

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 812 578**

51 Int. Cl.:

**G06K 9/00** (2006.01)

**G06T 7/73** (2007.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.05.2012 PCT/CH2012/000100**

87 Fecha y número de publicación internacional: **22.11.2012 WO12155279**

96 Fecha de presentación y número de la solicitud europea: **08.05.2012 E 12720395 (8)**

97 Fecha y número de publicación de la concesión europea: **24.06.2020 EP 2707834**

54 Título: **Estimación de una postura basada en la silueta**

30 Prioridad:

**13.05.2011 US 201161485919 P**

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

**17.03.2021**

73 Titular/es:

**VIZRT AG (100.0%)  
Technoparkstrasse 1  
8005 Zürich, CH**

72 Inventor/es:

**POPA, TIBERIU;  
GERMANN, MARCEL;  
KEISER, RICHARD;  
GROSS, MARKUS y  
ZIEGLER, REMO**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

ES 2 812 578 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

Estimación de una postura basada en la silueta

- 5 La invención se refiere al campo de procesamiento de imágenes por ordenador. Se refiere a un método y a un sistema para estimar una postura de un modelo de objeto articulado y/o para optimizar una estimación de una postura de un modelo de objeto articulado que representa un cuerpo que es observado por una o más cámaras, como se describe en el preámbulo de las correspondientes reivindicaciones independientes.

10 **Antecedentes**

1. Introducción

- 15 La estimación de postura o la captura de movimiento es un problema fundamental en la visión por ordenador y los gráficos [13, 14] con muchas aplicaciones como la animación de personajes en juegos y películas, interfaces libres de controlador para vigilancia de juegos [12]. Debido a la complejidad del problema, todavía no existe una solución universal para todas las aplicaciones. Las soluciones dependen en gran medida de las condiciones y de las restricciones impuestas en la configuración. En general, cuantas más restricciones se impongan en la configuración, cuanto más exactamente se puede calcular la estimación de postura. En escenarios del mundo real, a menudo es muy difícil imponer restricciones en la configuración. Sin embargo, muchas aplicaciones prácticas se basan en estos escenarios. Por ejemplo, Germann et al [11] muestran cómo se puede usar una estimación de postura precisa para la representación de alta calidad de jugadores desde un punto de vista arbitrario durante un juego deportivo utilizando solo imágenes de vídeo ya disponibles en transmisiones de TV. Además de las aplicaciones en renderizado, la estimación precisa de las posturas de los jugadores durante un juego también se puede utilizar para el análisis biomecánico y la síntesis, así como para las estadísticas del juego o incluso la transferencia de un juego real a un juego de ordenador.

- 20 M. Cloots et al.: Human Body Postura Recognition Using Spatio-Temporal Templates, Workshop on Modeling People and Human Interaction en el ICCV 2005, 1 enero 2005 (2005-01-01), XP55033869, Beijing, China, divulga un método para unir siluetas de una secuencia de imágenes con una secuencia de siluetas de plantilla.

2. Trabajo relacionado

- 35 Muchos sistemas actuales de captura de movimiento comercialmente disponibles [22] generalmente usan marcadores ópticos colocados en todo el cuerpo para rastrear el movimiento a lo largo del tiempo. Estos sistemas son muy precisos y pueden capturar todo tipo de posturas corporales, así como expresiones faciales. Sin embargo, son invasivos y funcionan en un entorno controlado. Por lo tanto, solo son adecuados para una gama específica de aplicaciones.

- 40 Los métodos de captura de movimiento sin marcadores han recibido mucha atención en la última década [13, 14]. Dependiendo del tipo de imágenes utilizadas, el problema de la reconstrucción de postura sin marcador (o captura de movimiento) puede clasificarse aproximadamente en dos grupos [24]: usando imágenes de vídeo de una cámara o usando imágenes de múltiples cámaras calibradas. La estimación de postura de secuencias de vídeo monoculares [2, 3, 24, 17, 1, 18] puede ser más conveniente para algunas aplicaciones, ya que impone menos restricciones al usuario, pero tiene una ambigüedad de profundidad inherente. Esta ambigüedad se puede resolver utilizando la estructura de los enfoques de movimiento, un problema muy difícil en la visión [13, 14]. La estructura de los algoritmos de movimiento generalmente se basa en escenas de alta resolución que contienen muchos detalles que normalmente no tenemos en nuestro escenario o configuración, es decir, en escenas deportivas Efron et al. [9] también procesa imágenes de fútbol. Aunque su trabajo se centra más en la detección de acciones, demostraron que incluso en datos de baja resolución se puede estimar una postura 2D aproximada.

- 50 Otro desafío importante en la estimación de la postura son las oclusiones. Si las imágenes provienen de una sola cámara, es muy difícil resolverlo. El uso de múltiples cámaras aumenta la probabilidad de tener una vista no ocluida del mismo sujeto. Cuanto mayor sea la cobertura espacial de las cámaras, menos ambigüedades permanecen. Además, las transmisiones deportivas ya usan múltiples cámaras en el campo. Por lo tanto, podemos aprovechar esta información para calcular una estimación de postura 3D más precisa.

- 55 La mayoría de los métodos para la estimación de postura en 3D de múltiples vistas utilizan algoritmos de seguimiento para reconstruir la postura en el tiempo  $t$  desde la postura en el tiempo  $t - 1$  [4]. El seguimiento se puede realizar utilizando flujo óptico [4] o emparejamiento estéreo [6]. Estos métodos pueden proporcionar una estimación de postura muy precisa, pero generalmente trabajan en un ambiente controlado, requieren una mayor cantidad de cámaras de alta resolución (generalmente al menos cuatro) y una buena cobertura espacial de la escena (generalmente cobertura circular) para resolver ambigüedades debido a oclusiones.

- 65 Otros métodos [21, 8, 23] construyen una geometría proxy usando siluetas de múltiples vistas o estéreo de múltiples vistas. El esqueleto se ajusta a esta geometría. Estos métodos proporcionan muy buenos resultados, pero imponen restricciones en la configuración. Requieren una configuración de estudio cuidadosamente construida, muchas

cámaras de alta resolución y muy buena cobertura espacial.

Otra clase de algoritmos se basa en el análisis y la segmentación de imágenes [15, 10]. Estos algoritmos utilizan métodos de aprendizaje automático para discriminar entre partes del cuerpo. Este análisis generalmente requiere imágenes de alta resolución, que no está disponible en nuestra configuración.

## Sumario

Por lo tanto, un objeto de la invención es crear un método y un sistema del tipo mencionado inicialmente que supere las desventajas mencionadas anteriormente.

Estos objetos se logran mediante un método y un sistema de acuerdo con las reivindicaciones independientes correspondientes.

El método implementado por ordenador para **estimar una postura de un modelo de objeto articulado**, en el que el modelo de objeto articulado es un modelo 3D basado en ordenador de un objeto del mundo real observado por una o más cámaras fuente, y el modelo de objeto articulado representa una pluralidad de articulaciones y enlaces que enlazan las articulaciones, y en el que la postura del modelo de objeto articulado se define por la ubicación espacial de las articulaciones, que comprende las etapas de

- obtener al menos una **secuencia de imágenes de origen** de una flujo de vídeo que comprende una vista del objeto del mundo real grabado por una cámara fuente;
- procesar las imágenes de origen de al menos una secuencia para extraer, para cada imagen, un segmento de imagen de origen correspondiente que comprende la vista del objeto del mundo real separado del fondo de la imagen, generando así al menos una **secuencia de segmentos de imagen de origen**;
- mantener, en una base de datos en forma legible por ordenador, un conjunto de **secuencias de siluetas de referencia**, cada silueta de referencia está asociada con un modelo de objeto articulado y con una **postura de referencia** particular de este modelo de objeto articulado;
- para cada secuencia de al menos una secuencia de segmentos de imagen de origen, **emparejar** esta secuencia con una pluralidad de secuencias de siluetas de referencia y determinar una o más **secuencias seleccionadas de siluetas de referencia** que mejor se emparejen con la secuencia de segmentos de imagen de origen;
- en el que el emparejamiento de las dos secuencias se realiza emparejando cada uno de los segmentos de la imagen de origen con la silueta de referencia que está en la misma posición dentro de su secuencia, calcular un **error de emparejamiento** que indica qué tan cerca se emparejan, y calcular a partir de los errores de emparejamiento de los segmentos de imagen de origen un **error de emparejamiento de secuencia**;
- para cada una de estas secuencias seleccionadas de siluetas de referencia, **recuperar una postura de referencia** que está asociada con una de las siluetas de referencia; y
- **calcular una estimación de la postura del modelo de objeto articulado a partir de la postura o posturas de referencia recuperada.**

En una realización,

- una imagen de la secuencia de imágenes de origen se designa un **cuadro de interés**, y el segmento de imagen de origen que se genera a partir de él se designa un **segmento de imagen de origen de interés**;
- el error de emparejamiento de secuencia es una **suma ponderada** de los errores de emparejamiento de la emparejamiento de las dos secuencias;
- la **ponderación** del error de emparejamiento para el segmento de imagen de origen de interés es **mayor y disminuye** con la distancia de los segmentos de imagen de origen (dentro de la secuencia) desde el segmento de imagen de origen de interés; y
- la postura de referencia se recupera para la silueta de referencia que se emparejó con el **segmento de interés de la imagen de origen.**

El resultado de estas etapas es una estimación inicial de postura, que luego se usa en otras etapas, por ejemplo, para mantener la consistencia local entre las estimaciones de postura de cuadros consecutivos y la consistencia global en una secuencia más larga de cuadros.

En resumen y en otras palabras, encontrando esta estimación inicial, es decir, la mejor emparejamiento o emparejamientos para una secuencia de segmentos de imagen de origen, puede hacerse mediante

- elegir repetidamente, del conjunto de secuencias de siluetas de referencia, una secuencia actual de siluetas de

referencia, y comparando la al menos una secuencia de segmentos de imagen de origen con esta secuencia actual de siluetas de referencia mediante

- 5 • comparar por pares cada segmento de imagen de origen con una silueta de referencia correspondiente y calcular un error de emparejamiento que indica qué tan cerca se emparejan;
- para uno de los segmentos de imagen de origen, viniendo de un **cuadro de interés**, calcular un **error de emparejamiento de secuencia** teniendo en cuenta los errores de emparejamiento de dos o más segmentos de imagen de origen adyacentes;
- 10 • seleccionar, del conjunto de secuencias de siluetas de referencia, un número predeterminado de secuencias que tienen los errores de emparejamiento de secuencia **más pequeños**;
- para cada una de estas secuencias seleccionadas de siluetas de referencia que **recuperan la postura de referencia** que está asociada con la silueta de referencia que a su vez corresponde al cuadro de interés; y
- 15 • calcular una estimación de la postura del modelo de objeto articulado en el cuadro de interés de las posturas de referencia recuperadas.

20 Las secuencias de siluetas de referencia pueden almacenarse como secuencias más largas y extraerse de secuencias más largas como subsecuencias. En otras palabras, los segmentos de imagen de origen se pueden comparar con la secuencia de referencia más larga deslizando una ventana (o "ventana de filtro") que tiene la longitud de la secuencia de imagen de origen a lo largo de la secuencia de referencia más larga. Para un segmento de imagen de origen dado, la silueta de referencia correspondiente es la que tiene la misma posición en la secuencia o subsecuencia que se está comparando actualmente con el segmento de la imagen de origen.

