Tema 5: MeanShift, CamShift y Template Matching

Motivación

Para esta práctica, se ha aprovechado el código para la detección de manos usado en la práctica con los filtros de Kalman y de partículas. Para ello, se vuelve a usar la librería "mediapipe". La idea es hacer el seguimiento de la mano por la pantalla. Esta vez, no se anota qué mano es la ganadora, ya que es complicado discernir qué mano es la izquierda y cuál la derecha, usando solamente uno de los tres métodos de seguimiento vistos. Esto es debido a que, al contrario que en el ejercicio anterior, no se va a aplicar la detección de manos en cada fotograma, sino solamente al principio, antes de pasar al seguimiento con las técnicas en cuestión.

La motivación para hacerlo de esta forma es conseguir un sistema de seguimiento de manos, de forma que luego es sencillo añadir una máscara para poner un objeto en la mano, lo cual puede tener diferentes aplicaciones, para juegos sencillos relacionados con realidad aumentada, por ejemplo.

Implementación

El seguimiento, para las tres técnicas, se divide en dos fases: en la primera, el seguimiento se hace usando la librería "mediapipe", que permite obtener la posición de las manos de la pantalla. En esta fase, se ve el esqueleto de la mano, y un recuadro marcando el fragmento de imagen que se usará como referencia para el seguimiento en la siguiente fase. Para empezar con el seguimiento, bastará pulsar la barra espaciadora, de forma que se puede elegir el trozo de imagen de referencia en las mejores condiciones posibles.

También es posible pasar como argumento al programa "--start_auto", de forma que en el primer fotograma que se detecte una mano, se seleccionará la región de imagen a usar automáticamente y se pasará directamente a hacer el seguimiento. Esta opción funciona bastante bien y es más rápida que la anterior, pero puede fallar en algunos casos, debido a la obtención de una mala captura de la mano.

Después de esta primera fase, empezará el seguimiento usando uno de los tres métodos: MeanShift, CamShift y Template Matching.

MeanShift y CamShift

En el caso de MeanShift y CamShift, el código usado es el mismo a excepción de la función que se llama, del mismo nombre que el método. En estos métodos, se calcula un histograma de colores en HSV para la región de la imagen donde se encuentra la mano.

La ventaja del CamShift sobre el MeanShift es que permite que el tamaño de la ventana se autoajuste, de forma que, si la mano se acerca a la cámara, el tamaño aumenta, y si se aleja, disminuye. Su mayor problema es que, durante el seguimiento de una mano, puede empezar a detectar también otra mano, o incluso la cara, añadiéndolo todo a la ventana usada, que se

vuelve demasiado grande, cosa que no ocurre con el MeanShift. Por tanto, el MeanShift permitiría jugar a robar la galleta al oponente, mientras que el CamShift no.

Aparte de estas diferencias, ambos métodos presentan el gran problema de no saber lidiar bien con las oclusiones: en todos los casos, si la oclusión es de toda la mano, se pierde la mano, y no se supera la oclusión (la ventana de detección se queda inmóvil mientras que la mano supera al objeto que genera la oclusión). Aun así, el seguimiento es muy bueno cuando no hay oclusión, aunque los movimientos sean rápidos y aleatorios.

Template Matching

El método de template matching simple es más sencillo que los anteriores: se usa como "template" la ventana de la imagen recogida en la primera fase, y luego se mide su similitud con las diferentes regiones de la imagen original. Como medida de similitud se han probado diferentes métricas, pero la que mejores resultados ha dado ha sido TM_SQDIFF_NORMED, que toma la diferencia de cuadrados normalizada.

Para visualizar el resultado de esta métrica, puede pulsarse la tecla "p" durante el seguimiento de la mano, para alternar entre ver los fotogramas RGB o el mapa de similitud calculado. El punto donde se detectará la mano será donde se encuentre el mínimo global.

Este método ha dado mejores resultados ante oclusiones, ya que encuentra rápidamente la mano cuando esta vuelve a estar visible. Su mayor problema, es que depende bastante de la iluminación, que puede hacer que el mínimo global se encuentre en otro sitio donde no haya ninguna mano, lo que puede llevar a que la detección se desplace a otro punto de la imagen, en algunos fotogramas.

Este método tampoco sirve cuando hay dos manos en la pantalla, ya que la detección oscilará entre una y otra dependiendo del momento.

Resultados

Los mejores resultados dependen de lo que busquemos:

- Para una sola mano, el método más robusto es el CamShift, que permite adaptar el tamaño de la ventana al tamaño de la mano en cada momento.
- Para usar más de una mano a la vez, el mejor método es el MeanShift, ya que el seguimiento se hará solamente sobre una mano, al usarse una ventana de tamaño fijo en la que cabrá solamente una. Esto permitirá poder "robar la galleta" a la otra mano.
- Si se encuentran muchas oclusiones en el vídeo, en cambio, el método de template matching da unos resultados bastante buenos, al volver a encontrar rápidamente dónde se encuentra la mano, al contrario de lo que ocurre con los otros dos métodos.

Análisis de Support Vector Tracking y Layered based Tracking

Si se utilizase el método de Support Vector Tracking, se podrían llegar a obtener resultados más estables para el seguimiento de una mano por la pantalla. En cambio, si tuviésemos dos manos, podría ocurrir que el seguimiento pasase de una mano a otra de forma

aparentemente aleatoria, dependiendo de la situación. Además, tal como se comenta en el artículo donde se publicó este método, esta técnica no funciona bien bajo condiciones de oclusión, ya sea parcial o total, aunque sea momentánea, ya que se puede perder el objeto seguido y que sea difícil volver a encontrarlo.

Por lo referente al Layered Based tracking, en cambio, sí que presenta muy buenos resultados ante oclusiones parciales o totales, al menos frente al resto de métodos vistos. Esto es gracias al cálculo y ubicación de los diferentes objetos de la escena en capas, que se ordenarán por orden de profundidad, y permitirán discernir si hay oclusiones, y qué capas las generan. Aun así, en el paper se comenta la dificultad para hacer el seguimiento de objetos articulados, con lo que es posible que se encuentren problemas de seguimiento debido al movimiento del brazo con el de la mano, ya que se encuentran relacionados.