

## Práctica 2: Control de reproducción de video mediante gestos con hand-pose-detection.

### GESTO 1: Mover el video adelante o atrás.

#### Implementación:

Se ha dividido la tarea en 3 partes:

- Detección de la orientación de la palma derecha a partir del vector normal a partir de landmarks3D: si el componente x del vector normal es mayor que 0.4, la palma está orientada hacia la izquierda.
- Detección de la palma derecha abierta con los dedos apuntando hacia arriba: si para cada dedo, la punta se encuentra encima de la base (`landmarks[tipIndex].y < landmarks[baseIndex].y`).
- Detección del movimiento horizontal de la muñeca derecha: se compara la distancia entre la posición actual de la muñeca en el eje X con la anterior, en caso de ser la diferencia mayor que 2, el vídeo avanzará 3 segundos, mientras que si es menor que -2, retrocederá 3 segundos.

Asimismo, se han utilizado los valores landmarks (2D) para definir la muñeca. Esto es debido a que, durante el experimento, se obtuvieron mejores precisiones del movimiento con landmarks (2D) en comparación con los valores de landmarks3D. Además, para el movimiento horizontal solo ha sido necesario utilizar el componente X.

#### Dificultades:

- Para la detección de la orientación de la palma, el umbral recomendado en la práctica y establecido en el enunciado para el componente x del vector normal es 1. Sin embargo, se ha tenido que ajustar acorde a los valores obtenidos durante el experimento, estableciendo un umbral de 0.4.

### GESTO 2: Cerrar el vídeo.

#### Implementación:

Al igual que para el gesto anterior, primero se comprueba que la mano detectada sea la derecha (`hand.handedness=='Right'`). A continuación, se verifica si la mano está

formando un puño, calculando el promedio de las distancias obtenidas para cada dedo. Dichas distancias se obtienen calculando la distancia entre la punta y la base de cada dedo. Es decir, cuanto mayor distancia exista entre la punta y la base de un dedo, significará que el dedo está más extendido y por lo tanto el promedio será mayor, y viceversa. En caso de que el promedio sea menor que un umbral establecido, en este caso 0.38, se detectará un puño.

#### Dificultades:

- La mayor dificultad para esta tarea ha sido encontrar el umbral óptimo para la detección del puño, ya que las distancias entre las puntas y bases de cada dedo varían muy poco entre sí. Es por ello por lo que se ha tenido que ir probando valores en unidades de centésima. Asimismo, la distancia varía también según el dedo. Por ejemplo, en caso de tener el meñique extendido y el resto de los dedos doblados, el promedio estaba más cerca del umbral que si hubiera sido el dedo medio.

### **GESTO 3: Subir o bajar el volumen.**

#### Implementación:

Se ha dividido la tarea en 2 partes:

- Mano izquierda:
  - Para detectar si la mano izquierda está orientada hacia la cámara, se ha calculado el vector normal (como en la tarea 1) y comprobado si el componente Z es positivo ( $>0$ ). En caso de ser positivo, la palma está orientada hacia el usuario, es decir, el dorso de la palma está orientada a la cámara (esto es lo que queremos), y viceversa para valores negativos.
  - Se ha reutilizado la función `areFingersExtendedUpwards` utilizada en la tarea 1, para detectar si los dedos están extendidos hacia arriba.
- Mano derecha:
  - Para detectar si el índice está apuntando hacia la izquierda, se ha comprobado que esté extendido, es decir que la punta del índice esté situado a la izquierda de la base de este (`indexTip.x < indexBase.x`), y que la distancia entre el índice y la muñeca en el eje X, sea mayor que 1 para asegurarnos que la mano está girada hacia la izquierda significativamente (`Math.abs(indexTip.x - wrist.x) > 1`).
  - Se ha comprobado que el resto de los dedos estén doblados (pulgar, medio, anular y meñique), comparando que la punta de cada dedo se encuentra a la

derecha de la base de este ( $\text{landmarks}[\text{tipIndex}].x > \text{landmarks}[\text{baseIndex}].x$ ). En caso del pulgar, se comprueba que la punta esté debajo de la base ( $\text{thumbTip}.y > \text{thumbBase}.y$ ).

A continuación, en caso de que ambas manos cumplan con las condiciones establecidas, se compara la posición Y de la muñeca derecha en el estado actual y del estado anterior. En caso de ser la diferencia positiva ( $>0$ ), el volumen baja 0.1, mientras que si es negativa ( $<0$ ), el volumen sube 0.1, siendo el mínimo 0 y el máximo 1.

#### Dificultades:

- Como se ha mencionado anteriormente, la detección del puño es muy sensible al tratarse de valores muy pequeños (diferencia de centésimas). Como consecuencia, durante la práctica de esta tarea, el vídeo se cerraba en caso de no sobrepasar el umbral, aun teniendo el índice extendido.
- Dependiendo de la luz de la sala, la detección de ambas manos a la vez era más complicado que detectar una sola mano en caso de ser muy brillante o muy oscura.

### **GESTO 4: Reproducir el vídeo a velocidad x2.**

#### Implementación:

Primero se ha comprobado que la mano sea la derecha, y posteriormente si esta está en forma de signo de paz, es decir, si los dedos índice y medio están extendidos hacia arriba, y el resto están doblados. Similar a la tarea 1, los dedos extendidos se detectan si la punta está encima de la base ( $\text{indexTip}.y < \text{indexBase}.y$ ). Mientras que los dedos doblados se identifican si la punta está debajo de la base ( $\text{ringTip}.y > \text{ringBase}.y$ ), excepto para el pulgar que se compara que la punta esté a la derecha de la base ( $\text{thumbTip}.x > \text{thumbBase}.x$ ). En caso de cumplirse dichas condiciones, la velocidad del vídeo aumenta a x2.

#### Dificultades:

- No se ha encontrado ninguna dificultad, aparte de las mencionadas en las tareas anteriores.

### **Reflexión sobre lo aprendido del proceso.**

Durante la práctica se ha aprendido lo siguiente:

- Diferencias entre los keypoints en 2D y 3D. El valor del eje Z ayuda a tener una mayor percepción de la imagen recibida por la cámara ya que nos permite conocer la profundidad de los objetos, si están más cerca o más lejos de la cámara. Asimismo, he descubierto que también es útil para identificar la orientación de dichos objetos, como es en esta práctica la palma de la mano.
- Importancia de una buena iluminación. Por ejemplo, para grabar el vídeo interaccionando con el sistema, he tenido que moverme a una sala con mayor iluminación.
- Mayores habilidades de programación en Javascript y CSS.

**Conexión de la práctica con otras áreas de formación o intereses personales.**

Esta práctica ha sido de gran utilidad para adentrarme al mundo de detección gestos. Lo cual será muy beneficiario para mi TFM, ya que una parte está relacionada con la detección de gestos que realiza el usuario en un entorno de realidad virtual.