25 El error de emparejamiento se puede adaptar o ponderar de acuerdo con un valor de confianza que indica la calidad del segmento de imagen de origen. A un segmento de imagen de origen que se sabe o que probablemente comprenda objetos superpuestos del mundo real se le asigna un valor de confianza más bajo que uno que comprende solo un objeto. Como resultado, si, por ejemplo, el segmento de la imagen de origen del cuadro de interés comprende jugadores superpuestos, el error de postura correspondiente tiene una ponderación menor que los errores de postura de los segmentos de imagen adyacentes en la secuencia, por lo que se mejora la estimación de postura final para el cuadro de interés.

El número de secuencias seleccionadas de siluetas de referencia se establece como una constante o puede ser elegido por el usuario. Habitualmente, son dos, pero puede ser tres o más también.

35 En una realización, La etapa de calcular la estimación de la postura del modelo de objeto articulado se basa en al menos dos secuencias de imágenes de origen que representan vistas de la misma escena tomadas desde diferentes puntos de vista al mismo tiempo, y utiliza al menos dos mejores emparejamientos y, en consecuencia, dos recuperados postura de referencia por secuencia de imágenes de origen. La estimación de la postura se calcula eligiendo una combinación de posturas de referencia recuperadas que se emparejen mejor en el espacio 3D (Germann et al. [11]).

40 La información de la silueta que se usa para el emparejamiento puede corresponder solo al contorno de un segmento. Esta información puede ser representada, por ejemplo, por un mapa de bits binario o por las coordenadas del esquema o por una secuencia vectorial que define el esquema, o por cualquier versión comprimida de dicha información. En todos estos casos, la información de la silueta solo define, para cada punto de un cuadro, si el punto está dentro o fuera de la silueta. En consecuencia, las etapas para emparejar un segmento de imagen de origen con una silueta de referencia solo tienen en cuenta esta información de esquema. Esto permite una implementación eficiente y un procesamiento rápido. Sin embargo, esto no excluye el uso de información de textura perteneciente al interior de la silueta en etapas adicionales, por ejemplo, al eliminar ambigüedades derecha/izquierda, optimizando posturas y secuencias estimadas de posturas, etc.

50 En consecuencia, la información de silueta también puede comprender información de textura perteneciente al interior de la silueta, y el emparejamiento de segmentos de imagen de origen con siluetas de referencia también usa esta información de textura. Por ejemplo, se pueden utilizar descriptores basados en puntos o características, o métodos de flujo óptico como en Efron [9].

55 Los términos "segmento de imagen de origen" y "silueta de referencia" describen el mismo tipo de información (ya sea solo información resumida, o que también comprende, por ejemplo, información de textura); se han elegido los diferentes términos para la legibilidad.

60 La configuración de la cámara que proporciona datos de imagen para el método actual es más flexible que las tradicionales, pero tiene que hacer frente a restricciones severas: estamos limitados por tener solo dos o tres cámaras que generalmente se colocan solo en un lado del campo, tienen líneas de base muy grandes y calibraciones imprecisas (la calibración comprende parámetros intrínsecos y extrínsecos de la cámara que describen la posición y orientación de la cámara en relación con la escena que se observa, y propiedades ópticas como la distancia focal, apertura, etc., que define la proyección de la escena a la imagen de la cámara). Asimismo, dado que la imagen generalmente cubre una gran parte del campo, cada jugador individual tiene una resolución muy baja. Lo más cercano a nuestro enfoque

es el método de Germann et al. [11]. Este método también utiliza una búsqueda de base de datos basada en silueta para recuperar la postura 3D. Sin embargo, se basa solo en un solo cuadro, sin ninguna comprobación de consistencia. Además, el resultado solo consiste en posturas ya disponibles en la base de datos, que a menudo no corresponden a la postura real. Por lo tanto, las posturas serán incorrectas y la corrección manual generalmente se requiere para todas las posturas en todas las vistas y todos los cuadros, haciendo que este enfoque sea muy tedioso e inviable para las secuencias. Sin embargo, nuestro método se empareja con secuencias enteras en lugar de un solo cuadro, asegura la consistencia para el pronóstico inicial y emplea un término de energía novedoso para calcular la postura 3D final, que puede diferir de la postura inicial de la base de datos.

Las siguientes etapas para establecer la **consistencia local** entre dos posturas se determinan a partir de al menos una secuencia de segmentos de imagen de origen consecutivos, cada postura está asociada con al menos un segmento de imagen de origen (desde al menos una vista) en el que los elementos, es decir, articulaciones y/o enlaces, del modelo de objeto articulado en una o ambas posturas corresponden a extremidades del objeto del mundo real que pueden etiquetarse de manera ambigua, se ejecutan después de la estimación de postura inicial:

- para cada uno de un par de segmentos de imagen de origen consecutivos, determinar, a partir de la postura asociada y para cada una de los posibles etiquetados ambiguos de cada postura, un etiquetado correspondiente, en el segmento de la imagen de origen, de puntos de imagen de extremidades;
- elegir, para el primero del par de segmentos de imagen de origen consecutivos, un etiquetado de la postura;
- calcular un flujo óptico entre el primero y el segundo de este par de segmentos de imagen de origen consecutivos;
- determinar, a partir del flujo óptico, ubicaciones en el segundo segmento de imagen a las que se han movido los puntos de imagen correspondientes a las extremidades del primer segmento de imagen, y etiquetar estas ubicaciones en el segundo segmento de imagen según el etiquetado de las extremidades en el primer segmento de imagen;
- elegir, de los posibles etiquetados ambiguos de la postura para el segundo segmento de imagen que se usaron para etiquetar los puntos de imagen de acuerdo con la postura, el etiquetado que es consistente con el etiquetado determinado a partir del flujo óptico.

Las etapas anteriores pueden repetirse, eligiendo otras alternativas para etiquetar la postura para el primer segmento de imagen. Después, el etiquetado de la postura del primer y del segundo segmento de imagen se elige de modo que el etiquetado sea coherente con el flujo óptico, resultando en un mejor etiquetado general.

Las etapas anteriores pueden repetirse para el siguiente par de imágenes consecutivas. Es decir, el segundo segmento de imagen toma el lugar del primero, y el siguiente segmento de imagen toma el lugar del segundo. Una consecuencia de esto puede ser que un etiquetado incorrecto se propaga a través de la secuencia de posturas emparejadas. Por este motivo, se puede realizar una revisión global posterior, asegurando la consistencia global volteando (o, más en general, cambiando), si es necesario, el etiquetado de toda la secuencia.

Las extremidades que se pueden etiquetar ambiguamente normalmente son pares de extremidades, pero en un entorno más general y con objetos del mundo real que no sean humanos, pueden ser conjuntos de extremidades con más de dos extremidades.

El método implementado por ordenador para **optimizar una estimación de una postura de un modelo de objeto articulado**, en el que el modelo de objeto articulado es un modelo 3D basado en ordenador de un objeto del mundo real observado por una o más cámaras fuente, y el modelo de objeto articulado representa una pluralidad de articulaciones y enlaces que enlazan las articulaciones, y en el que la postura del modelo de objeto articulado se define por la ubicación espacial de las articulaciones, comprendiendo el método las etapas de

- **dada una estimación de la postura** (o, más específico: posiciones de las articulaciones) del modelo de objeto articulado y un **conjunto de segmentos de imagen de origen**, cada uno de los cuales comprende una vista del objeto del mundo real representado por el modelo de objeto articulado observado al mismo tiempo desde diferentes puntos de vista
- **realizar un procedimiento de optimización** en el que se varía la postura (**o las posiciones de las articulaciones**) del modelo de objeto articulado para minimizar una **función de error**, en el que la función de error comprende un **error de emparejamiento de modelo** que se calcula sobre todas las vistas y penaliza las posturas cuyas posiciones de las articulaciones, cuando se proyecta en las vistas correspondientes al conjunto de segmentos de imagen de origen, no se emparejan con los segmentos de la imagen de origen.

Este método para optimizar la estimación de una postura se puede realizar en una estimación según lo encontrado por el método para estimar una postura como se describió inicialmente, o se puede implementar como un procedimiento o sistema independiente que comienza con una estimación determinada por otros medios.

En una realización, el producto de programa informático para el método para estimar una postura de un modelo de objeto articulado y/o para optimizar una estimación de una postura de un modelo de objeto articulado se puede cargar en una memoria interna de un ordenador digital o un sistema informático, y comprende instrucciones ejecutables para hacer que uno o más procesadores del ordenador o sistema informático ejecuten el método para estimar una

5 postura de un modelo de objeto articulado y/o para optimizar una estimación de una postura de un modelo de objeto articulado. En otra realización, el producto de programa de ordenador comprende un medio legible por ordenador que tiene las instrucciones ejecutables por ordenador registradas en él. El medio de almacenamiento legible por ordenador preferentemente es no transitorio; es decir, tangible. En otra realización más, el programa de ordenador está incorporado como una señal reproducible legible por ordenador, y por lo tanto puede transmitirse en forma de dicha señal.

Las realizaciones adicionales son evidentes a partir de las reivindicaciones dependientes de la patente.

## 10 Breve descripción de los dibujos

La materia objeto de la invención se explicará con más detalle en el siguiente texto con referencia a las realizaciones a modo de ejemplo que se ilustran en los dibujos adjuntos, que muestran:

- La figura 1 esquemáticamente una visión general sobre una escena del mundo real;
- La figura 2 un modelo de objeto articulado;
- La Figura 3a silueta típica en una imagen segmentada;
- La Figura 3b tres mejores posturas emparejadas de una base de datos;
- La figura 4 posturas estimadas en un juego de fútbol con una proyección de esqueletos 3D en una cámara fuente;
- La figura 5 muestra esquemáticamente una descripción general del método;
- La figura 6 estimación de postura 2D según la técnica anterior y según el presente método;
- La figura 7 un ejemplo para ambigüedades de postura;
- La figura 8 una ilustración de consistencia local;
- La figura 9 una postura estimada antes y después de la optimización de la postura;
- La figura 10 casos de fallo;
- La figura 11 secuencias de resultados con todas las vistas de cámara mostradas por cuadro.

15 Los datos de la imagen son cortesía de Teleclub y LiberoVision.

En principio, se proporcionan elementos idénticos con los mismos símbolos de referencia en las figuras.

## 20 Descripción detallada

Nos enfocamos en la estimación de postura basada en imágenes de transmisión de fútbol sin restricciones. Esto implica varios desafíos para las posiciones de la cámara, tamaño del objeto y coherencia temporal. Aunque la estimación de postura se puede calcular en función de una configuración de múltiples cámaras, solo hay pocas cámaras disponibles, que además cuentan con líneas de base amplias. Además, por lo general, las cámaras se colocan solo en un lado del campo, lo que proporciona una cobertura limitada de la escena. Las cámaras proporcionan imágenes de alta resolución, pero generalmente están configurados para ser de gran angular por razones editoriales. Por lo tanto, los jugadores generalmente cubren solo una altura entre 50 y 200 píxeles. Además, el movimiento de los jugadores puede ser muy complejo y, especialmente en deportes de contacto como el fútbol, hay mucha oclusión.

