Universidad Autónoma de Madrid

Escuela politécnica superior

Proyecto fin de carrera

Reconocimiento en tiempo real de gestos manuales a partir de vídeos capturados con cámara de profundidad

José Antonio Pajuelo Martín

Septiembre de 2010

Reconocimiento en tiempo real de gestos manuales a partir de vídeos capturados con cámara de profundidad

AUTOR: José Antonio Pajuelo Martín
TUTOR: Javier Molina Vela
PONENTE: José María Martínez Sánchez

Video Processing and Understanding Lab
Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Septiembre de 2010

Resumen

Resumen

El uso de gestos manuales ofrece una alternativa a los interfaces entre humano y máquina tradicionales, permitiendo que la comunicación tenga lugar de un modo mucho más intuitivo. Este proyecto presenta la evolución de un sistema de reconocimiento de gestos, destinado al control de aplicaciones multimedia. Partiendo de las imágenes capturadas por una cámara TOF (cámara que captura imágenes con una intensidad inversamente proporcional a la profundidad de los objetos presentes en la escena) el sistema realiza una segmentación y extracción de características que servirán para definir morfológicamente la silueta de la mano. Basándose en esas características se discrimina entre un conjunto de posibles posturas estáticas (SHPs), lo que sumado al patrón de movimiento estimado de la mano, da como resultado el reconocimiento de gestos dinámicos (DHGs). El sistema trabaja en tiempo real, lo que permite una práctica interacción entre el usuario y la aplicación.

Palabras Clave

Interfaces de usuario, Reconocimiento de objetos, Reconocimiento de patrones.

Abstract

The use of hand gestures offers an alternative to the commonly used human computer interfaces, providing a more intuitive way of navigating among menus and multimedia applications. This project presents the evolution of a system for hand gesture recognition devoted to control windows applications. Starting from the images captured by a time-of-flight camera (TOF, a camera that produces images with an intensity level inversely proportional to the depth of the objects observed) the system performs hand segmentation as well as a low-level extraction of potentially relevant features which are related to the morphological representation of the hand silhouette. Classification based on these features discriminates between a set of possible Static Hand Postures (SHPs) which results, combined with the estimated motion pattern of the hand, in the recognition of Dynamic Hand Gestures (DHGs). The whole system works in real-time, allowing practical interaction between user and application.

Key words

User interfaces, Object recognition, Pattern recognition.

Agradecimientos

Seré breve. Gracias a D. José María y a D. Jesús, y no sólo por darme la oportunidad de realizar este proyecto en el VPULab, sino por haber estado presentes a lo largo de toda la carrera para dar solución a mis dudas y ayudarme en todo lo posible.

Gracias a la gente del laboratorio, por "ofrecerse voluntarios" para pasar horas y horas delante de la dichosa camarita grabando vídeos. Al "portero" por proteger mi oreja cuando lo he necesitado, a Marcos por esos cruces matutinos de críticas, a "Alvarito" por tantos y tantos ratos a lo largo de la carrera, y cómo no, a mi tutor, porque "bajo su abrigo" he tenido la suerte, mejor dicho, la gran suerte de realizar este proyecto.

"Cocretando", gracias a todas esas personas que durante estos largos años han soportado (o no..) mi difícil carácter, porque sin ellos nada hubiese sido igual. Familia, compañeros, profesores,... Gracias a todos.

Y en especial, gracias a Lu, por haber tenido y tener, la paciencia necesaria para aguantarme y apoyarme cada día, cada hora, cada minuto (porque mira que soy pesado!). Por ser como es, y por tener siempre una sonrisa que alumbre el más oscuro de los días. Un millón de gracias.

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

1.1. Motivación.

Las nuevas tecnologías tienden al desarrollo de interfaces con alto grado de usabilidad.. La interacción entre humano y computadora *(HCI, Human-Computer Interaction)* es un campo de investigación que está en continua evolución. Las empresas tecnológicas, en particular las empresas punteras en el campo de los videojuegos están invirtiendo una parte importante de sus recursos en el desarrollo de nuevos interfaces que seduzcan al usuario con nuevas formas de comunicación con la máquina, aplicables tanto a videojuegos como al control de entornos multimedia.

A la hora de hablar de comunicación entre seres humanos los gestos manuales están a la orden del día, y muchos de ellos son entendidos por cualquiera independientemente de su cultura o nacionalidad (e.g. números o direcciones), lo que hace que esta vía de entendimiento resulte más intuitiva y efectiva que otras. Podemos decir por tanto que la utilización de gestos manuales como base de la comunicación usuario-máquina dota al sistema de una gran usabilidad. siendo la consecución de ésta un importante objetivo para cualquier aplicación de usuario.

Para la implementación de un interfaz gestual, necesitaremos cierto hardware: sensores capaces de captar la información relativa a la realización de los gestos, y el desarrollo de un sistema software: todo lo relativo a tratamiento de la información captada y diseño de aplicaciones.

Dado el caracter visual de los gestos que queremos captar, es indispensable contar con un sistema de captura de imágenes. Además, si dicho sistema es capaz de conseguir información de profundidad(e.g. visión estereoscópipca, cámara de profundidad), la colección de gestos detectables podrá ser mucho más amplia abriendo un abanico mucho más amplio de entornos de aplicación.

Una vez captadas las imágenes la información ha de ser procesada, buscando diferencias entre los distintos gestos que queden implicitamente reflejadas en la descripción que se elija para el modelado de la mano. De esta forma, entrenando un sistema convenientemente pueden ser distinguidos distintos gestos. Posteriormente, una máquina de estados o razonador servirá para definir las transiciones entre gestos detectados.

Poco a poco los avances en este campo de la ciencia, *HCI,* permitirán a cada usuario elegir la forma que le resulte más cómoda para interaccionar con las máquinas. Manejar un videojuego, elegir una película o canción desde el sofá, subir o bajar el volumen, realizar una presentación en público o el control de aplicaciones domóticas, son ejemplos que serían realizables mediante la inclusión de gestos manuales.

1.2. Objetivos.

El principal fin de este proyecto es el de contribuir al desarrollo de un interfaz gestual que, partiendo de capturas de vídeo con cámara de profundidad, discrimine entre distintos gestos manuales permitiendo así el control de aplicaciones. Más concretamente, la contribución de este trabajo, dentro del proyecto VISION[1], se enmarca en el procesamiento de lás características de bajo nível extraidas de la captura así como el estudio de su evolución, con el objetivo de hacer claramente diferenciables los gestos a detectar. Para ello se plantean distintas líneas de acción, cuya integración dará como resultado un sistema de detección robusto, todo ello en tiempo real:

- Selección, tratamiento previo y normalización de las características que describirán la mano, con el fin de separar las distintas posturas estáticas que se contemplan.

[1] http://vision.tid.es/

- Detección de patrones de movimiento presentes en los gestos. Valiéndose de las distintas técnicas de correspondencia entre curvas se discriminará entre los patrones de movimiento detectables.
- Finalmente se configurará una máquina de estados para gobernar las transiciones en gestos compuestos (i.e. gestos formados por más de una postura una postura estática), en presencia o ausencia de movimiento.

Con todo ello el sistema quedará preparado para funcionar como interfaz de cualquier aplicación que pueda beneficiarse de la comodidad de ser controlada por gestos.

1.3. Organización de la memoria.

A lo largo de este documento comenzaremos con el estudio del Estado del Arte (ver capítulo 2), para luego presentar el trabajo previo y mejoras realizadas en el capítulo 3. Se planteará el proceso para la integración de los cambios introducidos en el capítulo 4. Se evaluará el sistema presentando los resultados para una significativa colección de vídeos en el capítulo 5 para finálmente apuntar líneas de trabajo futuro fruto del análisis de los resultados en el capítulo 6.

Capítulo 2

Estado del arte

2.1. Introducción.

El estudio de la interacción entre humano y computadora (HCI, *Human Computer Interaction*) avanza en busca de una relación más intuitiva entre personas y ordenadores. Como consecuencia de esta búsqueda han surgido diferentes formas de comunicación hombre-máquina, como el control mediante voz (reconocimiento de voz), mediante el tacto (pantallas táctiles), o mediante gestos (reconocimiento gestual).

Dentro de todas estas innovaciones, el reconocimiento gestual se plantea como el más intuitivo y natural, por lo que su desarrollo está en continua evolución. Tal y como se menciona en [8] los interfaces de usuario 3D (3D UI) están ganando mucho protagonismo en el campo de las consolas de videojuegos. Compañías punteras en este mercado, como Nintendo, Sony o Microsoft, han presentado nuevos modos de interacción con sus videoconsolas: Nintendo con su consola Wii[1], Sony con su PlayStation-Eye[2] y Microsoft con su innovador proyecto Kinect[3].

El problema de diseñar un sistema capaz de reconocer gestos manuales que sirvan para el control de aplicaciones se puede plantear en las siguientes fases: Adquisición y procesamiento de datos (captura de cámara, procesado y

[1] http://wii.com
[2] http://www.us.playstation.com/News/PressReleases/396
[3] http://www.xbox.com/kinect/

segmentación), modelado de la mano (selección y tratamiento de características de la mano), definición de gestos (estáticos y dinámicos), detección de los mismos (separación de posturas y detección de movimiento) y control de transiciones (máquina de estados).

En este capítulo se presta especial atención al modelado de la mano, necesario para la separación en posturas tipo; técnicas de correspondencia entre curvas, necesarias para la detección de movimiento; diseño de máquinas de estado, necesarias para gobernar las transiciones y el funcionamiento global del detector; y finalmente, una serie de colecciones de gestos y aplicaciones controladas mediante la detección gestual.

