación 3D de la cabeza 102, descritas con detalle adicional a continuación, pueden distribuirse entre más de uno del dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, y realizarse de una manera colaborativa. En las figuras 5A-5C se ilustran ejemplos ilustrativos para distribuir las diferentes operaciones entre el

dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, de una manera colaborativa y se dilucidan con más detalle además a continuación.

5 A lo largo de esta divulgación, se describen realizaciones de la invención en relación con la generación de una representación 3D de una cabeza humana 102, que incluye el rostro. El rostro humano incluye los ojos, nariz, oídos, y boca. En las comunicaciones interhumanas, la estructura detallada y los movimientos del rostro y sus partes son importantes para transmitir emociones, por ejemplo, entre los participantes 101 y 103.

10 A continuación, y con referencia a la figura 6, se describen con más detalle realizaciones 600 del dispositivo informático para generar una representación 3D de una cabeza 102 de un participante 101 en una sesión de comunicación por vídeo (por brevedad también denominado como el “dispositivo informático”). Se hace referencia adicional a la figura 4, que ilustra
15 esquemáticamente el flujo de datos al generar una representación 3D de la cabeza de un participante en una sesión de comunicación por vídeo.

El dispositivo informático 600 comprende circuitería de procesamiento 602 que hace que el dispositivo informático 600 esté operativo para adquirir una representación 3D capturada de la cabeza 102. Esto puede, por ejemplo,
20 lograrse capturando 502 la representación 3D de la cabeza usando el sensor 3D 111. Opcionalmente, si el dispositivo informático 600 está realizado como el dispositivo informático emisor 110, el dispositivo informático 110 puede

comprender el sensor 3D 111. Alternativamente, el dispositivo informático 110 puede estar conectado operativamente al sensor 3D 111. Por ejemplo, el sensor 3D 111 puede ser una unidad separada que está conectada al dispositivo informático 110 a través de un circuito de interfaz (“interfaz de E/S”
5 en la figura 6) usando cualquier tecnología cableada o inalámbrica conocida en la técnica, tal como un Bus Universal en Serie (USB), Iluminación, Interfaz Multimedia de Alta Definición (HDMI), Bluetooth, o similares. Alternativamente, si el dispositivo informático 600 está realizado como el dispositivo informático de borde 120 o como el dispositivo informático receptor
10 130, o una combinación de los mismos, el dispositivo informático 120/130 puede estar operativo para adquirir la representación 3D capturada de la cabeza 102 mediante recibir 512 la representación 3D capturada a través de la red de comunicaciones 140, por ejemplo, como un flujo de datos directamente (o indirectamente a través del dispositivo informático emisor
15 110) desde el sensor 3D 111, por ejemplo, usando el Protocolo de Tiempo Real (RTP), el Protocolo de Transporte Seguro en Tiempo Real (SRTP), o cualquier otro protocolo adecuado.

El sensor 3D 111 puede comprender una o más de una cámara 3D (también conocida como cámara estéreo), un sensor óptico 3D, un LiDAR, y
20 una cámara 2D. Se pueden usar sensores ópticos 3D para capturar y reconstruir la profundidad 3D de objetos del mundo real, tales como la cabeza 102. Dependiendo de la fuente de radiación usada, los sensores ópticos 3D

se pueden dividir en dos categorías, pasivos y activos. Los sensores estereoscópicos, sensores de Conformación a partir de Siluetas (SfS), y sensores de Conformación a partir de Textura (SfT), son ejemplos de sensores 3D pasivos, que no emiten ningún tipo de radiación por sí mismos.

- 5 Los sensores 3D recolectan imágenes de la escena, por ejemplo, la cabeza 102, opcionalmente desde diferentes puntos de vista o con diferentes configuraciones ópticas. Luego las imágenes se analizan para calcular la profundidad 3D de los puntos en la escena capturada, por ejemplo, puntos que representan la superficie de la cabeza 102 y sus partes. Por el contrario,
- 10 los sensores 3D activos emiten radiación, por ejemplo, ondas electromagnéticas tales como luz, y la interacción entre el objeto, tal como la cabeza 102, y la radiación es capturada por el sensor. A partir del análisis de los datos capturados, y en base a las propiedades de la radiación emitida, se obtienen las coordenadas de los puntos en la escena capturada, por ejemplo,
- 15 puntos que representan la superficie de la cabeza 102 y sus partes. Los sensores de Tiempo de Vuelo (ToF), sensores de desplazamiento de fase, y sensores de triangulación activa, son ejemplos de sensores 3D activos. La salida de un sensor óptico 3D típicamente es una imagen de mapa de profundidad.

- 20 Se puede usar LiDAR (Detección Y Alcance de Luz) para medir distancias (también conocido como “alcance”) iluminando el objetivo, tal como la cabeza 102, con luz y luego midiendo los reflejos con un sensor óptico. Los

sensores LiDAR pueden operar en el espectro ultravioleta, visible, o infrarrojo. Dado que la luz láser, que se usa típicamente, está colimada, el sensor LiDAR necesita escanear la escena con el fin de generar una imagen con un campo de visión deseado. La salida de un sensor LiDAR es típicamente una nube de puntos que subsecuentemente puede enriquecerse con otros datos de sensor, tales como datos RGB desde una cámara convencional (2D) que puede estar comprendida en el sensor 3D 111.

