

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 558 558**

51 Int. Cl.:

G06T 7/00 (2006.01)

G06T 7/20 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **16.12.2011 E 11805777 (7)**

97 Fecha y número de publicación de la concesión europea: **04.11.2015 EP 2791903**

54 Título: **Procedimiento y dispositivo para la estimación de una postura**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.02.2016

73 Titular/es:

**GESTIGON GMBH (100.0%)
Maria-Goeppert-Str. 9a
23562 Lübeck, DE**

72 Inventor/es:

**MARTINETZ, THOMAS;
EHLERS, KRISTIAN;
TIMM, FABIAN;
BARTH, ERHARDT y
KLEMENT, SASCHA**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 558 558 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 dispositivo para la estimación de una postura

5 **Campo de la invención**

La invención se refiere a un procedimiento y a un dispositivo para el análisis que puede realizarse en tiempo real de una secuencia de imágenes electrónicas para la estimación de las posiciones y orientaciones de un objeto móvil registrado en la secuencia de imágenes, es decir la estimación de la postura del objeto móvil. Además, la invención se refiere también a la implementación de la invención en software y en este contexto en un medio legible por ordenador que almacena instrucciones, cuya realización provoca la realización del procedimiento de acuerdo con la invención.

15 **Antecedentes de la técnica**

La estimación de las poses humanas mediante un ordenador forma la base de una interacción ser humano-máquina con control de gestos. Los gestos del cuerpo o de las manos se registran por cámaras, las imágenes digitales registradas se procesan en el ordenador y se interpretan como instrucciones que realiza finalmente el ordenador o un aparato accionado por el mismo. El usuario humano ya no necesita ningún aparato de entrada separado cuando éste controla los gestos que dan instrucciones.

Los campos de aplicación especialmente interesantes del control de gestos son por un lado el sector de las intervenciones médicas, en las que el médico cirujano le gustaría controlar directamente los aparatos auxiliares (por ejemplo dispositivos de obtención de imágenes tales como ultrasonido o TRM), sin embargo para la conservación de la esterilidad de sus manos no puede tocar ningún dispositivo de control, y por otro lado el sector de los terminales de información públicos o máquinas expendedoras de billetes que actualmente están equipadas aún con los paneles táctiles absolutamente no higiénicos. Otro campo de aplicación, que se ha explorado ya comercialmente, es el sector de los juegos informáticos.

El fin de un procedimiento de control de gestos ha de observarse en proporcionar a la imagen óptica de un sujeto un significado que puede interpretarse mecánicamente. Para ello se necesita un dispositivo que reproduzca al sujeto de manera electrónicamente evaluable, que compacte esta imagen con respecto a su contenido de información y finalmente que traduzca la imagen compactada del sujeto en una emisión interpretable mecánicamente. La emisión del dispositivo puede estar constituida por instrucciones de control para dispositivos conectados posteriormente, que van a controlarse. Sin embargo puede comprender también únicamente la información de imagen compactada que se alimenta a una unidad conectada posteriormente para la interpretación de esta información.

Un ejemplo de una información de imagen compactada es por ejemplo la emisión continua de las coordenadas de posición de la mano derecha del sujeto en un sistema de coordenadas 3D. A este respecto es suficiente con frecuencia emitir solo coordenadas de un punto individual para la posición de la mano, por ejemplo cuando se reproduce todo el cuerpo del sujeto. Si se reproduce el sujeto en su movimiento mediante una secuencia de imágenes, entonces proporciona el dispositivo mencionado anteriormente por ejemplo las coordenadas 3D de partes del cuerpo predeterminadas, que se modifican con el tiempo (durante el movimiento). Las coordenadas pueden servir como entradas de datos variables en un programa, que controla según esto por ejemplo una posición del cursor en una pantalla.

En la segmentación de imagen se eliminan todos los datos de imagen registrados (valores de medición) que no pueden asignarse al sujeto reproducido, o sea en particular elementos de imagen que se refieren al fondo. Tales elementos de imagen han de eliminarse de la evaluación posterior.

La segmentación de imagen es difícil con datos bidimensionales sobre todo cuando el usuario se reproduce delante de un fondo complejo (por ejemplo se mueven otras personas en el fondo) o cuando éste realiza gestos, con los que éste mueve las extremidades hacia la cámara de modo que éstas ocultan partes de su torso. Dado que debe tener lugar un control de gestos en tiempo real y debe ser posible una estimación de la postura habitualmente con una velocidad de imagen de vídeo de 25 Hz o más, es necesario que la segmentación de imagen pueda realizarse en el intervalo de algunos milisegundos. Para ello pueden usarse cámaras de detector de profundidad que puedan medir no solo, como las cámaras habituales, una imagen de claridad, sino adicionalmente la distancia de la cámara al objeto.

Una cámara de detector de profundidad conocida se designa como cámara *Time-of-Flight* (TOF), tiempo de vuelo. Ésta emite luz infrarroja cuya intensidad se modula en forma sinusoidal. En cada píxel se mide el desplazamiento de fases entre la luz emitida y la luz reflejada por el objeto. A partir de este desplazamiento de fases puede calcularse el tiempo de vuelo ("Time of Flight") de la luz y con ello la distancia de la cámara al punto del objeto. Una cámara TOF proporciona un mapa de profundidad que está registrado con una imagen de claridad (designada en la nomenclatura de TOF con frecuencia como imagen de amplitud).

Otro procedimiento para la obtención simultánea de valores de medición de imagen y de distancia se basa en luz estructurada que se irradia en el objeto que va a medirse y se refleja por éste. Una cámara detecta la luz reflejada (habitualmente con un ángulo distinto del ángulo de irradiación) y registra la modificación de la estructura de un patrón proyectado debido a la posición o extensión de la superficie del objeto reflectante. Por ejemplo puede
 5 calcularse una curvatura de la superficie reflectante a partir de la curvatura de una línea reflejada detectada por la cámara, que se ha proyectado originariamente como una recta sobre el objeto, o sea por consiguiente puede medirse una variable de distancia con respecto a un proyector y/o cámara. De manera similar es adecuado un haz de rayos espacialmente divergente, que proyecta puntos en una escena tridimensional, mediante registro de las reflexiones de puntos y determinación de las distancias entre éstas. En una superficie situada más cerca del
 10 proyector son las distancias entre puntos más bajas que en una superficie en el fondo de la imagen. Esto se usa para la medición de la distancia de superficies o zonas de superficies hacia el proyector.

Una cámara de detector de profundidad es según esto un dispositivo que además de una imagen de claridad bidimensional proporciona también una información de distancia para cada punto de objeto reproducido, de modo
 15 que se mide adicionalmente la posición de todos los puntos de objeto reproducidos a lo largo de un eje de profundidad (que habitualmente coincide con el eje óptico de la cámara). La imagen electrónica registrada con una cámara de detector de profundidad con información de distancia se designa también como imagen de dos dimensiones y media ($2\frac{1}{2}D$) de la escena. Los dispositivos mencionados anteriormente son solo ejemplos de cómo pueden generarse imágenes $2\frac{1}{2}D$ y no representan necesariamente una lista concluyente.

David Simon *et al.*, "Real-time 3-D Pose Estimation Using a High-Speed Range Sensor", Proceedings of the IEEE International Conference on Robotics and Automation, 1994, describen un sistema para la detección de las poses 3-D de un cuerpo sólido conformado de manera discrecional con una velocidad de hasta 10 Hz. Un modelo de red triangular del objeto que va a seguirse se genera fuera de línea con detectores de recorrido convencionales. Los
 25 datos en tiempo real del objeto se registran con un detector de recorrido VLSI de alta velocidad de CMU. La detección de las poses se consigue por medio de la reproducción de los datos de recorrido en tiempo real en el modelo de red triangular por medio del algoritmo de punto más cercano iterativo, *iterative closest point* (ICP), que se propuso en 1992 por Besl y McKay.

Greg Turk *et al.*, "Zippered polygon meshes from range images", Proceedings of the 21st annual conference on Computer graphics and interactive techniques, 1994, describen un procedimiento para la combinación de un grupo de imágenes de profundidad en una red poligonal individual, que describe completamente la parte detectable desde
 30 fuera de un objeto. El procedimiento comprende las etapas: 1) alinear las redes por medio de un algoritmo ICP iterativo, 2) reunir las redes adyacentes para formar una superficie continua que reproduzca correctamente la superficie del objeto y 3) calcular valores promedio localmente ponderados de las posiciones de superficie en todas las redes para obtener una geometría de superficie concordante. Conforme a los autores de la redacción se diferencia el procedimiento de los procedimientos conocidos en su modo de funcionamiento incremental.