2.2. Técnicas de captura, procesado y segmentación.

El primer paso en cualquier sistema de detección de gestos manuales es la captura y segmentación de la mano. Este problema es resuelto en algunas aplicaciones con técnicas de sustracción de fondo, como en [9], donde mediante una cámara cenital y un fondo homogéneo se simplifica el problema de segmentar la mano. Sin embargo, muchas aplicaciones requieren gestos de realización vertical, en los que el movimiento del cuerpo provoca problemas en las técnicas tradicionales de sustracción del fondo. La segmentación basada en el color, que podemos ver en [10], también presenta problemas en este contexto debido a que la cara de la persona también está visible y es del mismo color que la mano. En [11] se enumeran algunos métodos de detección de la mano, buscando una caracterización del movimiento humano e interpretando sus acciones, en un campo mucho más genérico que el que nos interesa para este proyecto. La obtención de información de profundidad ha sido una línea de acción muy frecuente para mejorar la detección de gestos. En esta línea, algunas propuestas han sido publicadas: sistemas basados en estereo-visión como en [12], donde la sustracción del fondo y una reconstrucción 3D hace posible la segmentación de la mano y la detección de siete gestos estáticos diferentes, o en [13], donde los gestos son definidos por la dirección a la que apuntan manos y cabeza y son utilizados para el control remoto de robots. En [14], donde se detectan posturas de la mano, y en [15], donde se ajusta un modelo 3D a imágenes de entrada 2D. Una alternativa bastante reciente es el uso de cámaras *Time-Of-Flight* (TOF), que devuelven

información de profundidad por pixel[6] trabajando en tiempo real ya que no requieren de un elevado coste computacional. Buscando algún ejemplo del uso de esta tecnología nos encontramos con [16] donde una cámara TOF es utilizada para llevar a cabo el seguimiento de personas en una pequeña habitación. El uso de información de profundidad da como resultado un enriquecimiento de la comunicación entre hombre y máquina mediante interfaces gestuales. Siguiendo esta línea, en [17] se hace énfasis en algunas de las ventajas que conlleva: mejora de la robustez frente a cambios de iluminación y facilita la segmentación, incluso con la cámara en movimiento. En [18] la imagen de profundidad capturada por la cámara TOF es transformada en una nube de puntos a los cuales se ajusta un modelo de mano 3D. En [1], se llevan a cabo experimentos en torno a una colección de 12 gestos estáticos. [6] estudia la posibilidad de utilizar esta tecnologia para la navegación por imágenes médicas usando gestos manuales. En [19] se propone un método diferente para obtener información de profundidad. La escena es iluminada con un patrón de colores, y es capturada por una cámara RGB normal, siendo posteriormente procesada para inferir información de profundidad.

Centrándonos en detección de gestos manuales, en [20] las posturas estáticas de un diccionario son detectadas mediante imágenes 2D en escala de grises. Otra propuesta para este problema es la realizada en [15], donde la discriminación entre posturas de la mano tiene lugar mediante la extracción de características de la mano partiendo de imágenes 2D. En [14] se presenta un sistema completo en el que aparece el componente temporal y los gestos son detectados por su patrón de movimiento. Dichos gestos son números dibujados hacia una cámara 2D. En [21] se propone un vector de características de la mano aplicado a la identificación de personas.

2.3. Modelos para la descripción de manos.

El modelado de la mano se entiende como un paso posterior a la segmentación de la mano donde se seleccionan y tratan las características extraídas para definir y separar las diferentes posturas estáticas que se podrán detectar.

En [4] la mano es modelada mediante un vector que incluye como características las coordenadas del centro de la elipse, el tamaño de la misma, y su

ángulo de inclinación.

En [1], el modelado de la mano se realiza mediante una proyección de la misma sobre los ejes x e y, teniendo en cuenta puntos significativos, como son el más a la derecha y el más alto. La figura 2.1 muestra un ejemplo donde se ve dicha proyección, además de la corrección que realizan para evitar la aparición de antebrazo, asumiendo para ello una longitud media de la mano.

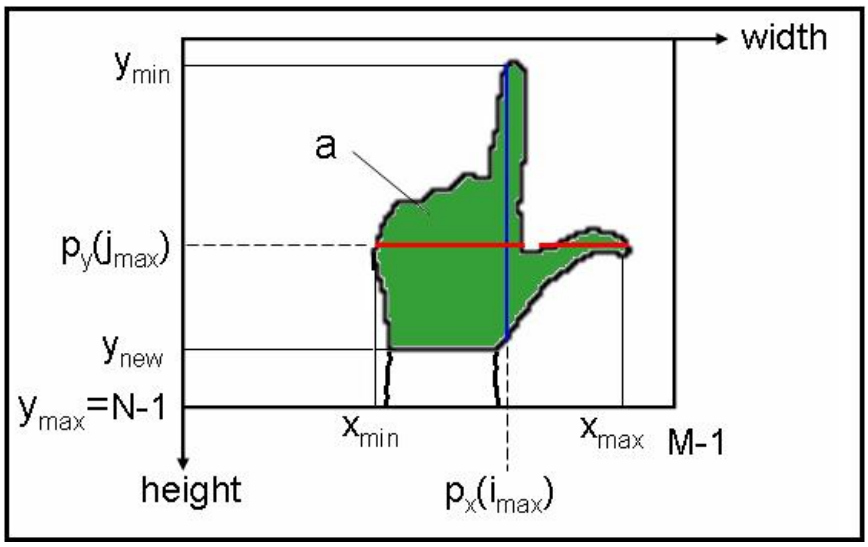

Figure 2.1: Modelado de la mano en [1].

Una forma más curiosa para describir la mano la encontramos en [2], con una solución basada en la detección de puntos característicos para su posterior unión, etiquetando los segmentos según su orientación. Esta puede verse en la figura 2.2.

Otra forma diferente de modelado, es encontrada en [22], donde el vector de características está formado por histograma de orientación local, fundamentando su elección en que la orientación local es menos sensible a cambios de luz. El uso del histograma permite detectar posturas en distintas zonas de la pantalla.

En [23] se utilizan proporciones y características geométricas de la mano para su descripción, y en [3] la silueta es transformada en una función de la distancia de la silueta a la base, como podemos ver en la figura 2.3, además de utilizar su ángulo de orientación.

En [7], se utilizan descriptores de Fourier modificados (MFD), calculados en

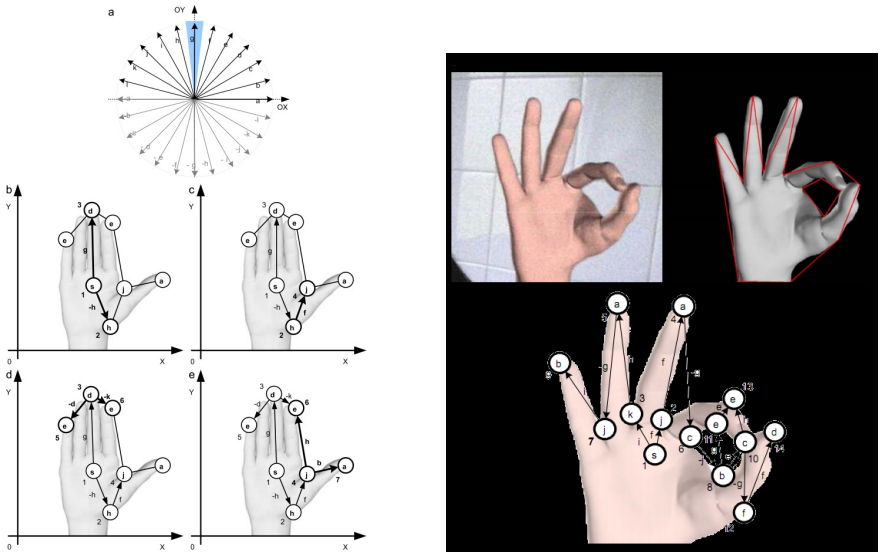

Figure 2.2: Modelado de la mano en [2]

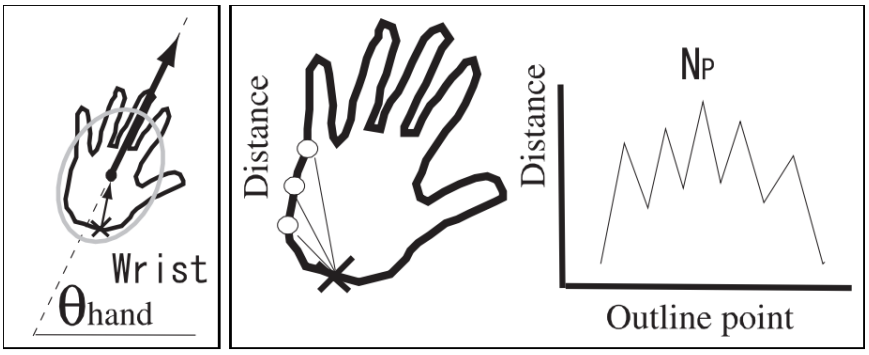

Figure 2.3: Descripción de la mano en [3].

los bordes de la mano. De los momentos de la imagen se extrae la dirección global y posición de la mano.

En [5], se utiliza un sistema de memoria visual (VMS) que almacena los distintos patrones estáticos de las posturas a reconocer, y usa la distancia Hausdorff para su separación.

2.4. Técnicas de clasificación

Una vez elegido el descriptor utilizado para la mano, pasamos a la detección de gestos. En [24] encontramos una separación de resultado binario medi-

ante el uso de Support Vector Machines (SVM). En [25] se utiliza una FSM para modelar estados basados en secuencias de posturas estáticas. Además de posturas estáticas, en [26] nos encontramos con la introducción de algo diferente, gestos dinámicos, a los que llama DHG (Dynamic hand gesture).

2.5. Detección de patrones de movimiento.

El estudio de las técnicas de correspondencia entre curvas permite decidir cuál es la mejor forma de comparar trayectorias, un paso crítico en la detección de movimiento.

La realización de gestos manuales puede ser muy variable en tiempo y espacio, esto es, la velocidad de realización no tiene por qué ser igual siempre y lo mismo sucede con la amplitud del movimiento. Por este motivo son necesarios algoritmos que sean robustos frente a estas circunstancias.

Algo similar a lo ya comentado con los gestos manuales sucede con los reconocedores de voz, donde es muy utilizado el algortimo *Dynamic Time Warping* (DTW), como puede verse en [27]. Se trata de un algoritmo para medir la similitud entre dos secuencias que pueden ser diferentes en longitud y velocidad.

Este mismo algoritmo es también utilizado en tratamiento de imágenes con distintos fines. En [28], por ejemplo, se le da uso para establecer similitudes en la textura de imágenes.

En [29] se usa este mismo algoritmo para el reconocimiento gestual, estableciendo comparativas entre hacerlo en muchas dimensiones o una sola, y también estudiando la trayectoria global o la pendiente local de la misma, esto es, la derivada.

En [30] podemos ver otro algoritmo planteado para la correspondencia entre curvas, el Knuth-Morri-Pratt (KMP), un algoritmo diseñado inicialmente para buscar la existencia de una palabra dentro de un texto. Encontramos este algoritmo modificado en [25], donde se usa para aumentar la velocidad en el reconocimiento gestual.

2.6. Máquinas de estado para el modelado de gestos.

Gobernar toda la información obtenida en las fases anteriores también es un problema que debe ser tratado con el suficiente detalle. La definición de las transiciones posibles, de los estados del sistema y sus salidas, es algo que puede ser realizado con una máquina de estados.