La circuitería de procesamiento 602 hace además que el dispositivo informático 600 esté operativo para identificar 503 posiciones de un conjunto de puntos de referencia faciales (también conocidos como “puntos clave faciales” o simplemente “puntos clave”) en la representación 3D capturada. El conjunto de puntos de referencia faciales comprende puntos de referencia faciales que son indicativos de un límite del rostro humano. En la técnica se usan diferentes conjuntos de puntos de referencia faciales. Como ejemplo, “Fast Facial Landmark Detection and Applications: A Survey” (por K. S. Khabarlak y L. S. Koriashkina, arXiv:2101.10808v2, arXiv, 2021) enumera diferentes conjuntos que comprenden entre 21 y 98 puntos de referencia faciales. Típicamente, un subconjunto de puntos de referencia faciales de cualquier conjunto determinado son indicativos de un límite del rostro humano. En la figura 3, se reproduce un conjunto de ejemplo de puntos de referencia faciales 1 hasta 27 que son indicativos de un límite del rostro humano, como se describe en “Facial Landmarks for Face Recognition with Dlib”

(<https://sefiks.com/2020/11/20/facial-landmarks-for-face-recognition-with-dlib/>, recuperado el 2022-03-11). Con propósitos ilustrativos, los puntos de referencia faciales 1 a 27 se superponen sobre un boceto de una representación 3D capturada 300 de la cabeza 102 en posiciones
 5 representativas.

Usando un modelo de ML tal como una red neuronal, se puede detectar el conjunto de puntos de referencia faciales en una imagen (2D) de un rostro y se pueden determinar las posiciones (3D) de los puntos de referencia faciales. Esto se puede lograr usando un algoritmo conocido de detección de
 10 puntos de referencia faciales, por ejemplo, como se describe en “Fast Facial Landmark Detection and Applications: A Survey”, usando la biblioteca Dlib (véase “Facial Landmarks for Face Recognition with Dlib”) o la biblioteca OpenCV (véase, por ejemplo, “Head Pose Estimation using Python”,
[https://towardsdatascience.com/head-pose-estimation-using-python-](https://towardsdatascience.com/head-pose-estimation-using-python-d165d3541600)
 15 [d165d3541600](https://towardsdatascience.com/head-pose-estimation-using-python-d165d3541600), recuperado el 2022-03-11).

La circuitería de procesamiento 602 hace además que el dispositivo informático 600 esté operativo para determinar 504 una pose de la cabeza 102 (también denominada como “pose de cabeza”). La pose determinada de la cabeza 102 puede expresarse en términos de ángulos de Euler, por
 20 ejemplo, cabeceo, guiñada, y balanceo, pero realizaciones de la invención también pueden depender de conjuntos de ángulos alternativos. La pose de la cabeza 102 puede, por ejemplo, determinarse 504 usando un enfoque

similar al descrito anteriormente en relación con identificar 503 posiciones del conjunto de puntos de referencia faciales. Por ejemplo, la pose de cabeza se puede determinar en base a puntos de referencia faciales usando la biblioteca OpenCV (véase “Head Pose Estimation using Python”). Este ejemplo ilustra

5 cómo se puede determinar la pose de cabeza usando solamente seis puntos de referencia faciales, identificando el borde de los ojos, la nariz, el mentón, y el borde de la boca. El conjunto de puntos de referencia faciales que se usa para determinar 504 la pose de cabeza puede ser diferente del conjunto de puntos de referencia faciales que son indicativos de un límite del rostro

10 humano. Alternativamente, el conjunto de puntos de referencia faciales que son indicativos de un límite del rostro humano puede comprender puntos de referencia que son indicativos de una pose de la cabeza humana. Se conocen en la técnica diferentes técnicas para determinar la pose de cabeza a partir de una imagen (2D) de una cabeza, véase, por ejemplo, “Fast Facial

15 Landmark Detection and Applications: A Survey”.

La circuitería de procesamiento 602 hace además que el dispositivo informático 600 esté operativo para determinar 505 un límite entre una parte interior y una parte exterior de la representación 3D capturada. La parte interior de la representación 3D capturada representa el rostro del participante

20 101, y también puede denominarse como la parte facial de la representación 3D capturada. El límite se determina 505 en base a las posiciones identificadas 503 del conjunto de puntos de referencia faciales, en particular

los puntos de referencia faciales que son indicativos de un límite del rostro humano. En la práctica, el límite entre la parte interior y la parte exterior de la representación 3D capturada se puede determinar 505 ajustando una conformación 2D, tal como una conformación ovalada, a las posiciones

5 identificadas del conjunto de puntos de referencia faciales que son indicativos de un límite del rostro humano. Como ejemplo, en la figura 3 se ilustra una conformación ovalada 310 que se ajusta al conjunto de puntos de referencia faciales 1 a 27. El límite 310 separa la parte interior (facial) 320 de la parte exterior 330 de la representación 3D capturada 300.

