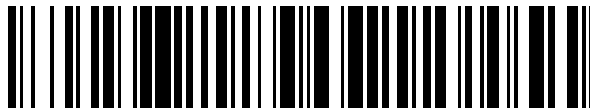


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 494 926**

51 Int. Cl.:

H04N 7/14 (2006.01)

H04N 7/15 (2006.01)

H04M 1/64 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **29.02.2008** **E 08003745 (0)**

97 Fecha y número de publicación de la concesión europea: **02.07.2014** **EP 1976291**

54 Título: **Procedimiento y sistema de video comunicación para el control en tiempo real basado en gestos de un avatar**

30 Prioridad:

02.03.2007 DE 102007010662

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

16.09.2014

73 Titular/es:

**DEUTSCHE TELEKOM AG (100.0%)
FRIEDRICH-EBERT-ALLEE 140
53113 BONN, DE**

72 Inventor/es:

**KASPAR, BERNHARD;
ENGLERT, ROMAN y
FELDES, STEFAN**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 494 926 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimiento y sistema de video comunicación para el control en tiempo real basado en gestos de un avatar

La invención se refiere a un procedimiento así como a un sistema de video comunicación basado en un avatar para el control en tiempo real basado en gestos de un modelo de cuerpo virtual. Por lo demás, la invención se refiere a una instalación terminal móvil para el empleo de un sistema de video comunicación basado en avatar de este tipo.

Se conoce el control en tiempo real basado en gestos de avatares sobre la base de ordenadores de altas prestaciones. En un avatar se trata de una figura representativa virtual, que representa una persona natural. A través de la representación avatar se pueden desplazar personas naturales implicadas a otros papeles. Los movimientos de la cabeza, de las mano y de los dedos de una persona natural son convertidor en este caso en tiempo real en movimientos de un modelo Avatar 3D. Con esta finalidad se representan secuencias de vídeo de la persona natural a través de una cámara. Por medio de análisis automático se extraen movimientos a partir de la secuencia de vídeo. Los movimientos son transmitidos codificados y se utilizan en el lado de recepción para la activación del avatar móvil. El control en tiempo real basado en gestos de avatares se conoce, por ejemplo, a partir del documento DE 10 2004 059 051 A1.

La tecnología de avatar encuentra cada vez más aplicación también en sistemas de vídeo de telecomunicaciones. Una arquitectura ejemplar para la video comunicación basada en avatar se representa en la figura 1. La figura 1 muestra un usuario A con un terminal 10' así como un usuario B con un terminal 30, que se pueden conectar entre sí, por ejemplo, a través de una red-IP 40, que puede ser Internet, Solamente con objeto de una representación sencilla, se muestra el terminal 10 como instalación de emisión, en cambio el terminal 30 del usuario B solamente se representa como aparato de recepción. Evidentemente, en la práctica, tanto el terminal 10 presenta una instalación de recepción como también el terminal 30 presenta una instalación de emisión correspondiente.

La instalación de emisión del terminal 10 está constituida por una cámara de vídeo 11 convencional, un seguidor de vídeo 15, un micrófono 12 y un codificador de audio 13. Las secuencias de vídeo suministradas por la cámara 11, analizadas con el seguidor de vídeo 15 y con un seguidor de móvil o un seguidor de movimiento, son convertidas en parámetros de movimiento en parámetros de textura. Unos procedimientos de estimación especiales, que son realizados por el seguidor de vídeo 15, encuentran la persona a registrar, siguen la posición de la persona sobre el tiempo, identifican partes del cuerpo como cabeza, cuello, hombros, ojos y manos y similares y siguen también su movimiento sobre el tiempo. La información obtenida es convertida entonces en una forma de parámetros correspondiente. Tales procedimientos de estimación han sido descritos, por ejemplo, por Eisert, Peter en la ponencia "MPEG 4 Facial Animation in Video Analysis and Synthesis", en International Journal of Imaging Systems and Technology Springer, vol. 13, no. 5, páginas 245-256, Marzo 2003 y por Schreer y col. en la ponencia "Vision-Based-Skin-Colour Segmentation of Moving Hands for Real-Time Applications", en Proc. of 1st European Conf. On Visual Media Production (CVMP 2004), Londres, Reino Unido, Marzo de 2004.