En [4] podemos ver un diagrama de las transiciones posibles entre los diferentes gestos dentro de su sistema (ver figura 2.4).

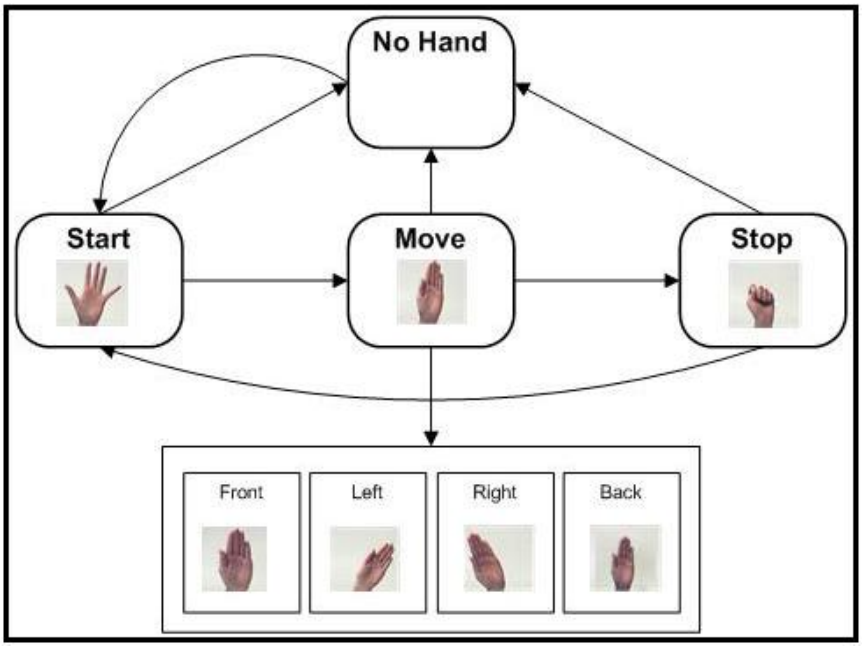

Figure 2.4: Diagrama de estados en [4]

Históricamente, los HMMs (Hidden Markov Models) han sido utilizados para el reconocimiento de voz, debido a su facilidad para reconocer patrones temporales. También tienen cabida en el reconocimiento gestual, como podemos ver en [31], donde los HMMs son utilizados para reconocer el movimiento del cuerpo humano, o en [32], usados para el reconocimiento del lenguaje de signos americano.

2.7. Colecciones de gestos y aplicaciones.

Muchas y muy diferentes son las colecciones de gestos que sirven para el control de aplicaciones, ajustándose a los requisitos de las mismas.

En [5] encontramos el alfabeto que vemos en la figura 2.5, el cual consta de 26 posturas estáticas diferentes.

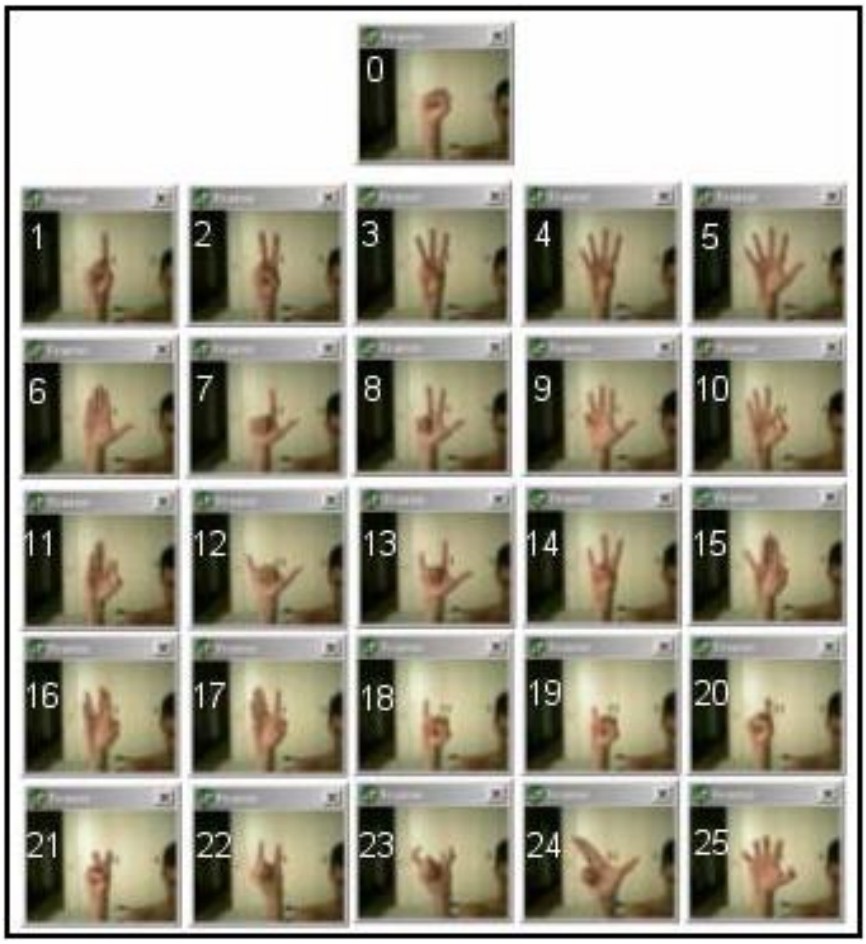

Figure 2.5: Alfabeto en [5].

Una colección de gestos más sencilla la encontramos en [4], donde se proponen los siguientes gestos (ver figura 2.6) para llevar a cabo el control de un videojuego.

El alfabeto propuesto por [1] contiene 12 gestos que se corresponden cada uno con una letra, como podemos ver en la figura 2.7, para cuyo reconocimiento,

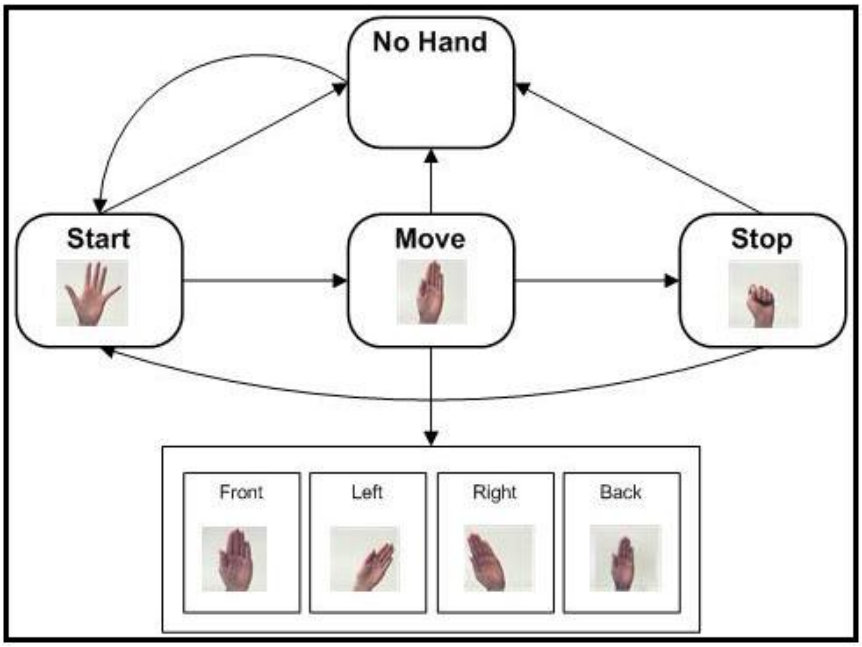

Figure 2.6: Colección de gestos en[4]

en diferencia con los anteriores, se utilizó una cámara con información de profundidad.

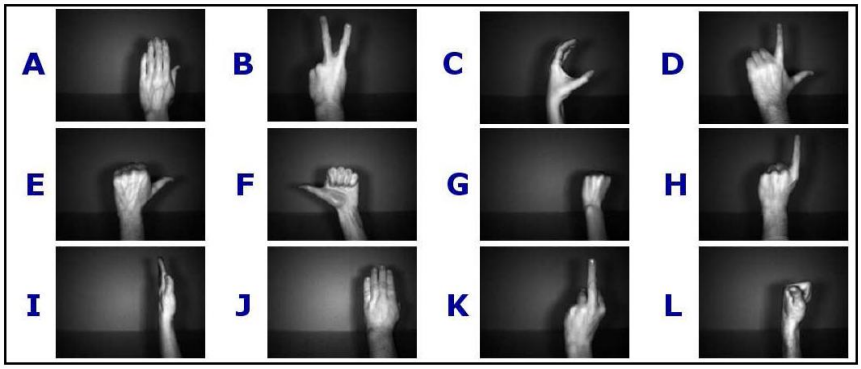

Figure 2.7: Alfabeto en [1].

En [6] nos encontramos con una colección de tan solo 5 gestos, los mostrados en la figura 2.8, cuyo fin es el de navegar en aplicaciones de imagen médica.

También en una aplicación de tratamiento de imágenes, en este caso, de vídeo[7], nos encontramos con la colección de gestos de la figura 2.9. Éstos son utilizados para el control de una aplicación de restauración de películas

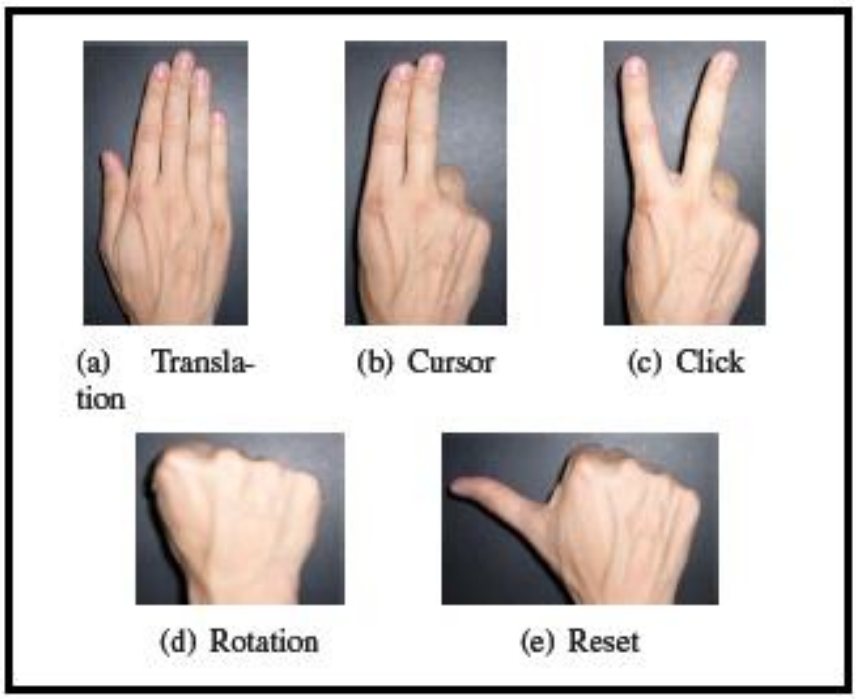

Figure 2.8: Colección en [6].

antiguas.

