

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 908 588**

51 Int. Cl.:

**G06T 7/10** (2007.01)

**G06T 19/00** (2011.01)

**G06T 7/30** (2007.01)

**G06T 7/73** (2007.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **09.09.2019** **E 19196193 (7)**

97 Fecha y número de publicación de la concesión europea: **01.12.2021** **EP 3789965**

54 Título: **Método para controlar una pantalla de presentación visual, programa informático y dispositivo de presentación visual de realidad mixta**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**03.05.2022**

73 Titular/es:  
**APOQLAR GMBH (100.0%)**  
**Raboisen 32**  
**20095 Hamburg, DE**

72 Inventor/es:

**PELZL, SIRKO y**  
**WOZNIAK, MICHAL**

74 Agente/Representante:

**LOZANO GANDIA, José**

ES 2 908 588 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

Método para controlar una pantalla de presentación visual, programa informático y dispositivo de presentación visual de realidad mixta

5 La invención se refiere a un método para controlar una pantalla de presentación visual de un dispositivo de realidad mixta.

10 Además, la invención se refiere a un programa informático que tiene medios de código de programa adaptados para realizar tal método.

Además, la invención se refiere a un dispositivo de presentación visual de realidad mixta con tal programa informático.

15 En general, la invención se refiere al campo de la visualización de información virtual en combinación con el entorno real. La información virtual se superpone a objetos reales en una pantalla de presentación visual de un dispositivo de presentación visual. Esta área se denomina habitualmente "realidad mixta".

20 El "continuo de la virtualidad" se extiende desde el entorno meramente real hasta el entorno meramente virtual, incluida la realidad aumentada y la virtualidad aumentada que oscila entre estos extremos. El término "realidad mixta" se define generalmente como "cualquier lugar entre los extremos del continuo de la virtualidad", es decir, la realidad mixta comprende generalmente el continuo de la virtualidad completo excepto la realidad pura y la virtualidad pura. En el contexto de la presente solicitud, el término "realidad mixta" puede referirse particularmente a "realidad aumentada".

25 La tecnología de realidad mixta es especialmente prometedora para aplicaciones médicas, por ejemplo para cirugía médica u otro tratamiento médico. Por ejemplo, pueden superponerse datos de formación de imágenes médicas (imágenes de TC, imágenes de IRM, etc.) que visualizan la anatomía y/o los procesos fisiológicos de un cuerpo humano o animal a la vista del mundo real del cuerpo por medio de un dispositivo de presentación visual de realidad mixta. De esta manera, por ejemplo, puede darse respaldo a los cirujanos durante la cirugía colocando de manera virtual tales datos de obtención imágenes médicas directamente sobre el objeto de tratamiento, es decir, sobre el cuerpo del paciente o una parte del mismo.

35 Uno de los desafíos más importantes en la realidad mixta es el problema del registro, es decir, el problema de alinear correctamente entre sí los objetos del mundo real y los objetos del mundo virtual. Sin un registro preciso, se verá comprometida la ilusión de que los dos mundos coexisten. Más gravemente, en aplicaciones médicas, pueden resultar riesgos para el éxito del tratamiento médico e incluso para la salud del paciente de una alineación imprecisa. Como resultado, es crucial que la información virtual que se visualiza, por ejemplo los datos de formación de imágenes médicas, coincida con precisión con el mundo real, es decir, el objeto de tratamiento, en su posición, tamaño y perspectiva.

45 A partir del documento EP 2 874 556 B1, se conocen un método y un sistema correspondiente basados en realidad aumentada que permiten el guiado de instrumentos en cirugía y otros procedimientos intervencionistas. Con este propósito, se obtiene una trayectoria de intervención para su uso en el procedimiento intervencionista, en la que la trayectoria de intervención se planifica basándose en datos de imag3D del interior de un paciente, y se obtiene una imagen de cámara del exterior del paciente durante el procedimiento intervencionista. Se establece una correspondencia espacial entre la imagen de cámara y los datos de imag3D y se calcula una vista de la trayectoria de intervención que se corresponde con la imagen de cámara. Finalmente, la vista de la trayectoria de intervención se combina con la imagen de cámara para obtener una imagen compuesta, que se presenta visualmente en una pantalla de presentación visual.

50 Además, a partir del documento US 9 892 564 B1, se conocen un método y un sistema correspondiente que permiten aumentar las vistas en tiempo real de un paciente con datos tridimensionales. Con este propósito, se registra una posición virtual de una capa exterior del paciente para alinearla con una posición en tiempo real de la capa exterior del paciente usando mediciones morfométricas virtuales y mediciones morfométricas en tiempo real.

60 Además, en Zaganidis, A. *et al.* "Integrating Deep Semantic Segmentation into 3D Point Cloud Registration", IEEE Robotics and Automation Letters, 2018, se describe un enfoque para el registro de nubes de puntos en aplicaciones robóticas autónomas que usa segmentación semántica.

Es un objeto de la presente invención proporcionar una técnica mejorada para visualizar información virtual en aplicaciones médicas que permita una alineación mejorada de objetos virtuales y reales.

65 El objeto de la invención se logra mediante un método para controlar una pantalla de presentación visual de un dispositivo de presentación visual de realidad mixta con las características de la reivindicación 1.

Según la invención, el método comprende al menos las siguientes etapas:

- 5 a) proporcionar un conjunto de datos de imagen que comprende una pluralidad de imágenes de un objeto de tratamiento, en el que el objeto de tratamiento es el cuerpo de un paciente o una parte del mismo y las imágenes representan el objeto de tratamiento desde diferentes perspectivas,
- 10 b) generar una nube de puntos objetivo 3D a partir del conjunto de datos de imagen, en el que la nube de puntos objetivo comprende una multitud de puntos definidos en un sistema de coordenadas tridimensional y los puntos representan la superficie del objeto de tratamiento,
- 15 c) determinar un número de máscaras de segmentación semántica en la nube de puntos objetivo aplicando la segmentación semántica,
- d) proporcionar un conjunto de datos de formación de imágenes médicas que comprende datos de formación de imágenes médicas del objeto de tratamiento,
- 20 e) generar una nube de puntos fuente 3D a partir del conjunto de datos de formación de imágenes médicas, en el que la nube de puntos fuente comprende una multitud de puntos definidos en un sistema de coordenadas tridimensional y los puntos también representan la superficie del objeto de tratamiento,
- 25 f) determinar un número de máscaras de segmentación semántica en la nube de puntos fuente aplicando la segmentación semántica,
- g) determinar una transformación entre la nube de puntos fuente y la nube de puntos objetivo usando las máscaras de segmentación de la nube de puntos fuente y las máscaras de segmentación de la nube de puntos objetivo, y
- 30 h) visualizar al menos una parte de los datos de formación de imágenes médicas en la pantalla de presentación visual, en el que los datos de formación de imágenes médicas se superponen sobre el objeto de tratamiento y se alinean con el objeto de tratamiento usando la transformación entre la nube de puntos fuente y la nube de puntos objetivo.

35 No es necesario ejecutar las etapas del método en el orden especificado y la invención no está limitada en consecuencia, es decir, el orden alfabético de las letras no implica una secuencia específica de etapas a) a h). Por ejemplo, por supuesto, las etapas a) a c) pueden ejecutarse después de las etapas d) a f), o algunas de las etapas del método pueden ejecutarse en paralelo.

40 La invención propone un método para controlar la pantalla de presentación visual del dispositivo de realidad mixta para superponer datos de formación de imágenes médicas sobre el objeto de tratamiento. La invención propone así un método para controlar la pantalla de presentación visual del dispositivo de realidad mixta para visualizar datos de formación de imágenes médicas sobre el objeto de tratamiento.

45 En el contexto de la presente solicitud, el término "objeto de tratamiento" se refiere al cuerpo de un paciente o a una parte del mismo. El paciente puede ser un ser humano o un animal, es decir, el objeto de tratamiento puede ser un cuerpo humano o animal o una parte del mismo.

50 Los términos "dispositivo de presentación visual de realidad mixta" y "dispositivo de realidad mixta" se usan indistintamente. El término "ordenador" se usa en su sentido más amplio, es decir, se refiere a cualquier dispositivo de procesamiento que puede ser instruido para llevar a cabo secuencias de operaciones aritméticas y/o lógicas.

El término "2D" se refiere a coordenadas bidimensionales. El término "3D" se refiere a coordenadas tridimensionales. El término "4D" se refiere a coordenadas de cuatro dimensiones.

55 Además de la pantalla de presentación visual, el dispositivo de realidad mixta puede comprender un ordenador y una memoria. El dispositivo de realidad mixta también puede comprender una pluralidad de ordenadores. Además, el dispositivo de realidad mixta puede comprender una cámara, en particular un sistema de cámara 3D, y/o varios sensores, en particular al menos un sensor de profundidad, por ejemplo un sensor de profundidad de tiempo de vuelo. El dispositivo de presentación visual de realidad mixta también puede comprender un sistema de posicionamiento y/o una unidad de medición inercial.

60

En la etapa a), se proporciona un conjunto de datos de imagen que comprende una pluralidad de imágenes del objeto de tratamiento, en el que las imágenes representan el objeto de tratamiento desde diferentes perspectivas. Estas imágenes representan una vista del mundo real del objeto de tratamiento. Pueden generarse, por ejemplo, por medio de una cámara y/o un sensor de profundidad, en particular por medio de una cámara y/o un sensor de profundidad del dispositivo de realidad mixta.

