

SISTEMA Y MÉTODO PARA DETECTAR LA POSTURA CORPORAL ARTICULADA

CAMPO TÉCNICO

[001] La presente divulgación se refiere en general a la detección de una postura de cuerpo articulado; y más específicamente, a sistemas para detectar posturas corporales articuladas a partir de contenidos de imágenes. Además, la presente divulgación también se refiere a métodos para detectar posturas corporales articuladas a partir de contenidos de imágenes.

ANTECEDENTES

[002] Con el avance de la tecnología, la estimación de postura está ganando una enorme importancia. La estimación de postura contribuye a diversas aplicaciones, como el análisis de las actividades humanas, el análisis de las actividades de los animales para la investigación, etc. Además, la estimación de postura proporciona asistencia en la videovigilancia al detectar actividades ilegales por parte de los atracadores de tiendas, tal como ladrones, y, por lo tanto, alertar al personal empleado en la tienda para evitar el robo. Además, la estimación de postura se emplea en sistemas inteligentes de asistencia al conductor, sistemas de vida asistida para humanos necesitados, videojuegos, fisioterapia, etc. Además, la estimación de postura se utiliza activamente en el campo de los deportes, el ejército, la medicina, la robótica, etc.

[003] Generalmente, la detección de posturas es una tarea desafiante ya que cada humano posee una estructura corporal diferente, una forma corporal diferente, un color de piel diferente, etc. Además, los diferentes tipos de ropa de los seres humanos aumentan la complejidad en la estimación de la postura. Convencionalmente, se utiliza un método de estimación de postura de una sola persona para la estimación de postura. El método de estimación de la postura de una sola persona comprende un detector de personas, que detecta a cada persona en la imagen una por una, lo que lo convierte en un proceso que requiere mucho tiempo. Además, la detección de varios humanos en la imagen es difícil, ya que segmentar a los humanos del fondo de la imagen es una tarea agotadora. En particular, a medida que aumenta el número de personas, aumenta la complejidad de una estimación en

tiempo real de la postura humana, lo que hace que el rendimiento en tiempo real de la estimación de la postura sea un gran desafío.

[004] Además, las técnicas de estimación de postura actualmente utilizadas pueden conducir a falsos positivos, es decir, pueden detectar a un no humano como una estatua humana como un ser humano. Normalmente, las técnicas de estimación de postura utilizadas emplean una configuración de cámara de línea de visión horizontal que proporciona una vista limitada del área para la que se necesita realizar la estimación de postura.

[005] Por lo tanto, a la luz de la discusión anterior, existe la necesidad de superar los inconvenientes mencionados anteriormente asociados con las técnicas de detección de postura.

RESUMEN

[006] De acuerdo con un primer aspecto de la presente divulgación, se proporciona un sistema para detectar la postura del cuerpo articulado a partir de un contenido de imágenes. El sistema puede incluir un módulo de imágenes para capturar el contenido de las imágenes y un procesador acoplado comunicativamente al módulo de imágenes. El procesador se puede operar para obtener una vista de arriba hacia abajo del contenido de las imágenes y procesar la vista de arriba hacia abajo para detectar la postura del cuerpo articulado correspondiente al contenido de las imágenes usando un algoritmo de aprendizaje de máquina. El procesamiento puede incluir crear un mapa de confianza de la parte correspondiente a cada articulación de la postura del cuerpo articulado y generar un mapa de calor proyectando el mapa de confianza de la parte en la vista de arriba hacia abajo del contenido de las imágenes. El procesamiento puede incluir además crear un mapa de afinidad de partes correspondiente a cada parte del cuerpo asociada con cada articulación de la postura corporal articulada, y generar un mapa vectorial proyectando el mapa de afinidad de partes en la vista de arriba hacia abajo del contenido de imágenes. El procesamiento puede incluir además generar un armazón corporal correspondiente a la postura del cuerpo articulado, usando el mapa de calor y el mapa vectorial, para detectar la postura del cuerpo articulado.