El documento US 5.715.166 se refiere a un dispositivo para la detección de superficies. El dispositivo determina los puntos de datos de una superficie y almacena un modelo matemático de la superficie de modelo. Un microcontrolador registra el modelo matemático y los puntos de datos mediante (i) determinación de una cantidad de puntos de datos que contiene distintos puntos de los puntos de datos registrados, (ii) detección de una cantidad de puntos más cercanos que comprende para cada punto de la cantidad de puntos de datos un punto más cercano en el primer modelo matemático, (iii) registro de la cantidad de puntos de datos y de la cantidad de puntos más cercanos para obtener una cantidad de puntos registrada, (iv) cálculo de la distancia cuadrática promedio de los puntos de la cantidad de puntos registrada y de la cantidad de puntos más cercanos, (v) comparación de la distancia cuadrática promedio con un valor umbral predeterminado y (vi) cuando la distancia cuadrática promedio sobrepasa el valor umbral, (vi.i) nueva definición de la cantidad de puntos de datos como la cantidad de puntos registrada y (vi.ii) repetición de las etapas (ii), (iii), (iv) y (v).
 40
 45
 50

Del documento WO 2010/130245 A1 puede deducirse entre otras cosas cómo puede realizarse la segmentación de imagen de imágenes $2\frac{1}{2}D$ de manera adecuada. La segmentación de imagen clasifica los valores de claridad registrados por los píxeles de la cámara por los valores de distancia medidos al mismo tiempo y registrados con los píxeles. Solo los valores de claridad del primer plano permanecen en evaluación posterior, basándose en que el sujeto que va a observarse permanece lo más cercano a la cámara con el fin de una mejor visibilidad. Los valores de claridad del primer plano resultan por consiguiente mediante la reproducción de la superficie del cuerpo del sujeto. Por medio de los parámetros de proyección de la cámara en sí conocidos puede atribuirse a los puntos de objeto reproducidos entonces en cada caso un juego de coordenadas 3D. Entonces se obtiene una lista con coordenadas 3D que comprende todos los puntos del sujeto directamente visibles para la cámara. Dentro de esta "nube" de puntos en el espacio 3D se encuentra la verdadera persona, y dentro esta nube de puntos 3D se encuentran también las coordenadas de interés de las partes del cuerpo predeterminadas que se quieren determinar para el fin de un control de gestos.
 55
 60

La segunda etapa parcial de la compactación de información ha de observarse por tanto en que a partir de la nube de puntos 3D determinada mediante segmentación de imagen, que representa el sujeto se determina un juego reducido de coordenadas de puntos que describe la postura total del sujeto a ser posible bien y es adecuada para la
 65

interpretación mecánica. Esta etapa se designa también como estimación de la postura. Un objetivo de la estimación de la postura es a este respecto la robustez del juego de datos reducido, es decir pequeñas modificaciones de la postura humana deben conducir también solo a pequeñas modificaciones en los juegos de datos que describen la postura. En particular deben moverse las coordenadas que describen las partes corporales humanas según la posibilidad en trayectorias que discurren temporalmente de manera continua, de modo que en cada momento se proporcione una asignación unívoca de las coordenadas a estas partes corporales.

Un enfoque conocido y generalmente aceptado es la definición de un modelo de esqueleto del sujeto que debe ajustarse en la nube de puntos 3D de la manera más rápida posible.

Del documento WO 2010/130245 A1 resulta un procedimiento para la estimación de la postura que puede realizarse en tiempo real de secuencias de imágenes 2½D, en el que se propone un modelo de esqueleto que se ha explicado como una topología de nodos y aristas. Las aristas que pueden describirse como pares de nodos, codifican una estructura adyacente entre los nodos. El ajuste de los nodos en la nube de puntos determinada previamente se realiza mediante el uso de una regla de aprendizaje para el entrenamiento de un mapa de auto-organización (“*self-organizing map*”, SOM).

En el ejemplo de realización del documento WO 2010/130245 A1 se modela el tronco humano con una topología de 44 nodos y 61 aristas. La nube de puntos 3D que representa el sujeto comprende aproximadamente 6500 puntos de datos (representado en el espacio 3D real, en el que el sujeto observado tiene un tamaño definido dependiendo de su distancia a la cámara), de los cuales se usan aproximadamente el 10 % para el entrenamiento de un SOM. La totalidad de los nodos de la topología puede considerarse directamente como un SOM, mientras que la especificación de las aristas puede interpretarse como un requerimiento o una limitación especial en la regla de aprendizaje.

La topología se entrena de manera separada para cada imagen individual de una secuencia de vídeo, sirviendo el resultado de entrenamiento de una imagen individual al mismo tiempo para la inicialización del entrenamiento de la siguiente imagen individual de la secuencia. En la inicialización de la primera imagen individual de una secuencia se adapta el tamaño de la topología preferentemente mediante una única escala al tamaño del sujeto delante de la cámara y su centro de gravedad se desplaza al centro de gravedad de la nube de puntos 3D. Si el tamaño de la topología se selecciona una vez correctamente, no necesita adaptarse más durante el procedimiento en desarrollo, ya que el procedimiento funciona con escala invariante. El entrenamiento de las imágenes individuales se realiza mediante uso de una regla de aprendizaje patrón por patrón, *pattern-by-pattern* con las siguientes etapas:

- a. seleccionar de manera aleatoria un punto de datos X de la nube de puntos 3D;
- b. determinar aquel nodo de la topología que presenta la distancia mínima a X;
- c. determinar todos los nodos adyacentes del nodo determinado en b. de acuerdo con la especificación de arista de la topología;
- d. desplazar los nodos determinados en b. y c. hacia X (véase con respecto a esto las ecuaciones (2) y (3) del documento WO 2010/130245 A1), en el que
- e. los vectores de desplazamiento se multiplican con tasas de aprendizaje que para los nodos determinados en c. son exactamente la mitad de grande que para los nodos determinados en b. (véase con respecto a esto el documento WO 2010/130245 A1, pág. 13, párrafo 4);
- f. repetir las etapas a. a e. por un número predeterminado de etapas de aprendizaje con reducción gradual de las tasas de aprendizaje.

La especificación de un número máximo de etapas de aprendizaje para cada imagen individual es conveniente para realizar la estimación de la postura (es decir en este caso el ajuste del modelo de esqueleto en la nube de puntos 3D y la lectura de todas las posiciones de nodos de interés) dentro de un intervalo de tiempo predeterminado. De esta manera pueden analizarse secuencias de imagen también con velocidad de imagen de vídeo o más alta.

Aunque el algoritmo del documento WO 2010/130245 A1 cumple bien el objetivo de la estimación de la postura en tiempo real, sin embargo éste presenta aún algunos puntos débiles que se mencionan en parte en el propio documento. En particular en el análisis de escenas en las que el sujeto junta o cruza sus brazos delante del cuerpo, pueden producirse mediante la regla de aprendizaje interpretaciones erróneas (que pueden corregirse en el transcurso de otras iteraciones) cuando los nodos individuales se retiran ampliamente de sus propios vecinos en la topología. Este efecto se encuentra en el documento WO 2010/130245 A1 con un punto de anclaje en el torso de modelo e impide una condición secundaria de la regla de aprendizaje, los desplazamientos de nodos desde el punto de anclaje más allá de un valor umbral predeterminado.

Además se muestran en la enseñanza del documento WO 2010/130245 A1 también dificultades en la determinación de la posición exacta de las articulaciones humanas, de los hombros y caderas, que pueden representarse respectivamente mediante varios nodos distintos. El modelo de esqueleto esbozado en el documento WO 2010/130245 A1 presenta relativamente muchos nodos, cuyo número no puede reducirse sin embargo sin más hasta 20 o inferior, sin asumir fallos considerables en la estimación de la postura. Los sistemas que pueden obtenerse actualmente en el mercado para el control de gestos por medio de cámaras de detector de profundidad

funcionan ya con modelos de esqueleto diseñados más bien según la anatomía humana con 15-20 nodos. Mediante una reducción del número de nodos puede conseguirse también una velocidad de procesamiento más alta de las imágenes de la cámara.

5 Los modelos de esqueleto justificados anatómicamente son adecuados además para recurrir a patrones de movimiento almacenados (plantillas) para el registro de movimientos que transcurren más rápidamente y más complejos (por ejemplo el movimiento oscilante de un palo de golf). En estos casos busca el software de control de gestos la coincidencia más probable de la modificación de la postura detectada con un desarrollo de movimiento almacenado de antemano y aprovecha esta plantilla conocida para el verdadero control. En juegos de ordenador se
10 usa ya esta técnica, sin embargo requiere muchos recursos. No en último lugar, la elaboración de los datos de movimiento almacenados origina ya costes considerables.

El control de gestos por medio de entrenamiento de SOM por el contrario prescinde completamente de plantillas y apuesta más bien solo por la capacidad de registro que puede realizarse en tiempo real de la continuidad de
15 movimiento. Éste tiene el potencial de detectar de manera fiable también movimientos humanos rápidos debido a reglas de aprendizaje que pueden transformarse de manera eficaz, y conservar a este respecto al mismo tiempo la capacidad de uso universal, de modo que resulte superfluo una adaptación eventualmente costosa del software al objetivo de medición.

20 **Sumario de la invención**

Es por tanto el objetivo de la invención compactar la información de imagen digital del registro de la cámara de un objeto en particular para la estimación de la postura de modo que pueda planearse una determinación de la postura del objeto, en particular de una postura humana, en tiempo real con poco esfuerzo de cálculo y con ello de manera
25 más rápida y/o de manera más exacta.

El objetivo se soluciona mediante el objeto de las reivindicaciones independientes. Ciertas formas de realización ventajosas de la invención son objeto de las reivindicaciones dependientes.

30 Un aspecto de la invención es por consiguiente la compactación de la información que se realiza normalmente en dos etapas parciales: segmentación de imagen y estimación de la postura. En este contexto, la invención se refiere en particular a una mejora de la estimación de la postura. De acuerdo con la invención se propone una estimación de la postura por medio de entrenamiento de SOM, que puede funcionar con un modelo de esqueleto inspirado en la anatomía del objeto observado y con número de nodos reducido, pudiéndose asignar a cada nodo de modelo de
35 manera fiable continuamente una parte del cuerpo predeterminada. La anatomía del objeto, cuya postura debe detectarse, se modela a este respecto como modelo de esqueleto.