30 Presentamos un método de estimación de postura basado en datos que puede funcionar en un entorno no controlado con cámaras calibradas libremente, jugadores de baja resolución y en presencia de oclusiones. El método y el sistema resultantes pueden usar tan solo dos cámaras para estimar la postura. No se hace una suposición restrictiva sobre la familia de posibles posturas o movimientos. Al utilizar la coherencia temporal para la estimación inicial de la postura y el refinamiento de la postura, la interacción del usuario se limita a unos pocos clics para invertir brazos y piernas en casos de fallo.

40 Muchos de los métodos más avanzados en la estimación de postura dependen del seguimiento o la segmentación de la imagen en 2D y el uso de información de calibración para extrapolar el esqueleto a 3D [4, 20]. Estos enfoques funcionan bien para imágenes de alta resolución, pero debido a la falta de información, a menudo fallan en imágenes de baja resolución y son sensibles a las condiciones de iluminación externa. Nuestro método funciona en configuraciones exteriores completamente incontroladas con bajas resoluciones, ya que solo se basa en siluetas gruesas y calibraciones gruesas.

45 Similar a Germann et al. utilizamos una base de datos de posturas y comparación de silueta para extraer candidatos de postura en 2D y utilizamos la información de calibración de la cámara para calcular el esqueleto 3D correspondiente. En contraste con su método, primero realizamos una nueva búsqueda basada en la silueta consistente en el tiempo en la base de datos para extraer el candidato de base de datos más cercano con coherencia temporal. Una etapa de coherencia de tiempo novedosa aplicada adicionalmente conduce a la estimación de postura inicial. Debido a que la postura real exacta generalmente no está en la base de datos, esto solo dará como resultado un emparejamiento más cercano, pero no en una postura precisa. Por lo tanto, desarrollamos una nueva técnica de optimización de espacio-

tiempo que aprovecha la información temporal para calcular automáticamente la postura 3D precisa.

Las contribuciones son:

- 5 • Una búsqueda de postura de base de datos basada en silueta consistente en el tiempo que proporciona una estimación de postura inicial
- Verificación de consistencia local y global para mejorar la estimación de postura inicial
- 10 • Una optimización de postura de espacio-tiempo basada en restricciones nuevas

En lugar de aprender un modelo estadístico para el esqueleto, nuestro método utiliza directamente una base de datos de posturas. Esto tiene ciertas ventajas. En primer lugar, dicho método basado en datos permite agregar fácilmente nuevas secuencias de postura para adaptarse a nuevas configuraciones o posturas previamente desconocidas. En segundo lugar, hay menos sesgo estadístico hacia posturas más comunes, ya que el método simplemente busca la postura más cercana en la base de datos. El uso de una base de datos con datos antropométricamente correctos siempre dará como resultado una postura plausible para la estimación inicial.

### 3. Visión de conjunto

**La Figura 1** muestra esquemáticamente una visión general sobre una escena del mundo real 8, la escena 8 que comprende un objeto 14 del mundo real, como un humano, siendo observado por dos o más cámaras fuente 9, 9', cada uno de los cuales genera un flujo de vídeo de imágenes de origen 10, 10'. El sistema y método de acuerdo con la invención genera una imagen virtual 12 que muestra la escena 8 desde el punto de vista de una cámara virtual 11 que es distinta de los puntos de vista de las cámaras fuente 9, 9'. Opcionalmente, a partir de una secuencia de imágenes virtuales 12 se genera un flujo de vídeo virtual. Un aparato según la invención comprende una unidad de procesamiento 15 que realiza los cálculos de procesamiento de imagen que implementan el método de la invención, dadas las imágenes de origen 10, 10' y generando una o más imágenes virtuales 12. La unidad de procesamiento 15 está configurada para interactuar con una unidad de almacenamiento 16 para almacenar imágenes de origen 10, imágenes virtuales 12 y resultados intermedios. La unidad de procesamiento 15 se controla por medio de una estación de trabajo 19 que normalmente comprende un dispositivo de visualización, un dispositivo de entrada de datos como un teclado y un dispositivo señalador como un ratón. La unidad de procesamiento 15 puede configurarse para suministrar un flujo de vídeo virtual a un transmisor de transmisión de TV 17 y/o dispositivos de visualización de vídeo 18.

**La figura 2** muestra esquemáticamente un modelo 1 3D de la escena 8, que comprende un modelo de objeto articulado 4 del objeto del mundo real 14. El modelo 3D 1 normalmente comprende además otros modelos de objetos, por ejemplo, representando a otros humanos, el terreno, edificios, etc. (no se muestran). El modelo de objeto articulado 4 comprende articulaciones 2 que están conectadas por enlaces 3, más o menos correspondiente a huesos o extremidades en el caso del modelo de un humano. Cada articulación 2 se define como un punto en el espacio 3D, y cada enlace 3 se puede representar mediante una línea recta que conecta dos articulaciones 2 a través del espacio 3D. Basándose en la postura, la forma 3D de un modelo de malla, o de otro modelo de representación de la forma 3D del objeto del mundo real, del objeto puede calcularse y su proyección en las imágenes 2D vistas por las cámaras fuente 9, 9'. Esto permite verificar y mejorar las posturas. Mientras que la postura se describe mediante enlaces unidimensionales, extremidades representadas por el modelo (malla) y colocadas de acuerdo con la postura, se extienden en tres dimensiones. Aunque la presente aplicación se explica en términos de formas humanas modeladas, También se cubre la aplicación a formas de animales o estructuras artificiales como robots.

En primer lugar, para la estimación de postura 2D en cada vista de entrada individual, utilizamos una base de datos de siluetas. Suponemos que está disponible una segmentación gruesa del sujeto 14 del fondo, por ejemplo, utilizando manipulación de croma o sustracción de fondo. **La Figura 3a** muestra un ejemplo típico de una imagen segmentada 13 en nuestro escenario de aplicación. La idea básica para calcular un pronóstico inicial de la postura de un sujeto, es decir, las posiciones 2D de las articulaciones esqueléticas 2, es compararlo con una base de datos de siluetas, para lo cual se conocen las posturas de esqueleto respectivas. Ver **Figura 3b** con diferentes siluetas de referencia 13', dos de ellos se reproducen a menor escala para ahorrar espacio.

**La Figura 4** muestra un ejemplo para un cuadro de una secuencia de imágenes de origen de una escena de fútbol, con posturas 3D encontradas por el método proyectado en la imagen y superpuestas en los jugadores.

El método, en una realización, comprende dos etapas como se ilustra en la **Figura 5**. En la primera etapa 62, calibraciones gruesas y siluetas 61 y siluetas de referencia de una base de datos de postura 66 dadas, el método extrae posturas 2D para cada vista de cámara individual utilizando una técnica de emparejamiento de silueta espacio-temporal, produciendo un pronóstico de postura en 3D triangulada 63 (etiquetada "Postura corporal estimada en 3D"). Esta detección de postura es inherentemente propensa a ambigüedades, a saber, volteos izquierdo derecho de partes simétricas. Aunque el esqueleto se empareja bastante bien con las siluetas, los brazos o las piernas del jugador todavía se pueden voltear. Debido a oclusiones y baja resolución, estas ambigüedades son a veces muy difíciles de detectar

incluso para el ojo humano. Por lo tanto, empleamos una técnica basada en el flujo óptico para detectar los casos en que ocurren los volteos y corregirlos para obtener una secuencia consistente. Es importante tener en cuenta que el flujo óptico en tales configuraciones no es lo suficientemente confiable como para seguir el movimiento completo de las partes del cuerpo de un jugador durante una secuencia completa, pero puede usarse para comparaciones locales como lo muestran Efros et al. [9]. La primera etapa usa una secuencia 51 de segmentos de imagen de origen 13, emparejando esta secuencia 51 con una pluralidad de secuencias 52 de siluetas de referencia 13', centrándose en un cuadro de imagen 56 de interés.

Sin embargo, en general, ninguna postura de la base de datos se emparejará exactamente con la postura real. Como consecuencia, en la segunda parte o etapa 64 (etiquetado como "Optimización de postura de espacio-tiempo") del método, esta postura 3D inicial 63 se refina mediante un procedimiento de optimización 64, que se basa en restricciones espacio-temporales. El esqueleto 3D optimizado 65 resultante (etiquetado como "Postura de cuerpo 3D optimizado") se empareja con las siluetas de todas las vistas y presenta una consistencia temporal en cuadros consecutivos.

#### 4. Estimación de postura inicial

La estimación de postura inicial se calcula recuperando primero la postura 2D de cada jugador y cada vista de cámara utilizando una nueva búsqueda basada en la silueta basada en datos espacio-tiempo. Una vez que encontremos las posturas 2D para cada jugador en cada cámara, podemos usar la información de calibración de las cámaras para colocar las posiciones de articulación 2D, como se observa en las imágenes, en el espacio 3D (esta etapa también se llama "levantar las posiciones 2D en 3D"). Calculamos la ubicación en 3D de cada articulación mediante la intersección de los rayos correspondientes a cada articulación 2D en cada vista de cámara. Los rayos no se cruzarán exactamente, por lo tanto, elegimos el punto más cercano a estos rayos en sentido de mínimos cuadrados. De esto obtenemos un error de triangulación  $E_t$  y un cambio de cámara inicial según lo descrito por Germann et al. [11].

Representamos el esqueleto 3D de una postura  $S$  en el espacio angular de la siguiente manera: Cada hueso  $i$  se representa en relación con su hueso padre utilizando dos ángulos  $\alpha_i$  y  $\beta_i$ , así como la longitud  $l_i$  del hueso. El hueso raíz se define por su orientación dada por tres ángulos  $\alpha_0$ ,  $\beta_0$ ,  $\gamma_0$  y por una posición global  $p_0$ . Las posiciones de las articulaciones  $j_i$  en el espacio euclidiano 3D se pueden calcular fácilmente a partir de esta representación del espacio angular y viceversa (teniendo en cuenta el bloqueo del cardán).

##### 4.1. Construcción de base de datos de postura

Una gran base de datos que muestrea todo el rango del movimiento humano es importante para nuestro método y es muy difícil de crear manualmente. Por lo tanto, utilizamos la base de datos de captura de movimiento CMU [7]. Una plantilla de malla aparejada con el mismo esqueleto se deforma usando un revestimiento de mezcla lineal para que se empareje con la postura de la base de datos. De esto, se toman instantáneas virtuales y se extrae la silueta. De esta forma creamos una base de datos de alrededor de 20000 siluetas.

Lamentablemente, la base de datos de CMU tiene solo un número limitado de tipos de posturas, principalmente de secuencias de correr y caminar. Por lo tanto, agregamos manualmente un conjunto de 900 siluetas de varias escenas de fútbol. Esto es significativamente menor que los generados automáticamente, pero lo suficiente como para ampliar el lapso de posturas de ejemplo para obtener buenos resultados. Es importante tener en cuenta que las posturas de ejemplo agregadas no se tomaron de las mismas secuencias que usamos para ajustar las posturas. La base de datos puede ampliarse continuamente mediante nuevas posturas generadas, resultando en una mejor estimación de postura inicial.

##### 4.2. Estimación de postura 2D

Similar a Germann et al. [11] aceptamos como entrada una máscara de silueta binaria gruesa para cada jugador, así como calibraciones de cámara gruesas. Comparamos estas siluetas con las siluetas de la base de datos utilizando la técnica presentada por Germann et al. que calcula la calidad de un emparejamiento entre la silueta de entrada y una silueta de base de datos en un tamaño de trama fijo (una cuadrícula con una altura = 40 y un ancho = 32 píxeles) que se ajusta a la segmentación.