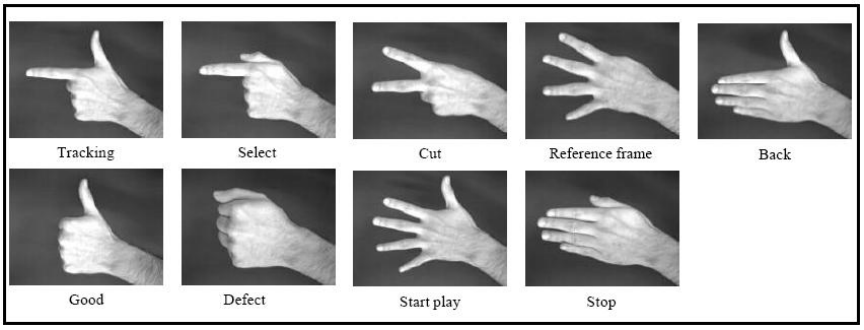

Figure 2.9: Colección de gestos en [7]

Una aplicación más común y que podria ser útil a una gran cantidad de usuarios la encontramos mencionada en [33] , donde una cámara se utiliza en el interior de un coche para controlar distintas aplicaciones como muestra la figura 2.10.

Figure 2.10: Aplicación en coche de reconocimiento gestual.

Capítulo 3

Diseño y desarrollo

3.1. Introducción. Conceptos básicos

3.1.1. Cámara de profundidad *(TOF Camera)*.

Tal y como se comentó en la redacción de Estado del Arte, el acrónimo TOF viene del inglés *Time-Of-Flight*, y es el nombre que reciben cámaras capaces de obtener información de profundidad basándose en la medición de la fase de una onda infrarroja reflejada en el objecto a caracterizar. Dicho de otra forma, el funcionamiento de esta cámara consiste en la emisión, durante un breve período de tiempo, de un pulso de luz que será reflejado por los objetos presentes en la escena. El sensor de la cámara capta la luz reflejada, que habrá experimentado un retardo dependiendo de la distancia a la que se encuentre el objeto.

La zona de interacción con la cámara se verá limitada por la duración del pulso de luz infrarroja. La distancia máxima se puede calcular mediante la siguiente fórmula: $D_{max} = \frac{c \cdot t_0}{2}$,donde c es la velocidad de la luz, y t_0, la anchura del pulso. Esta fórmula tiene su origen en el tiempo que necesita la luz para recorrer una distancia, en este caso de ida y vuelta entre cámara y objeto.

La información de profundidad obtenida será de gran utilidad, no sólo como información adicional por pixel, sino además para establecer la activación o desactivación de los gestos, entendiendo por activación un acercamiento a la

cámara a una distancia inferior a D_{max} y por desactivación el correspondiente alejamiento.

En la realización de este proyecto se han utilizado dos cámaras. La primera desarrollada por 3DV Systems[1]. La otra cámara, desarrollada por Mesa-Imaging[2], mejora la calidad y la cantidad de información obtenida. Además de tener un frame-rate superior, es capaz de capturar imágenes a una mayor profundidad. Ambas cámaras se muestran en la figura 3.1.

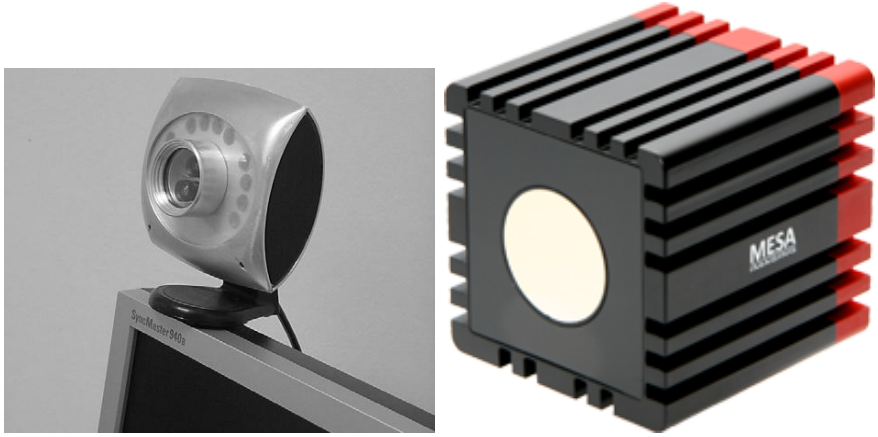

Figure 3.1: Cámaras de profundidad.

En el caso de la SR4000, la precisión de captura en profundidad es de ±1cm, siendo su rango de trabajo de 0,3m a 5m de distancia. En nuestro sistema, la cámara está configurada para trabajar en un rango de 3m (de 0,3m a 3,3m), con el objetivo de eliminar objetos del fondo. Además, tiene un frame-rate alrededor de los 30 fps.

Por su parte, la cámara 3DV presenta una precisión en profundidad de 1-2 cm y tiene un rango de trabajo inferior a la cámara anterior, el cuál va desde 0,5m a 2,5m. Su frame-rate es de 20 fps.

3.1.2. Análisis estático y dinámico: una diferenciación necesaria.

El reconocimiento de gestos manuales que se plantea puede dividirse en dos fases de detección claramente diferenciadas. Por un lado la detección de la

[1] http://www.3dvsystems.com/technology/tech.html
[2] http://www.mesa-imaging.ch/

postura de la mano y por el otro la del movimiento de la misma. Ha llegado el momento de introducir dos conceptos importantes que se repetirán a lo largo de esta memoria: *Static Hand Posture*(SHP) y *Dynamic Hand Gesture* (DHG).

- SHP, es un nivel intermedio para conseguir la detección de un DHG. Será el resultado de estudiar la postura estática de la mano en cada frame, derivada de las características instantáneas de la misma.
- DHG, se corresponde con la salida final del sistema como resultado de combinar la información de movimiento con la secuencia de SHPs detectados a lo largo de una ventana temporal.

Comentar, que no todos los gestos han de presentar un patrón concreto de movimiento, o una postura estática determinada. Habrá DHGs en los que solo tenga importancia la información estática, es decir, el SHP obtenido, siendo irrelevante el movimiento realizado. Asi como se darán DHGs que se vean definidos por patrones concretos de movimiento, sin tener en cuenta la postura de la mano.

3.2. Trabajo Previo.

Para el desarollo de este trabajo se parte de un sistema inicial del que cabe destacar algunos conceptos importantes para la comprensión de lo aquí expuesto.

3.2.1. Sistema inicial y limitaciones.

A continuación se trata con mayor detalle el estado inicial del sistema, previo al comienzo del presente proyecto. La figura 3.2 muestra el diagrama de bloques del sistema, dividido en cuatro etapas. La primera de ellas (etapa i) consiste en la captura de la imagen de profundidad. La imagen es procesada y umbralizada dando como resultado la segmentación de la mano. Esta etapa es necesaria para que se activen las etapas posteriores, y para ello la mano debe estar lo suficientemente cercana a la cámara. Si se detecta la mano, esta pasa a ser modelada mediante un vector de características (etapa ii),

permitiendo con ello la detección de SHPs. Usando esa detección y el análisis del movimiento de la mano se produce la detección del gesto, DHG (etapa iii). La predicción de DHG y las coordenadas 3D de la mano son enviadas a la GUI, donde controlan una serie de sencillas aplicaciones(etapa iv).

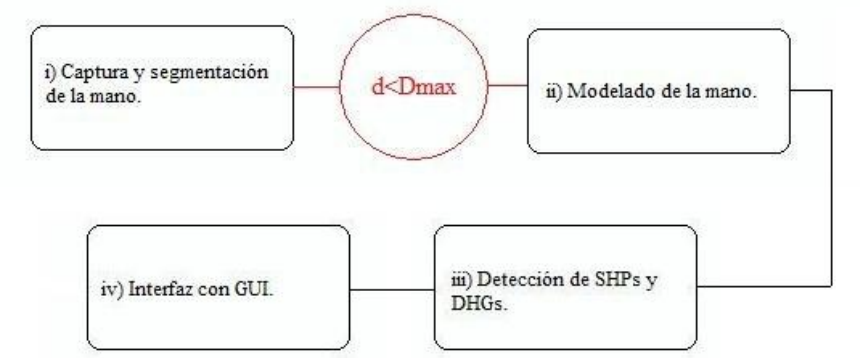

Figure 3.2: Diagrama de bloques del sistema inicial.

El sistema introducido presenta una serie de limitaciones cuya mejora será el principal objetivo del proyecto descrito en este documento:

- Dificultad en la inclusión de nuevos DHGs.
- Dificultad en la inclusión de nuevos patrones de movimiento.
- Baja tasa de acierto en la detección de los DHGs, especialmente los definidos por un patrón de movimiento.
- Evaluación a mejorar en en terminos de relevancia estadística.

3.2.2. Modelado de la mano.

Para poder diferenciar entre los distintos SHPs se antoja necesaria una descripción generalista de la mano. La aproximación tomada en este sentido responde a un modelado de la mano como una elipse con protuberancias caracterizadas con las coordenadas de puntos característicos de las mismas, así como algunos puntos más que describen la mano tal y como se aprecia en las figuras 3.3, 3.4 y 3.5. Concretando, hablamos de las siguientes características:

- Centro de gravedad de la elipse (CoGE).
- Dimensión de sus ejes y ángulo de inclinación.
- Centro de gravedad de la mano (CoG).
- Punto más cercano a la cámara (Zmin).
- Número de protuberancias, y por cada una de ellas su intensidad (longitud), amplitud (anchura), ángulo de inclinación, extremo de la protuberancia y puntos de la base.