65

En la etapa d), se proporciona un conjunto de datos de formación de imágenes médicas que comprende datos de formación de imágenes médicas del objeto de tratamiento. Estos datos de formación de imágenes médicas representan la información virtual que va a visualizarse en la pantalla de presentación visual. Los datos de formación de imágenes médicas pueden comprender, por ejemplo, imágenes en sección del objeto de tratamiento. Los datos de formación de imágenes médicas pueden generarse, por ejemplo, usando un método de formación de imágenes médicas como la formación de imágenes por resonancia magnética (IRM). Los datos de formación de imágenes médicas pueden generarse antes y/o durante el tratamiento médico, por ejemplo antes y/o durante la cirugía.

En las etapas b) y e), se generan nubes de puntos tridimensionales que representan el objeto de tratamiento'. Con este propósito, pueden usarse métodos de reconstrucción 3D.

En las etapas c) y f), se aplica la segmentación semántica para determinar un número de máscaras de segmentación en la nube de puntos objetivo y la nube de puntos fuente, respectivamente. La segmentación semántica (también denominada segmentación semántica de imágenes) puede definirse como la tarea de agrupar entre sí partes de imágenes que pertenecen a la misma clase de objeto. En el contexto de la presente solicitud, la clase de objeto puede ser, por ejemplo, una parte específica de la anatomía del objeto de tratamiento. Si el objeto de tratamiento es una cabeza humana, por ejemplo, las clases de objeto simples pueden incluir nariz, orejas, boca, ojos, cejas, etc. El término "máscara de segmentación semántica" se refiere a una parte (o segmento) de la imagen que se ha determinado, mediante la segmentación semántica, como perteneciente a la misma clase de objeto. La segmentación semántica puede realizarse en datos 2D, es decir basado en píxeles, o en datos 3D, es decir basado en vóxeles.

En la etapa g), se determina una transformación entre la nube de puntos fuente y la nube de puntos objetivo usando las máscaras de segmentación de la nube de puntos fuente y las máscaras de segmentación de la nube de puntos objetivo. En esta etapa, puede determinarse una transformación que, cuando se aplica a una de las nubes de puntos, alinea los puntos de ambas nubes de puntos entre sí. La transformación puede incluir traslación y/o rotación. Por ejemplo, la transformación puede tener la forma de una matriz de transformación, en particular la forma de una matriz de transformación 4x4. La transformación determinada puede transformar, en particular aproximadamente, la nube de puntos fuente en la nube de puntos objetivo o viceversa.

Además de las máscaras de segmentación, pueden usarse otros parámetros como entradas para determinar la transformación entre las nubes de puntos. En particular, la transformación entre la nube de puntos fuente y la nube de puntos objetivo puede determinarse usando las máscaras de segmentación de la nube de puntos fuente y las máscaras de segmentación de la nube de puntos objetivo y las coordenadas de los puntos de la nube de puntos fuente y las coordenadas de los puntos de la nube de puntos objetivo.

En la etapa h), los datos de formación de imágenes médicas se visualizan en la pantalla de presentación visual, en la que los datos de formación de imágenes médicas se superponen a la vista del mundo real del objeto de tratamiento y se alinean con la misma usando la transformación determinada en la etapa g). Al hacerlo, se crea una fusión virtual de los datos de formación de imágenes médicas visualizados y la vista del mundo real del objeto de tratamiento.

La invención se basa en el descubrimiento de que usando la segmentación semántica, el problema del registro puede resolverse de manera más precisa y fiable. Esto se logra determinando una transformación entre la nube de puntos fuente y la nube de puntos objetivo basándose en las máscaras de segmentación de ambas nubes de puntos.

A modo de ejemplo, puede diseñarse una función de optimización usada para determinar la transformación entre la nube de puntos fuente y la nube de puntos objetivo para favorecer transformaciones que coincidan con precisión con las máscaras de segmentación semántica correspondientes en ambas nubes de puntos, es decir, transformaciones que hacen coincidir con precisión las máscaras de segmentación semántica con clases de objeto idénticas y/o similares (por ejemplo, nariz, oreja, ojos, cejas). Esto puede realizarse, por ejemplo, mediante el uso de un algoritmo de optimización de cuatro dimensiones para determinar la transformación entre la nube de puntos fuente y la nube de puntos objetivo, en el que para cada punto de la nube de puntos respectiva, la clase de objeto (por ejemplo, nariz, oreja, boca) de la máscara de segmentación semántica correspondiente se interpreta como una cuarta dimensión del punto (además de las coordenadas 3D del punto). Por ejemplo, puede usarse una variante 4D del algoritmo de punto más cercano iterativo (ICP) con este propósito.

La invención permite alinear la información virtual de los datos de formación de imágenes médicas de manera muy precisa y fiable con la vista del mundo real del objeto de tratamiento. Como resultado, puede crearse una fusión virtual mejorada de los datos de formación de imágenes médicas visualizados y la vista del mundo real del objeto de tratamiento, y pueden evitarse riesgos para el éxito del tratamiento médico y para la salud del paciente.

Según una realización ventajosa de la invención, se propone que la pantalla de presentación visual esté configurada como pantalla de presentación visual transparente óptica, en particular como pantalla de presentación visual transparente óptica montada en la cabeza.

Tales realizaciones de la invención proporcionan la ventaja de una percepción realista e intuitiva del entorno real para el usuario, por ejemplo un cirujano.

5 Según otra realización ventajosa de la invención, se propone que el dispositivo de presentación visual de realidad mixta comprenda o consista en un dispositivo de presentación visual de realidad mixta montado en la cabeza y/o gafas inteligentes de realidad mixta. El dispositivo de presentación visual de realidad mixta puede comprender o consistir, por ejemplo, en un dispositivo HoloLens de Microsoft o un dispositivo HoloLens 2 de Microsoft o un dispositivo similar.

10 Tales realizaciones de la invención proporcionan la ventaja de que son cómodas de usar y al mismo tiempo ofrecen un potente hardware para visualizar información virtual para el usuario. En particular, muchos dispositivos de presentación visual de realidad mixta montados en la cabeza y gafas inteligentes de realidad mixta comprenden, además de la pantalla de presentación visual, componentes de hardware integrados potentes y versátiles que  
15 incluyen procesadores y memoria de alto rendimiento, sistemas de cámara 3D y sensores de profundidad de tiempo de vuelo, sistemas de posicionamiento y/o unidades de medición inercial.

Según otra realización ventajosa de la invención, el dispositivo de presentación visual de realidad mixta puede comprender un ordenador externo adicional, por ejemplo un servidor externo, que se conecta a la pantalla de  
20 presentación visual y adaptado para ejecutar al menos partes del método propuesto según la invención. El dispositivo de presentación visual de realidad mixta puede comprender, por ejemplo, un dispositivo de presentación visual de realidad mixta montado en la cabeza y/o gafas inteligentes de realidad mixta y un ordenador adicional, por ejemplo un servidor externo, que se conecta al dispositivo de presentación visual de realidad mixta montado en la cabeza y/o a las gafas inteligentes de realidad mixta, respectivamente, por medio de una conexión por cable o  
25 inalámbrica. El servidor externo puede estar diseñado como un servidor en la nube.

Tales realizaciones proporcionan la ventaja de potencia de cálculo adicional para operaciones complejas e intensas a nivel computacional, que son particularmente comunes en los campos relevantes de la visión por ordenador y los gráficos por ordenador.

30 Según otra realización ventajosa de la invención, todos los componentes del dispositivo de presentación visual de realidad mixta pueden estar integrados en el dispositivo de presentación visual de realidad mixta montado en la cabeza y/o en las gafas inteligentes de realidad mixta. Esto proporciona la ventaja de un dispositivo de presentación visual de realidad mixta compacto y, por tanto, altamente móvil.

35 Según otra realización ventajosa de la invención, se propone que los datos de formación de imágenes médicas se generen usando al menos uno de los siguientes métodos de formación de imágenes médicas: formación de imágenes por resonancia magnética nuclear (IRM), tomografía computarizada (TC), tomografía computarizada de haz cónico (TCHC), tomografía volumétrica digital (TVD), imágenes intraoperatorias fluoroscópicas, rayos X, radiografía, ecografía, endoscopia y/o formación de imágenes de medicina nuclear.

40 Esto proporciona la ventaja de que los resultados de los métodos de formación de imágenes médicas modernos, potentes y versátiles, pueden usarse de manera beneficiosa durante el tratamiento médico, por ejemplo durante la cirugía, por medio de visualización de realidad mixta. Esto permite, por ejemplo, visualizar los órganos internos del paciente y/o carcinomas y/u otros defectos internos del cuerpo del paciente.

Según otra realización ventajosa de la invención, se propone que las máscaras de segmentación semántica en la nube de puntos objetivo y/o las máscaras de segmentación semántica en la nube de puntos fuente se determinen mediante una red neuronal convolucional configurada para la segmentación semántica.

50 La segmentación semántica por medio de redes neuronales convolucionales (RNC) ha tenido un avance impresionante en los últimos años. Las RNC diseñadas y entrenadas adecuadamente permiten una segmentación semántica fiable, precisa y rápida de datos de imágenes 2D e incluso 3D. Las realizaciones de la invención que emplean una RNC para la segmentación semántica proporcionan, por tanto, la ventaja de que las potentes capacidades de las RNC pueden utilizarse para mejorar la visualización de realidad mixta. Por ejemplo, la arquitectura U-Net de RNC puede usarse con este propósito, es decir, la red neuronal convolucional configurada para la segmentación semántica puede diseñarse como una RNC U-NET.