[007] De acuerdo con un segundo aspecto de la presente divulgación, se proporciona un método para detectar una postura de cuerpo articulado a partir de un contenido de imágenes. El método puede incluir obtener una vista de arriba hacia abajo del contenido de las imágenes y procesar la vista de arriba hacia abajo para detectar la postura del cuerpo articulado correspondiente al contenido de las imágenes usando un algoritmo de aprendizaje de máquina. El procesamiento puede incluir crear un mapa de confianza de la parte correspondiente a cada articulación de la postura del cuerpo articulado y generar un mapa de calor proyectando el mapa de confianza de la parte en la vista de arriba hacia abajo del contenido de las imágenes. El procesamiento puede incluir además crear un mapa de afinidad de partes correspondiente a cada parte del cuerpo asociada con cada articulación de la postura corporal articulada y generar un mapa vectorial proyectando el mapa de afinidad de partes en la vista de arriba hacia abajo del contenido de imágenes. El procesamiento puede incluir además generar un armazón corporal correspondiente a la postura del cuerpo articulado, usando el mapa de calor y el mapa vectorial, para detectar la postura del cuerpo articulado.

[008] De acuerdo con un tercer aspecto de la presente divulgación, se proporciona un producto de programa informático para detectar una postura corporal articulada a partir de un contenido de imágenes. El producto programable por computadora comprende un conjunto de instrucciones, de modo que cuando lo ejecuta un procesador hace que el procesador obtenga una vista de arriba hacia abajo del contenido de la imagen y procese la vista de arriba hacia abajo para detectar la postura del cuerpo articulado correspondiente al contenido de la imagen. utilizando un algoritmo de aprendizaje de máquina, en el que la postura del cuerpo articulado incluye una pluralidad de articulaciones. El procesamiento puede incluir crear un mapa de confianza de la parte correspondiente a cada articulación de la postura del cuerpo articulado y generar un mapa de calor proyectando el mapa de confianza de la parte en la vista de arriba hacia abajo del contenido de las imágenes. El procesamiento puede incluir además crear un mapa de afinidad de partes correspondiente a cada parte del cuerpo asociada con cada articulación de la postura corporal articulada y generar un mapa vectorial proyectando el mapa de afinidad de partes en la vista de arriba hacia abajo del contenido de imágenes. El procesamiento puede incluir además generar un armazón corporal correspondiente a la postura del cuerpo articulado, usando el mapa de calor y el mapa vectorial, para detectar la postura del cuerpo articulado.

[009] Se apreciará que las características de la presente divulgación son susceptibles de combinarse en varias combinaciones sin apartarse del alcance de la presente divulgación tal como se define en las reivindicaciones adjuntas.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0010] El resumen anterior, así como la siguiente descripción detallada de las modalidades ilustrativas, se entienden mejor cuando se leen junto con los dibujos adjuntos. Con el fin de ilustrar la presente divulgación, en los dibujos se muestran construcciones ejemplares de la divulgación. Sin embargo, la presente divulgación no se limita a los métodos e instrumentos específicos descritos en el presente documento. Además, los expertos en la técnica comprenderán que los dibujos no están a escala. Siempre que ha sido posible, los elementos similares se han indicado con números idénticos.

[0011] Las modalidades de la presente divulgación se describirán ahora, solo a modo de ejemplo, con referencia a los siguientes diagramas en los que:

[0012] La FIG. 1 es un diagrama de bloques de un sistema para detectar una postura de cuerpo articulado a partir de un contenido de imágenes, de acuerdo con una modalidad de la presente divulgación;

[0013] La FIG. 2 es una ilustración de un contenido de imágenes obtenido por el sistema para detectar una postura de cuerpo articulado, de acuerdo con una modalidad ejemplar de la presente divulgación;