Un modelo de esqueleto que se ha descrito mediante solo un número bajo de puntos (a continuación nodos) en el espacio 3D, representa una buena compactación de la información de la información de imágenes, cuando las
40 coordenadas de los nodos describen en todo momento la posición de partes predeterminadas del objeto que se mueve. El modelo de esqueleto representa al mismo tiempo conocimientos previos sobre el objeto, estando definidos por ejemplo pares de nodos y eventualmente también tripletes de nodos en el modelo de esqueleto, que describen partes del objeto o eventualmente superficies del objeto continuas, que están contenidas en la información de imagen 2½D medida, o sea que son visibles para la cámara. El modelo de esqueleto ha de ajustarse de manera
45 rápida y precisa en la información de imágenes que corresponde al objeto. El ajuste se realiza entre dos imágenes de una secuencia de imágenes mediante el desplazamiento continuo de los nodos y la actualización del modelo de esqueleto en línea con la secuencia de imágenes. De las partes de objeto o eventualmente superficies de objeto continuas mencionadas anteriormente se parte de que éstas se mueven en su totalidad. De acuerdo con la invención, por tanto con determinadas condiciones previas se desplaza simultáneamente un par de nodos o
50 eventualmente un triplete de nodos. A este respecto ha de destacarse que las reglas de desplazamiento descritas a continuación no conservan necesariamente las distancias de los nodos de un par de nodos o eventualmente triplete de nodos uno con respecto a otro, sino que el desplazamiento puede conducir también a un aumento de las distancias de los nodos de un par de nodos o eventualmente de un triplete de nodos.

55 Una forma de realización de la invención se refiere a un procedimiento para la estimación de la postura de un objeto que se mueve (por ejemplo de un ser humano o de un robot) mediante cálculo por ordenador de desplazamientos de las coordenadas de posición en 3D de los nodos de un modelo de esqueleto, que se ajusta de manera continua en una secuencia de nubes de puntos 3D. Las coordenadas de nodos se encuentran tabuladas en una memoria de datos electrónicos y las nubes de puntos 3D se determinan a partir de imágenes electrónicamente registradas de
60 una cámara de detector de profundidad, que representan el ser humano que se mueve. El modelo de esqueleto usado en este procedimiento es una topología que presenta como elementos de topología N_1 nodos, N_2 aristas y N_3 triángulos con $N_1, N_2 > 0$ y $N_3 \geq 0$, en el que cada elemento de topología se describe mediante nodos, pares de nodos o tripletes de nodos y se asigna de manera fija a una parte del objeto (por ejemplo a una parte del cuerpo humano o a una parte del robot). El procedimiento se realiza por una unidad de cálculo y comprende las siguientes
65 etapas:

- a. seleccionar de manera aleatoria un punto de datos X de la nube de puntos 3D;
- b. calcular el punto de base de altura P con respecto a X con referencia a cada elemento de topología y determinar si P se encuentra respectivamente dentro del elemento de topología;
- 5 c. calcular la distancia de X a cada elemento de topología como norma del vector diferencial $X - P$;
- d. determinar aquel elemento de topología que presenta la distancia mínima a X de todos los elementos de topología, cuyo punto de base de altura P se encuentra dentro del elemento de topología;
- e. desplazar el elemento de topología determinado en la etapa d mediante desplazamiento de todos los nodos que determinan el elemento de topología en dirección del vector $X - P$, en el que el vector de desplazamiento para un nodo se multiplica por una tasa de aprendizaje y , en caso de que el elemento de topología sea una arista
- 10 o un triángulo, por un peso, que resulta del punto de base de altura P con respecto a X con referencia al elemento de topología determinado en la etapa d y
- f. repetir las etapas a. a e. por un número predeterminado de etapas de aprendizaje con reducción gradual de la tasa de aprendizaje;
- 15 g. actualizar las coordenadas de nodos en la tabla de la memoria de datos electrónicos tras K pases del número predeterminado de etapas de aprendizaje con $K \geq 1$;
- h. proporcionar al menos las coordenadas de nodos actualizadas en la tabla para el procesamiento posterior.

En otra forma de realización de la invención se representa en la etapa b el punto de base de altura P con referencia a un elemento de topología como una combinación lineal de los vectores de coordenadas de nodos que determinan el elemento de topología y a partir de los coeficientes de representación se determina si P se encuentra dentro del elemento de topología.

En otra forma de realización de la invención se calcula el peso en la etapa e a partir de los coeficientes de representación de P .

De acuerdo con una forma de realización de la invención, el número de repeticiones de las etapas a a e se encuentra entre 1000 y 5000, en particular entre 2000 y 3000. Ventajosamente puede encontrarse la tasa de aprendizaje entre el valor inicial 0,5 y el valor final 0,01.

Otra forma de realización de la invención propone un dispositivo para la estimación de la postura de un objeto que se mueve. Este dispositivo comprende una cámara de detector de profundidad, una memoria de datos electrónicos y una unidad de cálculo programable, en el que la memoria de datos almacena las imágenes electrónicas de la cámara de detector de profundidad y la unidad de cálculo está configurada para determinar nubes de puntos 3D que representan el objeto registrado a partir de estas imágenes temporalmente en línea con el registro de imágenes mediante la cámara. La memoria de datos almacena además una lista con coordenadas 3D para los nodos de un modelo de esqueleto, siendo el modelo de esqueleto una topología que presenta como elementos de topología $N1$ nodos, $N2$ aristas y $N3$ triángulos con $N1, N2 > 0$ y $N3 \geq 0$.

La unidad de cálculo puede leer y modificar estas coordenadas para nodos individuales, pares de nodos predeterminados y tripletes de nodos predeterminados como representación de los elementos de topología del modelo de esqueleto. La unidad de cálculo está configurada además para realizar, tras la determinación de una nube de puntos 3D, las siguientes etapas:

- a. seleccionar de manera aleatoria un **punto de datos X** de la nube de puntos 3D;
- b. calcular el punto de base de altura P con respecto a X con referencia a cada elemento de topología y determinar si P se encuentra respectivamente dentro del elemento de topología;
- c. calcular la distancia de X a cada elemento de topología como norma del vector diferencial $X - P$;
- d. determinar aquel elemento de topología que presenta la distancia mínima a X de todos los elementos de topología, en los que el punto de base de altura P se encuentra dentro del elemento de topología;
- 50 e. desplazar el elemento de topología determinado en la etapa d mediante desplazamiento de todos los nodos que determinan el elemento de topología en dirección del vector $X - P$, en el que el vector de desplazamiento para un nodo se multiplica por una tasa de aprendizaje y , en caso de que el elemento de topología sea una arista o un triángulo, por un peso, que resulta del punto de base de altura P con respecto a X con referencia al elemento de topología determinado en la etapa d y
- 55 f. repetir las etapas a. a e. por un número predeterminado de etapas de aprendizaje con reducción gradual de la tasa de aprendizaje;
- g. actualizar las coordenadas de nodos en la tabla de la memoria de datos electrónicos tras K pases del número predeterminado de etapas de aprendizaje con $K \geq 1$;
- 60 h. proporcionar al menos las coordenadas de nodos actualizadas en la tabla para el procesamiento posterior.

Además, otra forma de realización de la invención se refiere a un medio de almacenamiento legible por ordenador que está caracterizado por que almacena instrucciones que pueden realizarse por un microprocesador, que inducen a éste a realizar el procedimiento para la estimación de la postura según unas de las formas de realización de la invención descritas.

Descripción de las figuras

A continuación se describe más detalladamente la invención por medio de ejemplos de realización con relación a las figuras. Los elementos y detalles correspondientes de manera recíproca en las figuras están dotados de los mismos números de referencia. A este respecto muestra

- la figura 1: bocetos de modelos de esqueleto que pueden usarse: a) de acuerdo con el documento WO 2010/130245 A1, b) modelo de nodos y aristas, c) modelo de nodos, aristas y triángulos;
- la figura 2: a) ilustración de la regla de aprendizaje del documento WO 2010/130245 A1, b) interpretación geométrica de los pesos de la regla de aprendizaje e c) ilustración de la acción de la regla de aprendizaje para el caso de un desplazamiento de aristas;
- la figura 3: a) interpretación geométrica de los pesos de la regla de aprendizaje y b) representación de la acción de la regla de aprendizaje para el caso de un desplazamiento de triángulos;
- la figura 4: imágenes de ejemplo con nubes de puntos en 3D y modelos de esqueleto ajustados de acuerdo con la invención en estas nubes de puntos.

Descripción detallada de la invención

Un modelo de esqueleto con solo un número bajo de nodos sirve para una compactación eficaz de la información de imágenes, cuando las coordenadas de los nodos describen en cualquier momento la posición de partes predeterminadas del objeto que se mueve. El modelo de esqueleto se define aprovechando el conocimiento previo del objeto. Por ejemplo se conoce de antemano que partes del objeto o eventualmente superficies del objeto anexas son visibles para la cámara. Los elementos predeterminados del modelo de esqueleto, en particular pares de nodos o tripletes de nodos, pueden representar estas partes del objeto o superficies del objeto mencionadas anteriormente y pueden ajustarse en conjunto en las imágenes del objeto. El ajuste se realiza entre dos imágenes de una secuencia de imágenes mediante el desplazamiento continuo de los nodos y la actualización del modelo de esqueleto en línea con la secuencia de imágenes. Las partes del objeto o superficies del objeto anexas se mueven básicamente en conjunto y por tanto de acuerdo con la invención en determinadas condiciones previas se desplaza simultáneamente un par de nodos o eventualmente un triplete de nodos. Las reglas de desplazamiento no mantienen a este respecto necesariamente las distancias de los nodos de un par de nodos o eventualmente de un triplete de nodos uno con respecto a otro. Más bien puede conducir el desplazamiento también a un aumento de las distancias de los nodos de un par de nodos o eventualmente de un triplete de nodos. Prescindir de acuerdo con la invención del cumplimiento del mantenimiento de la distancia simplifica y acelera los cálculos necesarios de los vectores de desplazamiento y conduce no obstante a una buena estimación de la postura. La exactitud del ajuste del modelo de esqueleto en las imágenes del objeto aumenta con el número de iteraciones (etapas de aprendizaje).

La invención se describe a continuación de manera determinante conforme a la divulgación del documento WO 2010/130245 A1. Básicamente ha de mencionarse que la estimación de la postura descrita en el documento WO 2010/130245 A1 puede considerarse como punto de partida y al menos una forma de realización de la invención puede considerarse como un perfeccionamiento del procedimiento conocido a partir de este documento para la estimación de la postura.