La extracción de la silueta extiende el método presentado por Germann et al. aprovechando la información temporal. En lugar de confiar en un emparejamiento de un solo cuadro, nuestro enfoque considera una suma ponderada de diferencias entre los segmentos de imagen de origen 13 y las siluetas de referencia 13' sobre una secuencia de cuadros de imagen. Un error de pixel resultante  $E_q(s)$  de la imagen de la silueta de entrada binaria  $I$  con el índice  $t$  (de la trama de imagen 56 de interés) cuando se compara con la silueta imagen  $I'_s$  con el índice  $s$  de la base de datos se calcula como sigue: en la que  $n$  es el tamaño de la ventana del filtro, es decir, el número de fotogramas que preceden y siguen al cuadro de interés 56 que se consideran, y  $P$  es el conjunto de todas las posiciones de trama en las que el píxel correspondiente está en ambas imágenes que posiblemente no están ocluidas, es decir, no se espera que sea parte de la silueta de otro jugador.  $|P|$  denota el número de posiciones de trama en  $P$ . Una posición de trama puede corresponder a un píxel de hardware real de la cámara, o a un píxel de una imagen a escala calculada a partir de la

imagen de la cámara.

Las ponderaciones  $\theta(i)$  describen una función gaussiana normalizada 53 con el centro alrededor de 0. Para  $I'_{s+i}(p)$  no incluidos en la base de datos,  $\theta_s(i)$  se establece en 0 antes de la normalización. Comparar secuencias en lugar de imágenes individuales no solo agrega coherencia temporal que resulta en movimientos suaves, pero también mejora la estimación de postura. Incluso las partes de la imagen ocluidas en unos pocos cuadros se pueden ajustar de manera más robusta. En general, este enfoque ayuda a evitar el emparejamiento de una silueta que es similar pero se originó a partir de una postura completamente diferente. Esto se muestra en la **Figura 6**, que también muestra una comparación directa de nuestra estimación de postura inicial con la estimación de postura en el trabajo de Germann et al. **La Figura 6** muestra: (a) Posición 2D estimada comparando solo el cuadro actual con la base de datos como en Germann et al. [11]. (b) El elemento de la base de datos encontrado para la comparación de un solo cuadro. (c) Postura 2D estimada mediante la comparación de secuencias de siluetas. (d) La secuencia de base de datos encontrada con el segmento de imagen correspondiente en el medio.

Usando este error de píxel, buscamos en cada vista de cámara las mejores hipótesis de dos posturas y seleccionamos la mejor combinación de ellas eligiendo el error de triangulación más bajo resultante  $E_t$ . Por supuesto, se pueden usar más de dos hipótesis de postura de cada cámara para determinar la mejor combinación, o un número diferente por cámara, o solo una hipótesis de postura de al menos una cámara, suponiendo que sea el mejor sin considerar el error de triangulación.

#### 4.3. Consistencia de la postura

La **Figura 7** muestra un ejemplo de ambigüedades de postura: (a) Posibles etiquetados en primera cámara. (b) Vista esquemática desde la parte superior para ilustrar las dos posiciones posibles de las rodillas. (c) Posibles etiquetados en la segunda cámara.

La etapa de detección de postura 2D se basa en el emparejamiento de silueta y, por lo tanto, es propenso a las ambigüedades. Dada una silueta y una postura 2D inicial emparejada recuperada de la base de datos, no podemos decidir si los etiquetados como "izquierda" y "derecha" en brazos y piernas son correctos. Por este motivo, en una realización, información de la postura de la base de datos recuperada que define si una pierna o un brazo (o, por regla general, uno de un conjunto de cadenas simétricas de articulaciones) debe etiquetarse, por ejemplo, "izquierda" o "derecha" se ignora después de un emparejamiento de una silueta 2D. La ambigüedad restante se resuelve de la siguiente manera. La Figura 7(a) muestra un ejemplo de silueta con dos posibles etiquetados para las piernas. La posible posición de la rodilla derecha está marcada por una forma de diamante. Esta vista es de la cámara izquierda en el esquema de la **Figura 7(b)**. El mismo sujeto en el mismo cuadro pero en la otra cámara muestra la silueta en la **Figura 7(c)**, de nuevo con los dos posibles etiquetados de las piernas. Por lo tanto, tenemos cuatro posiciones posibles en 3D para la rodilla derecha después de levantar en 3D. Como se muestra en la **Figura 7(b)**. Si la rodilla derecha cae en una de las posiciones marcadas con una estrella, la rodilla izquierda caerá sobre la otra estrella. Si la rodilla derecha cae en una de las posiciones marcadas con un círculo, entonces la rodilla izquierda caerá sobre el otro círculo. Dejando que los círculos sean las posiciones correctas de las rodillas, entonces podemos tener dos tipos diferentes de falla: o las rodillas están mal etiquetadas en 3D pero en las posiciones correctas, o las rodillas están en las posiciones incorrectas (las estrellas).

Sin información adicional no podemos decidir en tal situación qué posiciones son correctas, es decir, seleccionar la única correcta entre las cuatro posibilidades, especialmente cuando solo hay dos cámaras disponibles. Un posible enfoque para desambiguar los casos invertidos consistiría en verificar todas las combinaciones posibles y mantener solo las anatómicamente posibles. Sin embargo, todavía es posible que varias configuraciones de volteos produzcan posturas anatómicamente correctas.

Para resolver correctamente estas ambigüedades, utilizamos un enfoque de dos etapas: primero, se establece la consistencia local entre cada par de tramas 2D consecutivas resultando en una secuencia completa de posturas 2D que son temporalmente consistentes. En segundo lugar, cualquier ambigüedad restante que atraviese toda esta secuencia se resuelve globalmente.

##### 4.3.1 Consistencia local

El objetivo de esta etapa es asegurarse de que las posturas 2D recuperadas de una cámara en los fotogramas  $k$  (**Figura 8(a)** y **8(b)**) y  $k+1$  (**Figura 8(c)**) sean consistentes, es decir, que no ocurran movimientos de brazos o piernas derechos/izquierdos entre cuadros consecutivos. En otras palabras, si un píxel en el cuadro  $k$  pertenece a la pierna derecha, debería pertenecer a la misma pierna en el cuadro  $k+1$  también. Aplicando esta suposición, asignamos a cada píxel en ambas imágenes en color  $I_{C(k)}$  e  $I_{C(k+1)}$  un hueso correspondiente, y calculamos el flujo óptico [5] (**Figura 8(d)**) entre los cuadros.

**La Figura 8** ilustra cómo se establece la coherencia local mediante un control de coherencia basado en el modelo. El modelo es un modelo basado en malla, pero el método también se puede realizar con otra representación de la forma 3D del objeto del mundo real. Las articulaciones asignadas a la pierna derecha están marcadas con diamantes: (a)

Marco anterior y (b) malla ajustada. (c) Piernas mal asignadas en el cuadro actual. (d) Flujo óptico. (e) Malla ajustada en el cuadro actual con emparejamientos correctos e incorrectos etiquetados. (f) Error de las piernas invertidas (correctas) en el cuadro actual. En (e) y (f), las formas blancas irregulares en la parte inferior de las piernas y los pies indican píxeles que han sido etiquetados como incorrectos 54 en (e) y correctos 55 en (f). Por casualidad no hay píxeles "correctos" visibles en (e) ni píxeles "incorrectos" en (f), pero podría haber.

La idea subyacente es que un píxel en el cuadro  $k$  y su píxel correspondiente en el cuadro  $k + 1$ , calculado usando flujo óptico, debe asignarse al mismo hueso. De lo contrario, podría haber un volteo como se muestra en la **Figura 8(c)**. Por lo tanto, calculamos esta asociación para todas las combinaciones de posibles etiquetados de articulaciones como izquierda o derecha (ver **Figura 7**), calcule la consistencia de los flujos de píxeles con cada una de las combinaciones y seleccione la combinación de etiquetas más consistente para el segundo cuadro. Para hacer el enfoque más robusto con respecto a los errores de flujo óptico, solo consideramos píxeles con buen flujo óptico y donde las etiquetas de píxeles correspondientes en ambos cuadros son del mismo tipo de hueso, es decir, ya sea ambos brazos o ambas piernas. Por ejemplo, si un píxel  $p$  pertenece al brazo izquierdo en el cuadro  $k$  y al torso en el cuadro  $k + 1$ , esto probablemente se deba a un flujo óptico inexacto basado en la oclusión y, por lo tanto, podemos omitir este píxel. Si el píxel pertenece a un hueso diferente del mismo tipo (brazo o pierna), entonces esta es una fuerte indicación de un volteo. Empleamos una estrategia de votación para seleccionar la configuración óptima de volteo: una extremidad se etiqueta "izquierda" o "derecha" de acuerdo con si comprende una mayoría de píxeles "izquierda" o "derecha".

Para hacer esto, cada píxel debe asignarse a su hueso correspondiente. Una asignación ingenua basada en la distancia al hueso no es óptima, ya que no tiene en cuenta las oclusiones. Por lo tanto, construimos la postura 3D usando la información de todas las cámaras como se describe en la sección de estimación de postura inicial. De nuevo, utilizamos una plantilla de malla deformada y renderizada para todos los cambios posibles en todas las cámaras, utilizando códigos de colores para todos los huesos. La malla renderizada lleva, para cada extremidad, la información ya sea "izquierda" o "derecha". Por tanto, la asignación de píxeles es una búsqueda simple que proporciona una asignación precisa a pesar de la auto oclusión: Para cada píxel, la malla renderizada en la misma posición indica si el píxel está "a la izquierda" o "a la derecha".

La comparación de este etiquetado con el etiquetado determinado por el flujo óptico permite que los píxeles se etiqueten como "correctos" o "incorrectos", dependiendo de si el etiquetado es consistente. **La Figura 8(e)** muestra un etiquetado incorrecto inicial que causa que ciertas áreas de las piernas, mostrado como blanco en (e), sean marcadas como "incorrectas". El etiquetado correcto provoca que otras áreas de las piernas, se muestra como blanco en (f), sean marcadas como "incorrectas".

Esto resuelve la mayoría de los volteos de brazos o piernas. Para casos de fallo, el usuario puede cambiar el volteo de todos los cuadros posteriores de una secuencia con un clic del ratón. Esta es la única interacción del usuario en nuestro sistema y para una vista de un jugador solo se necesita aproximadamente 1 clic por cada 10 fotogramas.