[0014] La FIG. 3 es una ilustración de una pluralidad de mapa de confianza parcial creado para detectar una postura corporal articulada, de acuerdo con una realización ejemplar de la presente divulgación;

[0015] Las Figs. 4A y 4B son ilustraciones de pluralidad de mapa de afinidad de partes creados para detectar una postura de cuerpo articulado, de acuerdo con una modalidad ejemplar de la presente divulgación;

[0016] La FIG. 5 es una ilustración de un marco corporal correspondiente a la postura del cuerpo articulado en el contenido de imágenes de la Fig. 2, de acuerdo con una modalidad ejemplar de la presente divulgación; y

[0017] La FIG. 6 ilustra los pasos de un método para detectar una postura de cuerpo articulado a partir de un contenido de imágenes, de acuerdo con una modalidad de la presente divulgación.

[0018] En los dibujos adjuntos, se emplea un número subrayado para representar un artículo sobre el que está colocado el número subrayado o un artículo al que está adyacente el número subrayado. Un número no subrayado se refiere a un artículo identificado por una línea que vincula el número no subrayado con el artículo. Cuando un número no está subrayado y está acompañado por una flecha asociada, el número no subrayado se utiliza para identificar un elemento general al que apunta la flecha.

DESCRIPCIÓN DETALLADA DE MODALIDADES

[0019] La siguiente descripción detallada ilustra modalidades de la presente divulgación y formas en las que se pueden implementar. Aunque se han divulgado algunos modos de llevar a cabo la presente divulgación, los expertos en la técnica reconocerán que también son posibles otras modalidades para llevar a cabo o poner en práctica la presente divulgación.

[0020] La presente divulgación proporciona un sistema y un método para detectar la postura del cuerpo articulado a partir de un contenido de imágenes que utiliza una vista de arriba hacia abajo del contenido de imágenes para detectar la postura del cuerpo articulado con precisión con la ayuda de una vista extendida proporcionada por la vista de arriba hacia abajo. Además, el sistema es capaz de detectar la postura corporal articulada de múltiples cuerpos humanos, animales o ambos, en el contenido de las imágenes, teniendo en cuenta el efecto de una estructura corporal diferente, una forma corporal diferente, un color de piel diferente asociado con cada cuerpo humano y / o animal en el contenido de las imágenes. Además, el sistema proporciona un proceso eficiente en el tiempo ya que la detección de múltiples posturas corporales articuladas ocurre simultáneamente. Además, el sistema permite reducir la complejidad a la que se enfrenta la detección de la postura del cuerpo

articulado en tiempo real al detectar con precisión cada cuerpo, incluso cuando el número de cuerpos aumenta en tiempo real.

[0021] Haciendo referencia a la FIG. 1, se muestra un diagrama de bloques de un sistema 100 para detectar una postura de cuerpo articulado a partir de un contenido de imágenes de acuerdo con la presente divulgación. El sistema 100 comprende un módulo 102 de imágenes para capturar el contenido de las imágenes. El contenido de imágenes comprende al menos uno de entre una imagen, un vídeo y un contenido basado en formato de intercambio de gráficos (GIF). El módulo de imágenes 102 está configurado para capturar el contenido de imágenes en forma de una o más imágenes, en el que la imagen incluye al menos un cuerpo cuya postura puede detectarse. Además, el contenido de las imágenes puede estar en forma de video que comprende una serie de cuadros que representan el cuerpo / cuerpos articulados en varias posturas. Además, el contenido de imágenes puede comprender un GIF que comprende una pluralidad de cuadros de naturaleza repetitiva, en el que la pluralidad de cuadros comprende al menos una postura corporal articulada.