En este contexto, en la siguiente descripción de las formas de realización de la invención se parte de que el registro de imágenes 2½D (individualmente o como secuencia) y la extracción de una nube de puntos 3D que representa al sujeto pueden realizarse tal como se describe en el documento WO 2010/130245 A1. En particular, la invención parte de que la facilitación de imágenes y nubes de puntos para la estimación de una postura movida pueda realizarse con una frecuencia superior a 25 Hz.

De acuerdo con la invención se usa ahora en lugar de la topología del documento WO 2010/130245 A1, que puede observarse para la comparación en la figura 1a), un modelo de esqueleto anatómicamente justificado. En particular es adecuado el modelo de la figura 1b), que asocia cada nodo de manera unívoca con un punto destacado de la anatomía humana (por ejemplo cabeza, hombro, codo, mano, cadera, pelvis, rodilla, pie). En la figura 1 c) está representada una variante del modelo de la figura 1 b), en la que se representa el torso mediante triángulos (respectivamente definidos por tres nodos que forman los vértices).

Mediante la identificación de los nodos con partes corporales humanas obtienen también las aristas de los modelos una interpretación anatómica. Así, por ejemplo en la figura 1 b), la arista que une los nodos 1 (mano derecha) y 2 (codo derecho) representa forzosamente el antebrazo derecho. Las aristas de la topología representan por consiguiente más de solo una relación de vecindad de los nodos. La aplicación de la regla de aprendizaje para el SOM del documento WO 2010/130245 A1 ya no puede hacer un buen ajuste del modelo en la nube de puntos 3D para un modelo de esqueleto de este tipo, con número de nodos claramente reducido, de modo que a continuación se describe un procedimiento para la estimación de la postura adaptado al uso de acuerdo con la invención de un modelo de esqueleto anatómicamente motivado.

En la figura 2 a) pueden observarse en el lado izquierdo dos nodos W_1 y W_2 que están unidos mediante una arista. El punto X designa un punto seleccionado de manera aleatoria de la nube de puntos 3D, en la que ha de ajustarse la

topología. Según la enseñanza del documento WO 2010/130245 A1 se determina en primer lugar el nodo del modelo próximo al punto X (en este caso: W_1) y se desplaza en $\varepsilon (X - W_1)$ hacia X . A este respecto es ε un número positivo real. El siguiente nodo adyacente en el sentido de la topología es el nodo W_2 unido con W_1 a través de la

arista. También éste se desplaza hacia X en $\frac{\varepsilon}{2}(X - W_2)$. Los nodos desplazados están mostrados en el lado

5 derecho. Mediante esta regla de aprendizaje se acorta continuamente la arista entre los nodos.

Si el modelo de esqueleto comprende muchos nodos como en el documento WO 2010/130245 A1, entonces se compensan en gran parte de nuevo los acortamientos en el transcurso de otras iteraciones y desplazamientos de nodos con el paso del tiempo. Sin embargo no es siempre posible de manera continua una atribución unívoca de

10 determinados nodos a determinadas partes del cuerpo.

Para el uso de las topologías anatómicas de la figura 1 b) o c) se formulan por tanto de acuerdo con la invención nuevas reglas de aprendizaje que permitan en cualquier momento la asociación de nodos, aristas y eventualmente triángulos de la topología con partes del cuerpo del sujeto.

15 Una topología anatómica (o también: modelo de esqueleto anatómico) en el sentido de la invención está constituido al menos por un primer número N_1 de nodos que se asocian con puntos corporales humanos, y un segundo número N_2 de aristas que están interpretados como una selección de pares de nodos mencionados anteriormente, que satisface circunstancias anatómicas.

20 Para la aclaración: en la figura 1 b), el modelo comprende los nodos W_j con $j = 1, \dots, 17$ así como 17 aristas adicionales que están representadas mediante una selección de pares de nodos. Por ejemplo, los pares (W_1, W_2) , (W_2, W_3) o (W_4, W_8) pertenecen a las aristas del modelo, mientras que por ejemplo los pares (W_2, W_6) o (W_{10}, W_{11}) no representan aristas.

25 En una configuración ventajosa de la invención, un tercer número N_3 de triángulos puede ser parte constituyente del modelo anatómico. Un triángulo se describe mediante un triplete de nodos, designando los nodos los vértices del triángulo.

30 En la figura 1 c), el modelo está constituido entre otras cosas por los tres triángulos descritos mediante el triplete de nodos (W_3, W_4, W_9) , (W_9, W_4, W_{12}) y (W_4, W_5, W_{12}) .

35 La inserción de triángulos en el modelo anatómico es ventajosa especialmente para la modelación de objetos (que corresponden por ejemplo a regiones corporales), que presentan solo una movilidad interior más baja y se mueven habitualmente en la imagen de modo que la posición relativa de los nodos que forman los vértices de los triángulos se modifica solo poco uno con respecto a otro. Esto se aplica por ejemplo al torso de un sujeto en un registro total, sin embargo puede referirse también por ejemplo en caso de registros grandes de una mano a la palma de la mano o al dorso de la mano. Eventualmente, el usuario del procedimiento de la estimación de la postura deba decidir siempre incluso qué objeto debe considerarse o qué topología es adecuada y/o parece especialmente favorable para

40 su fin concreto.

La presente invención proporciona de acuerdo con una forma de realización reglas de aprendizaje de SOM para topologías anatómicas de N_1 nodos, N_2 aristas y N_3 triángulos para el ajuste en nubes de puntos en 3D, aplicándose $N_1, N_2 > 0$ y $N_3 \geq 0$. Los nodos, las aristas y los triángulos se reúnen a continuación con el término elementos de

45 topología.

El SOM se identifica con los nodos de la topología. Las reglas de aprendizaje se refieren a desplazar las posiciones de nodos de modo que la topología se ajuste en una nube de puntos 3D. La especificación de aristas y eventualmente triángulos significa además que algunos pares y eventualmente algunos tripletes de nodos tienen como consecuencia respectivamente reglas de aprendizaje enlazadas entre sí.

50

Las reglas de aprendizaje de acuerdo con una forma de realización de la invención se forman según el siguiente concepto. Partiendo de la existencia de una secuencia de imágenes, para la que debe realizarse una estimación de la postura continua, se actualiza el modelo de esqueleto, que es una lista de posiciones de nodos (descritas en un sistema de coordenadas 3D), pares de nodos y eventualmente tripletes de nodos, respectivamente en caso de existencia de una nueva imagen de la secuencia. Tan pronto como se haya registrado la nueva imagen por la cámara de detector de profundidad y se haya transformado por medio de la segmentación de imágenes y proyección en una lista de coordenadas 3D (nube de puntos 3D) para puntos de la superficie del objeto considerado (por ejemplo de todo el sujeto, su tronco, su mano, etc.) se realiza el entrenamiento de SOM debido a que se selecciona de manera aleatoria un punto individual de la nube de puntos y se desplaza el elemento de topología próximo a este punto hacia el punto seleccionado. El desplazamiento se realiza mediante adición de vector en el espacio 3D, de manera que se modifican o se actualizan las posiciones de nodos individuales en la lista mencionada anteriormente de las posiciones de nodos.

60

Tras este desplazamiento se selecciona el siguiente punto de la nube de puntos 3D de manera aleatoria y se desplaza un elemento de topología (muy probablemente otro) próximo al punto ahora seleccionado a este punto. La selección de puntos y el desplazamiento se repite por un número predeterminado de etapas, reduciéndose en cada etapa la extensión de desplazamiento general. Tras un número de etapas predeterminado suficientemente alto, se actualiza finalmente el modelo de esqueleto para la nueva imagen.

Preferentemente se realizan al menos 1000 y como máximo 5000 desplazamientos de elementos de topología para el ajuste del modelo de esqueleto en una nube de puntos 3D. De manera especialmente preferente se encuentra el número de desplazamientos entre 2000 y 3000. Varios 1000 de desplazamientos de posiciones de nodos pueden realizarse con ordenadores actuales en el transcurso de algunos milisegundos.

Según esto es posible que el elemento de topología próximo a un punto seleccionado no sea un nodo. Las aristas o los triángulos se desplazan debido a que todos los nodos que determinan una arista o un triángulo se desplazan de manera enlazada entre sí. El desplazamiento puede afectar por tanto, dependiendo del elemento de topología que va a desplazarse, a dos o tres nodos al mismo tiempo.

A continuación se explica en más detalle la determinación del elemento de topología próximo, y se mencionan las reglas de desplazamiento concretas para los elementos de topología.

Con respecto a un punto X de la nube de puntos 3D (o sea con respecto a un punto de coordenadas en el espacio 3D, que tras la reproducción del objeto observado con una cámara de detector de profundidad y la posterior segmentación de imagen y proyección representa un punto de la superficie corporal el sujeto y se selecciona aleatoriamente de la totalidad de estos puntos) se determinan en primer lugar los puntos de base de altura de X con referencia a todos los elementos de topología. El punto de base de altura P de X con referencia a un elemento de topología es aquel punto del subespacio del espacio 3D fijado por el elemento de topología que se encuentra lo más próximo al punto X . A este respecto, los nodos fijan subespacios de dimensión cero que contienen solo los propios nodos. Las aristas fijan rectas en el espacio 3D que discurren por los nodos que definen las aristas. Los triángulos fijan planos en el espacio 3D que contienen los nodos que definen los triángulos.

El punto próximo al punto X de un subespacio fijado por un elemento de topología se calcula con ayuda de una dimensión de distancia. Las distancias en el espacio 3D pueden determinarse con normas discretionales. Preferentemente se usa la norma euclídea (también: norma L_2 o distancia de Pitágoras), sin embargo pueden usarse también otras dimensiones de distancia.