#### 4.3.2 Consistencia global

Después de la etapa de coherencia local, todos los cuadros consecutivos no deberían tener volteos entre ellos, lo que significa que toda la secuencia es consistente. Todavía existe la posibilidad de que toda la secuencia se invierta de la manera incorrecta. Sin embargo, este es un problema simple ya que solo tenemos una desambiguación binaria para hacer toda la secuencia. Por lo tanto, la consistencia global se verifica evaluando una función de los posibles etiquetados globales de los brazos y los posibles etiquetados globales de las piernas. El etiquetado final se selecciona eligiendo la combinación de etiquetado que minimiza el siguiente término de error, resumido en toda la secuencia:

$$E_g = \lambda_{DB} E_{DB} + \lambda_t E_t \quad (2)$$

Esta es una suma ponderada con parámetros constantes  $\lambda_{DB}$  y  $\lambda_t$ .  $E_{DB}$  es una "distancia a la base de datos" que asegura que el etiquetado/volteo seleccionado dé como resultado posturas plausibles a lo largo de la secuencia. Penaliza una postura por la distancia a la postura más cercana  $P$  en la base de datos:

$$E_{DB} = \min_P \frac{1}{2|J|} \sum_{i=0}^{|J|} (\alpha'_i - \alpha_i)^2 + (\beta'_i - \beta_i)^2 \quad (3)$$

donde  $\alpha$  y  $\beta$  son los ángulos de unión de las posiciones de unión trianguladas  $J$ .  $\alpha'$  y  $\beta'$  son las de la postura de base de datos  $P$ . El  $|J|$  es el número de articulaciones. Al buscar la postura de base de datos más cercana para cada postura a lo largo de la secuencia, se tiene en cuenta el etiquetado (derecha/izquierda) de las posturas en la base de datos. Es decir, las extremidades de las posturas de la secuencia solo se emparejan con las extremidades de las posturas de la base de datos que están etiquetadas de la misma manera. Como la base de datos contiene solo posturas antropométricamente correctas, esto penaliza las posturas no plausibles.

5. Optimización de postura

Las mejores posturas en 3D calculadas por la estimación de la postura hasta ahora todavía están limitadas al ajuste, en cada vista, a una postura que existe en la base de datos. Sin embargo, la base de datos comprende solo un subconjunto de todas las posturas posibles y, por lo tanto, a menudo no contiene la solución precisa.

Por este motivo, se aplica un método de optimización para recuperar una postura más precisa, como se muestra en la **Figura 9**. Para guiar este método de optimización, combinamos varias funciones de energía espacial y temporal y las minimizamos utilizando un método de optimización.

5.1. Función de energía (o función de error)

La función de energía o función de error se basa en nuestra representación del esqueleto  $S$  descrito en la sección sobre estimación de postura inicial. Todos los parámetros, excepto la longitud del hueso, son variables por cuadro. Las longitudes de los huesos también son variables, pero permanecen iguales en toda la secuencia y se inicializan como el promedio de las longitudes locales de todos los cuadros. Esto introduce automáticamente una restricción antropométrica, ya que los huesos no deberían encogerse ni crecer con el tiempo. Otra buena propiedad de la representación esqueleto elegida es que reduce significativamente el número de variables. Para hacer frente a los servidores de calibración, también optimizamos para el vector de cambio dimensional dado por sujeto, cámara y cuadro.

Definimos nuestra energía o error funcional por cuadro y sujeto como la siguiente suma ponderada de términos de error:

$$E(S) = \omega_s E_s + \omega_f E_f + \omega_{DB} E_{DB} + \omega_{roi} E_{roi} + \omega_p E_p \quad (4)$$

No todos los términos de error son necesarios para que la optimización arroje resultados útiles. Según la naturaleza de la escena y los objetos del mundo real, se pueden omitir uno o más de los términos de error. En una realización, para observar escenas deportivas, se utiliza al menos el error de relleno de silueta y, opcionalmente, la distancia al error de la base de datos.

El error funcional se puede minimizar en una optimización local en la que se varía la postura del sujeto como se ve en un cuadro o en un conjunto de cuadros desde el mismo instante en el tiempo. Como alternativa, el error funcional puede minimizarse en una secuencia más larga de cuadros en los que las posturas correspondientes a todos los cuadros varían para encontrar emparejamientos óptimos para toda la secuencia que sean consistentes entre sí (según los criterios de optimización que vinculan cuadros consecutivos).

**Término de error de emparejamiento de silueta  $E_s$**  Los huesos del esqueleto 3D correcto deben proyectarse sobre la silueta 2D en todas las cámaras. El término de error  $E_s$  penaliza las posiciones de las articulaciones cuyas proyecciones 2D están fuera de las siluetas:

$$E_s = \frac{1}{|C||J_+|} \sum_{c \in C} \sum_{j \in J_+} EDT_c(P_c(j)), \quad (5)$$

donde  $C$  es el conjunto de todas las cámaras que cubren una silueta de este tema.  $J_+$  es la unión del conjunto  $J$  de todas las articulaciones y los puntos que se encuentran en el medio de un hueso, o uno o más puntos colocados a lo largo del hueso. La transformada de distancia euclidiana normalizada  $EDT$  devuelve para cada punto 2D en la imagen de la cámara la distancia al punto más cercano dentro de la silueta dividida por el lado más grande del cuadro delimitador de siluetas. Esta normalización es importante para hacer que el error sea independiente del tamaño del sujeto en la imagen de la cámara, que puede variar según el zoom.  $P_c(j)$  es la proyección para transformar la articulación 3D  $j$  en el espacio de la cámara teniendo en cuenta el desplazamiento de la cámara como en [11], para corregir pequeños errores de calibración.

**Término de error de relleno de silueta  $E_f$**  Aunque el término de emparejamiento de silueta  $E_s$  penaliza las articulaciones fuera de la silueta, hasta el momento no hay restricciones sobre dónde se colocan dentro de la silueta.

El término de error de llenado  $E_f$  evita que las articulaciones se ubiquen cerca una de la otra en algún lugar dentro del torso y, por lo tanto, en otras palabras, asegura que haya articulaciones ubicadas en todas las extremidades:

$$E_f = \frac{1}{|C||J||R|} \sum_{c \in C} \min_{j \in J} \sum_{r \in R} \text{dist}(P_c^{-1}(r), j), \quad (6)$$

5 donde  $R$  es el conjunto de todos los puntos de cuadrícula de la sección en la estimación de postura 2D que se encuentran dentro de la silueta  $P_c^{-1}(r)$  transforma dicho punto de cuadrícula del espacio de la cámara  $c$  en un rayo en el espacio mundial, mientras que  $\text{dist}()$  es la distancia de un rayo a una articulación.

10 La intención de este término de error es penalizar las posturas en las que los elementos del modelo de objeto articulado (en este caso, articulaciones y enlaces, o solo enlaces) se colapsan dentro de la silueta. El término de error favorece las posturas en las que cada punto de trama o punto de cuadrícula dentro de la silueta está cerca de dicho elemento. En otras palabras, con una distancia creciente de un punto desde el elemento más cercano, el término de error de relleno de silueta aumenta. El elemento más cercano puede ser un enlace o una articulación o ambos.

15 **Distancia al término de error de postura de la base de datos  $E_{DB}$**  Esto ya fue definido por la ecuación 3. Asegura que la postura 3D final sea cinemáticamente posible (por ejemplo, la articulación de la rodilla se dobla correctamente) aprovechando la base de datos de posturas correctas. Implícitamente agrega restricciones antropométricas a nuestra optimización. La postura de base de datos más cercana utilizada aquí se encuentra mediante una nueva búsqueda a través de las posturas, ya que la postura estimada puede haber cambiado en el curso del proceso de optimización.

20 **Términos de error de suavidad  $E_{rot}$  y  $E_p$**  El movimiento humano es generalmente suave, de modo que los esqueletos de los cuadros adyacentes deben ser similares. Esto nos permite introducir coherencia temporal en la optimización de la postura. Por lo tanto, La  $E_{rot}$  penaliza grandes cambios de los ángulos internos del esqueleto de cuadros consecutivos y  $E_p$  penaliza el movimiento grande:

$$E_{rot} = \frac{1}{2|J|} \sum_{i=0}^{|J|} (\alpha'_i - \alpha_i)^2 + (\beta'_i - \beta_i)^2 \quad (7)$$

25

$$E_p = |p_0 - p'_0| \quad (8)$$

30 donde  $\alpha'$  y  $\beta'$  son los ángulos correspondientes del mismo sujeto en el cuadro anterior  $yp'_0$  es la posición global de la unión raíz en el cuadro anterior. La rotación del hueso de la raíz se puede tener en cuenta y restringir de manera similar.

35 **Término de error de longitud  $E_l$**  La inicialización de las longitudes de hueso (o longitudes de enlace) ya es una buena aproximación, cuando se maneja una secuencia de cuadros. Por lo tanto, Tratamos de mantener la postura optimizada cerca de estas longitudes:

$$E_l = \frac{1}{|J|} \sum_{i=0}^{|J|} (l_i - \hat{l}_i)^2 \quad (9)$$

40 donde  $l_i$  es la longitud del hueso final  $\hat{l}_i$ , es el hueso inicial o la longitud del enlace u otro hueso de referencia o longitud del enlace. También  $l_i$  puede considerarse como la longitud ósea verdadera y también se hace variable, y se puede estimar al realizar una optimización sobre las secuencias completas de cuadros.

## 45 5.2. El procedimiento de optimización

Para minimizar el término de energía en la ecuación 4, empleamos, por ejemplo, una estrategia de optimización local donde optimizamos iterativamente las variables una por una mediante la búsqueda de líneas a lo largo de direcciones elegidas al azar [19]. Para cada variable seleccionamos 10 direcciones aleatorias para la optimización y realizamos 20 iteraciones globales. Debido a la naturaleza no suave inherente de nuestras funciones objetivas, Este método funcionó mejor en la práctica que Levenberg-Marquardt [16].

55 El procedimiento de optimización puede implementarse independientemente del procedimiento de estimación de postura inicial descrito anteriormente, es decir, con cualquier otro procedimiento de estimación de postura o con posturas por defecto constantes como estimaciones iniciales. Sin embargo, usando el procedimiento de estimación de postura inicial, que proporciona una estimación inicial bastante buena, garantiza que es probable que el procedimiento de optimización encuentre una emparejamiento óptimo a nivel mundial, evitando mínimos locales de la función de error.

La **figura 9** ilustra el efecto del procedimiento de optimización. El ejemplo de la izquierda muestra la influencia del término de error de relleno de silueta: El brazo del jugador se puede subir o bajar para reducir el término de error de emparejamiento de silueta, pero el término de error de relleno de silueta solo se reduce al mover el brazo hacia arriba.

5 La **Figura 9** muestra una clara mejora sobre el método de Germann et al. [11] que no incluía una optimización de postura en absoluto y donde cada postura tenía que corregirse manualmente.

6. Resultados

10 Evaluamos nuestro sistema en cuatro secuencias de imágenes de TV de juegos de fútbol reales con dos o tres cámaras, produciendo aproximadamente 1500 posturas para procesar. Un subconjunto de resultados se ilustra en las **Figuras 4 y 11**.

15 Cada fila de la **Figura 11** muestra un conjunto de posturas consecutivas y cada elemento muestra la imagen del jugador respectivo en todas las cámaras disponibles. Incluso con solo dos cámaras e imágenes de muy baja resolución, Nuestro método puede recuperar buenas posturas en la mayoría de los casos.

Los valores de los parámetros que utilizamos para todos nuestros resultados, calculado utilizando nuestro sistema de ajuste automático de parámetros, son

20

Param.	$\omega_s$	$\omega_f$	$\omega_{db}$	$\omega_{rot}$	$\omega_p$	$\omega_l$	$\lambda_{DB}$	$\lambda_t$
	9	15	0,05	0,1	1	1	0,15	0,3

Para las funciones de optimización en las ecuaciones (4) y (2) utilizamos los parámetros que se muestran en la tabla 1 para todos nuestros resultados. Fueron encontrados por el siguiente procedimiento de ajuste de parámetros. Anotamos manualmente las posturas 2D en dos escenas. Luego se ejecutó el método y los resultados se compararon automáticamente con las anotaciones manuales. Usar esto como una función de error y los parámetros como variables permite una optimización automática de parámetros.