[0022] El módulo de formación de imágenes 102 comprende un dispositivo de formación de imágenes, un procesador y una memoria. Opcionalmente, el dispositivo de imágenes incluye, pero no se limita a , una cámara de circuito cerrado de televisión (CCTV), una cámara de alta definición (HD), una cámara que no es de HD, una cámara de mano, una cámara, una cámara de coche de policía y cámaras empleadas en vehículos aéreos no tripulados (UAV). En particular, pueden usarse uno o más dispositivos de formación de imágenes dentro del módulo de imágenes 102 para capturar y / o grabar el contenido de las imágenes. Opcionalmente, el módulo de imágenes 102 está acoplado comunicativamente a un dispositivo remoto configurado para capturar y / o grabar el contenido de las imágenes. El dispositivo remoto incluye, entre otros, un teléfono inteligente, una cámara digital, una computadora portátil, una computadora personal y una tableta. Además, el módulo de imágenes 102 comprende el procesador configurado para procesar el contenido de imágenes recibido y / o capturado por el módulo de imágenes 102. A lo largo de la presente divulgación, el término "procesador" se refiere a un elemento computacional que es operable para responder y procesar instrucciones. que manejan el sistema 100. Opcionalmente, el procesador incluye, entre otros, un microprocesador, un microcontrolador, un microprocesador de computación de conjunto de instrucciones complejas (CISC), un

microprocesador de conjunto de instrucciones reducido (RISC), un microprocesador de palabra de instrucción muy larga (VLIW) o cualquier otro tipo de circuito de procesamiento. Además, el término "procesador" puede referirse a uno o más procesadores individuales, dispositivos de procesamiento y varios elementos asociados con el módulo 102 de formación de imágenes que pueden ser compartidos por otros dispositivos de procesamiento. El contenido de imágenes procesadas se almacena además en la memoria del módulo de imágenes 102. El término "memoria", como se usa en este documento, se refiere a un medio volátil o persistente, como un circuito eléctrico, disco magnético, memoria virtual o disco óptico, en el que una computadora puede almacenar datos o software durante cualquier período de tiempo. Opcionalmente, la memoria incluye almacenamiento masivo no volátil, como medios de almacenamiento físicos.

[0023] La figura 2 ilustra un contenido 200 de imágenes a modo de ejemplo generado por el módulo 102 de imágenes al enfocarse en una línea de visión vertical, mientras se configura el módulo 102 de imágenes. El contenido 200 de imágenes puede obtenerse directamente mediante un dispositivo de imágenes tal como cámaras CCTV, empleadas en altura para capturar una vista de arriba hacia abajo, etc. En un ejemplo, las CCTV se emplean para la vigilancia en un área como el vestíbulo de un hotel. En otro ejemplo, las cámaras se emplean a una altura en un campo de béisbol para capturar la vista de arriba hacia abajo de cada jugador en el campo de béisbol. En una modalidad, la vista de arriba hacia abajo puede obtenerse procesando una pluralidad de vistas de un contenido de imágenes para obtener la vista de arriba hacia abajo desde el mismo. En una modalidad, la pluralidad de vistas del contenido de imágenes comprende una vista trasera, una vista frontal, una vista superior, una vista inferior, una vista lateral izquierda, una vista lateral derecha y una vista en perspectiva.

[0024] Haciendo referencia de nuevo a la figura 1, el sistema 100 comprende además un módulo 104 de generación de verdad terreno que genera verdad terreno (GT) para el contenido de imágenes tal como el contenido de imágenes 200. El módulo de generación de GT 104 se refiere a una estructura y / o módulo que incluye componentes programables y / o no programables configurados para almacenar, procesar y / o compartir información.

[0025] La primera parte de la etiqueta GT incluye mapas de confianza de partes (PCM), donde los PCM definen dónde se ubican las articulaciones en una postura de cuerpo

articulado del contenido de imágenes 200, y qué parte del área alrededor de la posición de la articulación se consideraría “GT”. En un ejemplo, los PCM incluyen anotaciones de dónde se encuentran realmente las posiciones de las articulaciones en el contenido de imágenes 200. Para cada tipo de articulaciones (por ejemplo, hombro izquierdo, codo derecho, cabeza, etc.), el módulo de generación de GT 104 puede generar una matriz de etiquetas. En un ejemplo, la matriz de etiquetas puede ser una matriz de etiquetas (46x46).