En los datos de movimiento se puede tratar de parámetros de animación de la cara y del cuerpo según la Norma-MPEG-4, que se conocen también en inglés como Body Animation Parameters (BAP) y Facial Animation Parameters (FAP). Los parámetros necesarios para la animación son transmitidos a través de la red de comunicaciones 40 hacia el terminal 30. Adicionalmente también la señal de voz registrada a través del micrófono del usuario A se puede transmitir al terminal 30. Los parámetros de voz pueden contener también informaciones, que describen los llamados visemas. Las informaciones de los visemas permiten la animación de la cara, en particular de la parte de la boca, de manera que se puede generar en el terminal 30 un movimiento de los labios adaptado a la señal de la voz. Las señales de vídeo y las señales de voz recibidas son reproducidas a través de un reproductor de avatar 32 conocido en sí en una pantalla 32, mientras que las señales de voz transmitidas pueden ser reproducidas acústicamente a través de una instalación de altavoz 33 correspondiente. El usuario B ve de esta manera en lugar del usuario real A su representante virtual.

Un inconveniente del sistema de video comunicación basado en avatar conocido se puede ver en que en los terminales deben estar implementados ordenadores de altas prestaciones y, por lo tanto, caros, que asumen los procesos costosos de cálculo y costosos de recursos, que son necesarios para el análisis de imágenes, es decir, segmentación, detección de objetos y seguimiento de objetos.

La invención tiene el cometido de proporcionar un procedimiento así como un sistema de video comunicación basado en avatar de un modelo de cuerpo virtual, que posibilita el empleo de terminales menos complejos y, por lo tanto, económicos.

La publicación US 2004/114731 A1 publica un sistema telefónico, en el que modelos de un usuario, de los que se sintetiza una secuencia de vídeo, son transmitidos a un receptor.

La publicación WO 99/57900 A publica un sistema de video teléfono con un sistema de reproducción de imágenes definido por el usuario, que recibe informaciones sensoriales desde los usuarios.

La publicación WO 03/058518 A publica un procedimiento para un sistema de interfaces de usuario de avatar, en el que un Avatar-Hosting-Server y un Session-Server están conectados a través de una red de comunicaciones.

5 La idea básica de la invención se puede ver en preparar una arquitectura distribuida para un sistema de video comunicación basado en avatar, en el que se lleva a cabo un análisis previo aproximado de la secuencia de vídeo registrada de una persona natural en un terminal, mientras que el análisis fino para la generación de los parámetros de movimiento y de los parámetros de textura, que son necesarios para la activación y animación de un modelo de cuerpo virtual, se realiza en una instalación de procesamiento central. En el terminal, la imagen actual de la secuencia de vídeo se divide en regiones globales. Solamente las regiones características de la imagen actual, las
10 llamadas Regiones de Interés (RoI), que son necesarias para la generación de datos de control, son seleccionadas a continuación y son transmitidas a la instalación de procesamiento central. El resto de la imagen se desecha. Las regiones características necesarias para el control de un avatar de una imagen contienen la cabeza y las dos manos. La cabeza y las manos se pueden identificar, por ejemplo, a través del reconocimiento del color de la piel. Segmentación global significa en este caso que no se describen los contornos de la mano o de la cabeza exactamente y entonces se transmiten los datos correspondientes. En su lugar, por ejemplo, se transmite un bloque rectangular, que contiene las manos. El análisis fino, que contiene una segmentación fina de las regiones transmitidas, se realiza en la instalación de procesamiento evacuada. Allí se realizan, además de la segmentación más fina, toda las otras etapas de análisis necesarias, como por ejemplo la descomposición de una zona transmitida en objetos parciales, por ejemplo los dedos de la mano, la determinación de la orientación del objeto (por ejemplo, la
20 alineación de la cabeza en dirección horizontal y en dirección vertical) y del estado de objetos parciales (por ejemplo, el grado de apertura de los ojos), de la orientación de objetos parciales (por ejemplo, la dirección de la mirada), del movimiento de objetos parciales (por ejemplo la boca o las cejas), así como la identificación de gestos de dirección (por ejemplo, la alineación de los dedos).

25 En este lugar hay que indicar todavía que o bien se transmite el contenido completo de las regiones halladas en la segmentación global, es decir, los puntos de la imagen de la región hacia la instalación de procesamiento o solamente se transmiten conjuntos de parámetros sobre el contenido de la imagen de la región respectiva. En el último caso, se puede reducir adicionalmente la velocidad de transmisión de los datos.