10 Como alternativa a ajustar una conformación 2D a las posiciones identificadas del conjunto de puntos de referencia faciales que son indicativos de un límite del rostro humano, el límite entre la parte interior y la parte exterior de la representación 3D capturada puede determinarse 505 ajustando primero una conformación 3D tal como un ovoide o un elipsoide a la representación

15 3D capturada de la cabeza 102. Subsecuentemente, las posiciones identificadas del conjunto de puntos de referencia faciales que son indicativos de un límite del rostro humano se proyectan sobre la conformación 3D ajustada, ya sea usando superficie normal o proyecciones al origen o un sistema de coordenadas usado para la representación 3D capturada, que

20 ventajosamente está cerca del centro de la cabeza 102. Las posiciones proyectadas de los puntos de referencia faciales son puntos en la superficie de la conformación 3D ajustada. Luego, se ajusta una conformación 2D, tal

como una conformación ovalada, a los puntos en la superficie de la conformación 3D ajustada.

Se apreciará que realizaciones de la invención no se limitan a usar conformaciones ovaladas para determinar 505 el límite entre la parte interior y la parte exterior de la representación 3D capturada. En particular, se pueden usar elipses o círculos, que son casos especiales de una conformación ovalada. Realizaciones de la invención también pueden depender de conformaciones estriadas.

La circuitería de procesamiento 602 provoca además que el dispositivo informático 600 esté operativo para generar 507 una representación de avatar que corresponde a la parte exterior 330 de la representación 3D capturada 300. La parte exterior 330 de la representación 3D capturada 300 está definida por el límite 310 determinado 505 entre la parte interior 320 y la parte exterior 330 de la representación 3D capturada 300. En la práctica, dado que el límite 310 está representado por una conformación 2D, tal como un óvalo, la representación de avatar que corresponde a la parte exterior 330 de la representación 3D capturada 300 se genera para partes de la cabeza 102 fuera de la conformación 2D que representa el límite 310 entre la parte interior 320 y la parte exterior 330 de la representación 3D capturada 300. Si la representación de avatar generada es una nube de puntos, los puntos generados están fuera de la conformación 2D que representa el límite 310.

La representación de avatar se genera 507 usando un modelo de ML que ha sido entrenado para cabezas humanas. La pose determinada 504 de la cabeza se usa como entrada para el modelo de ML. La representación de avatar generada 507 que corresponde a la parte exterior 330 de la representación 3D capturada 300 es una representación animada de la parte exterior de la cabeza 102, es decir, las partes de la cabeza 102 que están fuera del rostro, como se define por el límite 310 entre la parte interior 320 y la parte exterior 330 de la representación 3D capturada 300. Como ejemplo, la representación de avatar que corresponde a la parte exterior 330 de la representación 3D capturada 300 puede generarse usando una GAN, como se describe en "Normalized Avatar Synthesis Using StyleGAN and Perceptual Refinement". El modelo de ML puede ser un modelo de ML genérico que ha sido entrenado para cabezas humanas en general. Alternativamente, el modelo de ML puede ser un modelo de ML específico que haya sido entrenado para uno o más tipos específicos de cabezas humanas, incluyendo los tipos un género, una edad o rango de edad, color de piel, tipo de cabello, etc. Se apreciará que se pueden prever realizaciones de la invención para generar una representación 3D de la cabeza de un animal.

La circuitería de procesamiento 602 opcionalmente hace además que el dispositivo informático 600 esté operativo para extraer 506 la parte interior 320 de la representación 3D capturada 300, y para fusionar 508 la parte interior 320 extraída 506 de la representación 3D capturada 300 y la

representación de avatar generada 507 en una representación 3D fusionada de la cabeza 102. En otras palabras, la parte interior 320 de la representación 3D capturada 300, que es el subconjunto de los datos capturados por el sensor 3D 111 que representa el rostro de la cabeza 102, es decir, está dentro

5 del límite 310 determinado 505 entre la parte interior 320 y la parte exterior 330 (cuyo límite 310 está representado por una conformación 2D, tal como un óvalo), se fusiona 508 con la representación de avatar generada 507 que corresponde a la parte exterior 330 de la representación 3D capturada 300. En la práctica, esto equivale a reemplazar la parte exterior 330 de la

10 representación 3D capturada 300 con la representación de avatar generada, es decir, una representación animada (generada por ordenador) de la parte exterior de la cabeza 102, mientras que se mantiene la parte interior 320 de la representación 3D capturada 300, es decir, la representación en tiempo real capturada del rostro de la cabeza 102. Más específicamente, si la

15 representación 3D capturada y la representación de avatar generada son nubes de puntos, extraer la parte interior 320 de la representación 3D capturada 300 equivale a seleccionar puntos desde la nube de puntos que representa la representación 3D capturada 300 que tienen coordenadas que están dentro del límite determinado 310 entre la parte interior 320 y la parte