60 La RNC puede entrenarse para la segmentación semántica de un cuerpo o una parte específica del cuerpo usando conjuntos de datos de entrenamiento apropiados que incluyen máscaras de segmentación semántica etiquetadas con su clase de objeto respectiva (por ejemplo, nariz, oreja, ojos, cejas en el caso de un conjunto de entrenamiento para una cabeza humana).

Según la invención, se propone que la etapa c) comprenda lo siguiente:

65 – determinar un número de máscaras de segmentación semántica en las imágenes del conjunto de datos de

imagen aplicando la segmentación semántica a las imágenes del conjunto de datos de imagen, en particular usando una red neuronal convolucional configurada para la segmentación semántica, y

- 5           – determinar las máscaras de segmentación semántica en la nube de puntos objetivo usando las máscaras de segmentación semántica en las imágenes del conjunto de datos de imagen.

10 Por tanto, se propone que se determinen máscaras de segmentación semántica en las imágenes del conjunto de datos de imagen. Las imágenes pueden diseñarse particularmente como imágenes 2D, en particular imágenes RGB 2D. Basándose en estas máscaras de segmentación semántica en las imágenes, pueden determinarse máscaras de segmentación semántica 3D en la nube de puntos objetivo. Por ejemplo, cada punto de la nube de puntos objetivo 3D puede proyectarse en la imagen 2D para determinar la máscara de segmentación semántica en la imagen 2D que corresponde al punto respectivo en la nube de puntos objetivo 3D.

15 De este modo, la invención proporciona las ventajas de que facilita la segmentación semántica y permite usar métodos sofisticados disponibles para la segmentación semántica en imágenes 2D, en particular en imágenes RGB 2D, para determinar las máscaras de segmentación semántica. Por ejemplo, están disponibles potentes redes neuronales convolucionales y conjuntos de datos de entrenamiento correspondientes para la segmentación semántica en imágenes RGB 2D. El uso de su potencial permite una segmentación semántica rápida, precisa y fiable de las imágenes 2D. Estos resultados pueden transferirse a la nube de puntos 3D para utilizarlos con los propósitos del método reivindicado, es decir, para determinar una transformación precisa entre la nube de puntos fuente y la nube de puntos objetivo.

Alternativa o adicionalmente, según la invención, se propone que la etapa f) comprenda

- 25           – determinar un número de máscaras de segmentación semántica en los datos de formación de imágenes médicas del conjunto de datos de formación de imágenes médicas mediante la aplicación de la segmentación semántica a los datos de formación de imágenes médicas del conjunto de datos de formación de imágenes médicas, en particular usando una red neuronal convolucional configurada para la segmentación semántica y
- 30           – determinar las máscaras de segmentación semántica en la nube de puntos fuente usando las máscaras de segmentación semántica en los datos de formación de imágenes médicas del conjunto de datos de formación de imágenes médicas.

35 Por tanto, se propone que se determinen máscaras de segmentación semántica en los datos de formación de imágenes médicas del conjunto de datos de formación de imágenes médicas. Los datos de formación de imágenes médicas pueden comprender imágenes médicas 2D, en particular imágenes en sección 2D, y las máscaras de segmentación semántica pueden determinarse en estas imágenes médicas 2D. Basándose en estas máscaras de segmentación semántica en los datos de formación de imágenes médicas, pueden determinarse máscaras de segmentación semántica 3D en la nube de puntos fuente. Por ejemplo, cada punto de la nube de puntos fuente 3D puede proyectarse en la imagen médica 2D para determinar la máscara de segmentación semántica en la imagen médica 2D que corresponde al punto respectivo en la nube de puntos fuente 3D.

45 De esta manera, la invención proporciona las ventajas de que facilita la segmentación semántica y permite usar métodos sofisticados disponibles para la segmentación semántica médica en imágenes 2D para determinar las máscaras de segmentación semántica. Por ejemplo, están disponibles potentes redes neuronales convolucionales y conjuntos de datos de entrenamiento correspondientes para la segmentación semántica médica en imágenes médicas 2D. Un ejemplo incluye la arquitectura de red neuronal convolucional U-Net. El uso del potencial de estas RNC permite una segmentación semántica rápida y precisa de las imágenes médicas 2D. Estos resultados pueden transferirse a la nube de puntos 3D para utilizarlos con los propósitos del método reivindicado, es decir, para determinar una transformación precisa entre la nube de puntos fuente y la nube de puntos objetivo.

50 Los datos de formación de imágenes médicas también pueden comprender datos de formación de imágenes médicas 3D, por ejemplo, un modelo de formación de imágenes médicas 3D del objeto de tratamiento. Los datos de formación de imágenes médicas 3D pueden reconstruirse a partir de una pluralidad de imágenes médicas 2D, en particular imágenes en sección 2D, del objeto de tratamiento. Estas imágenes médicas 2D pueden generarse usando un método de formación de imágenes médicas. En tales realizaciones en las que los datos de formación de imágenes médicas también comprenden datos de formación de imágenes médicas 3D, las máscaras de segmentación semántica también pueden determinarse en los datos de formación de imágenes médicas 3D. Con este propósito, pueden usarse métodos de segmentación semántica 3D, en particular basados en redes neuronales convolucionales configuradas para la segmentación semántica 3D. Basándose en las máscaras de segmentación semántica en los datos de formación de imágenes médicas 3D, pueden determinarse máscaras de segmentación semántica 3D en la nube de puntos fuente.

65 Según otra realización ventajosa de la invención, se propone que el conjunto de datos de imagen comprenda varias

imágenes visuales y/o imágenes de profundidad del objeto de tratamiento y que la nube de puntos objetivo 3D se genere a partir de las imágenes visuales y/o las imágenes de profundidad, en particular usando un método de fotogrametría y/o un método de fusión de profundidad.

5 Las imágenes visuales pueden generarse por medio de una cámara, en particular un sistema de cámara 3D. Las imágenes visuales pueden diseñarse, por ejemplo, como imágenes RGB y/o imágenes en escala de grises. Las imágenes de profundidad pueden generarse por medio de un escáner 3D y/o un sensor de profundidad, en particular por medio de un escáner láser 3D y/o un sensor de profundidad de tiempo de vuelo. Las imágenes de profundidad pueden diseñarse como mapas de profundidad. La combinación de imágenes visuales e imágenes de profundidad puede diseñarse como datos de imagen RGB-D. Las imágenes del conjunto de datos de imagen pueden diseñarse así como imágenes RGB-D.

15 La nube de puntos objetivo 3D puede generarse a partir de las imágenes visuales y/o las imágenes de profundidad usando métodos de reconstrucción 3D, en particular métodos de reconstrucción 3D activos y/o pasivos.

20 La nube de puntos objetivo 3D puede generarse a partir de las imágenes visuales usando fotogrametría. En particular, la nube de puntos objetivo 3D puede generarse a partir de las imágenes visuales usando un proceso de estructura a partir de movimiento (SfM, *structure from motion*) y/o un proceso estéreo multivista (MVS, *multi-view stereo*). Por ejemplo, puede usarse la canalización de reconstrucción 3D COLMAP con este propósito.

25 La nube de puntos objetivo 3D puede generarse a partir de las imágenes de profundidad usando la fusión de profundidad, es decir, usando la reconstrucción 3D a partir de múltiples imágenes de profundidad. La fusión de profundidad puede basarse en la función de distancia con signo truncada (FDST). Por ejemplo, la biblioteca de nubes de puntos (BNP) puede usarse para generar imágenes de profundidad usando fusión de profundidad. Por ejemplo, puede usarse el escáner Kinect Fusion como método de fusión de profundidad. En particular, puede usarse una implementación de Kinect Fusion contenida en BNP con este propósito, por ejemplo KinFu.

30 Tales realizaciones de la invención que comprenden generar la nube de puntos objetivo 3D a partir de imágenes visuales y/o imágenes de profundidad, tal como se explicó anteriormente, proporcionan la ventaja de que puede lograrse una reconstrucción precisa y detallada de la superficie del objeto de tratamiento.

Según otra realización ventajosa de la invención, se propone que la etapa c) comprenda lo siguiente:

- 35 – determinar un número de máscaras de segmentación semántica en las imágenes visuales y/o en las imágenes de profundidad del conjunto de datos de imagen aplicando la segmentación semántica a las imágenes visuales y/o las imágenes de profundidad del conjunto de datos de imagen, en particular usando una red neuronal convolucional configurada para la segmentación semántica, y
- 40 – determinar las máscaras de segmentación semántica en la nube de puntos objetivo usando las máscaras de segmentación semántica en las imágenes visuales y/o las imágenes de profundidad del conjunto de datos de imagen.

45 Tales realizaciones proporcionan las ventajas de que facilitan la segmentación semántica y permiten usar métodos sofisticados disponibles para la segmentación semántica en imágenes 2D y/o imágenes en profundidad para determinar las máscaras de segmentación semántica. Las máscaras de segmentación semántica resultantes en las imágenes visuales y/o en las imágenes de profundidad pueden transferirse a la nube de puntos 3D para utilizarlas con los propósitos del método reivindicado, es decir, para determinar una transformación precisa entre la nube de puntos fuente y la nube de puntos objetivo. Al hacerlo, puede lograrse una reconstrucción rápida, precisa y detallada de la superficie del objeto de tratamiento.

50 Según otra realización ventajosa de la invención, se propone que el conjunto de datos de imagen comprenda una pluralidad de imágenes visuales e imágenes de profundidad del objeto de tratamiento y la etapa b) comprende lo siguiente:

- 55 – generar una primera nube de puntos 3D a partir de las imágenes visuales del conjunto de datos de imagen,
- generar una segunda nube de puntos 3D a partir de las imágenes de profundidad del conjunto de datos de imagen, y
- 60 – generar la nube de puntos objetivo 3D usando la primera nube de puntos 3D y la segunda nube de puntos 3D, en particular fusionando la primera nube de puntos 3D y la segunda nube de puntos 3D.

