

REIVINDICACIONES

1. Un sistema para reconocer la actividad humana a partir de una secuencia de video capturada por un dispositivo de imágenes, el sistema comprende:

una memoria para almacenar una o más instrucciones; y

un procesador acoplado comunicativamente a la memoria para ejecutar una o más instrucciones, en el que el procesador comprende:

un clasificador acoplado comunicativamente al dispositivo de imágenes y configurado para:

clasificar un cuadro de imagen de la secuencia de video en una o más clases de un conjunto de clases predefinidas, en el que el cuadro de imagen se clasifica basándose en la acción del usuario en una región de interés del cuadro de imagen; y

generar un vector de probabilidad de clase para el cuadro de imagen basado en la clasificación, en el que el vector de probabilidad de clase incluye un conjunto de probabilidades de clasificación del cuadro de imagen en cada clase predefinida;

un módulo de filtrado y binarización de datos configurado para filtrar y binarizar cada valor de probabilidad del vector de probabilidad de clase basado en un valor umbral de probabilidad predefinido;

un módulo de composición de palabras comprimidas configurado para:

determinar una o más transiciones de una o más clases en uno o más cuadros de imagen consecutivos de la secuencia de video, basándose en los correspondientes vectores de probabilidad binarizados; y

generar una secuencia de palabras comprimidas en base a una o más transiciones determinadas en uno o más cuadros de imagen consecutivos; y

un clasificador dependiente de la secuencia configurado para extraer una o más acciones del usuario analizando la secuencia de palabras comprimidas y reconocer la actividad humana de la misma.

2. El sistema de acuerdo con la reivindicación 1, en el que el clasificador es una red neuronal convolucional.
3. El sistema de acuerdo con lo reivindicado en la reivindicación 1, en el que el conjunto de clases predefinidas para una zona de escaneo de Self-check out (auto pago) (SCO) incluye clases como mano, objeto en mano, objeto, parte del cuerpo y escáner vacío.
4. El sistema de acuerdo con la reivindicación 1, en el que el módulo de filtrado y binarización de datos es además operable para eliminar errores de clasificador en los vectores de probabilidad de clase de uno o más cuadros de imagen consecutivos.
5. El sistema de acuerdo con la reivindicación 1, en el que el procesador comprende además un módulo de detección de intervalo silencioso, en el que el módulo de detección de intervalo silencioso está configurado para detectar uno o más intervalos silenciosos en la secuencia de video basándose en la no detección de actividad en la región de interés para una duración de umbral predefinida.
6. El sistema de acuerdo con la reivindicación 1, en el que una palabra comprimida se forma a partir de letras de un alfabeto que contiene un número de letras equivalente al doble del número de clases predefinidas.
7. El sistema de acuerdo con la reivindicación 1, en el que cada palabra comprimida de la secuencia de palabras comprimidas comprende al menos una trama de inactividad entre ellas.

8. El sistema de acuerdo con la reivindicación 1, en el que el clasificador dependiente de la secuencia es una red neuronal recurrente.

9. Un método para reconocer la actividad humana a partir de una secuencia de video, comprendiendo el método

clasificar mediante un clasificador, un cuadro de imagen de la secuencia de video en una o más clases de un conjunto de clases predefinidas, en el que el cuadro de imagen se clasifica basándose en la acción del usuario en una región de interés del cuadro de imagen;

generar un vector de probabilidad de clase para el cuadro de imagen basado en la clasificación, en el que el vector de probabilidad de clase incluye un conjunto de probabilidades de clasificación del cuadro de imagen en cada clase predefinida;

binarizar cada valor de probabilidad del vector de probabilidad de clase basándose en un valor umbral de probabilidad predefinido;

determinar una o más transiciones de una o más clases en uno o más cuadros de imagen consecutivos de la secuencia de video, basándose en los correspondientes vectores de probabilidad binarizados;

generar una secuencia de palabras comprimidas en base a una o más transiciones determinadas en uno o más cuadros de imagen consecutivos; y

extraer una o más acciones del usuario analizando la secuencia de palabras comprimidas mediante un clasificador dependiente de la secuencia y reconocer la actividad humana de las mismas.

10. El método de acuerdo con la reivindicación 9, en el que el clasificador es una red neuronal convolucional.

11. El método de acuerdo con la reivindicación 9, en el que el conjunto de clases predefinidas para una zona de exploración de Self-check out (auto pago) (SCO) incluye clases tales como mano, objeto en mano, objeto, parte del cuerpo y escáner vacío.

12. El método de acuerdo con la reivindicación 9, que comprende además la eliminación de errores de clasificador en los vectores de probabilidad de clase de uno o más cuadros de imagen consecutivos.

13. El método de acuerdo con la reivindicación 9, que comprende además detectar uno o más intervalos silenciosos en la secuencia de video basándose en la detección de ausencia de actividad en la región de interés durante una duración de umbral predefinida.

14. El método de acuerdo con la reivindicación 9, en el que una palabra comprimida se forma a partir de letras de un alfabeto que contiene un número de letras equivalente al doble del número de clases predefinidas.

15. El método de acuerdo con la reivindicación 9, en el que cada palabra comprimida de la secuencia de palabras comprimidas comprende al menos una trama de no actividad entre ellas.

16. El método de acuerdo con la reivindicación 9, en el que el clasificador dependiente de la secuencia es una red neuronal recurrente.

17. Un producto programable por computadora para reconocer la actividad humana a partir de una secuencia de video, el producto programable por computadora que comprende un conjunto de instrucciones, el conjunto de instrucciones cuando es ejecutado por un procesador hace que el procesador pueda:

clasificar un cuadro de imagen de la secuencia de video en una o más clases de un conjunto de clases predefinidas, en el que el cuadro de imagen se clasifica en función de la acción del usuario en una región de interés del cuadro de imagen;

generar un vector de probabilidad de clase para el cuadro de imagen basado en la clasificación, en el que el vector de probabilidad de clase incluye un conjunto de probabilidades de clasificación del cuadro de imagen en cada clase predefinida;

binarizar cada valor de probabilidad del vector de probabilidad de clase basándose en un valor umbral de probabilidad predefinido;

determinar una o más transiciones de una o más clases en una o más tramas de imagen consecutivas de la secuencia de video, basándose en los correspondientes vectores de probabilidad binarizados;

generar una secuencia de palabras comprimidas en base a una o más transiciones determinadas en uno o más cuadros de imagen consecutivos; y

extraer una o más acciones del usuario analizando la secuencia de palabras comprimidas y reconociendo la actividad humana a partir de ellas.

18. El producto programable por computador de acuerdo con la reivindicación 17, en el que una palabra comprimida se forma a partir de letras de un alfabeto que contiene un número de letras equivalente al doble del número de clases predefinidas.

19. El producto programable por computador de acuerdo con la reivindicación 17, en el que cada palabra comprimida de la secuencia de palabras comprimidas comprende al menos un cuadro de inactividad entre ellos.