20 exterior 330 de la representación 3D capturada 300, es decir, puntos que están dentro la conformación 2D que representa el límite 310. De manera correspondiente, los puntos de la nube de puntos que representan la

representación de avatar generada tienen coordenadas que están fuera del límite determinado 310 entre la parte interior 320 y la parte exterior 330 de la representación 3D capturada 300, es decir, estas son puntos que están fuera de la conformación 2D que representa el límite 310. De manera similar, la

5 fusión 508 de la parte interior extraída 320 de la representación 3D capturada 300 y la representación de avatar generada en una representación 3D fusionada de la cabeza equivale a combinar los diferentes conjuntos de puntos (representados por nubes de puntos separadas) en una única nube de puntos. Si la representación 3D capturada y la representación de avatar generada se

10 representan en un formato diferente que las nubes de puntos, tales como mallas e imágenes de mapas de profundidad, opcionalmente se pueden convertir en nubes de puntos antes de extraer 506 la parte interior 320 de la representación 3D capturada 300 y fusionar 508 la parte interior 320 extraída 506 de la representación 3D capturada 300 y la representación de avatar 507

15 generada en una representación 3D fusionada de la cabeza 102. Alternativamente, extraer 506 la parte interior 320 de la representación 3D capturada 300 y fusionar 508 la parte interior 320 extraída 506 de la representación 3D capturada 300 y la representación de avatar generada 507 en una representación 3D fusionada de la cabeza 102 se puede realizar en el

20 formato nativo de la representación 3D capturada y la representación de avatar generada, sin convertir los datos en nubes de puntos.

La circuitería de procesamiento 602 opcionalmente hace además que el dispositivo informático 600 esté operativo para mostrar 509 la representación 3D fusionada de la cabeza 102 usando el dispositivo de visualización 131. Opcionalmente, el dispositivo de visualización 131 puede estar comprendido en el dispositivo informático 600, si el dispositivo informático está realizado como el dispositivo informático receptor 130. El dispositivo de visualización 131 puede ser uno cualquiera de una pantalla de ordenador, un televisor, un dispositivo de AR, un dispositivo de VR, un dispositivo de MR, un dispositivo de XR, y un dispositivo de HMD.

10 Al reemplazar la parte exterior 330 de la representación 3D capturada 300 de la cabeza 102, que necesita transmitirse en tiempo real desde el dispositivo informático emisor 110 al dispositivo informático receptor 130 para representación en el dispositivo de visualización 131, con una representación de avatar generada, se puede reducir la cantidad de datos capturados que necesitan transmitirse a través de la red de comunicaciones 140 en tiempo real.

Una ventaja adicional surge del hecho de que la representación 3D capturada de la cabeza 102 puede ser incompleta, en particular dentro de la parte exterior 320 de la cabeza 102, es decir, fuera de la región facial. Esto puede suceder cuando la cabeza 102 se mueve, y debido a limitaciones en el campo de visión del sensor 3D 111. Tal situación se ilustra en la figura 2, que muestra los parches faltantes 211 y 221 en la representación 3D capturada

en diferentes poses 210 y 220 de la cabeza 102 en relación con el sensor 3D 111. Al reemplazar la parte exterior 330 de la representación 3D capturada 300 de la cabeza 102 con una representación de avatar generada, la representación 3D fusionada que se muestra usando el dispositivo de visualización 131 no sufre la falta de datos capturados dentro de la parte exterior 330 de la representación 3D capturada 300. De esa manera, se mejora la experiencia de usuario del participante que ve 103, ya que es menos probable que la representación 3D fusionada mostrada de la cabeza 102 sufra la falta de datos capturados.

10 Opcionalmente, el modelo de ML, que se usa para generar 507 una representación de avatar que corresponde a la parte exterior 330 de la representación 3D capturada 300, se entrena 510 para la cabeza 102 del participante 101. En otras palabras, el modelo de ML se entrena específicamente 510 para la cabeza 102 que se captura durante la sesión de comunicación por vídeo. De esa manera, se mejora la experiencia de usuario durante la sesión de comunicación por vídeo, en particular la del participante receptor 103 que ve la representación 3D fusionada de la cabeza 102.

La circuitería de procesamiento 602 opcionalmente hace además que el dispositivo informático 600 esté operativo para adquirir el modelo de ML desde un almacenamiento de datos asociado con el participante 101. Preferentemente, el modelo de ML adquirido se entrena para la cabeza 102 del participante 101, también denominado en este documento como un