Tener en cuenta que todos los puntos significativos (CoGE, CoG, Zmin...) estarán compuestos por tres coordenadas, excepto los puntos de la base de las protuberancias, de los cuales sólo se obtienen dos coordenadas, x e y. En total, disponemos de 63 puntos de información por cada frame capturado. No serán utilizadas todas las características recibidas. Dado que el objetivo es un sistema de detección en tiempo real, premiará la discriminación entre gestos empleando el menor número posible de coordenadas.

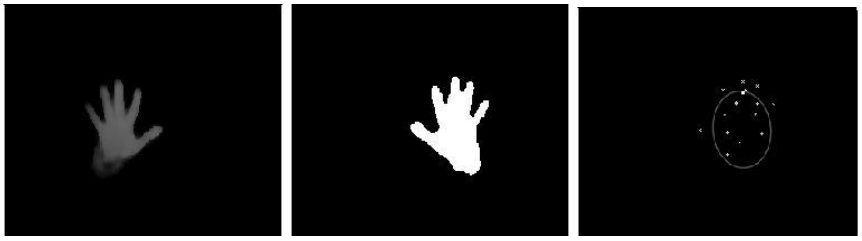

Figure 3.3: Elipse + 5 protuberancias.

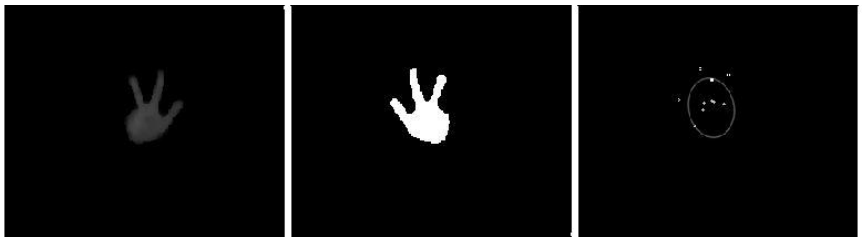

Figure 3.4: Elipse + 3protuberancias.

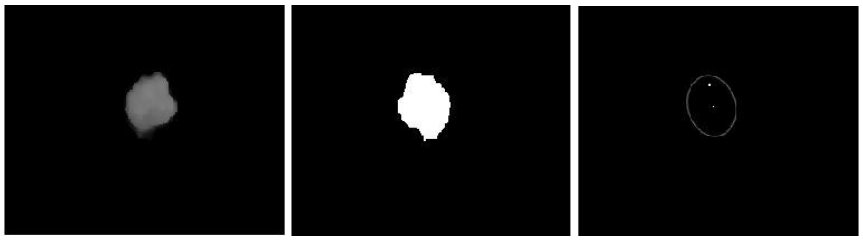

Figure 3.5: Elipse sin protuberancias.

Las características obtenidas tras la segmentación de la mano no variarán a lo largo del proyecto. Se trabajará realizando una selección y tratamiento previo de las mismas. El objetivo es poder separar las distintas posturas estáticas utilizando el menor número de coordenadas posible, buscando optimizar el tiempo necesario para su análisis.

3.2.3. Separación en posturas tipo.

El primer paso para desarrollar un sistema de reconocimiento de gestos manuales es la definición de las posturas tipo a detectar. Dicho de otra manera, concretar el diccionario de SHPs que permitirá al usuario comunicarse con las aplicaciones a controlar.

La elección de los SHPs tiene que responder a dos criterios:

- Usabilidad: el usuario ha de entender de forma intuitiva y natural el significado de las posturas manuales.
- Detectabilidad: un SHP ha de ser discriminable con un mínimo de tasa de acierto para hacer del mismo un gesto detectable.

Bajo estos dos criterios se encuentra la parte más compleja del desarrollo del sistema: alcanzar un equilibrio entre usabilidad y detectabilidad que satisfaga al usuario final. No todas las posturas son detectables, y son muchas las intuitivas que resultan de imposible discriminación.

3.2.3.1. Diccionario de SHPs.

En [34] se lleva a cabo un experimento realizado con usuarios reales con el objetivo de asociar a diferentes acciones el gesto más frecuente realizado por

los mismos. Pero la selección de diccionario no sólo debe resultar cómoda e intuitiva para el usuario, además hay que tener en cuenta las características de las que se dispone y las posibilidades que éstas pueden ofrecer para la discriminación.

Finalmente, la colección elegida en el desarrollo de este trabajo previo es mostrada en la figura 3.6, donde pueden verse las capturas en profundidad de los distintos SHPs, acompañados por el nombre con que se denotarán de ahora en adelante.

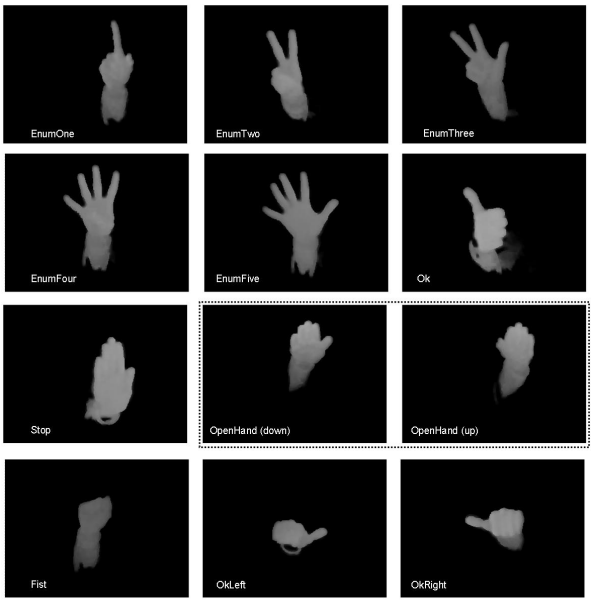

Figure 3.6: Diccionario de SHPs

3.2.3.2. Discriminación entre SHPs.

El modelado de la mano se traduce en un vector de características con un total de 63 coordenadas por imagen, de los cuales se seleccionan los más significativos o discriminantes con el objetivo de tener un vector de menor tamaño para que su análisis sea menos costoso en términos temporales pero

sin perder calidad en la detección. Recuérdese la restricción de tiempo real en el sistema de detección.

De toda la información obtenida en la captura, el vector de características seleccionado y utilizado para la discriminación entre los distintos SHPs esta formado por 19 coordenadas: Número de máximos o protuberancias (Nmax); Intensidad(In), amplitud(An) e inclinación(Angn) de cada una de ellas; Dimensiones de la elipse(DimMa y DimMe) y ángulo de inclinación(AngE). En caso de ausencia de alguno de los máximos el relleno de los parámetros derivados de él se realiza dándole valor '-1'.

Característica\SHP	Fist	EnumOne	EnumThree	EnumFive
Nmax	0	1	3	5
I1	-1	49	23	33
A1	-1	11.48	10.08	9.10
Ang1	-1	159.11	208.30	218.33
I2	-1	-1	43	26
A2	-1	-1	9.59	12.00
Ang2	-1	-1	174.29	192.23
I3	-1	-1	38	28
A3	-1	-1	11.52	11.24
Ang3	-1	-1	103.45	178.15
I4	-1	-1	-1	34
A4	-1	-1	-1	11.51
Ang4	-1	-1	-1	156.93
I5	-1	-1	-1	41
A5	-1	-1	-1	16.31
Ang5	-1	-1	-1	102.53
DimMa	30	47	35	44
DimMe	20	20	27	33
AngE	142.16	162.11	160.00	160.10

Table 3.1: Ejemplos de vectores de caracteristicas del trabajo previo.

Una vez definido el vector de características , el primer paso consiste en la normalización de cada coordenada para que quede con media 0 y varianza 1. La separación de SHPs se realiza mediante el entrenamiento de una SVM [35] por cada SHP. Para ello se utilizan vídeos grabados a 6 usuarios que realizaron durante 10 segundos cada SHP. Teniendo en cuenta que la cámara con que se grabaron (3DV) captura 20 fps, tendremos un total de 1200 frames por cada SHP, que serán utilizados como muestras positivas para el

entrenamiento de su SVM, tomando como negativas las 13200 muestras de los otros SHPs.

La función kernel RBF (*Radial Basis Function*) utilizada para la transformación de los datos en la SVM fue,

$$K(x_i, x_j) = C \cdot e^{-\gamma \|v^i - v^j\|^2}$$

, donde C y γ son los parámetros del kernel, y v^iy v^json dos vectores de características. Para conseguir la configuración óptima del kernel se utilizaron los siguientes rangos de valores: $C = 2^{-4,-3,\ldots,5}$combinado con $\gamma = 2^{-15,-13,\ldots,3}$. Finalmente se definió una función de validación cruzada para optimizar la F-score [36].

$F_\beta = \frac{(1+\beta^2)\cdot(precision \cdot recall)}{\beta^2 \cdot precision + recall}$, donde $precision = \frac{tp}{tp+fp}$ y $recall = \frac{tp}{tp+fn}$.

El valor de β era fijado a 0.5, dando con ello más peso a la precisión que al recall. Experimentalmente este valor mostraba unos buenos resultados de positivos detectados sin deteriorar en exceso la precisión.

3.2.3.3. Resultados de la estimación de SHPs.

La tabla 3.2 muestra los valores conseguidos de verdaderos positivos(tp), falsos positivos(fp), verdaderos negativos(tn) y falsos negativos(fn), asi como la evaluación de la F-score para β igual a 0.5 que fué optimizada para cada SHP. Se observa como los resultados obtenidos a nivel de frame no son lo suficientemente buenos para algunos SHPs, lo que motivó como veremos más adelante el uso del contexto temporal de cada frame (Vease 3.2.4.1).

Cabe destacar la particularidad que presentan los SHPs "OpenHandUp" y "OpenHandDown", difícilmente diferenciables con las características extraídas. Fueron fusionados y tratados como un solo SHP, "OpenHand".

SHP	id	$F_{0.5}$	tp	fp	tn	fn
EnumOne	1	0.82	933	194	13006	267
EnumTwo	2	0.97	1141	23	13177	59
EnumThree	3	0.96	1160	55	13145	40
EnumFour	4	0.95	1070	40	13160	130
EnumFive	5	0.97	1060	6	13194	140
Ok	6	0.63	759	413	12787	441
Stop	7	0.93	995	40	13160	205
Fist	8	0.52	640	648	12552	560
OpenHand	9	0.60	1281	831	11169	1119
OkLeft	10	0.87	964	116	13084	236
OkRight	11	0.84	860	110	13090	340

Table 3.2: Precisión en la estimación intraframe de SHPs.