30 El problema técnico mencionado anteriormente se soluciona, por una parte, por medio de un procedimiento para el control en tiempo real basado en gestos de un modelo de cuerpo virtual en un entorno de video comunicación. A continuación se registra una secuencia de vídeo de una persona natural en una instalación terminal. La imagen actual de la secuencia de vídeo se segmenta entonces en varias regiones de forma global. Segmentación global significa, por ejemplo, que no se describen exactamente las partes del cuerpo de la persona natural, sin que en su
35 lugar se considere una zona geométrica, por ejemplo un bloque rectangular, en el que se representa una parte característica del cuerpo. A partir de las regiones se determinan entonces regiones preferidas de la imagen actual de la secuencia de vídeo. Las regiones preferidas son transmitidas entonces desde la instalación terminal hacia una instalación de procesamiento central a través de una red de comunicaciones, de manera que al menos una parte de cada zona calculada s transmitida a través de los datos correspondientes. En la instalación de procesamiento se obtienen parámetros de movimiento como reacción a los datos de al menos una parte de cada región calculada. Opcionalmente, a partir de los datos, que representan al menos una parte de cada región calculada, se obtienen también datos de textura. Los parámetros de movimiento y, dado el caso, los parámetros de textura son transmitidos a través de la red de comunicaciones o a través de otra red de comunicaciones hacia una instalación de recepción. Utilizando los parámetros de movimiento y, dado el caso, los parámetros de textura se activa y se anima un modelo
45 de cuerpo virtual almacenado en la instalación de recepción.

En este lugar hay que indicar que gracias a la segmentación global realizada en el aparato terminal solamente deben transmitirse los datos de aquellas regiones, que contienen las partes relevantes del cuerpo de la persona natural. Éstas son, por ejemplo, la región de la cabeza y las regiones, que contienen las dos manos y/o brazos de la
50 persona natural. De esta manera se puede reducir claramente la potencia de cálculo en el terminal así como la velocidad de datos necesaria para la transmisión, puesto que las etapas necesarias del procedimiento para el análisis de la secuencia de vídeo registrada y la generación de parámetros de movimiento, que son necesarios para la activación de un avatar, se distribuyen sobre la instalación terminal y la instalación de procesamiento central.

55 Para conseguir una activación lo más natural posible del modelo de cuerpo virtual, se descomponen antes de la generación de parámetros de movimiento y, dado el caso, de parámetros de textura, los contenidos de la imagen contenidos en las regiones calculadas en objetos parciales y se determina la orientación de los objetos parciales y/o el estado de los objetos parciales.

60 Para poder establecer una video comunicación audiovisual basada en avatar se registran en la instalación terminal, además, las señales de voz de la persona natural y se convierten en parámetros de voz correspondientes. Los parámetros de voz se pueden transmitir a través de la instalación de procesamiento hacia la instalación de recepción. De manera alternativa, los parámetros de voz se pueden transmitid junto con un sello de tiempo sin intercalación de la instalación de procesamiento hacia la instalación de recepción. Los sellos de tiempo se utilizan en

la instalación de recepción para sincronizar la reproducción de los contenidos de voz y de los contenidos de imágenes.

5 El problema técnico mencionado anteriormente se soluciona de la misma manera a través de un sistema de comunicación basado en avatar para el control en tiempo real basado en gestos de un modelo de cuerpo virtual. Basado en gestos significa que los movimientos de un modelo de cuerpo virtual siguen esencialmente los movimientos de la persona natural.

10 El sistema de video comunicación basado en avatar presenta al menos una red de comunicación, que puede ser una red de comunicación por cable o sin hilos. En el caso de una red de comunicación por cable, se puede tratar de Internet, el ISDN o una red local. Las redes de comunicaciones sin hilos pueden ser redes públicas de telefonía móvil y redes locales sin hilos, las llamadas WLANs.

