

MÉTODO PARA PROBARSE PRENDAS DE VESTIR BASADO EN INTELIGENCIA ARTIFICIAL Y TÉCNICAS DE PROCESAMIENTO DE IMÁGENES EN 2D

SECTOR DE LA TÉCNICA

5

La presente invención hace parte del campo técnico de procesamiento de imágenes de prendas de vestir, específicamente en el área de herramientas para pruebas de prendas de vestir mediante procesamiento de imágenes e inteligencia artificial.

10

ANTECEDENTES DE LA INVENCION

Con el aumento exponencial de compras de ropa a través del comercio electrónico, también han aumentado las devoluciones, principalmente por problemas de tallas o simplemente porque no satisface las expectativas de los consumidores.

15

El problema se debe a la limitante que tienen las personas en probarse la ropa, en los canales digitales de las múltiples marcas o vendedores de ropa, lo cual genera incertidumbre al momento de la compra en cuanto a las tallas, medidas y aspecto.

20

Acertar en las tallas y como lucen las prendas de vestir de manera digital en las personas, es un reto muy grande para las empresas del sector de la moda, principalmente por el desconocimiento de las medidas corporales de las personas y se combina con la diversidad de patronajes y estilos de prendas de las diferentes marcas.

25

Actualmente, existen procedimientos que permiten probar prendas de vestir en imágenes de personas mediante inteligencia artificial y procesamiento de imágenes, mediante la segmentación de las mismas y ajuste automático de la prenda sobre la imagen objetivo.

30

En primera medida, la patente CN112598806 divulga un método de prueba virtual basado en inteligencia artificial. Se usa un modelo de red neuronal convolucional para obtener una imagen de ropa con una acción específica, y se obtiene una imagen de usuario de una acción específica a través de un modelo de red adversario generativo.

35

Las imágenes de efecto de prueba generadas posteriormente tienen un mayor grado de fusión, y la ropa en las imágenes de prueba tiene un mayor grado de ajuste con el usuario. Este modelo resulta relevante para fusionar una imagen segmentada de una prenda de vestir con una máscara de las partes del cuerpo para posteriormente

entrenar y clasificar un modelo de inteligencia artificial para escoger la mejor prenda de vestir, sin embargo, este método no cuenta con un proceso de deformación de la región de la prenda de vestir objetivo para ajustarla a un área que delimite esta prenda ni tampoco realiza una validación de la pose de la persona en la imagen de referencia mediante el entrenamiento de un modelo de inteligencia artificial, lo que resulta en una clasificación menos eficiente de la prenda correcta sobre la imagen de la persona al no tener un ambiente controlado en el procesamiento de las imágenes.

Por otra parte, la patente WO2021040256 divulga un método para recomendar ropa de un dispositivo electrónico que incluye recibir una imagen que incluye una prenda de vestir y una pluralidad de imágenes combinadas obtenidas combinando una pluralidad de prendas de vestir obtenidas de una pluralidad de proveedores de alimentación. Esta combinación de prendas correspondientes a cada uno de la pluralidad de proveedores de alimentación incluye la prenda de vestir incluida en la imagen de entrada ingresando la imagen de entrada a un modelo de inteligencia artificial que proporciona información sobre una prenda de vestir incluida en al menos una prenda combinación entre la pluralidad de combinaciones de ropa. En el caso de esta patente, igualmente no menciona un proceso de deformación de la región de la prenda de vestir objetivo para ajustarla a un área que delimite esta prenda ni tampoco realiza una validación de la pose de la persona en la imagen de referencia mediante el entrenamiento de un modelo de inteligencia artificial, lo que resulta al igual que la patente anterior, en una eficiencia menor en el proceso de clasificación de la prenda correcta sobre la imagen de la persona al no tener un ambiente controlado en el procesamiento de las imágenes.

Adicionalmente, la patente KR20220000123 presenta un método para sintetizar y generar una imagen fotográfica natural, como una fotografía tomada después de probarse la ropa seleccionada sin probarse realmente la imagen de la ropa seleccionada del cliente poniendo la imagen de la ropa seleccionada del cliente en la imagen de la fotografía del cliente (Gafas, zapatos, accesorios, etc.). Esta invención logra el ajuste virtual solo con una imagen 2D sin usar un modelo 3D. En este caso en particular, esta patente resulta igualmente relevante para la unión de dos regiones segmentadas de una prenda de vestir y las partes del cuerpo en una persona mediante procesamiento de imágenes, sin embargo, posee las limitantes de las patentes anteriormente mencionadas respecto a la deformación de la región de la prenda de vestir objetivo para ajustarla a un área que delimite esta prenda y la validación de la pose de la persona en la imagen de referencia mediante el entrenamiento de un modelo de inteligencia artificial, así como que tampoco

menciona el entrenamiento y clasificación de un modelo de inteligencia artificial para la escogencia de la mejor prenda de vestir sobre cada persona.