“modelo de ML específico de participante”. Por ejemplo, el modelo de ML específico de participante puede almacenarse en un dispositivo de usuario asociado con el participante 101, tal como el dispositivo informático emisor 110 o un almacenamiento en la nube. Cuando el participante 101 inicia o se une a una sesión de comunicación por vídeo, el modelo de ML específico de participante puede ser recuperado 511/522 por el dispositivo informático 600 y usado para generar 507 la representación de avatar que corresponde a la parte exterior de la representación 3D capturada. Por ejemplo, el modelo de ML específico de participante puede transmitirse 511/522 desde el dispositivo informático emisor 110, que es un dispositivo personal usado por el participante emisor 101, al dispositivo informático 600 realizado como el dispositivo informático de borde 120 y/o como el dispositivo informático receptor 130. Alternativamente, el dispositivo informático 600 puede recuperar, es decir, solicitar y recibir, el modelo de ML específico de participante desde un almacenamiento en la nube que está asociado con el participante 101 y accesible mediante el dispositivo informático 600 (no se ilustra en las figuras 5A–5C). Este último puede, por ejemplo, ser el caso si el modelo de ML específico de participante se almacena en un almacenamiento en la nube (tal como iCloud, One Drive, etc.) y está asociado con un identificador de usuario del participante 101 (tal como ID de Apple, dirección de correo electrónico del participante 101, etc).

La circuitería de procesamiento 602 opcionalmente hace además que el dispositivo informático 600 esté operativo para entrenar 510 el modelo de ML usando al menos la parte exterior 330 de la representación 3D capturada 300 y la pose determinada 504 de la cabeza 102. Es decir, la parte exterior

5 330 de la representación 3D capturada 300 se extrae, por ejemplo, simultáneamente con extraer 506 la parte interior 320 de la representación 3D capturada 300, y se usa como entrada para entrenar 510 el modelo de ML, junto con la pose determinada 504 de la cabeza 102. Opcionalmente, el dispositivo informático 600 puede estar operativo para entrenar 510 el modelo

10 de ML en base además a la parte interior 320 de la representación 3D capturada 300, es decir, usando la representación 3D capturada sustancialmente completa 300 de la cabeza 102. Esto es ventajoso ya que el modelo de ML que se usa para generar 507 la representación de avatar que corresponde a la parte exterior 320 de la representación 3D capturada 300

15 puede entrenarse 510 para la cabeza específica 102 del participante 101 mientras comienza la sesión de comunicación por vídeo. El modelo de ML puede, por ejemplo, ser un modelo de ML genérico que se entrena para cabezas humanas en general. Alternativamente, el modelo de ML puede ser un modelo de ML específico que esté entrenado para ciertos tipos de cabezas

20 humanas, como se describe anteriormente en la presente. Como aún otra alternativa, el modelo de ML puede ser un modelo de ML específico de participante que haya sido entrenado durante sesiones de comunicación por

vídeo previas, o durante un procedimiento de entrenamiento dedicado, y almacenado para uso posterior en un almacenamiento de datos asociado con el participante 101, por ejemplo, un almacenamiento de datos comprendido en el dispositivo informático emisor 110 o un almacenamiento en la nube.

5 La representación 3D capturada, la parte interior de la representación 3D capturada, la representación 3D fusionada, y la representación de avatar, pueden almacenarse, y transmitirse entre el dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, a través de la red de comunicaciones 140, usando cualquier
10 formato de datos adecuado, en particular formatos de medios inmersivos 3D, y/o protocolos. Más específicamente, la representación 3D capturada, las partes interior y exterior de la representación 3D capturada, la representación 3D fusionada, y la representación de avatar, pueden almacenarse y transmitirse como nubes de puntos, mallas, o imágenes de mapas de
15 profundidad. Una nube de puntos es un conjunto de puntos de datos en el espacio, representando los puntos un objeto 3D tal como la cabeza 102. Una malla, también denominada como malla poligonal, es una colección de vértices, aristas y caras que define la conformación de un objeto 3D tal como la cabeza 102. Una imagen de mapa de profundidad contiene información
20 relacionada con la distancia de las superficies de un objeto 3D, tal como la cabeza 102, desde un punto de vista, en particular el del sensor 3D 111. Protocolos usados para transmitir la representación 3D capturada, la parte

interior de la representación 3D capturada, la representación 3D fusionada, y la representación de avatar, entre, el dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, a través de la red de comunicaciones 140, incluyen, pero no se limitan a, RTP, 5 SRTP, Generación de Flujo Dinámico Adaptativo sobre HTTP (DASH), etc.

A continuación, y con referencia a las figuras 5A-5C, se ilustran diferentes realizaciones de la invención, con enfoque particular en cuál del dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, se pueden realizar las operaciones 10 involucradas en la generación de una representación 3D de una cabeza de un humano.

La realización ilustrada en la figura 5A se caracteriza por un procesamiento centrado en bordes. Esto es ventajoso ya que el dispositivo informático de borde 120 típicamente tiene una abundancia de recursos 15 informáticos, en términos de energía informática, memoria, y suministro de energía eléctrica, en comparación con el dispositivo informático emisor 110 y el dispositivo informático receptor 130, que pueden estar realizados como teléfonos inteligentes, tabletas, HMDs, u otros tipos de dispositivos informáticos móviles que a menudo son alimentados por batería y son menos 20 potentes en términos de energía de procesamiento.