25

Nuestro método de estimación de postura toma aproximadamente 40 segundos por jugador por cuadro en una configuración de dos cámaras y aproximadamente 60 segundos para una configuración de tres cámaras. Implementamos una versión paralela que ejecuta un hilo para cada jugador. En un sistema de 8 núcleos, esto dio una aceleración de aproximadamente un factor de 8.

30

Tenga en cuenta que la estimación de postura inicial no depende de la estimación de postura del cuadro anterior. Por tanto, no hay deriva y el proceso puede recuperarse de las malas suposiciones.

35

La **Figura 10** muestra casos de falla para el método descrito hasta ahora: (a) Una postura que está demasiado lejos de la base de datos y no se pudo estimar correctamente. (b) los brazos están demasiado cerca del cuerpo y no se pueden colocar correctamente.

40 La **Figura 11** muestra las secuencias de resultados con todas las vistas de cámara mostradas por cuadro.

Extensión usando flujo óptico

El método descrito hasta ahora puede fallar debido a la falta de información proporcionada solo por una silueta binaria, particularmente cuando los brazos están demasiado cerca del cuerpo como se ilustra en la **Figura 10(b)**. Es decir, varias posturas pueden tener siluetas binarias muy similares. Por tanto, solo usar la información de la silueta puede no ser suficiente para desambiguar las posturas. La incorporación del flujo óptico en el procedimiento de optimización puede resolver tales ambigüedades. Para más detalles, esto se hace determinando el flujo óptico entre dos imágenes sucesivas y determinando a partir de ellas, para uno o más huesos o articulaciones, su posición esperada. Esto se puede hacer para todas las articulaciones, o solo para las articulaciones que ocluyen (u obstruyen) otras partes del cuerpo. Por ejemplo, dada la posición de las articulaciones (o posiciones y orientaciones óseas) en un cuadro, el flujo óptico a un cuadro adyacente se utiliza para calcular las posiciones esperadas en el cuadro adyacente.

45

50

Sin embargo, el flujo óptico es más confiable si la parte del cuerpo no está ocluida en esta cámara. Por tanto, similar al método descrito anteriormente, una malla renderizada desde la vista de la cámara se usa para etiquetar cada píxel con la parte del cuerpo a la que pertenece también. Después, para propagar la posición de una articulación en el siguiente cuadro, solo se usa el flujo óptico de esos píxeles cerca (o sobre) la posición proyectada de la articulación que pertenece a la parte del cuerpo correspondiente de esta articulación.

55

Dado, por cada cámara disponible, estas posiciones esperadas, las posiciones de las articulaciones esperadas se calculan por triangulación. La distancia, también llamado término de error de flujo,  $E_{EX}$  a la postura esperada correspondiente (teniendo en cuenta el cuerpo completo o solo la extremidad o el hueso de interés se calcula esencialmente de la misma manera que la distancia a la base de datos  $E_{DB}$ , y se agrega un término ponderado

60

correspondiente  $\omega_{EX}E_{EX}$  a la función energética de la ecuación (4).

#### Actualización automática de la base de datos

5 Además, los resultados de nuestro método dependen en gran medida de la base de datos de postura. Una buena base de datos tendrá una amplia gama de movimientos, así como una amplia gama de vistas, de modo que el pronóstico inicial esté cerca de la postura correcta. **La figura 10(a)** muestra un ejemplo en el que no hay una postura similar en la base de datos y, por lo tanto, la estimación de la postura falla. Las buenas posturas se pueden seleccionar de forma manual o automática y luego agregarse a la base de datos, ampliando el espacio de posibles posturas.

10

Uso de restricciones de ángulo articular

Otra importante prioridad que se puede aprovechar aún más es la información cinemática del esqueleto humano: El método presentado hasta ahora ya utiliza algunas restricciones antropométricas implícitas. Sin embargo, restricciones específicas en ángulos articulares, es decir, explicando el hecho de que los ángulos articulares del cuerpo humano están limitados a un cierto rango de valores, Se puede incorporar en la optimización de postura.

15

#### Referencias

- 20 [1] A. Agarwal y B. Triggs. Recovering 3d human postura from monocular images. PAMI, 2006.  
 [2] M. Andriluka, S. Roth y B. Schiele. Pictorial Structures Revisited: People Detection and Articulated Postura Estimation. CVPR, 2009.  
 [3] M. Andriluka y SRB Schiele. Monocular 3d postura estimation and tracking by detection. ECCV, 2010.  
 25 [4] L. Ballan y GM Cortelazzo. Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes. En 3DPVT, 2008.  
 [5] J.-Y. Bouguet, Pyramidal implementation of the lucas kanade feature tracker: Descripción del algoritmo. Technical report, Intel Corp., Micropr. Research Labs, 1999.  
 [6] C. Choi, S.-M. Baek y S. Lee. Real-time 3d object postura estimation and tracking for natural landmark based visual servo. En IROS 2008.  
 30 [7] C. G. L. M. Database, 2010. <http://mocap.cs.cmu.edu>.  
 [8] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel y S. Thrun. Performance capture from sparse multi-view vídeo. En SIGGRAPH, 2008.  
 [9] A. A. Efros, A. C. Berg, G. Mori y J. Malik. Recognizing action at a distance. En ICCV, 2003.  
 [10] V. Ferrari, M. Marin-Jimenez y A. Zisserman. Progressive search space reduction for human postura estimation. En CVPR, 2008.  
 35 [11] M. Germann, A. Hornung, R. Keizer, R. Ziegler, S. Wurmlin y M. Gross. Articulated billboards for videobased rendering. Eurographics, 2010.  
 [12] Kinect, 2010. <http://www.xbox.com/en-US/kinect>.  
 [13] T. B. Moeslund y E. Granum. A survey of computer vision-based human motion capture. CVIU, 2001.  
 40 [14] T. B. Moeslund, A. Hilton y V. Kruger. A survey of advances in vision-based human motion capture and analysis. CVIU, 2006.  
 [15] G. Mori. Guiding model search using segmentation. En ICCV, 2005.  
 [16] J. Mor. The levenberg-marquardt algorithm: Implementation and theory. En Lecture Notes in Mathematics, volumen 630. 1978.  
 45 [17] D. Ramanan, D. Forsyth y A. Zisserman. Strike a postura: tracking people by finding stylized posturas. En CVPR, 2005.  
 [18] D. Ramanan, D. Forsyth y A. Zisserman. Tracking people by learning their appearance. PAMI, 2007.  
 [19] J. Schreiner, A. Asirvatham, E. Praun y H. Hoppe. Inter-surface mapping. En SIGGRAPH, 2004.  
 50 [20] A. Shahrokni, T. Drummond y P. Fua. Markov-based Silhouette Extraction for Three-Dimensional Body Tracking in Presence of Cluttered Background. En BMVC, 2004.  
 [21] C. Teobalto, E. de Aguiar, M. A. Magnor, H. Theisel y H.-P. Seidel Marker-free kinematic skeleton estimation from sequences of volume data. En VRST, 2004.  
 [22] Vicon, 2010. <http://www.vicon.com>.  
 55 [23] D. Vlastic, I. Baran, W. Matusik y J. Popovic. Articulated mesh animation from multi-view silhouettes. En SIGGRAPH, 2008.  
 [24] B. Zou, S. Chen, C. Shi y UM Providence. Automatic reconstruction of 3d human motion postura from uncalibrated monocular vídeo sequences based on markerless human motion tracking. Pattern Recognition, 2009.

## REIVINDICACIONES

1. Un método implementado por ordenador para estimar una postura de un modelo de objeto articulado (4), en el que el modelo de objeto articulado (4) es un modelo 3D (1) basado en ordenador de un objeto del mundo real (14) observado por una o más cámaras fuente (9), y el modelo de objeto articulado (4) representa una pluralidad de articulaciones (2) y de enlaces (3) que enlazan las articulaciones (2), y en donde la postura del modelo de objeto articulado (4) está definida por la ubicación espacial de las articulaciones (2), comprendiendo el método las etapas de

• obtener al menos una secuencia de imágenes de origen (10) de un flujo de vídeo que comprende una vista del objeto del mundo real (14) grabado por una cámara fuente (9);

• procesar las imágenes de origen de al menos una secuencia (10) para extraer, para cada imagen, un segmento de imagen de origen correspondiente (13) que es una silueta de entrada que comprende la vista del objeto del mundo real (14) separado del fondo de la imagen, generando así al menos una secuencia (51) de segmentos de imagen de origen (13);

• mantener, en una base de datos en forma legible por ordenador, un conjunto de secuencias (52) de siluetas de referencia (13'), siendo cada silueta de referencia (13') de un modelo de objeto articulado (4) y con una postura de referencia particular de este modelo de objeto articulado (4);

• para cada secuencia (51) de al menos una secuencia de segmentos de imagen de origen (13), emparejar esta secuencia (51) con una pluralidad de secuencias (52) de siluetas de referencia (13') y determinar una o más secuencias seleccionadas de siluetas de referencia (13') que mejor se emparejen con la secuencia (51) de segmentos de imagen de origen (13);

• en donde el emparejamiento de las dos secuencias (51, 52) se realiza emparejando cada uno de los segmentos de imagen de origen (13) con la silueta de referencia (13') cuya posición dentro de su secuencia es la misma que la posición del segmento de imagen de origen (13) dentro de la secuencia (51) de segmentos de imagen de origen (13), calculando un error de emparejamiento que indica hasta qué punto se emparejan, y calculando a partir de los errores de emparejamiento de los segmentos de imagen de origen un error de emparejamiento de secuencia;

• para cada una de estas secuencias seleccionadas de siluetas de referencia (13'), recuperar la postura de referencia que está asociada a la silueta de referencia (13') que se ha emparejado con un segmento de imagen de origen (13) para el cual se determinará la postura del modelo de objeto articulado (4); y

• para cada segmento de imagen de origen (13) para el cual se determinará la postura del modelo de objeto articulado (4), calcular una estimación de la postura del modelo de objeto articulado (4) a partir de la postura o posturas de referencia recuperada,

**caracterizado por que** el método comprende las siguientes etapas para establecer la consistencia local entre dos posturas determinadas a partir de al menos una secuencia de segmentos de imagen de origen consecutivos de acuerdo con las etapas anteriores, estando cada postura asociada a al menos un segmento de imagen de origen (13) en donde los elementos, es decir, articulaciones (2) y/o enlaces (3), del modelo de objeto articulado (4) en una o ambas posturas corresponden a extremidades del objeto del mundo real (14) que se pueden etiquetar de manera ambigua:

• para cada uno de un par de segmentos de imagen de origen consecutivos, determinar, a partir de la postura asociada y para cada una de los posibles etiquetados ambiguos de cada postura, un etiquetado ambiguo correspondiente de puntos de imagen de extremidades, en el segmento de la imagen de origen;

• elegir, para el primero del par de segmentos de imagen de origen consecutivos, uno de los etiquetados ambiguos de la postura;

• calcular un flujo óptico entre el primero y el segundo de este par de segmentos de imagen de origen consecutivos;

• determinar, a partir del flujo óptico, ubicaciones en el segundo segmento de imagen de origen a las que se han movido los puntos de imagen correspondientes a las extremidades del primer segmento de imagen de origen, y etiquetar estas ubicaciones en el segundo segmento de imagen de origen según el etiquetado de las extremidades en el primer segmento de imagen de origen;

• asignar, de los posibles etiquetados ambiguos de la postura para el segundo segmento de imagen de origen que se utilizaron para etiquetar los puntos de imagen de acuerdo con la postura, el etiquetado que es más consistente con el etiquetado determinado a partir del flujo óptico.