3.2.4. Detección de DHGs.

Combinando las detecciones de SHPs con la información de movimiento se puede obtener un diccionario de gestos más completo y enriquecido. Este sería el formado por la colección de DHGs que se pretenden detectar. Además, para los DHGs consistentes en activación-SHP-desactivación con la ejecución de un solo SHP, supondrá la consecución de un sistema más robusto, tal y como se explicará en el apartado correspondiente a la ventana temporal (3.2.4.1) dentro de esta misma sección.

Notar llegado este punto que la salida de las SVMs es binaria, obteniéndose '0' como resultado negativo y '1' como resultado positivo por cada SHP, pudiendo haber para un mismo frame más de un SHP detectado como positivo. La evaluación de las salidas de las SVMs para una ventana temporal, así como la detección de un patrón de movimiento, revierte en la detección de un DHG.

3.2.4.1. Ventana temporal.

Las dos principales limitaciones que presenta el reconocimiento intra-frame de SHPs son:

1) La capacidad de discriminación en base a las características de bajo nivel en consideración puede ser no aceptable para la total separación entre los SHPs del diccionario.

2) La gran variedad de posibles usuarios y escenarios hace que, ni la utilizada, ni cualquier otra colección de datos sea lo suficientemente representativa para modelarlos.

Estas dos limitaciones podrían llegar a afectar gravemente a las capacidades del sistema. Por este motivo los SHPs fueron seleccionados teniendo en cuenta no sólo la usabilidad, sino también la separabilidad de los mismos.

Retomemos algo ya introducido con aterioridad, la dificultad de separación entre posturas usando un único frame. Cuando un usuario realiza un SHP, es razonable esperar que lo mantenga durante varios frames. Con el objetivo de sacar ventaja de esta redundancia temporal, y basándose en las estadísticas obtenidas del entrenamiento de los SVMs se decide dar distinta fiabilidad a las detecciones dependiendo del SHP del que se trate, más concretamente, en base a la relación de tp, fp, tn y fn obtenidos en entrenamiento. A continuación se plantea la aproximación seguida:

Sea un vector de características v^j(la j nos está diciendo que SHP describe). Siendo $i \neq j$, la probabilidad de que la máquina entrenada para el SHP-i acierte en su predicción negativa es:

$$p^i(0/pred = 0) = \frac{tn^i}{tn^i + fn^i}$$

,y la probabilidad de que con una predicción positiva la máquina se equivoque es:

$$p^i(0/pred = 1) = \frac{fp^i}{tp^i + fp^i}$$

Los valores de estas probabilidades quedan reflejados en la tabla 3.3.

SHP/id	$p^i(0/pred = 0)$	$p^i(0/pred = 1)$
EnumOne/1	0,980	0,172
EnumTwo/2	0,996	0,020
EnumThree/3	0,997	0,045
EnumFour/4	0,990	0,036
EnumFive/5	0,989	0,006
Ok/6	0,967	0,352
Stop/7	0,985	0,039
Fist/8	0,957	0,503
OpenHand/9	0,909	0,393
OkLeft/10	0,982	0,107
OkRight/11	0,975	0,113

Table 3.3: Probabilidades de la correcta detección de un negativo y la incorrecta detección de un positivo.

Puntualizar que la fiabilidad de las predicciones para una SVM entrenada con los patrones del SHP-i será mejor para valores bajos de $p^i(0/pred = 1)$ y altos de $p^i(0/pred = 0)$. Dada la diferencia de los valores de las probabilidades entre distintos SHPs, parece lógico tratar de manera diferente las predicciones según el SHP del que se trate.

La predicción del SHP-i para el frame n se modela como la función $pred^i(n)$. Los posibles valores para esta función son '0' y '1' (negativos y positivos).

Empezando por considerar la ventana temporal definida por:

$$\Delta T_{n_0} \equiv \{n : n - n_0 < N\}$$

La probabilidad de que no haya positivos durante ΔT_{n_0}es:

$$p^i_{\Delta T_{n_0}}(\#pos = 0) = p^i_{\Delta T_{n_0}}(\#neg = |\Delta T_{n_0}|) = \prod_{\nabla n \epsilon \Delta T_{n_0}/pred^i(n)=1} p^i(0/pred = 1) * \prod_{\nabla n \epsilon \Delta T_{n_0}/pred^i(n)=0} p^i(0/pred = 0)$$

La expresión anterior corresponde a un producto en el cual hay dos grupos diferenciados de factores: por un lado, las probabilidades de fallar prediciendo un positivo, y por otro, las probabilidades de acertar prediciendo negativo. En consecuencia, el producto global es la probabilidad de que no haya tenido lugar la ejecución del SHP-i en la ventana ΔT_{n_0}.

Por tanto, la probabilidad de que sí se haya dado en uno o más frames es:

$$p^i_{\Delta T_{n_0}}(\#pos \geq 1) = 1 - p^i_{\Delta T_{n_0}}(\#pos = 0)$$

En conclusión, partiendo de las predicciones binarias de SHP-i en una ventana temporal, se puede estimar la probabilidad de incidencia de uno o más frames en los que tenga lugar la realización del SHP-i. Calculando las probabilidades para cada SHP y comparándolas podemos estimar aquel que tiene mayor probabilidad de haber sido realizado durante la ventana temporal.

3.2.4.2. Colección de gestos.

La mayoría de los DHGs pueden ser entendidos como una secuencia de SHPs donde la mano no se mueve o no sigue ningún patron de movimiento establecido. Su reconocimiento vendrá dado por el análisis de las salidas de las SVMs dentro de la ventana temporal. En la figura 3.7 se muestran algunos de los DHGs que pertenecen a este grupo, los gestos estáticos, y la tabla 3.4 muestra el diccionario de estos DHGs simples, que permitirían el control de numerosas aplicaciones complejas de un modo simple e intuitivo.

DHG	SHP Id	Patrón de movimiento
Select	1	Total o parcialmente estático
EnumTwo	2	Total o parcialmente estático
EnumThree	3	Total o parcialmente estático
EnumFour	4	Total o parcialmente estático
EnumFive	5	Total o parcialmente estático
Accept	6	Total o parcialmente estático
Cancel	7	Total o parcialmente estático
MenuRight	10	Total o parcialmente estático
MenuLeft	11	Total o parcialmente estático
Fist	12	Total o parcialmente estático

Table 3.4: DHGs simples sin patrón de movimiento específico.

Además se proponen otros dos DHGs simples listados en la tabla 3.5, cuya evolución se puede ver en la figura 3.8. Estos DHGs, cuya misión podría ser, por ejemplo, la de abrir o cerrar un menú, contenían información de SHP, la correspondiente a una mano extendida, además de seguir un patrón específico de movimiento, elevación de la mano (move up) en el caso de 'MenuOpen', y el movimiento análogo descendente de la misma (move down) para 'MenuClose'.

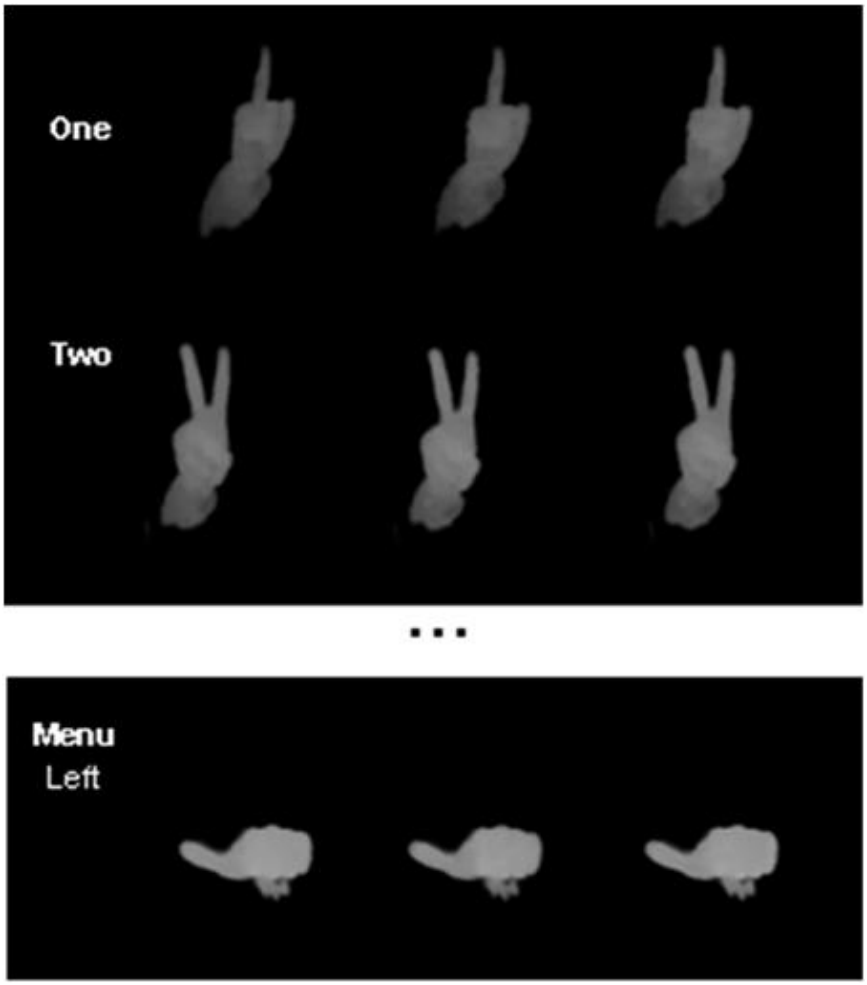

Figure 3.7: Tres ejemplos de DHGs cuya detección se basa en el reconocimiento de SHPs.

Como respuesta al requisito de coger, arrastrar y soltar objetos en una aplicación, se definieron dos gestos compuestos: 'Catch' y 'Release', ambos resultado de una combinación de dos DHGs simples. En primer lugar el usuario mueve un 'EnumFive' a lo largo y ancho de la pantalla, hasta llegar al objeto que desea coger, cerrando su mano sobre él, realizando el SHP 'Fist'. En ese momento el objeto es seleccionado y el gesto 'Catch' detectado. Sin abrir la mano el usuario puede realizar un desplazamiento del objeto y soltarlo cuando desee, esto es, volver a abrir la mano volviendo a la posición de 'EnumFive' y detectándose con ello 'Release'. El proceso completo resulta un gesto natural e intuitivo, 'Take', que puede verse ilustrado en la figura

DHG	SHP Id	Patrón de movimiento
MenuOpen	9	move up
MenuClose	9	move down

Table 3.5: DHGs simples con patrones de movimiento específicos.