Más específicamente, si el dispositivo informático 600 para generar una representación 3D de una cabeza 102 de un participante 101 en una sesión

de comunicación por vídeo está realizado como el dispositivo informático de borde 120, el dispositivo informático de borde 120 está operativo para adquirir una representación 3D capturada de la cabeza al recibir 512 la representación 3D capturada de la cabeza 102 desde el dispositivo informático emisor 110.

5 El dispositivo informático de borde 120 es además operativo para identificar 503 posiciones de un conjunto de puntos de referencia faciales en la representación 3D capturada, cuyo conjunto de puntos de referencia faciales comprende puntos de referencia faciales indicativos de un límite del rostro humano. El dispositivo informático de borde 120 es además operativo para
10 determinar 504 una pose de la cabeza 102. El dispositivo informático de borde 120 es además operativo para determinar 505 un límite 310 entre una parte interior 320 y una parte exterior 330 de la representación 3D capturada 300, en base a las posiciones identificadas 503 del conjunto de puntos de referencia faciales. La parte interior 320 de la representación 3D capturada
15 300 representa el rostro del participante 101. El dispositivo informático de borde 120 es además operativo para generar 507 una representación de avatar que corresponde a la parte exterior 330 de la representación 3D capturada 300, usando un modelo de ML entrenado para cabezas humanas, con la pose determinada 504 de la cabeza 102 como entrada.

20 Opcionalmente, el dispositivo informático de borde 120 puede estar operativo además para extraer 506 la parte interior 320 de la representación 3D capturada 300, y para fusionar 508 la parte interior extraída 320 de la

representación 3D capturada 300 y la representación de avatar generada en una representación 3D fusionada de la cabeza 102.

El dispositivo informático de borde 120 puede además estar operativo para transmitir 521 la representación 3D fusionada de la cabeza 102 al
5 dispositivo informático receptor 130, donde se muestra 509 usando el dispositivo de visualización 131.

Opcionalmente, el dispositivo informático de borde 120 puede estar operativo para entrenar 510 el modelo de ML usando al menos la parte exterior 330 de la representación 3D capturada 300 y la pose determinada 504 de la
10 cabeza. Además opcionalmente, el dispositivo informático de borde 120 puede estar operativo para entrenar 510 el modelo de ML en base además a la parte interior 320 de la representación 3D capturada 300.

En comparación con la realización centrada en borde ilustrada en la figura 5A, en la realización ilustrada en la figura 5B y descrita a continuación,
15 parte del procesamiento se ha movido desde el dispositivo informático de borde 120 al dispositivo informático receptor 130. En este caso, el dispositivo informático de borde 120 y el dispositivo informático receptor 130 en combinación implementan realizaciones de la invención. En otras palabras, la invención se realiza como un sistema de dispositivos informáticos para
20 generar una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo.

Más específicamente, el dispositivo informático de borde 120 está operativo para adquirir una representación 3D capturada de la cabeza al recibir 512 la representación 3D capturada de la cabeza 102 desde el dispositivo informático emisor 110. El dispositivo informático de borde 120 es

5 además operativo para identificar 503 posiciones de un conjunto de puntos de referencia faciales en la representación 3D capturada, cuyo conjunto de puntos de referencia faciales comprende puntos de referencia faciales indicativos de un límite del rostro humano. El dispositivo informático de borde 120 es además operativo para determinar 504 una pose de la cabeza 102, y

10 para transmitir 523 la pose determinada de la cabeza 102 al dispositivo informático receptor 130. El dispositivo informático de borde 120 es además operativo para determinar 505 un límite 310 entre una parte interior 320 y una parte exterior 330 de la representación 3D capturada 300, en base a las posiciones identificadas 503 del conjunto de puntos de referencia faciales. La

15 parte interior 320 de la representación 3D capturada 300 representa el rostro del participante 101. El dispositivo informático de borde 120 es opcionalmente además operativo para extraer 506 la parte interior 320 de la representación 3D capturada 300, y para transmitir 524 la parte interior extraída 320 de la representación 3D capturada 300 al dispositivo informático receptor 130.

20 El dispositivo informático receptor 130 está operativo para generar 507 una representación de avatar que corresponde a la parte exterior 330 de la representación 3D capturada 300, usando un modelo de ML entrenado para

cabezas humanas, con la pose determinada 504 de la cabeza 102 que el dispositivo informático receptor 130 ha recibido 523 como entrada. Opcionalmente, el dispositivo informático receptor 130 está operativo para fusionar 508 la parte interior 320 recibida 524 de la representación 3D capturada 300 y la representación de avatar generada 507 en una representación 3D fusionada de la cabeza 102. El dispositivo informático receptor 130 puede estar operativo además para visualizar 509 la representación 3D fusionada de la cabeza 102 usando el dispositivo de visualización 131.

10 Opcionalmente, el dispositivo informático de borde 120 puede estar operativo además para entrenar 510 el modelo de ML usando al menos la parte exterior 330 de la representación 3D capturada 300 y la pose determinada 504 de la cabeza. Además opcionalmente, el dispositivo informático de borde 120 puede estar operativo para entrenar 510 el modelo de ML en base además a la parte interior 320 de la representación 3D capturada 300, es decir, usando la representación 3D capturada sustancialmente completa de la cabeza 102. En este caso, el dispositivo informático de borde 120 está operativo para transmitir 525 el modelo de ML actualizado al dispositivo informático receptor 130.