[0026] En una realización de la presente divulgación, para generar los PCM, el módulo de generación de GT 104 está configurado para determinar primero (x, y) ubicaciones de cada una de las anotaciones conjuntas, -, escalar las ubicaciones determinadas a un valor de una resolución más baja y luego aplique una distribución normal, como la distribución de Gauss alrededor de la ubicación determinada. . La longitud de la distribución alrededor de la ubicación determinada se considera "GT" y se indica mediante un valor, sigma.

[0027] La FIG. 3 es una ilustración de los mapas de confianza de partes ejemplares primero a catorce (PCMs) 302a a 302h (en lo sucesivo denominados colectivamente PCM 302) generados para el contenido de imágenes 200. Cada PCM 302 se refiere a una representación gráfica de una ubicación de un punto clave anatómico bidimensional para cada articulación de las posturas corporales articuladas del contenido de imágenes 200. El cuerpo articulado se segmenta del fondo para calcular cada PCM 302, y cada articulación del cuerpo articulado se identifica para crear un PCM 302 correspondiente. En un ejemplo, cuando el contenido de imágenes 200 incluye dos jugadores que juegan al fútbol, se identifica cada articulación de cada uno de los jugadores, como una rodilla, un codo, un hombro, una muñeca, etc., y se puede crear el PCM para cada articulación. Una vez que se generan los PCM 302, se puede crear un mapa de calor proyectando los PCM 302 en la vista de arriba hacia abajo del contenido de imágenes correspondiente 200. En el mapa de calor, las uniones detectadas en los PCM 302 pueden superponerse en una región de las respectivas articulaciones del cuerpo articulado. La región puede incluir una ubicación exacta de las articulaciones respectivas.

[0028] La segunda parte de la etiqueta GT incluye Campos de afinidad de partes (Part Affinity Fields) (PAF), donde cada PAF define una conexión articulada de la postura del cuerpo articulado del contenido de imágenes 200. Ejemplos de conexión articulada incluyen, pero no se limitan a, un conexión cabeza-cuello, conexión codo derecho-hombro derecho y conexión codo derecho-hombro derecho. Para cada conexión articulada, el módulo de

generación de GT 104 está configurado para generar dos matrices de etiquetas, una matriz de etiquetas para la dirección X y otra matriz de etiquetas para la dirección Y. En un ejemplo, cada matriz de etiquetas es una matriz de etiquetas (46x46).

[0029] Para generar las matrices de etiquetas, el módulo de generación de GT 104 toma dos posiciones de articulación, por ejemplo, posiciones de articulación de cabeza y cuello, y traza una línea entre las dos posiciones de articulación para generar una conexión articulada. A continuación, el módulo 104 de generación de GT calcula un conjunto de puntos en cada una de las direcciones X e Y, que corresponden a la conexión articulada. El conjunto de puntos incluye uno o más puntos que se encuentran dentro de un umbral de distancia del segmento de línea entre las dos articulaciones.

[0030] Las Figs. 4A y 4B son ilustraciones de PAF ejemplares 400a y 400b generados para el contenido de imágenes 200, de acuerdo con una modalidad ejemplar de la presente divulgación. El PAF, como se usa en este documento, se refiere a un campo vectorial bidimensional asociado con cada conexión articulada del cuerpo articulado. En un ejemplo, en un cuerpo humano, un codo y una muñeca están conectados a través de un antebrazo. Así, el PAF corresponde a un antebrazo creado para detectar la postura del cuerpo articulado. Para crear un PAF para una conexión de articulación (como codo-muñeca, rodilla-tobillo, etc.), se pueden generar dos matrices correspondientes a los ejes x e y. A continuación, se calcula una línea entre la conexión de la articulación, en la que la línea puede corresponder a una región (como el antebrazo) que une la conexión de la articulación.