15 A través de la red de comunicaciones se pueden conectar entre sí varias instalaciones terminales, que pueden ser accionadas como instalación de emisión y/o instalación de recepción. Además, se puede conectar un servidor de procesamiento con la red de comunicación. El servidor de procesamiento se puede conectar directamente en la red de comunicación o a través de una red de comunicación separada con la red de comunicación. Al menos una de las instalaciones terminales está configurada como instalación de emisión. En este caso, la instalación terminal presenta una instalación para el registro de una secuencia de vídeo de una persona natural. Además, está prevista una
 20 instalación para la segmentación global de la imagen actual de la secuencia de vídeo en varias regiones y para el cálculo de regiones preferidas de la imagen actual de la secuencia de vídeo así como una instalación de transmisión, que transmite datos de al menos una parte de cada zona calculada hacia el servidor de procesamiento a través de la red de comunicación. Las funciones de segmentación global y cálculo de regiones preferidas se pueden realizar también en instalaciones separadas. El servidor de procesamiento presenta una instalación para la
 25 generación de parámetros de movimiento y, dado el caso, de parámetros de textura y, en concreto, como reacción a los datos de al menos una parte de cada región recibida. Además, el servidor de procesamiento dispone de una instalación para la transmisión de los parámetros de movimiento y, dado el caso, de los parámetros de textura a través de la red de comunicación o a través de otra red de comunicación hacia una instalación de recepción. La
 30 instalación de recepción presenta una instalación para la activación de un modelo de cuerpo virtual utilizando los parámetros de movimiento y, dado el caso, los parámetros de textura. El modelo de cuerpo virtual está depositado, por ejemplo, en un reproductor de avatar de la instalación de recepción.

35 Para tener en cuenta exclusivamente las partes del cuerpo características para la obtención de parámetros de movimiento y, dado el caso, de parámetros de textura de la persona natural, la instalación para la segmentación global está configurada para el cálculo de la región de la cabeza y de las regiones de las manos de la persona registrada.

40 Para conseguir una activación y animación lo más natural posible del modelo de cuerpo virtual, el servidor de procesamiento está equipado con una instalación de descomposición, que descompone el contenido de la imagen contenido en las regiones calculadas en objetos parciales y determina la orientación de los objetos parciales y/o del estado de los objetos parciales.

45 Para posibilitar una video comunicación audiovisual, la instalación de emisión de cada aparato terminal presenta un micrófono y un codificador de audio para la generación de parámetros de voz. La instalación de recepción de cada instalación terminal presenta una instalación de reproducción audiovisual, de manera que la instalación de activación de la instalación de recepción puede ser un reproductor de avatar convencional.

50 En este lugar hay que indicar que evidentemente se pueden accionar varias instalaciones terminales tanto como instalación de emisión y como instalación de recepción, para posibilitar una conexión dúplex completa entre los dos usuarios.

55 Gracias a la arquitectura distribuida del sistema de video comunicación basada en avatar, es posible el empleo de un entorno móvil. En este caso, las instalaciones terminales pueden representar aparatos terminales móviles, como por ejemplo PDAs, Smartphones y similares. Los aparatos terminales móviles se pueden comunicar en este caso entre sí a través de una red de telefonía móvil o una red de radio local sin hilos (WLAN).

60 El problema técnico mencionado anteriormente se soluciona de la misma manera a través de una instalación terminal móvil para el empleo en un sistema de video comunicación basado en Avatar de este tipo. La instalación terminal móvil presenta una instalación para el registro de una secuencia de vídeo de una persona natural, una instalación para la segmentación global de la imagen actual de la secuencia de vídeo en varias regiones, una instalación para el cálculo de regiones predeterminadas de la imagen actual de la secuencia de vídeo así como una instalación para la transmisión de datos de al menos una parte de cada región calculada hacia el servidor de procesamiento a través de la red de comunicación.

A continuación se explica en detalle la invención con la ayuda de un ejemplo de realización en combinación con los dibujos adjuntos. Los componentes que tienen la misma función se identifican por los mismos signos de referencia. En este caso:

5 La figura 1 muestra la representación esquemática de un sistema de video comunicación basado en avatar conocido, y

La figura 2 muestra la representación esquemática de una arquitectura distribuida de un sistema de video comunicación basado en avatar de acuerdo con la invención.

10 La figura 2 muestra una representación esquemática de un sistema de video comunicación basado en avatar, que está instalado, por ejemplo, en un entorno de telefonía móvil. Solamente para simplificación de la representación, la figura 2 muestra solamente un terminal móvil 10, que está conectado a través de una red de telefonía móvil 40 con un terminal 30. A través de la red de telefonía móvil 40 se puede acceder, además, a un servidor de procesamiento 20. Hay que indicar que la arquitectura distribuida representada en la figura 2 es adecuada para una video conferencia con más de dos interlocutores de la comunicación implicados.