20 En la figura 5C se ilustra una realización adicional de un sistema de dispositivos informáticos para generar una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo. En comparación

con la realización ilustrada en la figura 5B, las operaciones adicionales involucradas en la generación de una representación 3D de la cabeza 102 se han movido desde el dispositivo informático de borde 120 al dispositivo informático receptor 130.

5 Más específicamente, el dispositivo informático de borde 120 está operativo para adquirir una representación 3D capturada de la cabeza al recibir 512 la representación 3D capturada de la cabeza 102 desde el dispositivo informático emisor 110. El dispositivo informático de borde 120 es además operativo para identificar 503 posiciones de un conjunto de puntos de
10 referencia faciales en la representación 3D capturada, cuyo conjunto de puntos de referencia faciales comprende puntos de referencia faciales indicativos de un límite del rostro humano. El dispositivo informático de borde 120 es además operativo para determinar 504 una pose de la cabeza 102, y para transmitir 523 la pose determinada de la cabeza 102 al dispositivo
15 informático receptor 130. El dispositivo informático de borde 120 es además operativo para determinar 505 un límite 310 entre una parte interior 320 y una parte exterior 330 de la representación 3D capturada 300, en base a las posiciones identificadas 503 del conjunto de puntos de referencia faciales, y para transmitir 527 el límite determinado entre la parte interior y la parte
20 exterior al dispositivo informático receptor 130. La parte interior de la representación 3D capturada representa el rostro del participante 101.

El dispositivo informático receptor 120 es opcionalmente operativo para extraer 506 la parte interior 320 de la representación 3D capturada 300, usando el límite 310 recibido 527 entre la parte interior 320 y la parte exterior 330 de la representación 3D capturada 300.

- 5 El dispositivo informático receptor 130 está operativo para generar 507 una representación de avatar que corresponde a la parte exterior 330 de la representación 3D capturada 300, usando un modelo de ML entrenado para cabezas humanas, con la pose determinada 504 de la cabeza 102 que el dispositivo informático receptor 130 ha recibido 523 como entrada.
- 10 Opcionalmente, el dispositivo informático receptor 130 está operativo para fusionar 508 la parte interior 320 recibida 524 de la representación 3D capturada 300 y la representación de avatar generada 507 en una representación 3D fusionada de la cabeza 102. El dispositivo informático receptor 130 puede estar operativo además para visualizar 509 la
- 15 representación 3D fusionada de la cabeza 102 usando el dispositivo de visualización 131.

- Opcionalmente, el dispositivo informático receptor 130 puede estar operativo además para entrenar 510 el modelo de ML usando al menos la parte exterior 330 de la representación 3D capturada 300 y la pose
- 20 determinada 504 de la cabeza. Además opcionalmente, el dispositivo informático de borde 120 puede estar operativo para entrenar 510 el modelo de ML en base además a la parte interior 320 de la representación 3D

capturada 300, es decir, usando la representación 3D capturada sustancialmente completa 300 de la cabeza 102.

Aunque se han descrito realizaciones de la invención con distribuciones específicas de las operaciones involucradas en la generación de una
5 representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo entre el dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, respectivamente, como se ilustra en las figuras 5A-5C, la persona experta en la técnica puede imaginar fácilmente alternativas para distribuir las
10 operaciones involucradas en la generación de una representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo entre el dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130.

A continuación, se describen realizaciones de la circuitería de
15 procesamiento 602 que está comprendida en el dispositivo informático 600 para generar representación 3D de una cabeza de un participante en una sesión de comunicación por vídeo con referencia a la figura 6. Realizaciones de la circuitería de procesamiento 600 pueden estar comprendidas en uno o más del dispositivo informático emisor 110, el dispositivo informático de borde
20 120, y el dispositivo informático receptor 130.

La circuitería de procesamiento 602 puede comprender uno o más procesadores 603, tales como Unidades Centrales de Procesamiento (CPUs),

microprocesadores, procesadores de aplicaciones, procesadores específicos de aplicaciones, Unidades de Procesamiento de Gráficos (GPUs), y Procesadores de Señales Digitales (DSPs) incluyendo procesadores de imágenes, o una combinación de los mismos, y una memoria 604 que

5 comprende un programa de ordenador 605 que comprende instrucciones. Cuando se ejecutan por los procesadores 603, las instrucciones hacen que el dispositivo informático 600 se vuelva operativo de acuerdo con realizaciones de la invención descrita en este documento. Si las operaciones involucradas en la generación de una representación 3D de una cabeza de un participante

10 en una sesión de comunicación por vídeo se distribuyen entre dos o más del dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, sus respectivas instrucciones hacen que los dos o más del dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, cuando se

15 ejecutan por sus respectivos procesadores 603, se vuelvan operativos de acuerdo con realizaciones de la invención descrita en este documento de una manera colaborativa. La memoria 604 puede, por ejemplo, ser una Memoria de Acceso Aleatorio (RAM), una Memoria de Solo Lectura (ROM), una memoria flash, o similares. El programa de ordenador 605 se puede descargar

20 a la memoria 604 por medio de la circuitería de interfaz de red 601, como una señal portadora de datos que porta el programa de ordenador 605. La circuitería de procesamiento 602 puede comprender alternativa o

adicionalmente uno o más Circuitos Integrados de Aplicación Específica (ASICs), Redes de Puertas Programables en Campo (FPGAs), o similares, que son operativos para hacer que el dispositivo informático 600 se vuelva operativo de acuerdo con realizaciones de la invención descrita en este documento.