65 Tal realización proporciona la ventaja de que puede mejorarse la precisión y la integridad de la reconstrucción 3D y, por tanto, puede mejorarse la precisión y la integridad de la nube de puntos objetivo resultante. Esto se logra fusionando la primera nube de puntos 3D, que puede ser una nube de puntos basada en RGB, con la segunda nube

de puntos 3D, que es una nube de puntos basada en profundidad.

Según otra realización ventajosa de la invención, se propone que

- 5 - la etapa c) comprende determinar las máscaras de segmentación semántica en la nube de puntos objetivo aplicando directamente la segmentación semántica a la nube de puntos objetivo, en particular usando una red neuronal convolucional configurada para la segmentación semántica, y/o
- 10 - la etapa f) comprende determinar las máscaras de segmentación semántica en la nube de puntos fuente aplicando directamente la segmentación semántica a la nube de puntos fuente, en particular usando una red neuronal convolucional configurada para la segmentación semántica.

Según otra realización ventajosa de la invención, se propone que la transformación entre la nube de puntos fuente y la nube de puntos objetivo se determine en la etapa g) por medio de un algoritmo de punto más cercano iterativo (ICP, *iterative closest point*) usando las coordenadas de los puntos de la nube de puntos fuente y las máscaras de segmentación de la nube de puntos fuente y las coordenadas de los puntos de la nube de puntos objetivo y las máscaras de segmentación de la nube de puntos objetivo.

Según esta realización, se propone que se use un algoritmo de punto más cercano iterativo (ICP) para determinar la transformación entre la nube de puntos fuente y la nube de puntos objetivo. En particular, puede usarse una variante 4D del algoritmo ICP con este propósito. Las coordenadas de los puntos y las máscaras de segmentación de las nubes de puntos fuente y de puntos objetivo se usan como entradas para el algoritmo ICP. Esto puede realizarse, por ejemplo, usando una variante 4D del algoritmo ICP para determinar la transformación entre la nube de puntos fuente y la nube de puntos objetivo, en la que para cada punto de la nube de puntos respectiva, la clase de objeto (por ejemplo, nariz, oreja, boca) de la máscara de segmentación semántica correspondiente se interpreta como una cuarta dimensión del punto (además de las coordenadas 3D del punto). Al incluir las máscaras de segmentación semántica en el algoritmo ICP tal como se explicó anteriormente, el algoritmo ICP tiene en cuenta no sólo la información espacial de los puntos de las nubes de puntos (es decir, sus coordenadas), sino también su semántica, es decir, su significado.

Las máscaras de segmentación semántica también pueden usarse para determinar una estimación inicial de la transformación entre la nube de puntos fuente y la nube de puntos objetivo para el algoritmo ICP, es decir, para determinar una alineación inicial.

Las realizaciones explicadas anteriormente proporcionan la ventaja de que mejoran el guiado del algoritmo ICP para hallar una transformación óptima entre la nube de puntos de la nube de puntos fuente y la nube de puntos objetivo. Por ejemplo, mediante el uso de máscaras de segmentación semántica para determinar la transformación, puede evitarse que el algoritmo ICP halle un resultado óptimo local como solución. Por tanto, puede mejorarse la precisión de la transformación y, por tanto, la alineación de los datos de formación de imágenes médicas con el objeto de tratamiento.

Según otra realización ventajosa de la invención, se propone que

- 45 - la etapa a) comprende eliminar un fondo y/u otras partes irrelevantes de las imágenes del objeto de tratamiento usando la segmentación semántica, en particular usando una red neuronal convolucional configurada para la segmentación semántica, y/o
- 50 - la etapa d) comprende eliminar un fondo y/u otras partes irrelevantes de los datos de formación de imágenes médicas del objeto de tratamiento usando la segmentación semántica, en particular usando una red neuronal convolucional configurada para la segmentación semántica.

Las partes irrelevantes pueden ser partes de las imágenes o los datos de formación de imágenes médicas, respectivamente, que no son relevantes para los propósitos de la formación de imágenes médicas y/o no son relevantes para superponer los datos de formación de imágenes médicas sobre el objeto de tratamiento, tal como el cabello de una cabeza humana.

Tales realizaciones proporcionan la ventaja de que facilitan la generación de las nubes de puntos, la determinación de las máscaras de segmentación semántica y la determinación de la transformación entre las nubes de puntos.

Según otra realización ventajosa de la invención, se propone que el conjunto de datos de formación de imágenes médicas comprenda un modelo de formación de imágenes médicas 3D del objeto de tratamiento, en el que el modelo de formación de imágenes médicas 3D se reconstruye a partir de una pluralidad de imágenes en sección 2D del objeto de tratamiento generadas mediante un método de formación de imágenes médicas.

Tales realizaciones proporcionan la ventaja de que pueden facilitar la generación de la nube de puntos fuente 3D,

mientras que al mismo tiempo puede mejorarse la precisión de la nube de puntos fuente 3D resultante.

Según otra realización ventajosa de la invención, se propone que en la etapa a), el conjunto de datos de imagen que comprende las imágenes del objeto de tratamiento se cree por medio de una cámara, en particular un sistema de cámara 3D, del dispositivo de presentación visual de realidad mixta, y/o por medio de un sensor de profundidad, en particular por medio de un sensor de profundidad de tiempo de vuelo, del dispositivo de presentación visual de realidad mixta.

De esta manera, el conjunto de datos de imagen, que es necesario para generar la nube de puntos objetivo, puede crearse y proporcionarse con los propósitos de la invención de una manera muy sencilla y fácil de usar. Por ejemplo, usando la cámara y/o el sensor de profundidad del dispositivo de presentación visual de realidad mixta, las imágenes del conjunto de datos de imagen pueden crearse escaneando automáticamente el objeto de tratamiento. En particular, si el dispositivo de presentación visual de realidad mixta comprende o consiste en un dispositivo de presentación visual de realidad mixta montado en la cabeza y/o gafas inteligentes de realidad mixta, la cámara y/o el sensor de profundidad pueden integrarse en el dispositivo de presentación visual de realidad mixta montado en la cabeza y/o las gafas inteligentes de realidad mixta, respectivamente. En este caso, el usuario puede simplemente dirigir su cabeza hacia la parte del objeto de tratamiento que va a escanearse.

Según otra realización ventajosa de la invención, se propone que

- en la etapa a), cuando las imágenes se crean por medio de la cámara de la pantalla de presentación visual de realidad mixta, la posición de la cámara se determina en un sistema de coordenadas tridimensional para cada imagen y se almacena como posición de cámara 3D y en la etapa b), la nube de puntos objetivo se genera usando las posiciones de cámara 3D, y/o
- en la etapa a), cuando las imágenes se crean por medio de la cámara de la pantalla de presentación visual de realidad mixta, la orientación de la cámara se determina en un sistema de coordenadas tridimensional para cada imagen y se almacena como orientación de cámara 3D y en la etapa b), la nube de puntos objetivo se genera usando las orientaciones de cámara 3D.

Tales realizaciones proporcionan la ventaja de que permiten una generación especialmente precisa y fiable de la nube de puntos objetivo a partir del conjunto de datos de imagen.

Según otra realización ventajosa de la invención, se propone que cualquier posición, cualquier orientación y cualquier transformación determinadas en las etapas a) a h) se determina sin marcadores y/o usando localización y mapeos simultáneos (SLAM, *simultaneous localization and mapping*).

Tales realizaciones proporcionan la ventaja de que son particularmente cómodas para el usuario, ya que no tienen que proporcionarse marcadores ópticos u otros marcadores de referencia.

El objeto de la invención se logra además mediante un programa informático que tiene medios de código de programa adaptados para realizar un método tal como se describió anteriormente cuando el programa informático se ejecuta en un ordenador.

El objeto de la invención se logra además mediante un dispositivo de presentación visual de realidad mixta que tiene una pantalla de presentación visual, un ordenador y una memoria, en el que el programa informático descrito anteriormente se almacena en la memoria y el ordenador está adaptado para ejecutar el programa informático.

El programa informático puede diseñarse como un programa informático distribuido. El ordenador y la memoria pueden diseñarse como un ordenador distribuido y una memoria distribuida, respectivamente, es decir, el ordenador que está adaptado para ejecutar el programa informático puede comprender dos o más ordenadores. La memoria distribuida puede comprender múltiples memorias, en las que cada memoria puede almacenar al menos una parte del programa informático distribuido, y cada uno de los dos o más ordenadores puede estar adaptado para ejecutar una parte del programa informático distribuido.

Además, el dispositivo de presentación visual de realidad mixta puede tener una interfaz que esté adaptada para recibir los datos de formación de imágenes médicas desde una fuente externa. La interfaz puede estar adaptada, por ejemplo, para conectarse a un dispositivo de formación de imágenes médicas, por ejemplo un dispositivo de formación de imágenes por resonancia magnética (IRM) y/o un dispositivo de tomografía computarizada (TC), y/o a una memoria que almacena los datos de formación de imágenes médicas. Esto puede incluir una conexión por cable y/o inalámbrica.