2. El método de la reivindicación 1, en el que

• una imagen de la secuencia de imágenes de origen (10) se designa un cuadro de interés (56), y el segmento de imagen de origen (13) que se genera a partir del mismo se designa un segmento de imagen de origen de interés;

• el error de emparejamiento de secuencia es una suma ponderada de los errores de emparejamiento del emparejamiento de siluetas de las dos secuencias (51, 52);

• la ponderación del error de emparejamiento para el segmento de imagen de origen de interés es mayor y disminuye con la distancia de los segmentos de imagen de origen desde el segmento de imagen de origen de interés; y

• la postura de referencia se recupera para la silueta de referencia que se emparejó con el segmento de interés de

la imagen de origen.

3. El método de la reivindicación 1 o la reivindicación 2, en el que se obtienen y procesan al menos dos secuencias de imágenes de origen (10) grabadas por al menos dos cámaras fuente (9, 9') al mismo tiempo, y en donde se calcula la estimación de la postura del modelo de objeto articulado (4) a partir de las posturas de referencia recuperadas, determinadas a partir de al menos dos secuencias de imágenes de origen eligiendo la combinación de posturas de referencia recuperadas que mejor se emparejen en el espacio 3D.
4. El método de la reivindicación 1, en el que la etapa de etiquetado, en el segmento de la imagen de origen, de los puntos de imagen de las extremidades es realizado mediante
- para cada uno de un par de segmentos de imagen de origen consecutivos, determinar, a partir de la postura asociada y para cada una de los posibles etiquetados ambiguos de cada postura, una proyección de un modelo del objeto del mundo real (14) en la imagen de origen, y desde allí etiquetar los puntos de imagen del segmento de imagen de origen de acuerdo con la extremidad proyectada que es visible en el punto de imagen de ubicación.
5. El método de la reivindicación 1 o la reivindicación 4, en el que se da una secuencia de posturas, estando cada postura asociada a una de una secuencia de segmentos de imagen de origen, y en donde existe una ambigüedad con respecto al etiquetado de uno o más conjuntos de extremidades, que comprende las siguientes etapas para establecer la consistencia global de posturas combinadas con una secuencia de imágenes de origen consecutivas
- para cada segmento de imagen de origen (13) de la secuencia de imágenes de origen,
    - recuperar la postura asociada y el etiquetado de los elementos del modelo que fueron determinados por las etapas anteriores;
    - determinar la postura en la base de datos que tiene la menor distancia de la postura recuperada, teniendo en cuenta el etiquetado de las posturas de la base de datos;
    - calcular un término de error de consistencia que expresa la diferencia entre las dos posturas;
  - calcular, a partir de estos términos de error de consistencia, un error de consistencia total para toda la secuencia de imágenes de origen;
  - repetir las etapas anteriores para calcular el error de consistencia total para todas las variantes de posibles marcas globales de los conjuntos de miembros ambiguos;
  - elegir la variante de etiquetado global para la cual el error de consistencia total es el más pequeño.
6. El método de la reivindicación 5, en el que el error de consistencia total es la suma de todos los términos de error de consistencia sobre la secuencia.
7. El método de una de las reivindicaciones 1-6, en el que el objeto del mundo real (14) es observado por dos o más cámaras fuente (9), que comprende, además, las etapas de:
- dada la estimación de la postura del modelo de objetos articulados (4) y un conjunto de segmentos de imagen de origen (13) que comprenden cada uno una vista del objeto del mundo real (14) representado por el modelo de objetos articulados (4), observado al mismo tiempo desde diferentes puntos de vista,
  - realizar un procedimiento de optimización en el que se varía la postura del modelo de objeto articulado (4) para minimizar una función de error, en donde la función de error comprende un error de emparejamiento de modelo que se calcula sobre todas las vistas y penaliza las posturas cuyas posiciones de las articulaciones, cuando se proyecta en las vistas correspondientes al conjunto de segmentos de imagen de origen, no se emparejan con los segmentos de la imagen de origen.
8. El método de la reivindicación 7, en el que el error de emparejamiento del modelo comprende un término de error de relleno de silueta que penaliza las posiciones de las articulaciones que, cuando se proyectan en las vistas correspondientes a los segmentos de imagen de origen, hacen que los puntos dentro de la silueta se alejen del elemento más cercano del modelo de objeto articulado (4).
9. El método de la reivindicación 8, que comprende la etapa de calcular el término de error de relleno de silueta al encontrar, para cada punto de trama dentro de la silueta, la distancia al elemento más cercano de la proyección del modelo de objeto articulado (4), de acuerdo con su postura, en la vista que se está considerando, sumando estas distancias sobre todos los puntos de trama dentro de la silueta, y normalizando.
10. El método de cualquiera de las reivindicaciones 7 a 9, en el que el error de emparejamiento del modelo comprende un término de error de emparejamiento de silueta que penaliza las posiciones de articulación que, cuando se proyectan en las vistas correspondientes a los segmentos de imagen de origen, caen fuera de una silueta del segmento de imagen de origen respectivo.
11. El método de la reivindicación 10, que comprende la etapa de calcular el término de error de emparejamiento de

silueta proyectando cada articulación (2) desde su ubicación espacial en la vista 2D de cada una de las cámaras fuente correspondientes a los segmentos de imagen de origen, calcular, para cada articulación (2) y en cada vista 2D, una distancia de la articulación proyectada (2) desde la silueta del segmento de la imagen de origen, normalizando cada distancia de acuerdo con el tamaño de la silueta correspondiente, y sumando las distancias normalizadas.

- 5
12. El método de cualquiera de las reivindicaciones 7 a 11, en el que la función de error comprende un término de error de distancia a la base de datos que representa la desviación de la postura con respecto a la postura más cercana en la base de datos.
- 10
13. Un **medio legible por ordenador** no transitorio que comprende un código de programa legible por ordenador que codifica un programa informático que, cuando se carga y ejecuta en un ordenador, hace que el ordenador realice el método de una de las reivindicaciones anteriores.
- 15
14. Un **método para fabricar** un medio legible por ordenador no transitorio, que comprende la etapa de almacenar, en el medio legible por ordenador, instrucciones ejecutables por ordenador que cuando son ejecutadas por un procesador de un sistema informático, hacen que el sistema informático realice las etapas del método de una de las reivindicaciones 1 a 12.
- 20
15. Un **sistema informático** que comprende una unidad de procesamiento (15) configurada para interactuar con la unidad de almacenamiento (16) para almacenar al menos una de las imágenes de origen (10) y segmentos de imagen de origen (13), estando programada la unidad de procesamiento (15) para realizar las etapas del procedimiento de una de las reivindicaciones 1 a 12.