El punto de base de altura P se encuentra en un nodo cuando el elemento de topología es un nodo y éste coincide a este respecto con justamente este nodo.

El punto de base de altura P se encuentra en una recta en el espacio 3D cuando el elemento de topología es una arista.

$$(1) \quad P = W + \alpha \Delta W$$

A este respecto designa W un primer nodo (discrecional) de la arista, ΔW el vector diferencial entre el segundo y el primer nodo de la arista y α un número real. Para la aclaración: la arista se describe mediante el par de nodos $(W, W + \Delta W)$.

El punto de base de altura P se encuentra en un plano en el espacio 3D cuando el elemento de topología es un triángulo. El punto P puede representarse como

$$(2) \quad P = W + \sigma_1 \Delta W_1 + \sigma_2 \Delta W_2$$

con W como un primer nodo (discrecional) del triángulo, ΔW_1 , ΔW_2 como vectores diferenciales entre el segundo o el tercer nodo y el primer nodo del triángulo y σ_1 , σ_2 como números reales. Para la aclaración: el triángulo se describe mediante el triplete de nodos $(W, W + \Delta W_1, W + \Delta W_2)$.

Los coeficientes α , σ_1 , σ_2 de las ecuaciones (1) y (2) se designarán a continuación como "coeficientes de representación conformes a la topología" de un punto de base de altura. A este respecto se trata formalmente de los componentes del vector P con referencia a una base no normalizada y eventualmente de ángulos oblicuos del subespacio del espacio 3D fijado por el elemento de topología para el que se determina P . También puede decirse que el punto de base de altura P con referencia a un elemento de topología se representa como una combinación lineal de los vectores de coordenadas de nodos que establecen el elemento de topología.

Los puntos de base de altura de un punto X con referencia a los elementos de topología se someten a prueba en

consecuencia para determinar si se encuentran *dentro* de estos elementos de topología.

Todos los puntos de base de altura con referencia a nodos se encuentran por definición *dentro* de los nodos. Igualmente por definición es el coeficiente de representación conforme a la topología de un punto de base de altura con referencia a un nodo siempre uno.

Un punto de base de altura con referencia a una arista se encuentra exactamente entonces *dentro* de la arista, cuando éste se encuentra entre los nodos que definen la arista. Esto significa que un punto de base de altura entonces y solo entonces se encuentra dentro de la arista cuando su coeficiente de representación α se encuentra entre 0 y 1.

Un punto de base de altura con referencia a un triángulo se encuentra exactamente entonces *dentro* del triángulo cuando éste se encuentra en el interior de la superficie del triángulo que se define por los nodos. Este es entonces y solo entonces el caso, para un punto de base de altura con referencia a un triángulo, cuando sus coeficientes de representación σ_1 , σ_2 y su suma $\sigma_1 + \sigma_2$ se encuentran entre 0 y 1.

La distancia de un punto X de la nube de puntos 3D a un elemento de topología puede calcularse como una norma del vector diferencial D entre X y el punto de base de altura P , o sea en este caso $D := X - P$. Preferentemente se usa la norma euclídea. Con ello se calcula la distancia

$$(3) \quad d = \|D\|_2 = \|X - P\|_2 = \sqrt{(x_1 - p_1)^2 + (x_2 - p_2)^2 + (x_3 - p_3)^2}$$

con $x_1, x_2, x_3, p_1, p_2, p_3$ como componentes de valor real de los **vectores** X y P con referencia al sistema de coordenadas 3D convencional, en el que se han descrito tanto la nube de puntos 3D como el modelo de esqueleto, para cada uno de los elementos de topología.

A continuación se usa además la distancia del punto X a un elemento de topología solo entonces cuando el punto de base de altura P con referencia a este elemento de topología se encuentra *dentro* del elemento de topología. Si esto no es cierto, se desestima el punto de base de altura y o bien no tiene lugar ningún cálculo de la distancia o se ignora a continuación la distancia calculada.

Esta selección garantiza que se consideren en el procesamiento posterior solo aquellas distancias que sirvan realmente para el fin del ajuste del modelo de esqueleto. Por ejemplo, el punto X podría encontrarse exactamente en un plano en el espacio 3D, que se fija por un triángulo que representa el tronco del sujeto. El punto de base de altura P coincide entonces con X y la distancia d es cero. Cuando el punto se encuentra, sin embargo, al mismo tiempo en la proximidad de la posición del nodo que representa la mano derecha, y el brazo derecho se mantiene extendido alejándose del cuerpo, entonces debe usarse la regla de desplazamiento en el nodo y no en el triángulo del tronco, también cuando la distancia numérica de su plano a X es más pequeña que la del nodo a X .

De la totalidad de todas las distancias calculadas (y que permanecen en la evaluación) se identifica mediante consulta del mínimo aquel elemento de topología que se encuentra lo más próximo al punto X , o sea presenta la distancia más pequeña a X .

El elemento de topología identificado se desplaza ahora de acuerdo con la invención, desplazándose los nodos que definen el elemento de topología eventualmente de manera conjunta.

Cuando el elemento de topología próximo al punto X es un nodo, entonces se realiza el desplazamiento de acuerdo con

$$(4) \quad W \xrightarrow{\text{actualización}} W + \varepsilon_t D$$

tal como se conoce ya por el estado de la técnica. En el documento WO 2010/130245 A1 (allí la ecuación (4)) se explica también la tasa de aprendizaje dependiente de la etapa de iteración

$$(5) \quad \varepsilon_t = \varepsilon_i \left(\frac{\varepsilon_f}{\varepsilon_i} \right)^{\frac{t}{t_{\max}}}$$

A este respecto son ε_i y ε_f valores iniciales y finales predeterminados de la tasa de aprendizaje y t_{\max} es el número máximo predeterminado de etapas de aprendizaje. El índice de recorrido t incrementa las etapas de aprendizaje (iteraciones) hasta t_{\max} . Para cada punto X nuevamente seleccionado de la nube de puntos 3D se eleva el numerador t en uno hasta que alcanza t_{\max} . Las determinaciones preferentes para la tasa de aprendizaje son $\varepsilon_i = 0,5$ y $\varepsilon_f = 0,01$.

Cuando el elemento de topología próximo al punto X es una arista, entonces se realiza el desplazamiento de

acuerdo con

$$(6) \quad \mathbf{W} \xrightarrow{\text{actualización}} \mathbf{W} + \varepsilon_t(1 - \alpha)\mathbf{D}$$

$$(7) \quad \mathbf{W} + \Delta\mathbf{W} \xrightarrow{\text{actualización}} \mathbf{W} + \Delta\mathbf{W} + \varepsilon_t\alpha\mathbf{D}$$

5

En la figura 2 b) están mostrados la arista ($\mathbf{W}, \mathbf{W} + \Delta\mathbf{W}$), un punto \mathbf{X} de la nube de puntos 3D y su correspondiente punto de base de altura \mathbf{P} con referencia a la arista así como el vector diferencial $\mathbf{D} = \mathbf{X} - \mathbf{P}$. El coeficiente de representación conforme a la topología α ha de considerarse como aquella proporción de la longitud de arista $\|\Delta\mathbf{W}\|_2$, por la que ha de irse partiendo del nodo \mathbf{W} hacia $\Delta\mathbf{W}$ para alcanzar el punto \mathbf{P} . Evidentemente en la figura 2 b) es $0 < \alpha < 1$, y el punto de base de altura se encuentra por consiguiente dentro de la arista.

La figura 2 c) esboza el desplazamiento de los dos nodos. Se mueven ambos en dirección del vector $\mathbf{D} = \mathbf{X} - \mathbf{P}$, o sea no (tal como en el estado de la técnica) directamente hacia \mathbf{X} . Si el punto \mathbf{X} se encuentra más próximo al nodo \mathbf{W} que al nodo $\mathbf{W} + \Delta\mathbf{W}$, entonces se aplica esto también para el punto de base de altura \mathbf{P} , y como consecuencia es $\alpha < 0,5$. El nodo \mathbf{W} se desplaza entonces más que el nodo $\mathbf{W} + \Delta\mathbf{W}$. De esta manera se aproxima toda la arista al punto \mathbf{X} de manera ponderada según la distancia de \mathbf{X} o \mathbf{P} a los nodos.

Cuando el elemento de topología próximo al punto \mathbf{X} es un triángulo, entonces se realiza el desplazamiento de acuerdo con

$$(8) \quad \mathbf{W} \xrightarrow{\text{actualización}} \mathbf{W} + \varepsilon_t(1 - \sigma_1 - \sigma_2)\mathbf{D}$$

$$(9) \quad \mathbf{W} + \Delta\mathbf{W}_1 \xrightarrow{\text{actualización}} \mathbf{W} + \Delta\mathbf{W}_1 + \varepsilon_t\sigma_1\mathbf{D}$$

$$(10) \quad \mathbf{W} + \Delta\mathbf{W}_2 \xrightarrow{\text{actualización}} \mathbf{W} + \Delta\mathbf{W}_2 + \varepsilon_t\sigma_2\mathbf{D}$$

La figura 3 ilustra la regla de desplazamiento para triángulos.

En la figura 3 a) están representados el triángulo de partida del modelo de esqueleto y el punto \mathbf{X} seleccionado de manera aleatoria. El punto de base de altura \mathbf{P} que pertenece a \mathbf{X} se encuentra en el plano del triángulo y puede representarse como una combinación lineal de los vectores de coordenadas de nodos de acuerdo con la ecuación (2). Los coeficientes de representación σ_1, σ_2 se han descrito a modo de explicación en los lados del triángulo. Éstos pueden interpretarse de manera análoga al coeficiente α en la figura 2 b) como proporciones de las longitudes de lados del triángulo. El punto de base de altura se encuentra dentro del triángulo, de modo que la distancia de \mathbf{X} al triángulo se determina como norma del vector $\mathbf{D} = \mathbf{X} - \mathbf{P}$ y se usa en el cálculo posterior. Cuando resulta esta distancia como la distancia más pequeña de \mathbf{X} a todos los elementos de topología del modelo de esqueleto, entonces se desplaza el triángulo.