Tal como se mencionó anteriormente, el dispositivo de presentación visual de realidad mixta puede comprender o consistir en un dispositivo de presentación visual de realidad mixta montado en la cabeza y/o gafas inteligentes de realidad mixta. El dispositivo de presentación visual de realidad mixta puede comprender o consistir, por ejemplo, en un dispositivo HoloLens de Microsoft o un dispositivo HoloLens 2 de Microsoft o un dispositivo similar.

Tal como se mencionó anteriormente, el dispositivo de presentación visual de realidad mixta puede comprender un ordenador externo adicional, por ejemplo un servidor externo, que se conecta a la pantalla de presentación visual y adaptado para ejecutar al menos partes del método propuesto según la invención. El dispositivo de presentación visual de realidad mixta puede comprender, por ejemplo, un dispositivo de presentación visual de realidad mixta montado en la cabeza y/o gafas inteligentes de realidad mixta y un ordenador adicional, por ejemplo un servidor externo, que se conecta al dispositivo de presentación visual de realidad mixta montado en la cabeza y/o gafas inteligentes de realidad mixta, respectivamente, por medio de una conexión por cable y/o inalámbrica. El servidor externo puede estar diseñado como un servidor en la nube.

Tal como se mencionó anteriormente, todos los componentes del dispositivo de presentación visual de realidad mixta pueden integrarse en el dispositivo de presentación visual de realidad mixta montado en la cabeza y/o en las gafas inteligentes de realidad mixta.

A continuación, la invención se explicará con más detalle usando las realizaciones a modo de ejemplo que se muestran esquemáticamente en los dibujos adjuntos. Los dibujos muestran lo siguiente:

Figura 1 - una representación esquemática de un dispositivo de presentación visual de realidad mixta según la invención;

Figura 2 - una representación esquemática de un método para controlar un dispositivo de presentación visual de realidad mixta según la invención;

Figura 3 - una representación esquemática de una nube de puntos objetivo 3D con máscaras de segmentación semántica;

Figura 4 - una representación esquemática de una nube de puntos fuente 3D con máscaras de segmentación semántica.

La figura 1 muestra una representación esquemática de un dispositivo 1 de presentación visual de realidad mixta, que comprende un dispositivo de presentación visual de realidad mixta montado en la cabeza en forma de un par de gafas 5 inteligentes de realidad mixta. En esta realización a modo de ejemplo, las gafas 5 inteligentes de realidad mixta son del tipo HoloLens de Microsoft. Las gafas 5 inteligentes tienen una memoria 13a y un ordenador 11a conectado con la memoria 13a, en el que el ordenador 11a comprende varias unidades de procesamiento, concretamente, una CPU (unidad de procesamiento central), una GPU (unidad de procesamiento de gráficos) y una HPU (unidad de procesamiento holográfico).

Además, las gafas 5 inteligentes del dispositivo 1 de presentación visual de realidad mixta tienen una cámara 9 en forma de un sistema de cámara 3D. La cámara 9 está adaptada para crear imágenes visuales de un objeto 15 de tratamiento desde diferentes perspectivas. Además, las gafas 5 inteligentes comprenden una pluralidad de sensores 7, incluido un sensor de profundidad de tiempo de vuelo que está adaptado para crear imágenes de profundidad del objeto 15 de tratamiento desde diferentes perspectivas. En esta realización a modo de ejemplo, el objeto 15 de tratamiento es la cabeza de un paciente 17 humano.

Además, las gafas 5 inteligentes del dispositivo 1 de presentación visual de realidad mixta tienen una pantalla 3 de presentación visual, que está diseñada como una pantalla de presentación visual transparente óptica montada en la cabeza en esta realización a modo de ejemplo. La pantalla 3 de presentación visual transparente está adaptada para visualizar información virtual, en esta realización a modo de ejemplo datos de formación de imágenes médicas, superponiendo la información virtual sobre la vista real del objeto 15 de tratamiento.

Además, la figura 1 ilustra que el dispositivo 1 de presentación visual de realidad mixta comprende un servidor 21, que se conecta a las gafas 5 inteligentes a través de una conexión inalámbrica y/o por cable. El servidor 21 comprende una memoria 13b y un ordenador 11b conectado con la memoria 13b, en el que el ordenador 11b comprende una CPU y una GPU. El servidor 21 tiene una interfaz 19 que está adaptada para recibir datos de formación de imágenes médicas desde una fuente externa, concretamente desde una memoria que almacena los datos de formación de imágenes médicas, a través de una conexión por cable y/o inalámbrica.

La figura 2 muestra una representación esquemática de un método a modo de ejemplo para controlar un dispositivo de presentación visual de realidad mixta según la invención.

En la etapa a), se proporciona un conjunto de datos de imagen que comprende una pluralidad de imágenes del objeto 15 de tratamiento, en el que las imágenes representan el objeto de tratamiento desde diferentes perspectivas. En esta realización a modo de ejemplo, el conjunto de datos de imagen comprende una pluralidad de imágenes visuales en forma de imágenes RGB 2D y una pluralidad de imágenes de profundidad en forma de mapas de profundidad. Las imágenes visuales se crean por medio de la cámara 9 del dispositivo 1 de presentación visual de realidad mixta, y las imágenes de profundidad se crean por medio del sensor 7 de profundidad del dispositivo 1 de

presentación visual de realidad mixta (véase la figura 1). Dicho de otro modo, esto significa que se crea un conjunto de datos de imagen que contiene datos de imagen RGB-D. Con este propósito, el usuario que lleva puestas las gafas 5 inteligentes, por ejemplo un cirujano, puede simplemente dirigir su cabeza hacia el objeto 15 de tratamiento en diferentes perspectivas y el objeto 15 de tratamiento se escanea automáticamente desde diferentes perspectivas usando la cámara 9 y el sensor 7 de profundidad. Cuando se crean las imágenes, la posición y la orientación de la cámara 9 se determinan en un sistema de coordenadas tridimensional para cada imagen y se almacenan como posición de cámara 3D y orientación de cámara 3D, respectivamente. Con este propósito, se usa la localización y mapeo simultáneos (SLAM). La posición de cámara 3D y la orientación de cámara 3D para cada imagen pueden denominarse parámetros de cámara extrínsecos. El conjunto de datos de imagen que incluye los parámetros extrínsecos de la cámara se transfiere después desde las gafas 5 inteligentes al servidor 21.

En la etapa b), el servidor 21 genera una nube de puntos objetivo 3D a partir del conjunto de datos de imagen, en el que la nube de puntos objetivo comprende una multitud de puntos definidos en un sistema de coordenadas tridimensional y los puntos representan la superficie del objeto 15 de tratamiento. Con este propósito, se usan los parámetros extrínsecos de la cámara, es decir, las posiciones de cámara 3D y las orientaciones de cámara 3D.

La nube de puntos objetivo 3D se genera a partir de las imágenes visuales y las imágenes de profundidad. Con este propósito, se genera una primera nube de puntos 3D a partir de las imágenes visuales del conjunto de datos de imagen usando un método de fotogrametría, concretamente, usando la canalización de reconstrucción 3D COLMAP que incluye un proceso de estructura a partir del movimiento (SfM) y un proceso estéreo multivista (MVS). Además, se genera una segunda nube de puntos 3D a partir de las imágenes de profundidad del conjunto de datos de imagen usando la fusión de profundidad, es decir, usando la reconstrucción 3D a partir de múltiples imágenes de profundidad. Con este propósito, por ejemplo, se usa la biblioteca de nubes de puntos (BNP), es decir, la implementación de Kinect Fusion KinFu contenida en BNP. Después de eso, la nube de puntos objetivo 3D se genera fusionando la primera nube de puntos 3D y la segunda nube de puntos 3D.

En la etapa c), se determina un número de máscaras de segmentación semántica en la nube de puntos objetivo aplicando la segmentación semántica.

En primer lugar, se determina un número de máscaras de segmentación semántica en las imágenes visuales 2D del conjunto de datos de imagen aplicando la segmentación semántica a estas imágenes RGB 2D. Las máscaras de segmentación semántica definen una clase de objeto para cada píxel de cada imagen visual 2D, en las que las clases de objeto pueden incluir una clase de objeto "otro", "vacío" o similar para áreas de la imagen que no pueden hacerse coincidir de otro modo. Para determinar las máscaras de segmentación semántica en las imágenes visuales 2D, se usa una red neuronal convolucional (RNC) configurada para la segmentación semántica, concretamente una RNC basada en la arquitectura de RNC U-Net (RNC U-Net). La RNC U-Net se entrena para la segmentación semántica del objeto 15 de tratamiento, es decir, para la segmentación semántica de una cabeza humana, usando conjuntos de datos de entrenamiento apropiados que incluyen máscaras de segmentación semántica etiquetadas con su clase de objeto respectiva (por ejemplo, nariz, oreja, ojos, cejas).

En segundo lugar, las máscaras de segmentación semántica en la nube de puntos objetivo 3D se determinan usando las máscaras de segmentación semántica determinadas previamente en las imágenes visuales 2D del conjunto de datos de imagen. Con este propósito, cada punto de la nube de puntos objetivo 3D se proyecta en varias imágenes RGB 2D para determinar la máscara de segmentación semántica en la imagen 2D respectiva que corresponde al punto en la nube de puntos objetivo 3D.

La figura 3 ilustra esquemáticamente los resultados de determinación de las máscaras de segmentación semántica en la nube de puntos objetivo. La figura muestra una nube 23 de puntos objetivo 3D que se ha generado a partir del conjunto de datos de imagen. En la nube 23 de puntos objetivo 3D, se han determinado varias máscaras 25a, 25b, 25c de segmentación semántica. Dos máscaras 25a de segmentación semántica representan el ojo derecho y el ojo izquierdo del paciente, respectivamente. Otra máscara 25b de segmentación semántica representa la nariz del paciente y otra máscara 25c de segmentación semántica representa la boca del paciente.