El terminal 10 del usuario A presenta una cámara de vídeo 11, que puede registrar movimientos del usuario A. Los datos de vídeo generados por la cámara de vídeo 11 son transmitidos a una instalación de segmentación global 14. La instalación de segmentación global extrae a partir de la imagen actual de la secuencia de vídeo registrada solamente las regiones, que contienen las partes preferidas del cuerpo del usuario A, que son necesarias para la activación de un modelo de cuerpo virtual. En las partes preferidas o características del cuerpo se trata con preferencia de la cabeza y de las dos manos del usuario natural A. Para posibilitar una comunicación audiovisual, el terminal 10 presenta, además, un micrófono 12, cuya salida está conectada con un codificador de audio 13.

25 El terminal 30 del usuario B presenta un reproductor de avatar 31 conocido en sí, en el que está depositado un modelo de cuerpo virtual a activar. En el lado de salida, el reproductor de avatar 31 está conectado con una pantalla 32 y con un altavoz 33, que proporcionan la reproducción visual y acústica de los datos de salida del reproductor de avatar 31.

30 Hay que indicar que en el caso de una comunicación dúplex completa, el terminal 10 contiene de la misma manera un reproductor de avatar, una pantalla y uno o varios altavoces. El terminal 30 puede presentar en este caso los componentes mostrados con relación al terminal 10.

35 En el servidor de procesamiento 20 está implementado un llamado seguidor de vídeo 21, que lleva a cabo un análisis fino de la secuencia de vídeo registrada. En este caso hay que tener en cuenta que las informaciones esenciales para el seguidor de vídeo 21 permanecen en el terminal 10 también después de la segmentación global y son transmitidas hacia el servidor de procesamiento 20.

40 A continuación se explica en detalle el modo de funcionamiento del sistema de video comunicación basado en avatar representado en la figura 2.

Se supone que los usuarios A y B quieren establecer una comunicación video telefónica.

45 Para una explicación más sencilla, solamente se considera un flujo de datos desde el terminal 10 hacia el terminal 30.

50 La cámara 11 del terminal 10 registra el usuario A y conduce las secuencias de imagen o de vídeo generadas a la entrada de la instalación de segmentación global 14. La instalación de segmentación global 14 lleva a cabo aplicando un algoritmo de estimación conocido en sí una segmentación global de la imagen actual de la secuencia de vídeo, analizando, por ejemplo, la información de color de los puntos individuales de la imagen. De esta manera, se determinan las regiones de la imagen con color y textura similar y son delimitadas de otras regiones de la imagen. Para hallar las regiones, que contienen la cabeza y las manos del usuario A, en la imagen actual, se pueden buscar las zonas que contienen un grado alto de color de la piel. Se pueden emplear algoritmos alternativos, que tienen en cuenta, por ejemplo, la consistencia en el movimiento, el tamaño mínimo y la posición relativa de las regiones de la imagen entre sí. Las regiones halladas por la instalación de segmentación global 14, llamadas también Regiones de Interés, contienen en primer lugar solamente regiones globales, por ejemplo rectangulares, en las que se encuentran la cabeza o bien las manos. La instalación de segmentación global 14 acondiciona en el lado de salida un conjunto de datos, que representa al menos una parte de cada una de las regiones determinadas.

60 A través del micrófono 12 se pueden registrar señales de voz del usuario A y se pueden alimentar al codificador de audio 13. El codificador de audio suministra en el lado de salida parámetros de voz, que pueden contener, por ejemplo, también visemas, de manera que se puede generar un movimiento de los labios adaptado a la señal de voz en el reproductor de avatar 31 del terminal 30.

5 En el presente ejemplo los datos de vídeo, que representan las regiones calculadas a través de la segmentación global, y los datos de voz de acuerdo con un protocolo de comunicación son empaquetados en paquetes de datos y son transmitidos a través de la red de comunicación 40 hacia el servidor de procesamiento 20. Los paquetes de datos contienen adicionalmente la dirección del servidor de procesamiento 20 como también la dirección de destino del terminal 30.