El triángulo desplazado está esbozado en la figura 3 b). También en este caso se desplazan todos los nodos (y con ellos también todos los puntos en los lados del triángulo y en la superficie del triángulo) de manera correspondientemente ponderada con la distancia original de los nodos a los puntos \mathbf{X} y \mathbf{P} . Esta ponderación con la distancia es importante para la eficacia de la estimación de la postura, dado que evita fallos innecesarios. Justo en la figura 3 b) puede reconocerse fácilmente la ventaja: si concretamente el punto \mathbf{P} se encuentra muy próximo a uno de los tres nodos que establecen el triángulo, entonces se realiza el desplazamiento del triángulo de manera que se mueva casi solo este nodo próximo, mientras que los otros dos mantienen esencialmente su posición. Por consiguiente existe un "paso fluido" entre el desplazamiento de nodos, aristas y triángulos en el procedimiento descrito en el presente documento.

Puede distinguirse bien en la figura 2 c) que una arista ($\mathbf{W}, \mathbf{W} + \Delta\mathbf{W}$) mediante uso de las reglas de aprendizaje (6) y (7) no puede acortarse de ningún modo, de todas formas puede sin embargo alargarse. Lo mismo se aplica para los lados de un triángulo ($\mathbf{W}, \mathbf{W} + \Delta\mathbf{W}_1, \mathbf{W} + \Delta\mathbf{W}_2$) con el uso de las reglas de aprendizaje de las ecuaciones (8) a (10).

Para no obtener en el transcurso del entrenamiento longitudes de aristas y longitudes de lados de triángulo desbordantes, se introduce en otra forma de realización de la invención un "parámetro de contracción" δ . Éste puede

fijarse por ejemplo en $\delta= 0,05$. Las reglas de aprendizaje se modifican al mismo tiempo de modo que los nodos se muevan durante el desplazamiento de manera insignificante uno con respecto a otro.

En el lugar de las reglas (6) y (7) se usan preferentemente

5

$$(11) \quad \mathbf{W} \xrightarrow{\text{actualización}} \mathbf{W} + \delta\Delta\mathbf{W} + \varepsilon_t(1 - \alpha)\mathbf{D}$$

$$(12) \quad \mathbf{W} + \Delta\mathbf{W} \xrightarrow{\text{actualización}} \mathbf{W} + (1 - \delta)\Delta\mathbf{W} + \varepsilon_t\alpha\mathbf{D}$$

Y en el lugar de las ecuaciones de las reglas (8) a (10) se usan

$$(13) \quad \mathbf{W} \xrightarrow{\text{actualización}} \mathbf{W} + \delta\Delta\mathbf{W}_1 + \delta\Delta\mathbf{W}_2 + \varepsilon_t(1 - \sigma_1 - \sigma_2)\mathbf{D}$$

$$(14) \quad \mathbf{W} + \Delta\mathbf{W}_1 \xrightarrow{\text{actualización}} \mathbf{W} + (1 - \delta)\Delta\mathbf{W}_1 + \delta(\Delta\mathbf{W}_2 - \Delta\mathbf{W}_1) + \varepsilon_t\sigma_1\mathbf{D}$$

$$(15) \quad \mathbf{W} + \Delta\mathbf{W}_2 \xrightarrow{\text{actualización}} \mathbf{W} + (1 - \delta)\Delta\mathbf{W}_2 + \delta(\Delta\mathbf{W}_1 - \Delta\mathbf{W}_2) + \varepsilon_t\sigma_2\mathbf{D}$$

10

El procedimiento descrito puede usarse para la estimación de la postura humana mediante el cálculo de desplazamientos de nodos de un modelo de esqueleto, que se inspira como en la figura 1 en la anatomía humana. Sin embargo es evidente que el mismo procedimiento pueda usarse también en los movimientos de un animal o de un robot que se mueve. La invención debe comprender la estimación de la postura de todos los objetos que están constituidos por partes unidas entre sí y móviles una contra otra y para los que puede definirse un modelo de esqueleto de nodos, aristas y eventualmente triángulos de manera conveniente.

15

El modelo de esqueleto es una tabla con coordenadas espaciales 3D para los nodos de la topología así como una lista con pares de nodos predeterminados y eventualmente tripletes de nodos para la determinación de aristas y triángulos. La lista de los pares de nodos y tripletes de nodos es invariable, únicamente las coordenadas de nodos pueden variarse durante la realización de la estimación de la postura. La tabla con coordenadas de nodos está a disposición en una memoria de datos electrónicos y puede leerse por una unidad de cálculo. La unidad de cálculo, por medio de una imagen igualmente almacenada de la cámara de detector de profundidad y de una nube de puntos 3D determinada a partir de la misma determina los desplazamientos de las posiciones de nodos para el ajuste continuo del modelo de esqueleto en las nubes de puntos.

20

25

El ajuste del modelo anatómico en la nube de puntos (por consiguiente el entrenamiento del SOM) se realiza de acuerdo con una forma de realización de la invención en resumen mediante:

30

- a. seleccionar de manera aleatoria un punto de datos \mathbf{X} de la nube de puntos;
- b. calcular el punto de base de altura \mathbf{P} con respecto a \mathbf{X} con referencia a un elemento de topología y determinar su al menos un coeficiente de representación conforme a la topología;
- c. desestimar un punto de base de altura cuando éste no se encuentre dentro del elemento de topología;
- d. calcular la distancia de \mathbf{X} al elemento de topología como una norma del vector diferencial $\mathbf{X} - \mathbf{P}$,
- e. repetir las etapas b. a d. para todos los elementos de topología del modelo anatómico;
- f. determinar aquel elemento de topología, que presenta la distancia mínima con respecto a \mathbf{X} ;
- g. desplazar el elemento de topología determinado en f. mediante desplazamiento de todos los modos que determinan el elemento de topología en dirección del vector $\mathbf{X} - \mathbf{P}$, multiplicándose los vectores de desplazamiento por una tasa de aprendizaje y por pesos que resultan de los coeficientes de representación conformes a la topología del punto de base de altura \mathbf{P} con respecto a \mathbf{X} con referencia al elemento de topología determinado en e. y
- h. repeticiones de las etapas a. a g. por un número predeterminado de etapas de aprendizaje con reducción gradual de la tasa de aprendizaje.

35

40

45

El modo exacto de cómo resultan las ponderosidades mencionadas en la etapa g a partir de los coeficientes de representación conformes a la topología, puede deducirse de las ecuaciones (6) a (10) o como alternativa (11) a (15), comprendiendo estas últimas un parámetro adicional, predeterminado δ . Para el desplazamiento de elementos de topología que son nodos, de acuerdo con la ecuación (4) son las ponderosidades por definición uno.

50

Si se alcanza el número predeterminado de etapas de aprendizaje, ha calculado la unidad de cálculo para todos los nodos del modelo de esqueleto coordenadas desplazadas. Las coordenadas desplazadas se diferencian por regla

general para todos los nodos de las coordenadas de nodos originariamente tabuladas, sin embargo pueden ser idénticas a éstas en casos excepcionales también para nodos individuales.

5 Las coordenadas de nodos desplazadas se escriben por la unidad de cálculo en la memoria de datos electrónicos, en la que las entradas de coordenadas originales o bien i) se sobre-escriben o ii) se caracterizan como desfasadas y ya no se usan en otras etapas de aprendizaje. Las coordenadas de nodos se actualizan así en la tabla de la memoria de datos electrónicos. Es posible ventajosamente realizar la etapa de la actualización de las coordenadas de nodos en la tabla de la memoria de datos electrónicos tras solo un pase del número predeterminado de etapas de aprendizaje y entonces consultar la siguiente imagen de la secuencia de imágenes de la cámara de detector de profundidad así como la nube de puntos determinada a partir de ésta, para comenzar el siguiente pase de las etapas a a h.

15 Igualmente pueden realizarse sin embargo también las etapas de aprendizaje para una y la misma nube de puntos múltiples veces una detrás de otra, para calcular así una multiplicidad de desplazamientos de las coordenadas de nodos. Esta multiplicidad de desplazamientos puede determinarse entonces por ejemplo aritméticamente y las coordenadas de nodos se actualizan en la tabla de la memoria de datos electrónicos solo con el desplazamiento determinado. Este modo de proceder requiere esfuerzo computacional y por consiguiente es más lento, sin embargo puede suprimir de manera eficaz desplazamientos desfavorables que se producen de manera aleatoria (por ejemplo cuando la nube de puntos comprende también puntos con posición excéntrica que se seleccionan de manera aleatoria).

Muy generalmente se realiza por tanto la actualización de las coordenadas de nodos tras K pases del número predeterminado de etapas de aprendizaje, siendo $K \geq 1$.

25 Un ejemplo de realización del procedimiento de estimación de la postura de acuerdo con la invención con el modelo de esqueleto de la figura 1 b) se representa por medio de imágenes de ejemplo de secuencias de vídeo en la figura 4. Las imágenes de vídeo de una cámara de detector de profundidad proporcionan continuamente nubes de puntos 3D por medio de la segmentación de imagen en sí conocida, que representan un ser humano que se mueve. El ajuste del modelo de esqueleto anatómico se realiza con las reglas de aprendizaje descritas anteriormente en tiempo real, y la asignación de los nodos de modelo a las distintas partes del cuerpo del ser humano se conserva correctamente de manera continua, tal como puede deducirse bien de la figura 4.