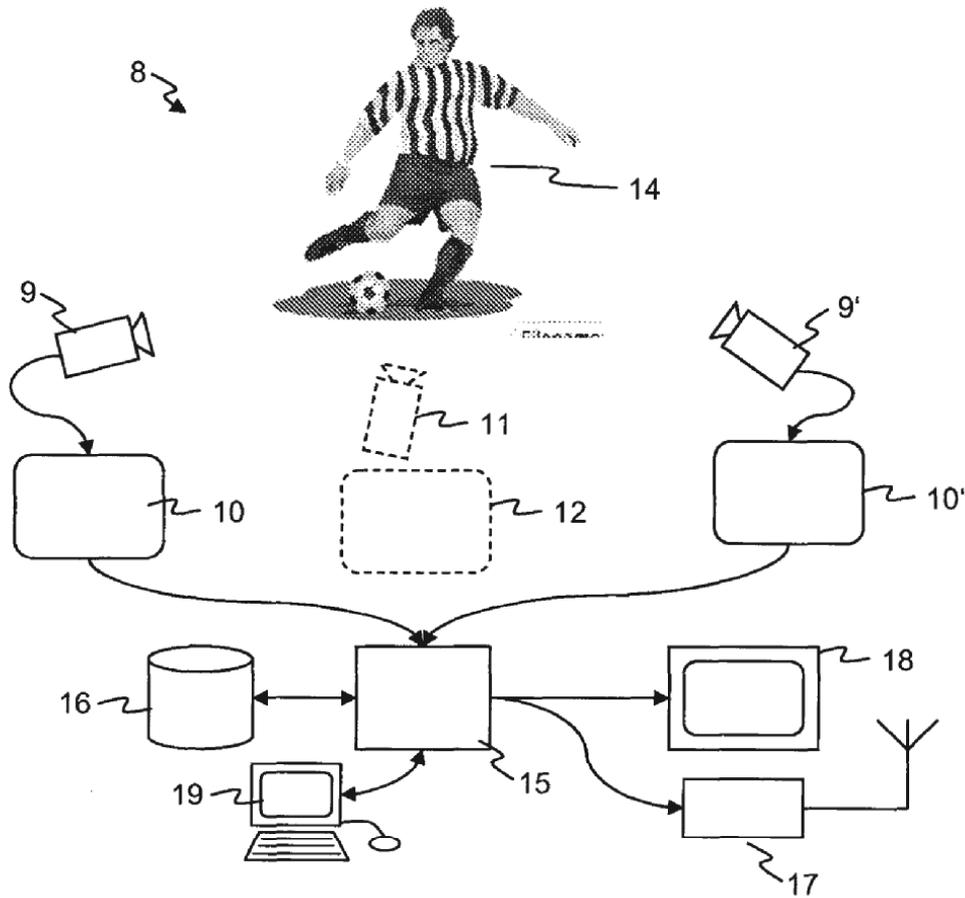


Fig. 1

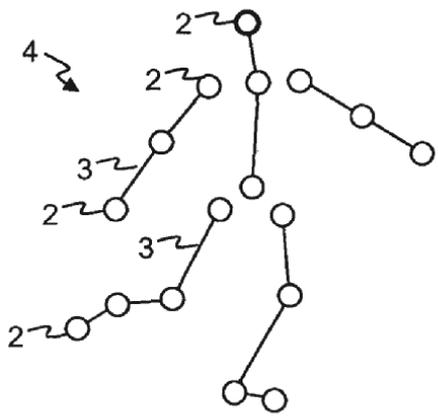


Fig. 2

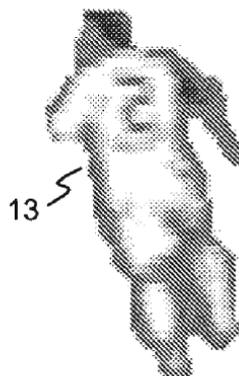


Fig. 3a

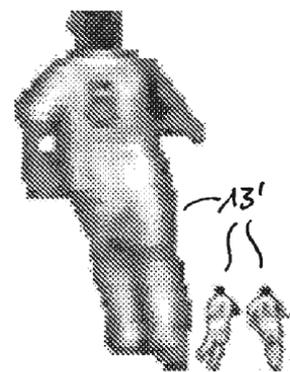


Fig. 3b

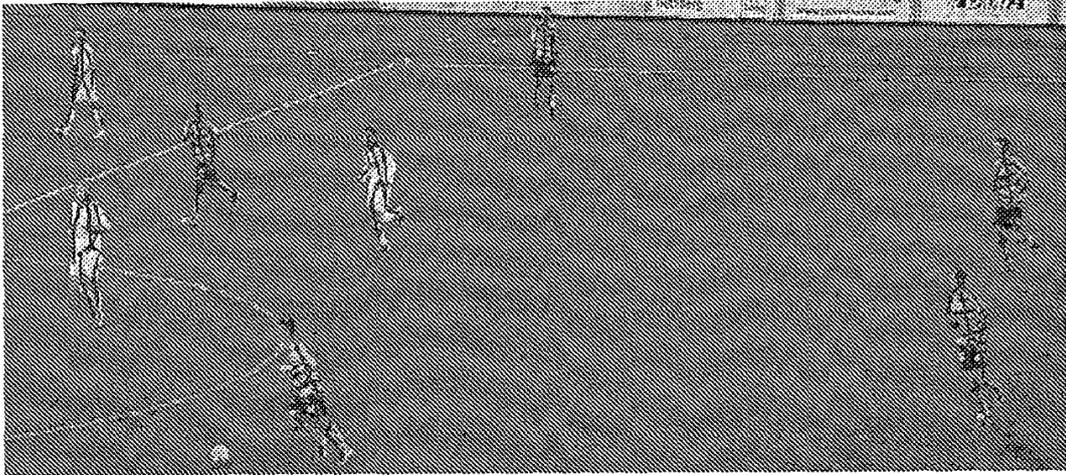
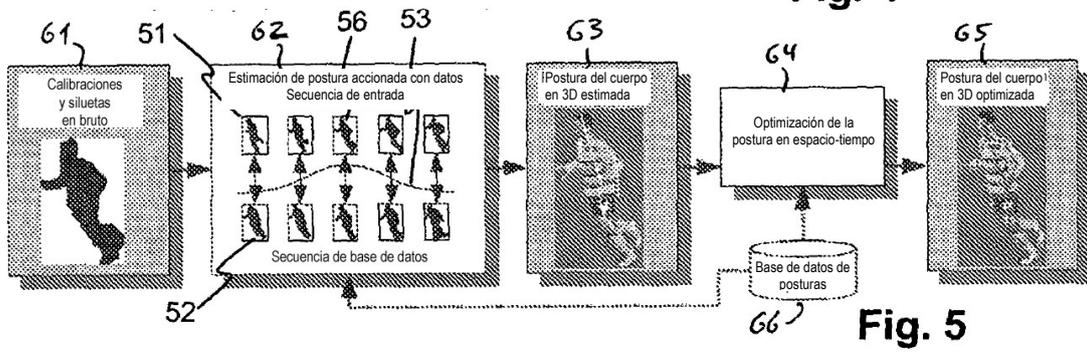


Fig. 4



(a)

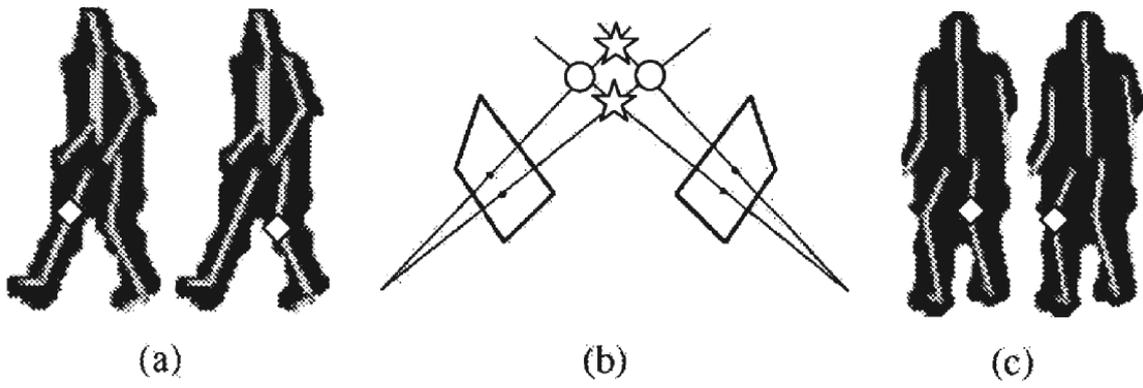
(b)

(c)

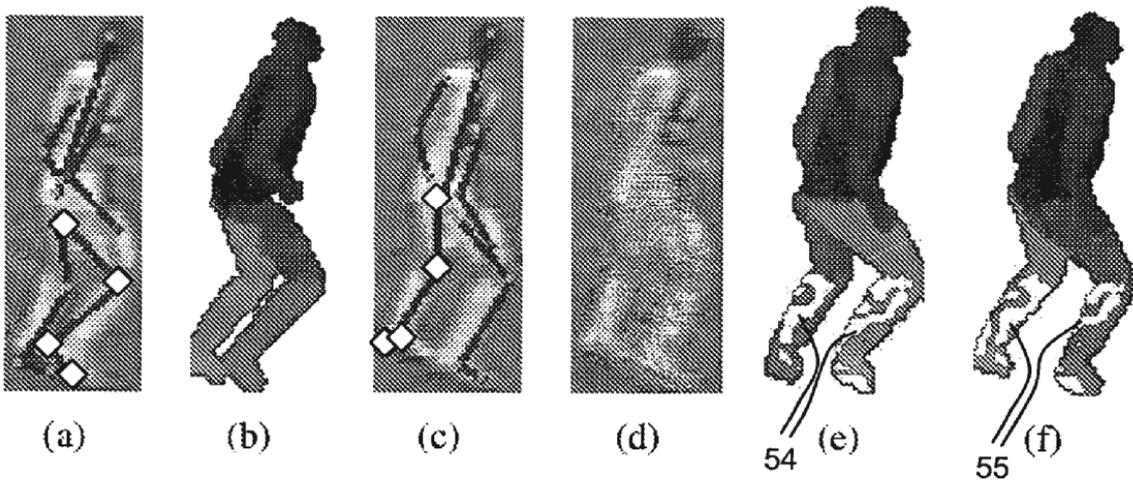
Fig. 6



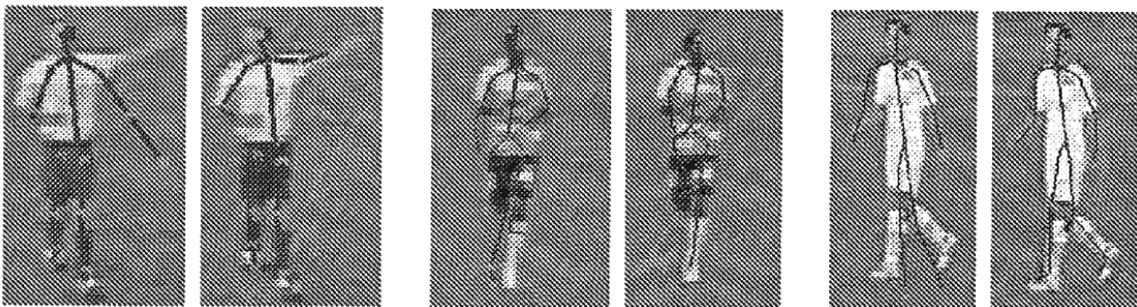
(d)



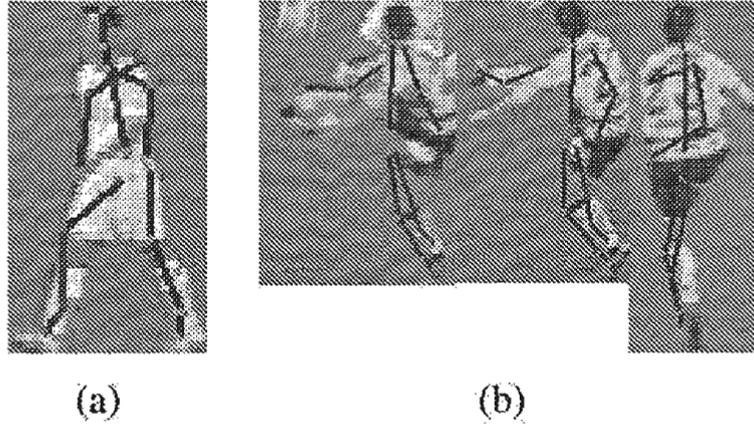
**Fig. 7**



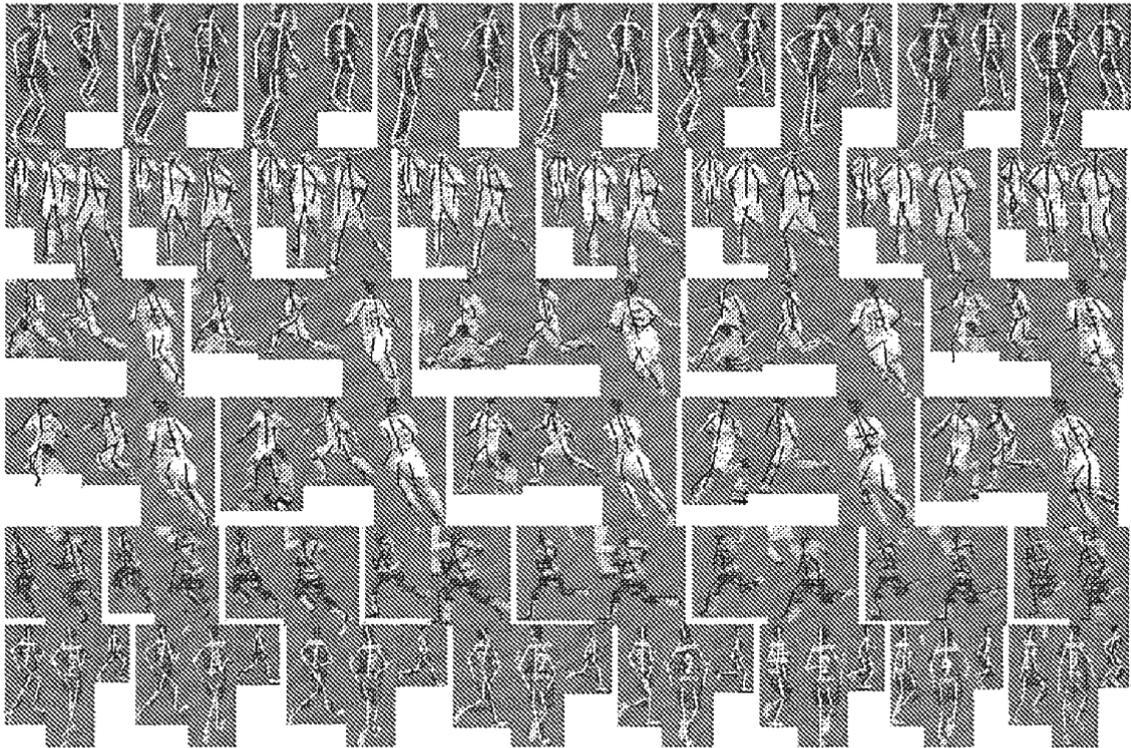
**Fig. 8**



**Fig. 9**



**Fig. 10**



**Fig. 11**