10 De acuerdo con la implementación, el sistema de video comunicación se ocupa de que, si se desea una comunicación basada en avatar del usuario A, se transmitan al menos los datos de vídeo en primer lugar hacia el servidor de procesamiento 20 y luego hacia el terminal 30. Los datos de voz pueden, pero no deben ser transmitidos a través del servidor de procesamiento 20 hacia el terminal 30. La transmisión de datos de vídeo hacia el servidor de procesamiento 20 se puede controlar desde el terminal 10 que llama o desde una instalación central (no representada) de la red de comunicación.

15 El seguidor de vídeo 21 realiza un análisis fino, descomponiendo las regiones calculadas durante la segmentación global en objetos parciales. Por ejemplo, las regiones, que representan la mano, se descomponen en otras regiones parciales más pequeñas, que representan los dedos de una mano. De manera similar, la región, que representa la cara, se puede descomponer en subregiones, que representan la nariz, los ojos, la barba y, por ejemplo, las orejas. Además, el seguidor de vídeo 21 calcula a partir de los datos de vídeo recibidos la orientación del objeto, por ejemplo la alineación de la cabeza. Además, el seguidor de vídeo 21 puede calcular el estado de objetos parciales, por ejemplo el grado de apertura de los ojos o de la boca. Además, también se puede calcular el movimiento de los objetos parciales, como por ejemplo el movimiento de la boca y el movimiento de los párpados. A partir de las informaciones acumuladas de la imagen actual, el seguidor de vídeo 21 genera parámetros de movimiento y, dado el caso, parámetros de textura. Los parámetros de movimiento y los parámetros de textura son transmitidos entonces en el presente ejemplo junto con los parámetros de voz en un formato normalizado, por ejemplo un formato compatible con MPEG a través de la red de comunicación 40 hacia el terminal 30. La dirección del terminal 30 es tomada por el servidor de procesamiento 20 con preferencia a partir de los paquetes de datos que proceden desde el terminal 10.

30 Los parámetros de movimiento y de textura son transferidos junto con los parámetros de voz al reproductor de avatar 31 del terminal 30, que puede procesar datos en el formato-MPEG. El reproductor de avatar 31 separa los parámetros de control recibidos en informaciones de imagen e informaciones de voz y los transfiere a la pantalla 32 y al altavoz 33. De esta manera, se puede conducir una comunicación audiovisual basada en Avatar entre el terminal 10 y el terminal 30.