Volviendo ahora a la figura 2, en la etapa d), se proporciona un conjunto de datos de formación de imágenes médicas que comprende datos de formación de imágenes médicas del objeto de tratamiento. En esta realización a modo de ejemplo, los datos de formación de imágenes médicas comprenden una pluralidad de imágenes en sección 2D del objeto 15 de tratamiento que se crearon por medio de formación de imágenes por resonancia magnética (IRM) antes de una cirugía. Estas imágenes de IRM 2D, en forma de datos DICOM, las recibe el servidor 21 a través de la interfaz 19 del servidor. El servidor 21 reconstruye un modelo de formación de imágenes médicas 3D del objeto 15 de tratamiento a partir de la pluralidad de imágenes de IRM 2D del objeto 15 de tratamiento.

Los datos de formación de imágenes médicas también pueden comprender metadatos, por ejemplo en forma de atributos almacenados como etiquetas en datos DICOM. Los ejemplos de tales metadatos incluyen información sobre el grosor de corte de imágenes en sección y/o información sobre la separación entre píxeles. Tales metadatos pueden usarse para reconstruir el modelo de formación de imágenes 3D y/o para generar la nube de puntos fuente 3D.

5 En la etapa e), se genera una nube de puntos fuente 3D a partir del conjunto de datos de formación de imágenes médicas, en la que la nube de puntos fuente comprende una multitud de puntos definidos en un sistema de coordenadas tridimensional y los puntos también representan la superficie del objeto 15 de tratamiento. Con este propósito, se usa la biblioteca de nubes de puntos (BNP).

En la etapa f), se determina un número de máscaras de segmentación semántica en la nube de puntos fuente aplicando la segmentación semántica.

10 En primer lugar, se determina un número de máscaras de segmentación semántica en las imágenes de IRM 2D del conjunto de datos de formación de imágenes médicas aplicando la segmentación semántica a estas imágenes. Las máscaras de segmentación semántica definen una clase de objeto para cada píxel de cada imagen de IRM 2D, en el que las clases de objeto pueden incluir una clase de objeto "otro", "vacío" o similar para áreas de la imagen que no pueden coincidir de otro modo. Para determinar las máscaras de segmentación semántica en las imágenes de IRM 2D, se usa una red neuronal convolucional (RNC) configurada para la segmentación semántica, concretamente una RNC basada en la arquitectura de RNC U-Net (RNC U-Net). La RNC U-Net se entrena para la segmentación semántica del objeto 15 de tratamiento, es decir, para la segmentación semántica de una cabeza humana, usando conjuntos de datos de entrenamiento apropiados que incluyen máscaras de segmentación semántica etiquetadas con su clase de objeto respectiva (por ejemplo, nariz, oreja, ojos, cejas).

20 En segundo lugar, las máscaras de segmentación semántica en la nube de puntos fuente 3D se determinan usando las máscaras de segmentación semántica determinadas previamente en las imágenes de IRM 2D del conjunto de datos de formación de imágenes médicas.

25 La figura 4 ilustra esquemáticamente los resultados de determinación de las máscaras de segmentación semántica en la nube de puntos fuente. La figura muestra una nube 27 de puntos fuente 3D que se ha generado a partir del conjunto de datos de formación de imágenes médicas. En la nube 27 de puntos fuente 3D, se han determinado varias máscaras 29a, 29b, 29c de segmentación semántica. Dos máscaras 29a de segmentación semántica representan el ojo derecho y el ojo izquierdo del paciente, respectivamente. Otra máscara 29b de segmentación semántica representa la nariz del paciente y otra máscara 29c de segmentación semántica representa la boca del paciente.

35 Volviendo ahora a la figura 2, en la etapa g), se determina una transformación entre la nube de puntos fuente y la nube de puntos objetivo usando las máscaras de segmentación de la nube de puntos fuente y las máscaras de segmentación de la nube de puntos objetivo. En esta realización a modo de ejemplo, se determina una matriz de transformación 4x4 que incluye traslación y rotación y, cuando se aplica a la nube de puntos fuente, alinea los puntos de la nube de puntos fuente con los puntos de la nube de puntos objetivo. Por tanto, la matriz de transformación 4x4 determinada sirve para la transformación entre las diferentes poses de vista (posición y orientación) en las que se han adquirido la nube de puntos fuente y la nube de puntos objetivo.

40 En esta realización a modo de ejemplo, se usa una variante 4D del algoritmo de punto más cercano iterativo (ICP) para determinar la matriz de transformación 4x4. Las coordenadas de los puntos y las máscaras de segmentación de las nubes de puntos fuente y de puntos objetivo se usan como entradas para el algoritmo, en la que para cada punto de la nube de puntos respectiva, la clase de objeto (por ejemplo, nariz, oreja, boca) de la máscara de segmentación semántica correspondiente se interpreta como una cuarta dimensión del punto (además de las coordenadas 3D del punto). La función de optimización de ICP 4D está diseñada para favorecer transformaciones que hagan coincidir con precisión las máscaras de segmentación semántica correspondientes en ambas nubes de puntos (nariz y nariz, oreja y oreja, boca y boca, etc.).

50 La matriz de transformación 4x4 y los datos de formación de imágenes médicas se transfieren desde el servidor 21 a las gafas 5 inteligentes de realidad mixta.

55 En la etapa h), al menos una parte de los datos de formación de imágenes médicas, es decir, los datos de formación de imágenes IRM se visualizan en la pantalla 3 de presentación visual transparente óptica de las gafas 5 inteligentes (véase la figura 1), en la que los datos de formación de imágenes médicas se superponen sobre la vista real del objeto 15 de tratamiento y se alinean con el objeto 15 de tratamiento usando la transformación entre la nube de puntos fuente y la nube de puntos objetivo, es decir, usando la matriz de transformación 4x4 determinada en la etapa g).

60 De esta manera, puede darse respaldo a un cirujano que use un dispositivo 1 de presentación visual de realidad mixta durante la cirugía al visualizar de manera virtual la anatomía del objeto 15 de tratamiento, tal como lo indican las imágenes de IRM, alineadas con precisión con la vista del mundo real del objeto 15 de tratamiento.

65 En esta realización a modo de ejemplo, se ejecutan las etapas a) y h) mediante gafas 5 inteligentes de realidad mixta, mientras que se ejecutan las etapas b) a g) mediante el servidor 21. En otras realizaciones, pueden ejecutarse etapas adicionales o todas las etapas del método mediante gafas 5 inteligentes de realidad mixta.

**Lista de símbolos de referencia**

5	1	dispositivo de presentación visual de realidad mixta
	3	pantalla de presentación visual
	5	gafas inteligentes de realidad mixta
10	7	sensores
	9	cámara
15	11a, 11b	ordenador
	13a, 13b	memoria
	15	objeto de tratamiento
20	17	paciente
	19	interfaz
25	21	servidor
	23	nube de puntos objetivo
	25a, 25b, 25c	máscaras de segmentación de la nube de puntos objetivo
30	27	nube de puntos fuente
	29a, 29b, 29c	máscaras de segmentación de la nube de puntos fuente

**REIVINDICACIONES**

1. Método para controlar una pantalla de presentación visual de un dispositivo (1) de presentación visual de realidad mixta, comprendiendo el método al menos las siguientes etapas:
- 5
- a) proporcionar un conjunto de datos de imagen que comprende una pluralidad de imágenes de un objeto (15) de tratamiento, en el que el objeto (15) de tratamiento es el cuerpo de un paciente o una parte del mismo y las imágenes representan el objeto (15) de tratamiento desde diferentes perspectivas,
- 10
- b) generar una nube (23) de puntos objetivo 3D a partir del conjunto de datos de imagen, en el que la nube (23) de puntos objetivo comprende una multitud de puntos definidos en un sistema de coordenadas tridimensional y los puntos representan la superficie del objeto de tratamiento,
- 15
- c) determinar un número de máscaras (25a, 25b, 25c) de segmentación semántica en la nube (23) de puntos objetivo aplicando la segmentación semántica,
- 20
- d) proporcionar un conjunto de datos de formación de imágenes médicas que comprende datos de formación de imágenes médicas del objeto (15) de tratamiento,
- 25
- e) generar una nube (27) de puntos fuente 3D a partir del conjunto de datos de formación de imágenes médicas, en el que la nube (27) de puntos fuente comprende una multitud de puntos definidos en un sistema de coordenadas tridimensional y los puntos también representan la superficie del objeto de tratamiento,
- 30
- f) determinar un número de máscaras (29a, 29b, 29c) de segmentación semántica en la nube (27) de puntos fuente aplicando la segmentación semántica,
- 35
- g) determinar una transformación entre la nube (27) de puntos fuente y la nube (23) de puntos objetivo usando las máscaras (29a, 29b, 29c) de segmentación de la nube (27) de puntos fuente y las máscaras (25a, 25b, 25c) de segmentación de la nube (23) de puntos objetivo, y
- h) visualizar al menos una parte de los datos de formación de imágenes médicas en la pantalla (3) de presentación visual, en el que los datos de formación de imágenes médicas se superponen sobre el objeto (15) de tratamiento y se alinean con el objeto (15) de tratamiento usando la transformación entre la nube (27) de puntos fuente y la nube (23) de puntos objetivo,
- en el que la etapa c) comprende
- determinar un número de máscaras de segmentación semántica en las imágenes del conjunto de datos de imagen aplicando la segmentación semántica a las imágenes del conjunto de datos de imagen, y
  - determinar las máscaras (25a, 25b, 25c) de segmentación semántica en la nube (23) de puntos objetivo usando las máscaras de segmentación semántica en las imágenes del conjunto de datos de imagen.
- 45
- y/o la etapa f) comprende
- determinar un número de máscaras de segmentación semántica en los datos de formación de imágenes médicas del conjunto de datos de formación de imágenes médicas aplicando la segmentación semántica a los datos de formación de imágenes médicas del conjunto de datos de formación de imágenes médicas, y
  - determinar las máscaras (29a, 29b, 29c) de segmentación semántica en la nube (27) de puntos fuente usando las máscaras de segmentación semántica en los datos de formación de imágenes médicas del conjunto de datos de formación de imágenes médicas,
- 50
- 55
- en el que cada máscara de segmentación semántica se refiere a un segmento de la imagen que se ha determinado, usando la segmentación semántica, como perteneciente a la misma clase de objeto.
- 60
2. Método según la reivindicación 1, caracterizado porque la pantalla (3) de presentación visual está diseñada como una pantalla (3) de presentación visual transparente óptica, en particular como una pantalla (3) de presentación visual transparente óptica montada en la cabeza, y/o el dispositivo (1) de presentación visual de realidad mixta comprende o consiste en un dispositivo (1) de presentación visual de realidad mixta montado en la cabeza y/o gafas (5) inteligentes de realidad mixta.
- 65
3. Método según la reivindicación 1 ó 2, caracterizado porque los datos de formación de imágenes médicas se