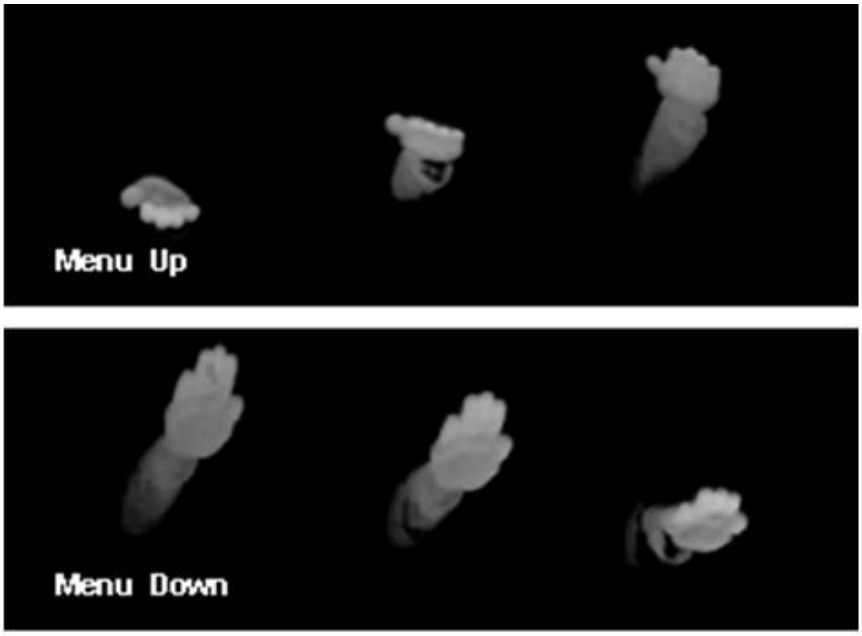

Figure 3.8: DHGs simples cuya detección esta basada en el reconocimiento de SHPs y un patrón de movimiento

3.9 y descrito en la tabla 3.6.

DHG compuesto	DHGs componentes	Patron de movimiento
Catch	EnumFive-Fist	Cualquiera
Release	Fist-EnumFive	Cualquiera
Take	Catch-Release	Cualquiera

Table 3.6: DHGs compuestos.

3.2.4.3. Estudio del movimiento.

Dentro del diccionario de DHGs nos encontramos con dos gestos, 'MenuOpen' y 'MenuClose', que tienen en común el SHP a detectar, 'OpenHand'. La diferencia entre ellos reside en el patrón de movimiento seguido, concretamente el desplazamiento ascendente o descendente sobre el eje vertical

Figure 3.9: Evolución de SHPs en la realización del DHG Take.

('move up' o 'move down').

Se estudió la evolución de la coordenada 'y' de tres de los puntos característicos extraídos de la mano: el centro de gravedad de la mano (CoG), el centro de la elipse (CoGE), y el punto más cercano a la cámara (Zmin). Las figuras 3.10 y 3.11 muestran la evolución de dichas coordenadas para 5 realizaciones distintas de los gestos dinámicos 'MenuOpen' y 'MenuClose' respectivamente.

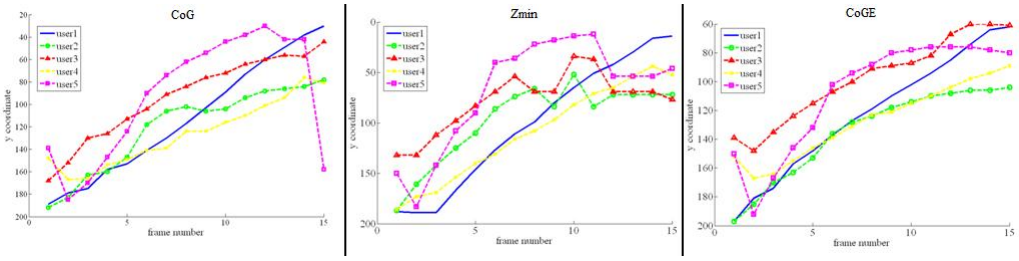

Figure 3.10: Evolución de la coordenada Y en los primeros 15 frames para 5 realizaciones distintas del DHG 'MenuOpen'.

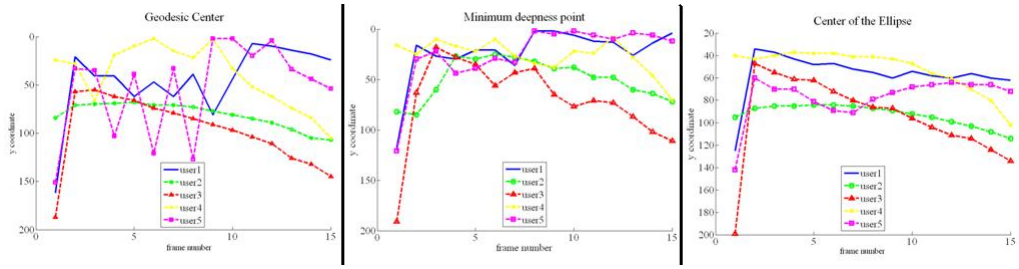

Figure 3.11: Evolución de la coordenada Y en los primeros 15 frames para 5 realizaciones distintas del DHG 'MenuClose'.

Analizando las gráficas se puede concluir que las coordenadas verticales de los puntos estudiados siguen el patrón esperado: suave movimiento descendente para 'MenuClose' y movimiento ascendente más acentuado para 'MenuOpen'. Realizando una comparación entre los distintos puntos significativos, el CoGE parece gozar de mayor estabilidad que el CoG o el Zmin, motivo para su selección como punto para la estimación del movimiento y para añadir información correspondiente a la posición de la mano en cada frame.

Tras extraer los valores de la coordenada 'y' en cada ventana de análisis, estos se filtran con un filtro de mediana, caracterizado por ser menos sensible que otros a valores extremos, buscando con ello eliminar posibles saltos bruscos. Una vez realizado el filtrado, se analiza la pendiente de las curvas, tanto global, como local.

Si la pendiente local no tiene el mismo signo durante los frames que dura la ventana temporal, y la variación en la magnitud de la misma es pequeña, el movimiento es considerado estático.

Por otro lado, si la pendiente local es negativa en todos los frames de la ventana, el patrón de movimiento es definido como descendente.

Y por el contrario, si la pendiente local mantiene un valor positivo a lo largo de los frames que dura la ventana temporal, o si la pendiente global supera un determinado umbral (100 pixels), el patrón de movimiento asignado era ascendente.

En cualquier caso no definido anteriormente, el patron de movimiento es considerado irrelevante.

Resumiendo, se podían detectar cuatro patrones distintos de movimiento: estático, ascendente, descendente o irrelevante.

La estimación del movimiento junto con la salida del análisis en la ventana temporal son las entradas de una máquina de estados que gobierna la salida del sistema, esto es, la detección de DHGs.

3.2.4.4. Máquina de estados.

La máquina de estados es la encargada de tomar decisiones que llevan a la salida final del sistema, esto es, el DHG detectado(ver apartado 3.2.4.2).

Para ello utiliza como entradas la detección de SHPs y la estimación del movimiento, ya comentadas en apartados anteriores.

La máquina de estados, FSM (Finite-State Machine), desarrollada para este sistema, combina ambas entradas, introduce información de activación y evita que se produzcan transiciones prohibidas, es decir, todas aquellas que difieren de lo definido en el apartado 3.2.4.2.

Esta FSM actúa como supervisor del sistema y sus funciones específicas son:

- Controlar que solo un DHG sea detectado en cada etapa de activación.
- Aplicar restricciones de acuerdo con el patrón de movimiento estimado.
- Modelar las transiciones en la ejecución de los DHGs 'Catch' y 'Release'.
- Desestimar aquellos DHGs que no se correspondan con las definiciones de gestos contempladas.

De aquí en adelante es importante tener claro que excepto para los casos de 'Catch' y 'Release', uno y sólo uno de los posibles DHGs es devuelto por el sistema como salida, inmediatamente después de su detección y en cada etapa de activación.

Ningún gesto puede ser detectado mientras el sistema esté desactivado, esto es, mientras la mano esté fuera de la zona de interacción. Esto lo controla la máquina de estados manteniéndose en estado de desactivación hasta que recibe la señal de activación por parte del módulo de captura y adquisición. Si el estado de desactivación es alcanzado durante el análisis de datos entrantes sin haber obtenido un DHG resultante, el sistema da como salida 'Unknown'.

Las restricciones derivadas de los patrones de movimiento, comienzan por la imposibilidad de detectar cualquier DHG si el movimiento que recibe la FSM es irrelevante, independientemente del SHP detectado mediante el análisis de la ventana temporal. Esta restricción encuentra su motivación en evitar que haya salida cuando movimientos continuos de la mano hacen que el SHP cambie de un frame a otro. Se hizo una excepción con los DHGs 'Catch' y 'Release', por razones obvias, ya que es intrínseco de ambos el poder conllevar cualquier tipo de movimiento.

Cuando el patrón de movimiento es estático, la salida del sistema puede ser cualquiera de los DHGs descritos en la tabla 3.4, a excepción del 'Fist', ya que se decidió no detectar este gesto por si mismo, sino como parte de los DHGs 'Catch' y 'Release'.

Para la detección de los gestos dependientes de la información de movimiento, la máquina de estados umbraliza la detección del SHP 'OpenHand', colocando dicho umbral en el 90% de la probabilidad más alta dentro de la ventana temporal. Superado este umbral, el movimiento actúa como decisor siendo posibles los DHGs 'MenuOpen', 'MenuClose' o 'Cancel'. Éste último puede confundirse con facilidad con algunos frames de la realización de 'MenuOpen' y 'MenuClose'(ver figura 3.8), y es por tanto la salida del sistema cuando se supera el umbral anterior y el movimiento estimado es estático.

La detección de los DHGs compuestos 'Catch' y 'Release' es también tarea de la FSM. La complejidad de estos gestos radica en la necesidad de mantener a la espera gestos que por sí solos tienen significado. Nótese también que el patrón de movimiento estimado no tenía relevancia para estos gestos, ya que como se dijo anteriormente pueden moverse libremente sin necesidad de hacerlo siguiendo un patrón fijo.