[0031] Una vez que se generan los PAF 400a y 400b por un contenido de imágenes, se puede generar un mapa vectorial proyectando los PAF 400a y 400b en la vista de arriba hacia abajo del contenido de imágenes 200. Por lo tanto, las conexiones articuladas detectadas en cada uno de los PAF 400a y 400b está superpuesto en una ubicación exacta de la conexión articulada respectiva del cuerpo articulado.

[0032] Para el contenido de imágenes 200, los PCM y PAF correspondientes se combinan para formar una etiqueta asociada a la imagen. La etiqueta es la verdad terreno para el contenido de imágenes 200.

[0033] En una modalidad de la presente divulgación, el módulo de generación de GT 104 está configurado para generar la verdad terreno para la red neuronal de estimación de postura 106 usando una pluralidad de conjuntos de datos que incluyen contenido de imagen similar al contenido de imágenes 200. Para múltiples tipos de conjuntos de datos, el módulo 104 de generación de GT puede configurarse para definir una estructura de esqueleto común. Además, los conjuntos de datos se validan de antemano para construir la mejor verdad terreno posible. Además, el contenido de la imagen de los conjuntos de datos se puede preprocesar para ajustar el contraste y el brillo. Además, el contenido de la imagen de los conjuntos de datos se puede multiplicar aplicando aumentos como rotaciones, traslaciones, escalado y zoom.

[0034] Haciendo referencia de nuevo a la figura 1, el sistema 100 incluye además una red neuronal 106 de estimación de postura que es una red neuronal convolucional para procesar las vistas de arriba hacia abajo de un contenido de imágenes. El término "red neuronal", como se usa en este documento, puede incluir una red altamente interconectada de elementos de procesamiento, cada uno asociado opcionalmente con una memoria local. En un ejemplo, la red neuronal puede ser un mapa de Kohonen, un perceptrón multicapa, etc. Además, los elementos de procesamiento de las redes neuronales pueden ser "unidades neuronales artificiales", "neuronas artificiales", "unidades neuronales", "neuronas", "nodos" y similares. Además, la neurona puede recibir datos de una entrada o de una o más neuronas, procesar los datos y enviar los datos procesados a una salida o incluso a una o más neuronas. La red neuronal o una o más neuronas de la misma se puede generar en hardware, software o una combinación de hardware y software, y la red neuronal se puede entrenar posteriormente. Se apreciará que la red neuronal convolucional consta de una capa de entrada, una pluralidad de capas ocultas y una capa de salida. Además, la pluralidad de capas ocultas de la red neuronal convolucional consiste típicamente en capas convolucionales, capas agrupadas, capas completamente conectadas y capas de normalización. Opcionalmente, se utiliza un modelo de Visual Geometry Group 19 (VGG 19) como arquitectura de red neuronal convolucional. El modelo VGG 19 está configurado para clasificar objetos en los datos de imágenes que se les alimentan. En un ejemplo, una imagen que comprende un automóvil, un ser humano sentado en el césped y un perro se alimenta a la red neuronal convolucional VGG 19. El VGG 19 identifica y clasifica el automóvil, el humano y el perro a partir de la imagen que se le alimenta. De manera similar, el modelo VGG 19 está entrenado para

identificar el cuerpo articulado en el contenido de las imágenes para la detección de la postura del cuerpo articulado. En particular, se pueden identificar múltiples cuerpos articulados y se pueden detectar las posturas asociadas con cada uno de los cuerpos articulados. El modelo VGG 19 está configurado para recibir el contenido de las imágenes a través de las capas de entrada del mismo. Además, el contenido de las imágenes se pasa a las capas ocultas para su posterior procesamiento. Se apreciará que las capas ocultas comprenden una pluralidad de conjuntos de capas de convolución.