generan usando al menos uno de los siguientes métodos de formación de imágenes médicas: formación de imágenes por resonancia magnética (IRM), tomografía computarizada (TC), radiografía, ecografía, endoscopia y/o formación de imágenes de medicina nuclear.

- 5 4. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque las máscaras (25a, 25b, 25c) de segmentación semántica en la nube (23) de puntos objetivo y/o las máscaras (29a, 29b, 29c) de segmentación semántica en la nube (27) de puntos fuente se determinan usando una red neuronal convolucional configurada para la segmentación semántica.
- 10 5. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque la etapa c), las máscaras de segmentación semántica en las imágenes del conjunto de datos de imagen se determinan usando una red neuronal convolucional configurada para la segmentación semántica.
- 15 6. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque el conjunto de datos de imagen comprende una pluralidad de imágenes visuales y/o imágenes de profundidad del objeto (15) de tratamiento y la nube (23) de puntos objetivo 3D se genera a partir de las imágenes visuales y/o las imágenes de profundidad, en particular usando un método de fotogrametría y/o un método de fusión de profundidad.
- 20 7. Método según la reivindicación 6, caracterizado porque la etapa c) comprende lo siguiente:
- determinar un número de máscaras de segmentación semántica en las imágenes visuales y/o en las imágenes de profundidad del conjunto de datos de imagen mediante la aplicación de la segmentación semántica a las imágenes visuales y/o las imágenes de profundidad del conjunto de datos de imagen, en particular usando una red neuronal convolucional configurada para la segmentación semántica, y
  - determinar las máscaras (25a, 25b, 25c) de segmentación semántica en la nube (23) de puntos objetivo usando las máscaras de segmentación semántica en las imágenes visuales y/o las imágenes de profundidad del conjunto de datos de imagen.
- 30 8. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque el conjunto de datos de imagen comprende una pluralidad de imágenes visuales e imágenes de profundidad del objeto (15) de tratamiento y la etapa b) comprende lo siguiente:
- generar una primera nube de puntos 3D a partir de las imágenes visuales del conjunto de datos de imagen,
  - generar una segunda nube de puntos 3D a partir de las imágenes de profundidad del conjunto de datos de imagen, y
  - generar la nube (23) de puntos objetivo 3D usando la primera nube de puntos 3D y la segunda nube de puntos 3D, en particular fusionando la primera nube de puntos 3D y la segunda nube de puntos 3D.
- 45 9. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque en la etapa f), las máscaras de segmentación semántica en los datos de formación de imágenes médicas del conjunto de datos de formación de imágenes médicas se determinan usando una red neuronal convolucional configurada para la segmentación semántica.
- 50 10. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque
- la etapa c) comprende determinar las máscaras (25a, 25b, 25c) de segmentación semántica en la nube (23) de puntos objetivo aplicando directamente la segmentación semántica a la nube (23) de puntos objetivo, en particular usando una red neuronal convolucional configurada para la segmentación semántica, y/o
  - la etapa f) comprende determinar las máscaras (29a, 29b, 29c) de segmentación semántica en la nube (27) de puntos fuente aplicando directamente la segmentación semántica a la nube (27) de puntos fuente, en particular usando una red neuronal convolucional configurada para la segmentación semántica.
- 60 11. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque la transformación entre la nube (27) de puntos fuente y la nube (23) de puntos objetivo se determina por medio de un algoritmo de punto más cercano iterativo (ICP) usando las coordenadas de los puntos de la nube (27) de puntos fuente y las máscaras (29a, 29b, 29c) de segmentación de la nube (27) de puntos fuente y las coordenadas de los puntos de la nube (23) de puntos objetivo y las máscaras (25a, 25b, 25c) de segmentación de la nube (23) de puntos objetivo.
- 65

12. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque
- 5 - la etapa a) comprende eliminar un fondo y/u otras partes irrelevantes de las imágenes del objeto (15) de tratamiento usando la segmentación semántica, en particular usando una red neuronal convolucional configurada para la segmentación semántica, y/o
- 10 - la etapa d) comprende eliminar un fondo y/u otras partes irrelevantes de los datos de formación de imágenes médicas del objeto (15) de tratamiento usando la segmentación semántica, en particular usando una red neuronal convolucional configurada para la segmentación semántica.
13. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque el conjunto de datos de formación de imágenes médicas comprende un modelo de formación de imágenes médicas 3D del objeto (15) de tratamiento, en el que el modelo de formación de imágenes médicas 3D se reconstruye a partir de una pluralidad de imágenes en sección 2D del objeto (15) de tratamiento generadas mediante un método de formación de imágenes médicas.
14. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque en la etapa a), el conjunto de datos de imagen que comprende las imágenes del objeto (15) de tratamiento se crea por medio de una cámara (9), en particular un sistema de cámara 3D, del dispositivo (1) de presentación visual de realidad mixta y/o por medio de un sensor (7) de profundidad, en particular por medio de un sensor (7) de profundidad de tiempo de vuelo, del dispositivo (1) de presentación visual de realidad mixta.
15. Método según la reivindicación 14, caracterizado porque
- 25 - en la etapa a), cuando las imágenes se crean por medio de la cámara (9) de la pantalla de presentación visual de realidad mixta, la posición de la cámara (9) se determina en un sistema de coordenadas tridimensional para cada imagen y se almacena como posición de cámara 3D y en la etapa b), la nube (23) de puntos objetivo se genera usando las posiciones de cámara 3D, y/o
- 30 - en la etapa a), cuando las imágenes se crean por medio de la cámara (9) de la pantalla de presentación visual de realidad mixta, la orientación de la cámara (9) se determina en un sistema de coordenadas tridimensional para cada imagen y se almacena como orientación de cámara 3D y en la etapa b), la nube (23) de puntos objetivo se genera usando las orientaciones de cámara 3D.
16. Método según cualquiera de las reivindicaciones anteriores, caracterizado porque cualquier posición, cualquier orientación y cualquier transformación determinadas en las etapas a) a h) se determina sin marcador y/o usando localización y mapeo simultáneos (SLAM).
17. Programa informático que tiene medios de código de programa adaptados para realizar un método según cualquiera de las reivindicaciones anteriores cuando el programa informático se ejecuta en un ordenador (11a, 11b).
18. Dispositivo (1) de presentación visual de realidad mixta que tiene una pantalla (3) de presentación visual, un ordenador (11a, 11b) y una memoria (13a, 13b), en el que el programa informático según la reivindicación anterior se almacena en la memoria (13a, 13b) y el ordenador (11a, 11b) está adaptado para ejecutar el programa informático.