35

REIVINDICACIONES

- 1.- Procedimiento para el control en tiempo real basado es gestos de un modelo de cuerpo virtual en un entorno de video comunicación, con las etapas:
- a) registro de una secuencia de vídeo de una persona en una instalación terminal (10);
 - 5 b) segmentación global de la imagen actual de la secuencia de vídeo en varias regiones;
 - c) cálculo de regiones preferidas de la imagen actual de la secuencia de vídeo;
 - d) transmisión de datos de al menos una parte de cada región calculada desde la instalación terminal (10) hacia una instalación de procesamiento (20) a través de una red de comunicación (40);
 - 10 e) generación de parámetros de movimiento y, en concreto, como reacción a los datos de al menos una parte de cada región calculada;
 - f) transmisión de los parámetros de movimiento desde la instalación de procesamiento (20) a través de la red de comunicación o de otra red de comunicación hacia una instalación de destino (30);
 - g) activación de un modelo de cuerpo virtual utilizando los parámetros de movimiento en la instalación de destino (30).
- 15 2.- Procedimiento de acuerdo con la reivindicación 1, caracterizado por que la etapa c) comprende la determinación de la región de la cabeza y de las regiones de la mano y/o de los brazos y la etapa e) comprende la generación de parámetros de textura.
- 3.- Procedimiento de acuerdo con una de las reivindicaciones anteriores, caracterizado por que antes de la realización de la etapa e), se descompone el contenido de la imagen, contenido en las regiones preferidas transmitidas hacia la instalación de procesamiento, en objetos parciales y se determina la orientación de los objetos parciales y/o el estado de los objetos parciales.
- 20 4.- Procedimiento de acuerdo con una de las reivindicaciones anteriores, caracterizado por las etapas: registro de señales acústicas; generación de parámetros de voz; transmisión de los parámetros de voz a través de la instalación de procesamiento (20) hacia la instalación de destino (30) o transmisión de los parámetros de voz sin intercalación de la instalación de procesamiento hacia la instalación de destino-
- 25 5.- Sistema de video comunicación basado en avatar para la activación en tiempo real basada en gestos de un modelo de cuerpo virtual, que comprende al menos una red de comunicación (40), varias instalaciones terminales (10, 30), que pueden ser accionadas como instalación de emisión y/o instalación de recepción y que se pueden conectar entre sí a través de la red de comunicación (40), y un servidor de procesamiento (20), que se puede conectar con la red de comunicación (40), en el que al menos una de las instalaciones terminales (10) presenta una instalación de emisión con las siguientes características: una instalación (11) para el registro de una secuencia de vídeo de una persona natural; una instalación (14) para la segmentación global de la imagen actual de la secuencia de vídeo en varias regiones y para la determinación de regiones preferidas de la imagen actual de la secuencia de vídeo; una instalación para la transmisión de datos al menos de una parte de cada región determinada hacia el servidor de procesamiento (20) al menos a través de la red de comunicación (40), y en el que el servidor de procesamiento (20) presenta: una instalación (21) para la generación de parámetros de movimiento y, en concreto, como reacción a los datos de al menos una parte de cada región recibida; una instalación para la transmisión de los parámetros de movimiento a través de la red de comunicación o a través de otra red de comunicación hacia una instalación de destino (30), que presenta una instalación (31) para la activación de un modelo de cuerpo virtual utilizando los parámetros de movimiento.
- 30 35 40 6.- Sistema de video comunicación basado en avatar de acuerdo con la reivindicación 5, en el que la instalación (14) para la segmentación global está configurada para la determinación de la región de la cabeza y de las regiones de las manos y/o de los brazos de la persona registrada y la instalación de generación está configurada para la generación de parámetros de textura.
- 45 7.- Sistema de video comunicación basado en avatar de acuerdo con la reivindicación 5 ó 6, en el que el servidor de procesamiento (20) presenta una instalación que está configurada para la descomposición del contenido de la imagen contenido en las regiones determinadas en objetos parciales y para la determinación de la orientación de los objetos parciales y/o del estado de los objetos parciales.
- 50 8.- Sistema de video comunicación basado en avatar de acuerdo con una de las reivindicaciones 5 a 7, en el que la al menos una instalación de emisión es un micrófono (12) y un codificador de audio (13) para la generación de parámetros de voz y la instalación de destino (30) presenta una instalación de reproducción audiovisual (32, 33), de

manera que la instalación de activación (31) de la instalación de destino (30) es un reproductor de avatar.

9.- Sistema de video comunicación basado en avatar de acuerdo con una de las reivindicaciones 5 a 8, en el que la instalación terminal que puede ser activada como instalación de emisión comprende un terminal móvil y la red de comunicación (40) comprende una red de telefonía móvil o una red local sin hilos.

- 5 10.- Instalación terminal móvil (10) para el empleo en un sistema de video comunicación basada en avatar, que comprende una instalación (11) para el registro de una secuencia de vídeo de una persona natural; una instalación (14) para la segmentación global de la imagen actual de la secuencia de vídeo en varias regiones y para la determinación de regiones predeterminadas de la imagen actual de la secuencia de vídeo; una instalación para la transmisión de datos de al menos una parte de cada región determinada hacia un servidor de procesamiento (20) a
- 10 través de al menos una red de comunicación (40).