[0035] La red neuronal de estimación de postura 106 está configurada para generar etiquetas de predicción 108 para el contenido de imágenes 200 en base al contenido de imágenes 200, y datos GT del contenido de imágenes 200. Para el contenido de imágenes 200, las etiquetas de predicción 108 incluyen PCM en múltiples valores sigma y PAF en múltiples valores de umbral.

[0036] En una modalidad de la presente divulgación, la red neuronal de estimación de postura 106 está configurada para crear el PCM para el contenido de imágenes 200 para una pluralidad de valores sigma. El PCM para cada articulación (como el hombro izquierdo, el codo derecho, la cabeza, etc.) se crea generando una matriz. Además, para generar la matriz, se toma una ubicación x-y para cada articulación. Además, se aplica una distribución normal (como una distribución de Gauss) alrededor de la articulación. Además, se asigna un valor "sigma" al PCM dependiendo de la cantidad de distribución normal alrededor de la articulación. En particular, cuanto mayor sea la distribución normal alrededor de la articulación, mayor será el valor de sigma.

[0037] En otra modalidad de la presente divulgación, la red neuronal de estimación de postura 106 está configurada para crear PAF para una pluralidad de valores umbral. El PAF para cada conexión articulada se crea generando dos matrices. Además, para generar las dos matrices, una matriz para la dirección X y la otra matriz para la dirección Y, se consideran dos conexiones articulares escaladas (como una cabeza-cuello). Además, se calcula la línea entre las conexiones articuladas y se calcula un conjunto de puntos que corresponden a la conexión articulada. Además, se asigna un valor de "umbral" al PAF dependiendo de la distancia de cada uno del conjunto de puntos desde la línea (es decir, la línea que conecta las dos articulaciones) en el PAF.

[0038] El sistema 100 incluye además un módulo 110 de extracción de articulaciones que está configurado para extraer una pluralidad de posiciones de articulaciones (x, y) para cada articulación del contenido de imágenes 200, de las etiquetas de predicción 108. La pluralidad de posiciones de articulaciones corresponde a la pluralidad de valores sigma y umbral de PCM y PAF respectivamente.

[0039] El sistema incluye además un módulo 112 de construcción de estructura de esqueleto que está configurado para construir una pluralidad de estructuras de esqueleto (en lo sucesivo también denominadas esqueletos inferidos) en base a la pluralidad correspondiente de posiciones conjuntas del contenido 200 de imágenes. La pluralidad de estructuras de esqueleto corresponde a la pluralidad de valores sigma y umbral de PCM y PAF respectivamente.

[0040] Haciendo referencia a la FIG. 5, se muestra una ilustración de un armazón corporal 500 generado por el módulo 112 de construcción de estructura de esqueleto para el contenido de imágenes 200, de acuerdo con una modalidad ejemplar de la presente divulgación. El armazón corporal 500 corresponde a una estructura de esqueleto construida a partir de articulaciones detectadas y conexiones articulares detectadas. Las posturas corporales articuladas detectadas pueden mostrarse como el armazón corporal 500 superpuesto a los cuerpos articulados.

[0041] Haciendo referencia de nuevo a la figura 1, el sistema 100 puede incluir además un dispositivo de visualización 114 para permitir al espectador ver la postura de cuerpo articulado detectada en tiempo real. Los ejemplos del dispositivo de visualización 114 incluyen, entre otros, dispositivos de visualización de cristal líquido (LCD), pantallas basadas en diodos emisores de luz (LED), dispositivos de visualización basados en LED orgánicos (OLED) y dispositivos de visualización basados en micro OLED.

[0042] En una modalidad de la presente divulgación, las estructuras de esqueleto generadas por el módulo 112 de construcción de estructura de esqueleto son utilizadas además por un módulo de entrenamiento 116 para el entrenamiento de la red neuronal 106 de estimación de postura. El módulo de entrenamiento 116 compara los esqueletos inferidos con los esqueletos GT y determina una serie de puntos de articulación coincidentes y una serie de esqueletos coincidentes. Se apreciará que el proceso de entrenamiento de la red neuronal 106

de estimación de postura se realiza hasta que sea capaz de generar la estructura o estructuras de esqueleto para el contenido de imágenes 200 con precisión durante un número predefinido de veces.