Por consiguiente, en la actualidad no existen métodos que permitan realizar una clasificación de prendas de vestir de acuerdo con diferentes imágenes de personas por medio de modelos de procesamiento de imágenes e inteligencia artificial con alta precisión, reduciendo tiempos de cálculos y respuesta, así como una reducción de carga computacional, generando mejores resultados en el proceso de inferencia.

DESCRIPCIÓN BREVE DE LA INVENCION

Un método implementado por computador para probar prendas de vestir basado en inteligencia artificial y procesamiento de imágenes que permite generar una nueva imagen de manera automática tomando como referencia la imagen individual de cada persona y la imagen individual de cada prenda de vestir, permitiendo el efecto visual de lucir cualquier prenda seleccionada sobre el cuerpo de las personas sin contacto físico con las prendas.

La solución plantea una serie de etapas del proceso que permite el cálculo automático de tallas y ver una aproximación de como luciría la prenda en la silueta de las personas, utilizando como base una fotografía real tanto de la prenda como del usuario, para mejorar la experiencia de compra por el comercio electrónico frente a la limitación de probarse las prendas de vestir de manera online y reducir las devoluciones por problemas de tallas.

El método recolecta y almacena imágenes individuales de prendas de vestir y personas, las cuales son transformadas y pre procesadas mediante la segmentación semántica de las mismas que permite extraer un mapa de las partes del cuerpo de cada imagen de personas y una región de la prenda de vestir objetivo, para posteriormente fusionarlas y ejecutar un modelo de inteligencia artificial que permita probar cualquier prenda de vestir sobre una fotografía de una personas todo esto en 2D.

DESCRIPCIONES DE LAS FIGURAS

La Figura 1 ilustra un diagrama de flujo del método implementado por computador para probar prendas de vestir basado en inteligencia artificial.

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

A continuación, se describirá de manera detallada la presente invención que se expone en la figura adjunta.

5

La figura 1 presenta el diagrama de flujo del método implementado por computador para probar prendas de vestir basado en inteligencia artificial (100), el cual se compone de una primera etapa de recolección y almacenamiento de imágenes de prendas de vestir y personas (101) en un repositorio o base de datos.

10

La recolección de imágenes corresponde a las actividades de construcción del conjunto de datos requerido para entrenamiento, pruebas y validaciones de los modelos de machine learning, mediante la toma de fotografías y/o generación de imágenes digitales que se clasifican y se almacenan en tres grupos: personas, prendas y prendas sobre personas.

15

Luego de recolectar las imágenes y de almacenarlas en algún repositorio de datos se requerirá aplicarles procesos de transformación, es decir, etiquetar o caracterizar las imágenes, para que estén preparados para el entrenamiento a partir de los modelos disponibles de Inteligencia Artificial.

20

En primera medida, se realizar una validación de la pose de la persona en la imagen de referencia mediante el entrenamiento de un modelo k-means (102) con las longitudes de los brazos y los cuatro ángulos que se crean entre los vértices del codo y muñeca con su horizontal perpendicular. El modelo k-means es un modelo de aprendizaje no supervisado en el cual no se realiza un proceso de etiquetado de los datos de entrenamiento y agrupa estos datos en grupos que se generan automáticamente a partir de similitudes.

25

Limitar la pose de la persona en la imagen de referencia, permite generar mejores resultados y tener un ambiente más controlado del mismo. Para obtener las coordenadas de los puntos claves del cuerpo, se utiliza el modelo PoseNet en una aplicación web, el cual genera 17 puntos claves del cuerpo, lo cual reduce cálculos computacionales, tiempos de respuesta y carga computacional de los servicios.

30

35

Posteriormente, se seleccionan solo las coordenadas de los brazos, para luego calcular las distancias de los mismos usando una distancia euclidiana y los ángulos son calculados usando la ley de coseno y la distancia de los brazos.

La pose de referencia que se utiliza es una persona en posición frontal, con las extremidades rectas, los brazos ligeramente separados del cuerpo, totalmente extendidos y las piernas semiabiertas.