La circuitería de interfaz de red 601 puede comprender uno o más de un módem celular (por ejemplo, GSM, UMTS, LTE, 5G, o generación superior), un módem WLAN/Wi-Fi, un módem Bluetooth, una interfaz Ethernet, una interfaz óptica, o similares, para intercambiar datos entre el dispositivo informático 600 y otros dispositivos informáticos, en particular entre el dispositivo informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, y la red de comunicaciones 140, que puede comprender el Internet y una o más RANs.

A continuación, se describen realizaciones del procedimiento 700 de generación de una representación 3D de una cabeza 102 de un participante 101 en una sesión de comunicación por vídeo con referencia a la figura 7.

El procedimiento 700 se realiza mediante un dispositivo informático 600 y comprende adquirir 701 una representación 3D capturada de la cabeza 102, e identificar 702 posiciones de un conjunto de puntos de referencia faciales en la representación 3D capturada. El conjunto de puntos de referencia faciales comprende puntos de referencia faciales que son indicativos de un límite del rostro humano. El procedimiento 700 comprende además determinar

703 una pose de la cabeza 102, y determinar 704 un límite entre una parte interior y una parte exterior de la representación 3D capturada. El límite se determina 704 en base a las posiciones identificadas 702 del conjunto de puntos de referencia faciales. La parte interior de la representación 3D capturada representa el rostro del participante. El procedimiento 700 comprende además generar 705 una representación de avatar que corresponde a la parte exterior de la representación 3D capturada. La representación de avatar se genera 705 usando un modelo de ML entrenado para cabezas humanas, con la pose determinada 703 de la cabeza 102 como entrada. El modelo de ML puede entrenarse opcionalmente para la cabeza 102 del participante 101.

El procedimiento 700 comprende además opcionalmente extraer 707 la parte interior de la representación 3D capturada, y fusionar 708 la parte interior extraída 707 de la representación 3D capturada y la representación de avatar generada 705 en una representación 3D fusionada de la cabeza 102.

El procedimiento 700 comprende además opcionalmente mostrar 709 la representación 3D fusionada de la cabeza 102 usando un dispositivo de visualización 131. El dispositivo de visualización 131 puede ser uno cualquiera de una pantalla de ordenador, un televisor, un dispositivo de AR, un dispositivo de VR, un dispositivo de MR, un dispositivo de XR, y un dispositivo de HMD.

La adquisición 701 de una representación 3D capturada de la cabeza 102 puede comprender capturar la representación 3D de la cabeza 102

usando un sensor 3D 111. El sensor 3D 111 puede comprender uno o más de una cámara 3D, un LiDAR, y un sensor óptico 3D.

El procedimiento 700 comprende además opcionalmente adquirir el modelo de ML desde un almacenamiento de datos asociado con el
5 participante 101.

El procedimiento 700 comprende además opcionalmente entrenar 706 el modelo de ML usando al menos la parte exterior de la representación 3D capturada y la pose determinada 703 de la cabeza. El modelo de ML opcionalmente se entrena además en base a la parte interior de la
10 representación 3D capturada.

Se apreciará que el procedimiento 700 puede comprender etapas adicionales, alternativas, o modificadas de acuerdo con lo que se describe a lo largo de esta divulgación. El procedimiento también puede ser realizado por más de un dispositivo informático, por ejemplo, dos o más del dispositivo
15 informático emisor 110, el dispositivo informático de borde 120, y el dispositivo informático receptor 130, de una manera colaborativa.

Una realización del procedimiento 700 puede implementarse como el programa de ordenador 605 que comprende instrucciones que, cuando el programa de ordenador 605 es ejecutado por el dispositivo informático 600
20 hacen que el dispositivo informático 600 lleve a cabo el procedimiento 700 y se vuelva operativo de acuerdo con realizaciones de la invención descrita en este documento. El programa de ordenador 605 puede almacenarse en un

portador de datos legible por ordenador, tal como la memoria 604. Alternativamente, el programa de ordenador 605 puede portarse mediante una señal portadora de datos, por ejemplo, descargarse a la memoria 604 a través de la circuitería de interfaz de red 601.

- 5 La persona experta en la técnica se da cuenta de que la invención no se limita de ninguna manera a las realizaciones descritas anteriormente. Por el contrario, son posibles muchas modificaciones y variaciones dentro del alcance de las reivindicaciones anexas.