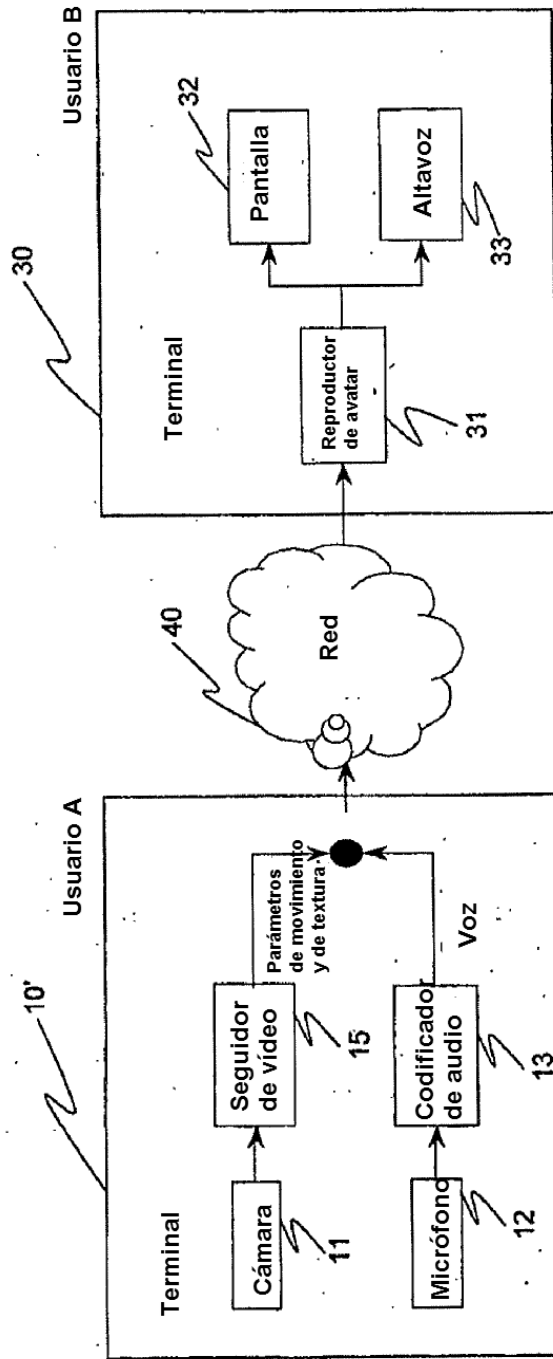


Fig. 1 (Estado de la técnica)

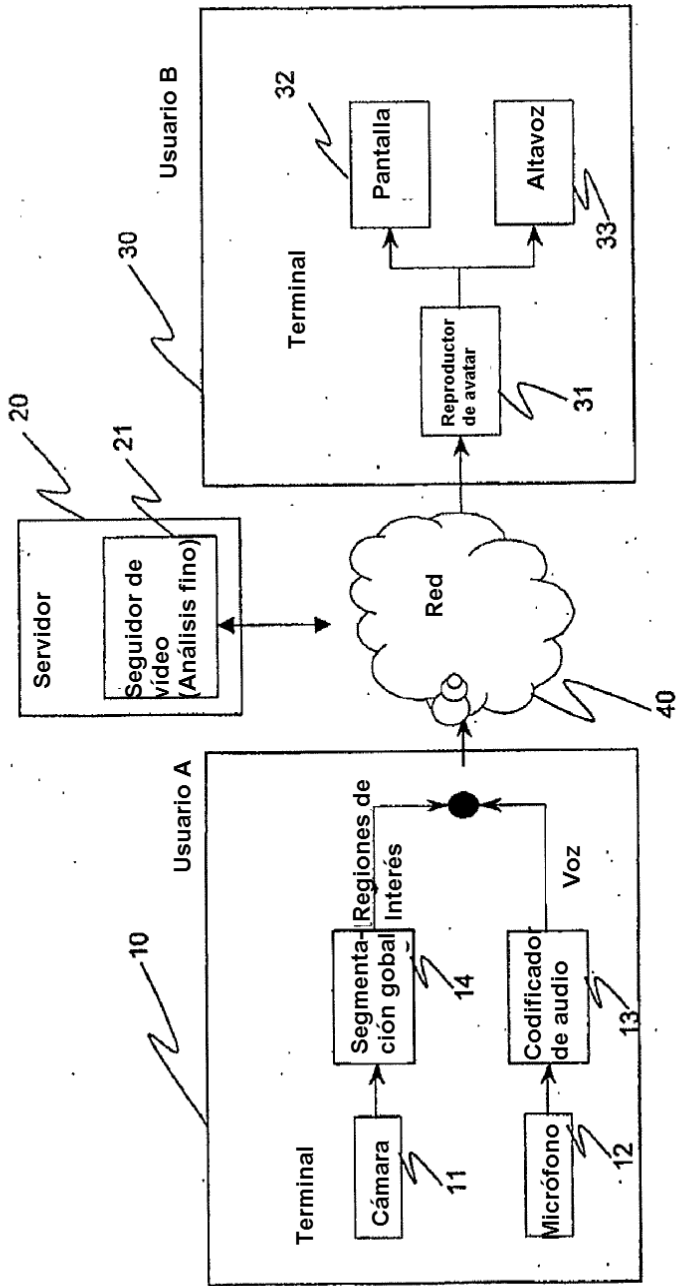


Fig. 2