[0043] En una modalidad de la presente divulgación, para el contenido de imágenes 200, el módulo de entrenamiento 116 está configurado para comparar el PCM definido (es decir, el PCM de la verdad terreno) con cada uno de la pluralidad de PCM de predicción generados por la red neuronal de estimación de postura 106. Además, se selecciona el PCM de predicción que mejor coincide con el PCM de verdad terreno. Además, el valor sigma (es decir, el valor sigma verdadero) correspondiente al mapa de confianza de la parte PCM seleccionada se asigna al contenido de imágenes 200. En otra modalidad de la presente divulgación, para el contenido de imágenes 200, el sistema 100 es operable para comparar el PAF definido (es decir, el PAF del GT), con cada uno de los PAF predichos para seleccionar un valor umbral verdadero de la pluralidad de valores umbrales. Además, se selecciona el campo de afinidad de la parte que mejor coincide con la verdad terreno. Además, el valor umbral correspondiente al campo de afinidad de la parte seleccionada (es decir, el valor umbral verdadero) se asigna al contenido de imágenes 200.

[0044] En un ejemplo, los mapas de calor generados por el módulo 110 de extracción de articulaciones se comparan con los mapas de calor de verdad del suelo, comparando puntos (el centro de cada círculo en mapas de calor) y emparejándolos. Se considera que los dos puntos coinciden si están en la misma ubicación o si hay una diferencia de como máximo un píxel en el eje x o en el eje y, pero no en ambos. Luego, la salida de la inferencia, es decir, los esqueletos generados por el módulo 112 de construcción de la estructura esquelética, se comparan con los esqueletos de verdad terreno. Se emparejan dos esqueletos maximizando el número de puntos emparejados entre ellos y minimizando la distancia entre los puntos restantes. Los puntos se igualan como se especifica arriba. Con base en esta comparación, se determinan métricas como el número de esqueletos emparejados y el número restante de esqueletos (sin emparejar).

[0045] Con referencia a la figura 6, se ilustran pasos del método para detectar la postura del cuerpo articulado a partir del contenido de imágenes, de acuerdo con una modalidad de la presente divulgación. En un paso 602, se obtiene una vista de arriba hacia abajo del contenido de imágenes. En un paso 604, la vista de arriba hacia abajo se procesa para detectar

la postura del cuerpo articulado correspondiente al contenido de las imágenes usando un algoritmo de aprendizaje de máquina. El paso 604 comprende además crear el mapa de confianza de la parte correspondiente a cada articulación de la postura del cuerpo articulado, generando el mapa de calor proyectando el mapa de confianza de la parte en la vista de arriba hacia abajo del contenido de la imagen, creando el campo de afinidad de la parte correspondiente a cada parte del cuerpo asociada con cada articulación de la postura del cuerpo articulado, generando el mapa vectorial proyectando el campo de afinidad de la parte en la vista de arriba hacia abajo del contenido de la imagen y generando un armazón corporal correspondiente a la postura del cuerpo articulado, usando el mapa de calor y el vector mapa, para detectar la postura del cuerpo articulado.

[0046] Las modificaciones a las modalidades de la presente divulgación descritas anteriormente son posibles sin apartarse del alcance de la presente divulgación tal como se define en las reivindicaciones adjuntas. Expresiones como "que incluye", "que comprende", "que incorpora", "que consiste en", "tiene", "es", que se utilizan para describir y reclamar la presente divulgación, deben interpretarse de manera no exclusiva, es decir, permitiendo artículos, componentes o elementos no descritos explícitamente también deben estar presentes. También se debe interpretar que la referencia al singular se refiere al plural.