5

Posteriormente, cada imagen se transforma mediante una segmentación semántica (103) que genere máscaras de las partes del cuerpo de la persona y una región de la prenda de vestir objetivo. Para esto, se utiliza la segmentación semántica de las partes del cuerpo y la ropa para generar progresivamente la máscara de las partes del cuerpo expuestas (es decir, la máscara de partes del cuerpo sintetizada) y la máscara de las regiones de la ropa deformada. Una segmentación semántica consiste en asociar una etiqueta o categoría a cada píxel de la imagen que se está analizando mediante un modelo de aprendizaje profundo.

10

15

La segmentación semántica es un problema del campo de la visión artificial y consiste en dividir una imagen digital en diferentes regiones denominadas segmentos. Para este caso en concreto, se realiza la segmentación semántica de la fotografía de la persona, clasificando cada píxel de la imagen en 20 posibles clases: 'Background', 'Hat', 'Hair', 'Glove', 'Sunglasses', 'Upper-clothes', 'Dress', 'Coat', 'Socks', 'Pants', 'Jumpsuits', 'Scarf', 'Skirt', 'Face', 'Left-arm', 'Right-arm', 'Left-leg', 'Right-leg', 'Left-shoe', 'Right-shoe', obteniendo así una imagen de la segmentación semántica de la fotografía de la persona.

20

25

Para obtener la imagen de la segmentación semántica, se utiliza un modelo de inteligencia artificial entrenado específicamente para realizar segmentación semántica de fotografía de personas. Este modelo recibe una imagen de entrada de una persona y retorna otra imagen de las mismas dimensiones, con la clasificación de cada píxel usando un número entre 0 y 19, donde cada número corresponde a una de las 20 clases mencionadas anteriormente.

30

Con base en lo anterior, se generan máscaras semánticas en dos etapas para generar primero las partes del cuerpo y sintetizar progresivamente la máscara de la ropa. Adicionalmente, se utiliza la segmentación semántica de la imagen de referencia (103) para predecir progresivamente la disposición semántica deseada después del proceso de "try-on". Este proceso separa la región de la prenda de vestir objetivo, preserva las partes del cuerpo (brazos) de la persona, sin cambiar la pose y el resto de los detalles del cuerpo. En este módulo se adopta el mecanismo de generación de

35

máscaras para generar la segmentación semántica de las partes del cuerpo y la región de la ropa objetivo con precisión.

Esta máscara del cuerpo se genera mediante el entrenamiento de una red generativa antagónica (GAN) a partir de la información de segmentación semántica de las partes del cuerpo de la persona de la imagen, la segmentación semántica de la prenda de vestir y un mapa de la pose de la persona en la imagen. Utilizando la información generada de las partes del cuerpo, su correspondiente mapa de pose y la imagen de la ropa objetivo, es posible obtener la región de la ropa estimada.

En un siguiente paso, se combina la máscara de las partes del cuerpo con el mapa de pose y la imagen de la ropa objetivo para generar la máscara sintetizada de la ropa mediante dos generadores de una red condicional generativa antagónica (cGAN), en la que se utiliza una estructura de red neuronal convolucional U-Net como generador, y un discriminador para distinguir las máscaras generadas de sus máscaras originales. Para cada una de las etapas, la función de pérdida de la red cGAN puede formularse como:

$$L_1 = E_{x,y}[\log \log (D(x,y))] + E_{x,z}[\log \log (1 - D(x,G(x,z)))]$$

Donde x indica la entrada, y es la máscara verdadera y z es el ruido, el cual es un canal adicional de entrada muestreado a partir de una distribución normal estándar. La función objetivo global para cada etapa del módulo de generación de máscaras de prueba propuesto se formula como L_m :

$$L_m = \lambda_1 L_1 + \lambda_2 L_2$$

Donde L_2 es la pérdida de entropía cruzada por píxel y λ_1 y λ_2 son los parámetros de compensación para cada término de pérdida.