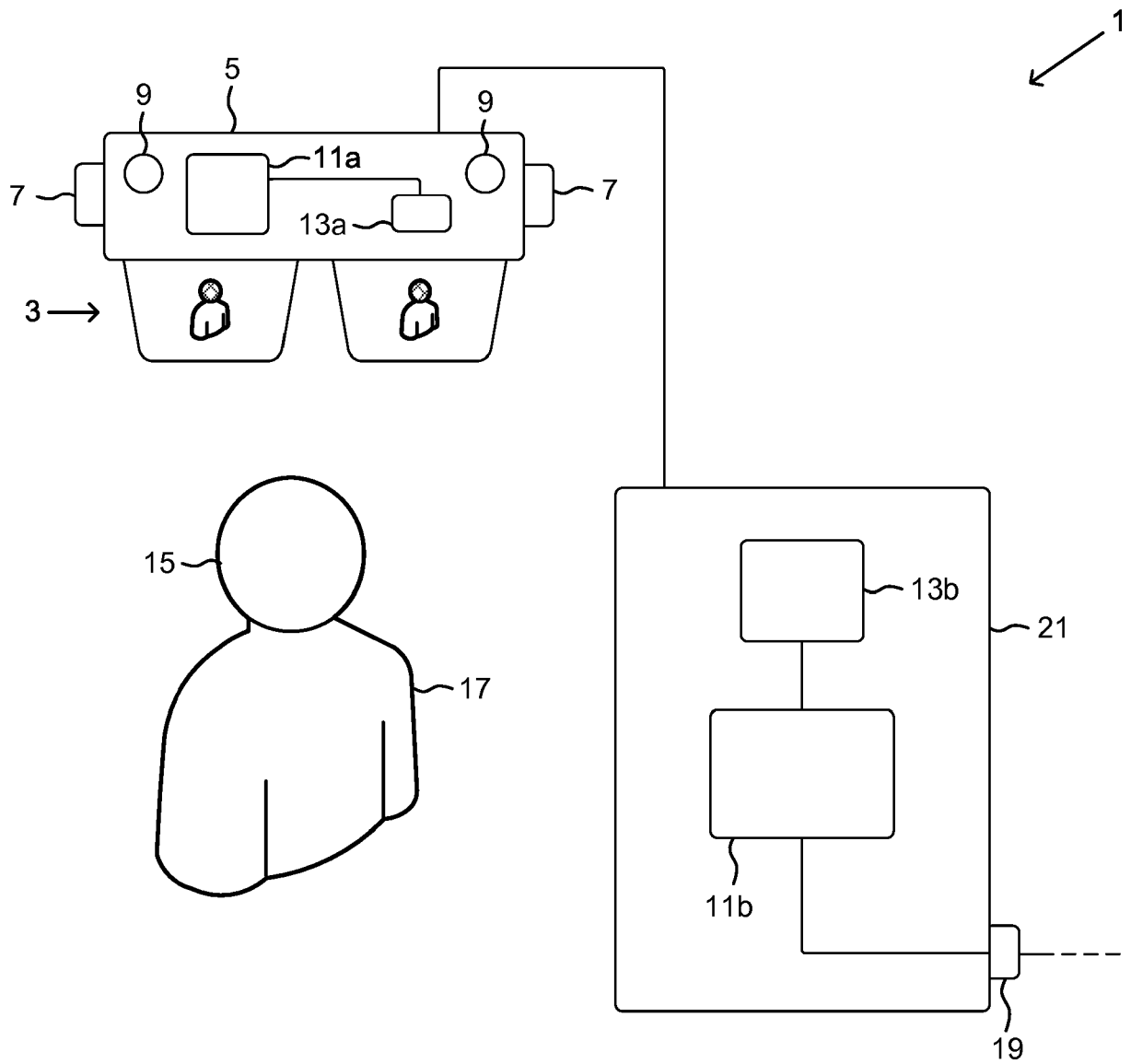
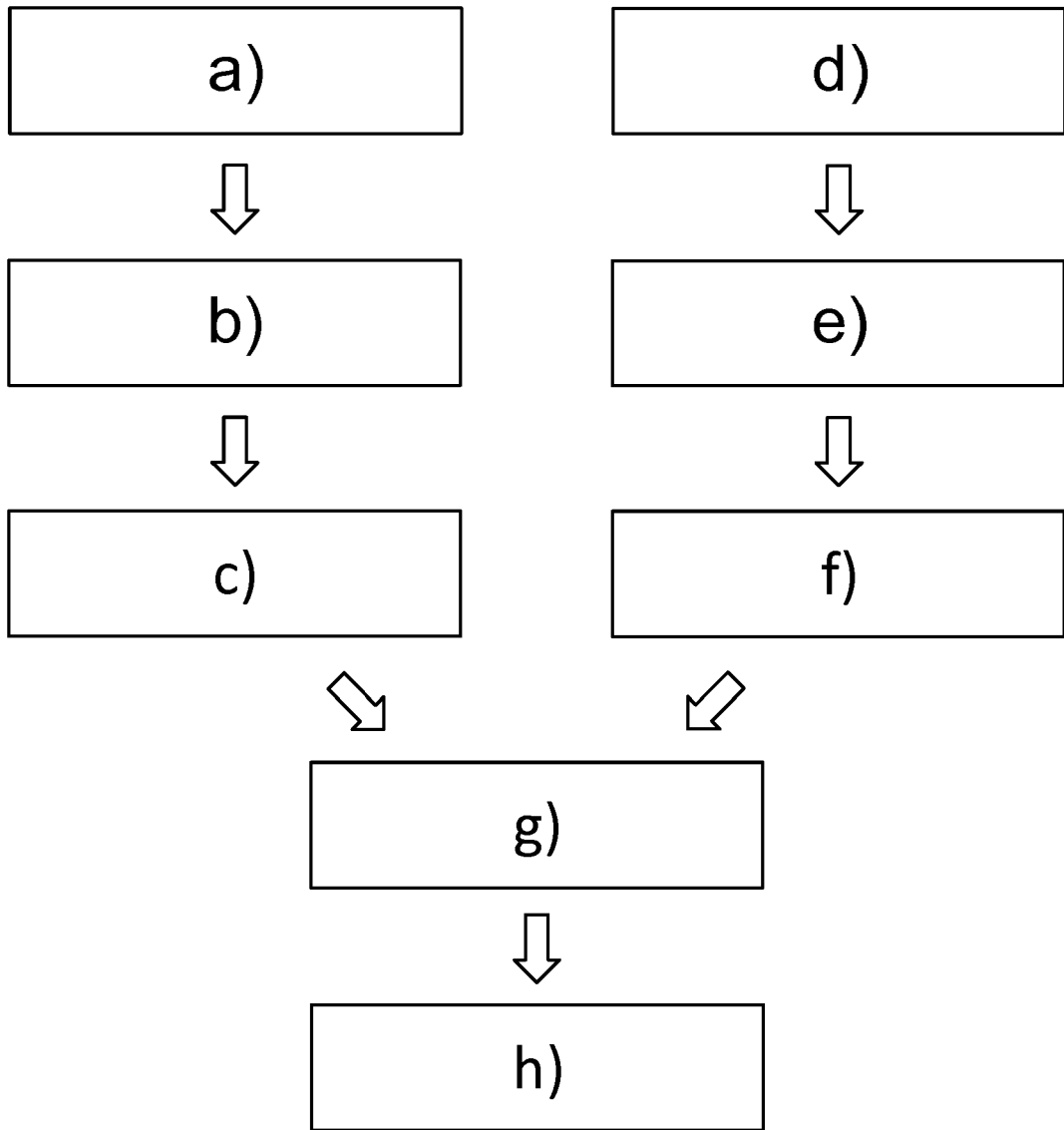


Fig. 1



**Fig. 2**

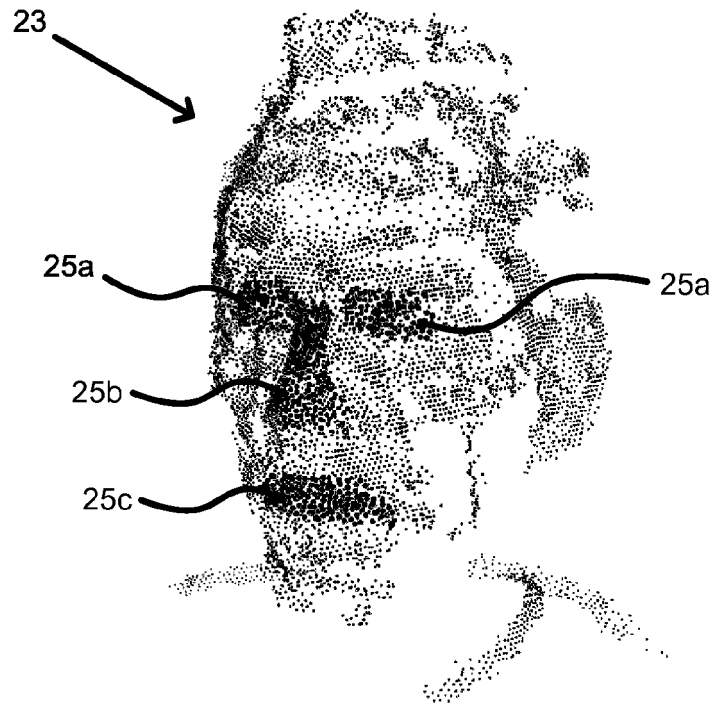


Fig. 3

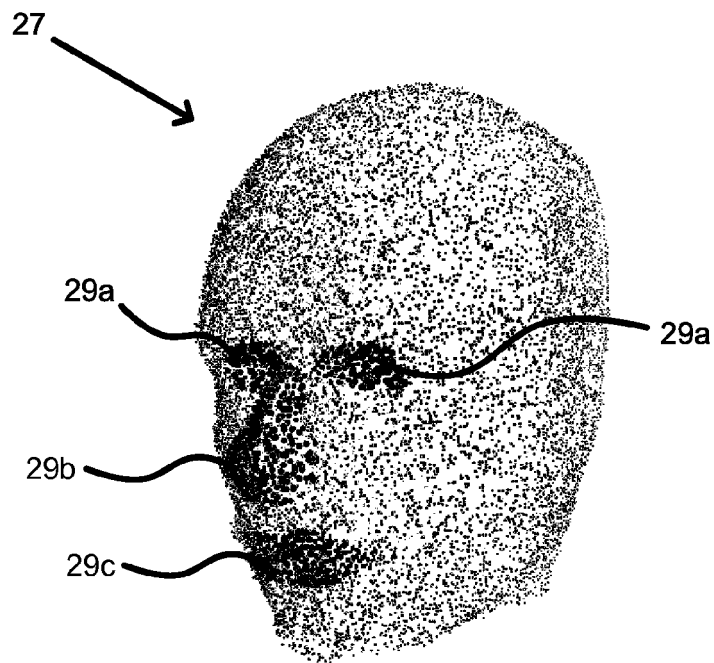


Fig. 4