35 En particular, la asignación nodo-parte del cuerpo alcanzada ahora robusta permite una entrada de instrucciones claramente más estable por medio del control de gestos, por ejemplo a través del seguimiento únicamente de los movimientos de la mano derecha. Los demás nodos de modelo pueden ignorarse según las necesidades o también valorarse como entrada adicional por una máquina que interpreta las posiciones de nodos del esqueleto. Así es posible por ejemplo realizar entradas más complejas debido a que se siguen y se interpretan por separado al mismo tiempo los movimientos de la mano derecha y de la mano izquierda. Por ejemplo, una elevación de la mano izquierda a la altura de la cabeza puede ser un equivalente a la presión de una determinada tecla (por ejemplo tecla de mayúsculas) en un teclado o puede activar otro juego de interpretaciones de instrucciones para las entradas de la mano derecha.

45 El procedimiento descrito requiere además de un registro y facilitación rápidos de imágenes 2½D por medio de una cámara de detector de profundidad también una memoria de datos electrónicos para el almacenamiento al menos durante cierto tiempo de los datos de medición y para el almacenamiento de las coordenadas de posición de los elementos de topología del modelo de esqueleto anatómico así como una unidad de cálculo electrónica que se comunica con la memoria de datos, que realiza los cálculos descritos y en particular ordena la actualización continua de las coordenadas de posición de los elementos de topología en la memoria de datos. Además, la misma unidad de cálculo o bien puede interpretar incluso el modelo de esqueleto respectivamente actual temporalmente en línea y traducirlo por ejemplo en instrucciones de control para dispositivos subordinados o únicamente puede emitir posiciones de nodos eventualmente seleccionadas para la interpretación posterior.

55 Tal como se ha descrito al principio, puede realizarse el procedimiento descrito para la estimación de la postura mediante un dispositivo, en particular una unidad de cálculo. La unidad de cálculo puede ser un microprocesador habitual en el comercio, programable, sin embargo también es posible el uso de FPGA o ASIC. Otra forma de realización de la invención se refiere a un medio de almacenamiento que mediante una unidad de cálculo almacena instrucciones realizables, que inducen a la unidad de cálculo a realizar los cálculos descritos en el presente documento para el fin de la estimación de la postura.

60 En particular, una cámara de detector de profundidad, que comprende una unidad de cálculo para la determinación de las coordenadas de distancia, puede estar configurada también directamente para realizar el procedimiento de acuerdo con la invención. Una cámara de este tipo en conjunto constructivo con al menos una unidad de cálculo configurada de acuerdo con la invención es correspondientemente adecuada para traducir directamente la imagen de un ser humano que se mueve en coordenadas 3D de sus partes corporales esenciales. Esto es comparable con un dispositivo de captura de movimiento, *Motion-Capture*, en el que puede prescindirse sin embargo de los marcadores habituales hasta ahora en el cuerpo del ser humano.

REIVINDICACIONES

1. Procedimiento para la estimación de la postura de un objeto que se mueve mediante cálculo por ordenador de desplazamientos de las coordenadas de posición 3D de los nodos de un modelo de esqueleto, que se ajusta de manera continua en una secuencia de nubes de puntos 3D, en el que las coordenadas de nodos se encuentran tabuladas en una memoria de datos electrónicos y las nubes de puntos 3D se determinan a partir de imágenes electrónicamente registradas de una cámara de detector de profundidad, que representan el objeto que se mueve, en donde el modelo de esqueleto es una topología que presenta como elementos de topología N_1 nodos, N_2 aristas y N_3 triángulos con $N_1, N_2 > 0$ y $N_3 \geq 0$, y cada elemento de topología se describe mediante nodos, pares de nodos o tripletes de nodos y se asigna de manera fija a una parte del objeto que se mueve y en donde una unidad de cálculo realiza las siguientes etapas:
 - a. seleccionar de manera aleatoria un punto de datos X de la nube de puntos 3D;
 - b. calcular el punto de base de altura P con respecto a X con referencia a cada elemento de topología y determinar si P se encuentra respectivamente dentro del elemento de topología;
 - c. calcular la distancia de X a cada elemento de topología como norma del vector diferencial $X - P$;
 - d. determinar aquel elemento de topología que presenta la distancia mínima a X de todos los elementos de topología, cuyo punto de base de altura P se encuentra dentro del elemento de topología;
 - e. desplazar el elemento de topología determinado en la etapa d mediante desplazamiento de todos los nodos que determinan el elemento de topología en dirección del vector $X - P$, en donde el vector de desplazamiento para un nodo se multiplica por una tasa de aprendizaje y, en caso de que el elemento de topología sea una arista o un triángulo, por un peso que resulta de la distancia del nodo al punto de base de altura P con respecto a X con referencia al elemento de topología determinado en la etapa d y
 - f. repetir las etapas a a e por un número predeterminado de etapas de aprendizaje con reducción gradual de la tasa de aprendizaje;
 - g. actualizar las coordenadas de nodos en la tabla de la memoria de datos electrónicos tras K pases del número predeterminado de etapas de aprendizaje con $K \geq 1$;
 - h. proporcionar al menos las coordenadas de nodos actualizadas en la tabla para el procesamiento posterior.
2. Procedimiento según la reivindicación 1, en el que en la etapa b se representa el punto de base de altura P con referencia a un elemento de topología como una combinación lineal de los vectores de coordenadas de nodos que determinan el elemento de topología y a partir de los coeficientes de representación se determina si P se encuentra dentro del elemento de topología.
3. Procedimiento según las reivindicaciones 1 o 2, en el que se calcula el peso en la etapa e a partir de los coeficientes de representación de P .
4. Procedimiento según una de las reivindicaciones 1 a 3, en el que el número de repeticiones de las etapas a a e se encuentra entre 1000 y 5000, en particular entre 2000 y 3000.
5. Procedimiento según una de las reivindicaciones 1 a 4, en el que la tasa de aprendizaje se encuentra entre el valor inicial 0,5 y el valor final 0,01.
6. Procedimiento según una de las reivindicaciones 1 a 5, que comprende además las etapas de generación de instrucciones de control para un dispositivo que va a controlarse basándose en informaciones de la tabla de las coordenadas de nodos actualizadas y de control del dispositivo que va a controlarse por medio de las instrucciones de control.
7. Dispositivo para la estimación de la postura de un objeto que se mueve que comprende:
 - una cámara de detector de profundidad para el registro de imágenes electrónicas del objeto que se mueve,
 - una memoria de datos electrónicos para almacenar las imágenes electrónicas de la cámara de detector de profundidad y
 - una unidad de cálculo que está configurada para determinar una nube de puntos 3D que representa el objeto a partir de las imágenes electrónicas temporalmente en línea con el registro de imágenes por la cámara y en donde la memoria de datos almacena además una lista con coordenadas 3D para los nodos de un modelo de esqueleto, en donde el modelo de esqueleto es una topología que presenta como elementos de topología N_1 nodos, N_2 aristas y N_3 triángulos con $N_1, N_2 > 0$ y $N_3 \geq 0$ y la unidad de cálculo puede leer y modificar estas coordenadas para
 - nodos individuales, pares de nodos predeterminados y tripletes de nodos predeterminados como representación de los elementos de topología del modelo de esqueleto,
 - en donde la unidad de cálculo está configurada además para realizar, tras la determinación de la nube de puntos 3D que representa el objeto, las siguientes etapas:
 - a. seleccionar de manera aleatoria un punto de datos X de la nube de puntos 3D;
 - b. calcular el punto de base de altura P con respecto a X con referencia a cada elemento de topología y

determinar si P se encuentra respectivamente dentro del elemento de topología;

c. calcular la distancia de X a cada elemento de topología como norma del vector diferencial $X - P$;

d. determinar aquel elemento de topología que presenta la distancia mínima a X de todos los elementos de topología, en los que el punto de base de altura P se encuentra dentro del elemento de topología;

5 e. desplazar el elemento de topología determinado en la etapa d mediante desplazamiento de todos los nodos que determinan el elemento de topología en dirección del vector $X - P$, en donde el vector de desplazamiento para un nodo se multiplica por una tasa de aprendizaje y, en caso de que el elemento de topología sea una arista o un triángulo, por un peso que resulta de la distancia del nodo al punto de base de altura P con respecto a X con referencia al elemento de topología determinado en la etapa d y

10 f. repetir las etapas a a e por un número predeterminado de etapas de aprendizaje con reducción gradual de la tasa de aprendizaje;

g. actualizar las coordenadas de nodos en la tabla de la memoria de datos electrónicos tras K pases del número predeterminado de etapas de aprendizaje con $K \geq 1$;

15 h. proporcionar al menos las coordenadas de nodos actualizadas en la tabla para el procesamiento posterior.

8. Medio de almacenamiento legible por ordenador que almacena instrucciones que, cuando las ejecuta una unidad de cálculo, inducen a la unidad de cálculo a realizar el procedimiento según una de las reivindicaciones 1 a 6.