Seguida a la etapa anterior, se realiza un proceso de deformación de la región de la prenda de vestir objetivo (104) para ajustarla a un área que delimite esta prenda mediante procesos de STN (Spatial Transformation Network) y curvas diferenciables definidas en porciones (splines) para la interpolación y suavizado de imágenes. Esta interpolación y suavizado de imágenes se implementa mediante la técnica de splines de placa delgada (TPS). Un proceso de STN consiste en un módulo de aprendizaje que se puede añadir a una red neuronal convolucional (CNN) para aumentar la invariancia espacial de manera eficiente, mientras que una curva diferenciable definida en porciones (spline) consiste en una curva polinomial para generar interpolaciones de datos y ajustes de curvas para reducción de ruido (suavizado), tal

como el caso de la técnica de splines TPS, el cual es método de interpolación que encuentra una superficie lisa que pasa por todos los puntos dados.

Adicionalmente, en este proceso de deformación se introduce una restricción de diferencia de segundo orden para estabilizar el proceso de deformación durante el entrenamiento en casos complejos como ropa con texturas complejas y colores vivos, evitando así una mala alineación y resultados borrosos. Esta restricción es un valor que se utiliza durante el proceso de deformación de la prenda, el cual limita o coloca un umbral de deformación y de esta manera se puede estabilizar el proceso de deformación y evitar tener deformaciones poco realistas.

A continuación, se recorta la parte frontal del cuello de la región de la prenda de vestir (105), logrando obtener mejores inferencias en los detalles y forma del cuello de la prenda, esto gracias a que se realiza una modificación a la máscara de la prenda objetivo (prenda que se desea medir la persona).

Luego, se une en una sola máscara las partes que componen el cuerpo de la persona de la imagen (106), donde posteriormente se rellena la parte cambiante del cuerpo utilizando las máscaras y las imágenes generadas en los pasos anteriores mediante una red GAN, que fusiona la región deformada de la prenda de vestir con la máscara de las partes del cuerpo (107) para determinar de forma adaptativa la generación o conservación de las distintas partes humanas en la imagen sintetizada.

Finalmente, se realiza el entrenamiento de un modelo de inteligencia artificial (108) basado en métricas de similitud estructural (SSIM), que se utiliza para medir la similitud entre las imágenes sintetizadas y la verdadera, y la medición de calidad visual de las imágenes sintetizadas (IS). Con este proceso de entrenamiento (108) completado, se realizan inferencias sobre el modelo de inteligencia artificial previamente entrenado para validación y prueba (109). Respecto a la validación, se utiliza la técnica de puntaje inicial (Inception Score) para determinar la calidad visual de las imágenes sintetizadas y además se realiza una evaluación perceptual de los resultados, utilizando calificadores humanos, los cuales califican en una escala de 1 a 5 la calidad visual de los resultados, donde 1 es muy malo, 2 es malo, 3 es regular, 4 es bueno y 5 muy bueno, además que para la fase de prueba se utilizan los calificadores humanos usando la misma metodología de calificación mencionada anteriormente.

EJEMPLO DE REALIZACIÓN DE LA INVENCIÓN

A continuación, se detallará la realización preferida para la realización y ejecución de la presente invención.

5

El método (100) ejecuta una herramienta que facilita el recorte de la imagen de referencia por parte del usuario. Esta herramienta fue desarrollada puesto que se requiere que el cuerpo de la persona ocupe el mayor porcentaje de la fotografía, lo cual permitirá generar mejores resultados de inferencia.

10

Posteriormente, se ejecuta el proceso de deformación de la prenda que se realiza utilizando un proceso de STN, el cual tiene como entrada la fotografía de la prenda y la máscara de la región donde debe estar puesta dicha prenda. De esta manera la red neuronal que está compuesta de capas convoluciones (Convolutional neural network) aprenderá a deformar la prenda a color basado en la máscara de la región donde debe colocar la misma. Luego de tener la prenda deformada, se realiza un proceso de refinamiento, en el cual se utiliza una red neuronal tipo U-Net, la cual genera la prenda deformada con la forma de la máscara de la región en la cual debe adaptarse.

20

Posteriormente, el proceso de transformación y entrenamiento se realiza de acuerdo al tipo de prenda, es decir, un modelo para prendas de manga corta y cuello redondo y en "V", y otro modelo para prendas manga larga, cuello redondo y en "V". De esta manera, se logran conseguir modelos más precisos disminuyendo la complejidad de la tarea a realizar.

25

Adicionalmente, las inferencias contienen mejores resultados respecto a la talla del pecho, donde se utilizan diversas versiones de una misma prenda en tallas de busto diferente. Para esto, se incluye un subproceso adicional en el procesado de las prendas, donde se aumenta o disminuye de manera sintética la talla del busto. De esta manera, se consiguen resultados visualmente más atractivos, ya que se conserva la talla del busto.

30