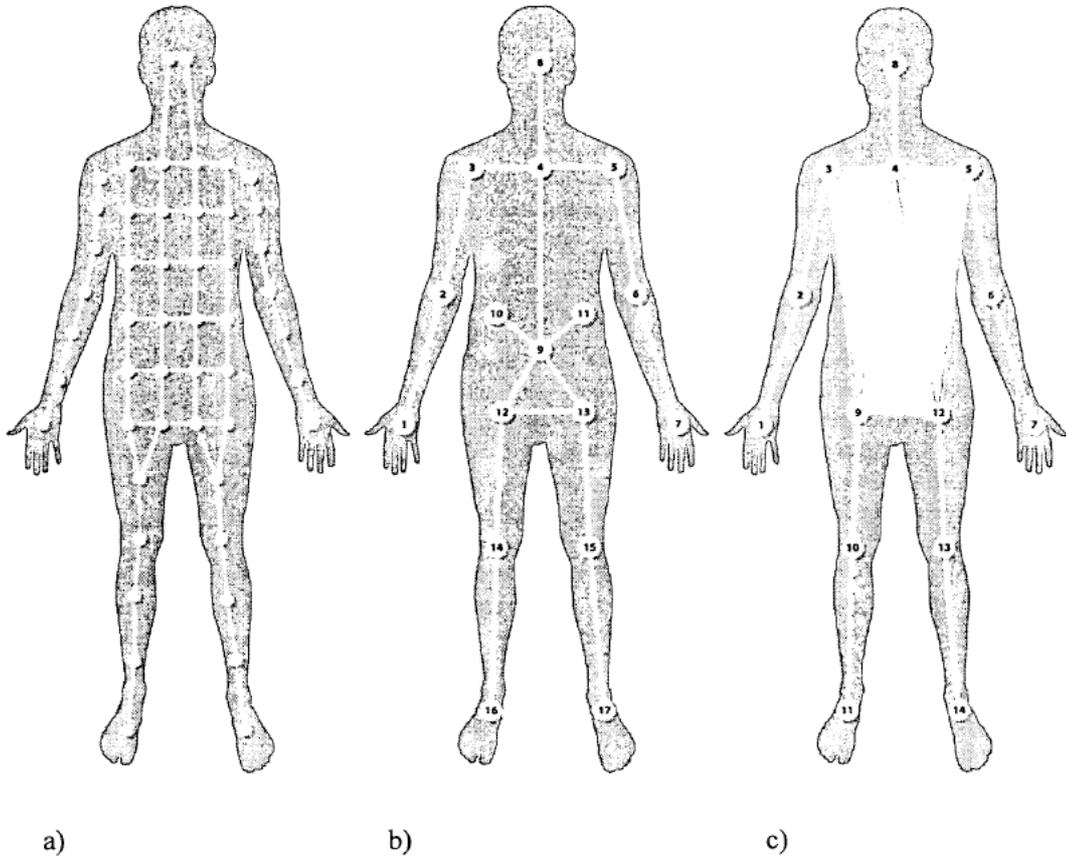
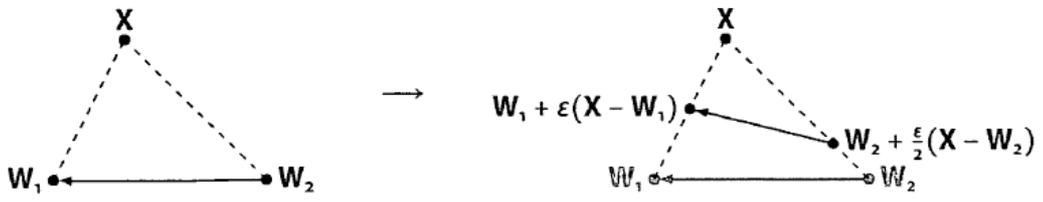
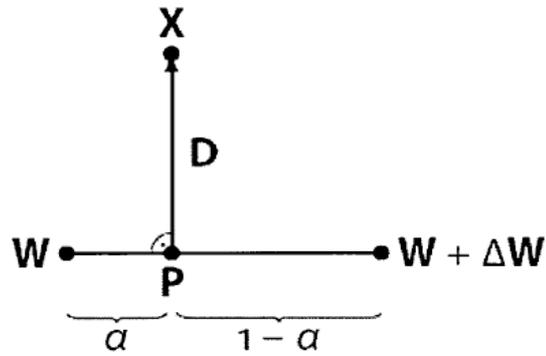


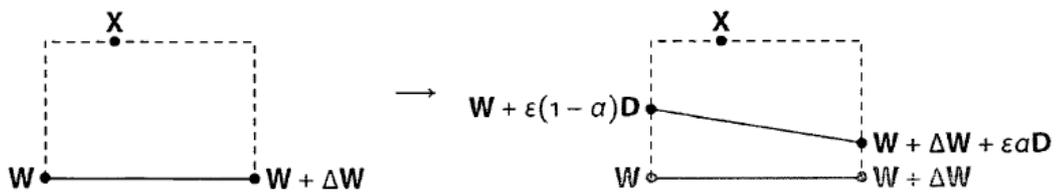
Fig. 1



a)

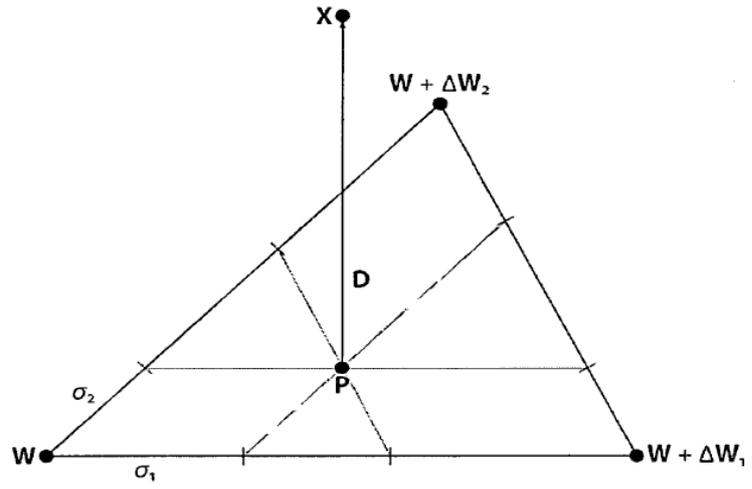


b)

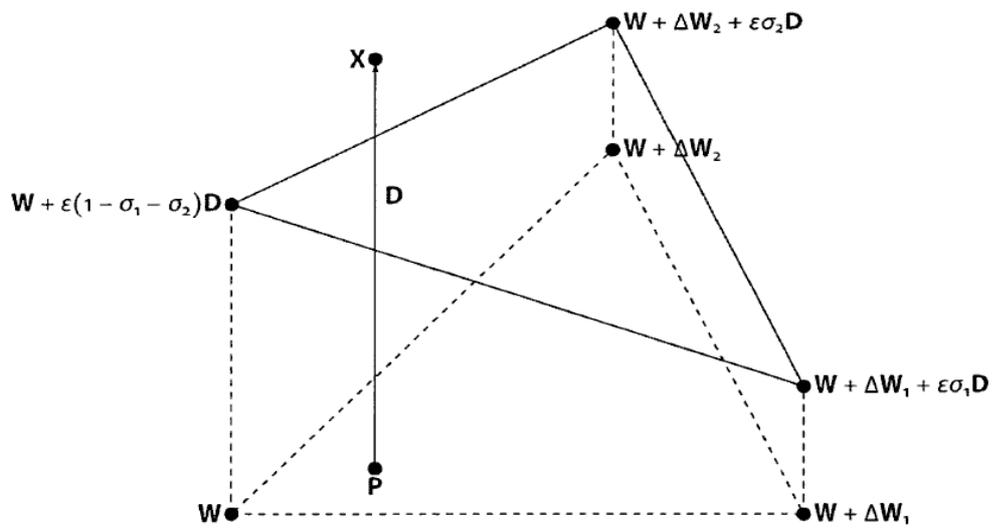


c)

Fig. 2



a)



b)

Fig. 3

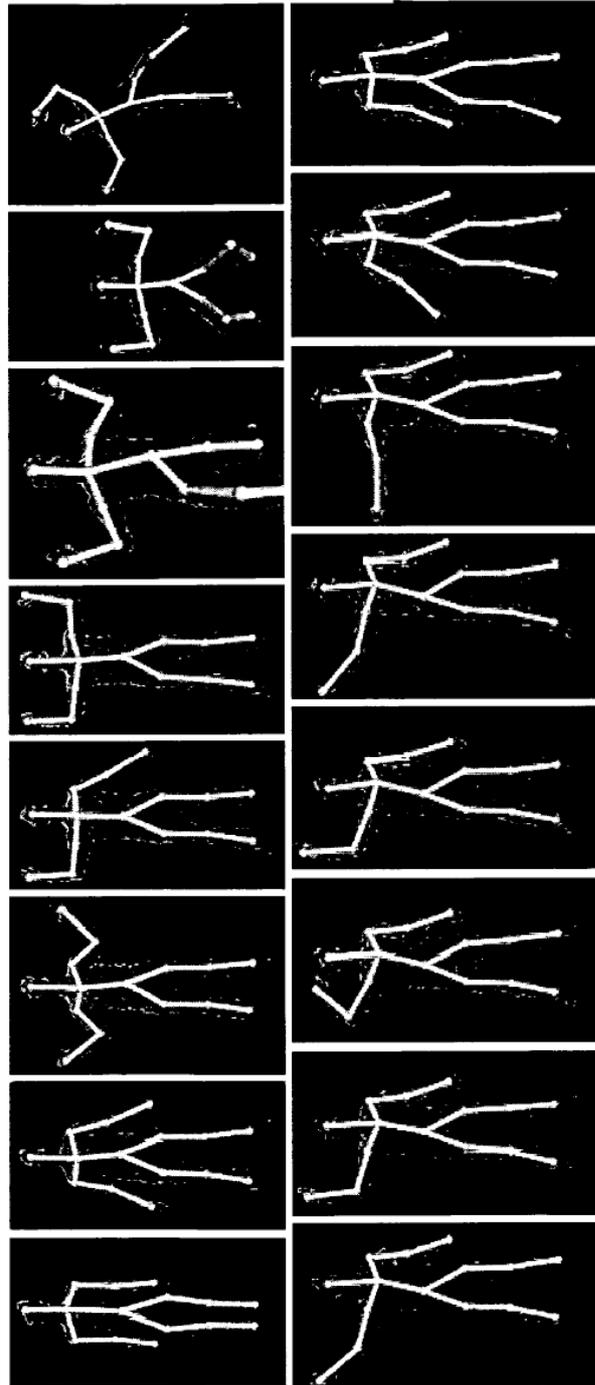


Fig. 4