Las transiciones de la FSM para una completa ejecución del gesto 'Take', se muestran en la figura 3.12, en la cual se puede observar el problema ya mencionado. 'EnumFive' es un DHG por sí solo, recordemos la tabla 3.4, pero también es una parte de los DHGs 'Catch' y 'Release', y en consecuencia, cuando es detectado, el sistema retrasa su salida a la espera de la detección o no del DHG 'Fist'. Si se detecta dicho gesto, inmediatamente se daba salida como gesto detectado a 'Catch'. El sistema devuelve 'EnumFive' si se dan las siguientes situaciones:

1) Se produce una desactivación antes de detectar 'Fist'.

2) Se detecta un SHP distinto de 'Fist' o 'EnumFive'.

3) Se detecta repetidamente 'EnumFive' hasta llegar a superar un tiempo de espera definido para ello.

En relación con el punto (3), se define un tiempo de espera máximo con el objetivo de no forzar al usuario a realizar una desactivación para reconocer

el gesto 'EnumFive'. Dicho tiempo es definido como el equivalente a tres ventanas de análisis sin solapamiento, tiempo que resulta adecuado según los usuarios preguntados al grabar los videos.

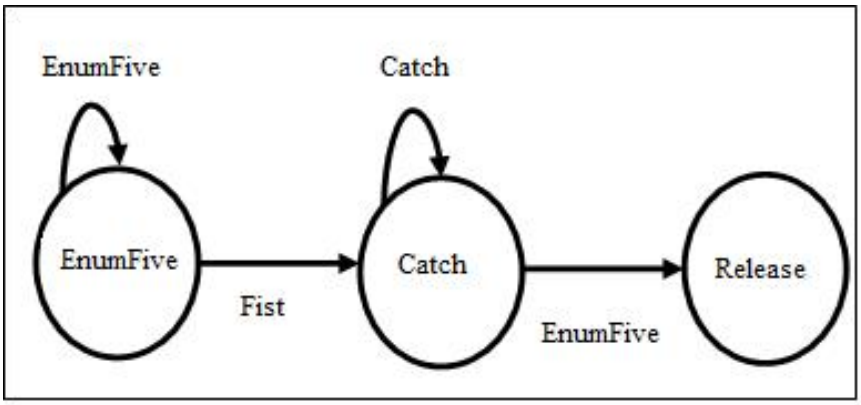

Figure 3.12: Transiciones de la FSM en la ejecución de 'Catch' y 'Release'.

Cuando el sistema detecta el gesto 'Catch', el usuario puede mover libremente el objeto cogido por la pantalla, manteniéndose la FSM en este estado hasta que el usuario vuelve a realizar 'EnumFive', llevando con ello al estado 'Release' y esperando a la desactivación. Si la desactivación se produce antes de llegar a detectar el segundo 'EnumFive', sólo se detecta el DHG 'Catch', y su interpretación ha de ser tratada en niveles superiores.

Finalmente, la configuración de la máquina de estados fuerza al sistema a devolver 'Unknown' cuando el usuario realiza gestos prohibidos, como 'Fist', que no tiene significado por sí mismo.

3.2.4.5. Datos para evaluación.

Para evaluar el resultado de todo lo detallado anteriormente se utilizan videos grabados por 6 usuarios, en los que cada usuario grababa un vídeo con cada uno de los 11 DHGs objeto de detección. Esto es, un total de 66 vídeos, algo que se queda corto si se quiere hacer una evaluación exhaustiva y relevante estadísticamente. Esto llevo a incluir la evaluación como una de las líneas de mejora en este proyecto.

3.3. Cambios introducidos.

Habiendo ya descrito el sistema inicial y puntualizado sus limitaciones, ahora pasamos a describir las mejoras introducidas por el presente proyecto para la consecución de un sistema más acertado en la deteccion y más flexible a la hora de introducir nuevos DHGs. Clasificaremos las mejoras propuestas en tres subsecciones:

- Separación en posturas tipo. Cambios en la normalización, nueva selección y tratamiento previo de las características que componen el vector, y cambios en el entrenamiento de SVMs.
- Detección de DHGs. Cambios en la estimación del movimiento y nueva máquina de estados.
- Ampliación del diccionario de gestos. Nuevos gestos detectables por el sistema.

3.3.1. Separación en posturas tipo.

Para mejorar la separación en posturas tipo y conseguir con ello una detección de SHPs más eficiente y robusta se plantean las siguientes líneas de trabajo:

- Cambios en la normalización. Búsqueda de una normalización no generalista.
- Selección y tratamiento previo de las caracteristicas que conformarán el vector de características final. Tratando de mejorar la precisión en la detección de SHPs.
- Metodología de entrenamiento de SVMs. Convirtiéndolo en válido para cualquier prueba que se quiera realizar.

3.3.1.1. Mejoras y cambios en la normalización.

En un primer intento de mejora, la propuesta era conseguir una normalización adaptada al sistema desarrollado, en lugar de normalizar cada coordenada independientemente del resto, lo que provoca deformaciones en la mano.

Entrando más en detalle, al normalizar la intensidad y amplitud de las protuberancias sin tener en cuenta unas proporciones a mantener, éstas se desvincularán de sus respectivas elipses, además de hacerlo también entre ellas, dando lugar a protuberancias irreales. Se propuso por tanto la siguiente normalización.

Declaración de algunas variables necesarias:

- Sean$I^{j,i}$y $A^{j,i}$la intensidad y amplitud (longitud y anchura) de la protuberancia i dentro de la mano j. Consideremos ambas como variables aleatorias.
- Sean a^jy b^jlos semiejes mayor y menor de la mayor elipse circunscrita en la silueta de la mano j detectada. Nuevamente, serán consideradas variables aleatorias.
- Sea $E^j = \pi \cdot a^j \cdot b^j$el área de la elipse ya mencionada, consideremos que es una variable aleatoria.
- Definimos también $\xi^{j,i} = \frac{I^{j,i} \cdot A^{j,i}}{2 \cdot E^j}$, que responde a la relación de tamaño entre protuberancias y elipse. Consideramos que se trata también de una variable aleatoria.

Para cada mano queremos normalizar $I^{j,i}$y $A^{j,i}$ para cada protuberancia, además de a^jy b^j. Buscamos por tanto obtener $I^{j,i}_{norm}$ y $A^{j,i}_{norm}$para cada par i,j , y a^j_{norm}y b^j_{norm} para cada mano j.

Comenzamos por normalizar E, con media 0 ($mean(E) = 0$) y varianza 1 ($std(E) = 1$). En consecuencia, para la mano j tendremos $E^j_{norm} = \pi \cdot a^j_{norm} \cdot b^j_{norm}$, donde además $E^j_{norm} = \frac{E^j - mean(E)}{std(E)}$, entonces tenemos que

$\pi \cdot a^j_{norm} \cdot b^j_{norm} = \frac{E^j - mean(E)}{std(E)}$ (1)

Continuamos con la normalización de ξ, consiguiendo media $0(mean(\xi) = 0)$ y varianza $1(std(\xi) = 1)$. Entonces, para cada par i, j tendremos $\xi^{j,i}_{norm} = \frac{I^{j,i}_{norm} \cdot A^{j,i}_{norm}}{2 \cdot E^j_{norm}}$.Además, $\xi^{j,i}_{norm} = \frac{\xi^{j,i} - mean(\xi)}{std(\xi)}$, obteniendo por tanto

$\frac{I^{j,i}_{norm} \cdot A^{j,i}_{norm}}{2 \cdot \pi \cdot a^j_{norm} \cdot b^j_{norm}} = \frac{\xi^{j,i} - mean(\xi)}{std(\xi)}$ (2)

Por otro lado asumimos las siguientes restricciones: Relación de aspecto constante para cada protuberancia,

$\frac{I^{j,i}}{A^{j,i}} = \frac{I^{j,i}_{norm}}{A^{j,i}_{norm}}(3)$

y relación de aspecto constante para la elipse,

$\frac{a^j}{b^j} = \frac{a^j_{norm}}{b^j_{norm}}(4)$

Juntando (1), (2), (3) y (4) tenemos 4 ecuaciones linealmente independientes y 4 incógnitas por despejar: $I^{j,i}_{norm}$, $A^{j,i}_{norm}$, a^j_{norm}y b^j_{norm}, siendo los demás datos de entrada.

El resultado de este intento de normalización no generalista, fue el esperado en cuanto a la no deformación de la mano, como se puede apreciar en la figura 3.13. Nótese que para esta representación los ángulos no han sido normalizados, con el objetivo de ver si se mantenían o no las proporciones de una forma más clara.

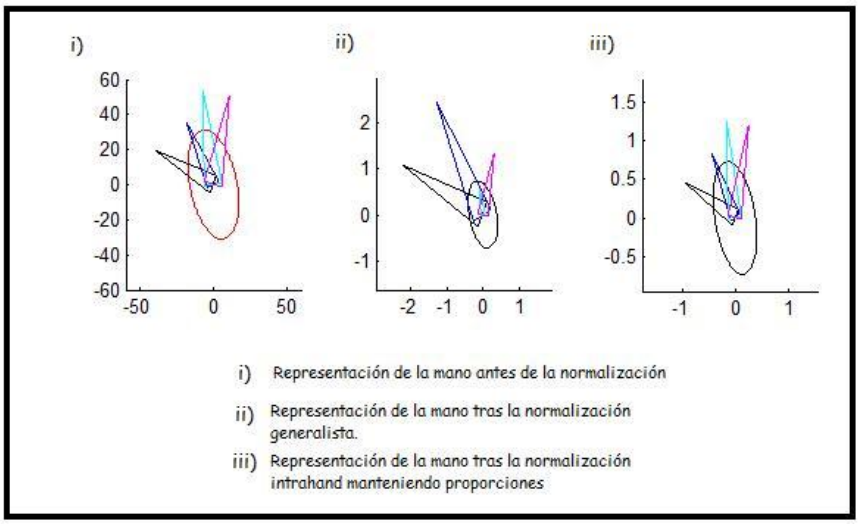

Figure 3.13: Representación de la mano antes y después de la normalización "intramano".

Los resultados obtenidos en el entrenamiento con esta nueva normalización, no mejoraron lo existente, como veremos más adelante en el capítulo de resultados, por lo que esta normalización fue descartada.

Al margen de esta normalización no generalista, y previendo los posibles cambios en el vector de carácteristicas, no sólo en los componentes del mismo, sino también en el número de ellos, se decidió añadir a la normalización una división por el número de coordenadas del vector. Esto no supondrá una

mejora cuantitativa en los resultados, pero sí mantiene constante el rango de distancias entre los distintos vectores, independientemente del número de coordenadas del vector. Con ello facilita el entrenamiento de las SVMs sin la necesidad de variar continuamente los valores de los parámetros utilizados para el mismo, y pudiendo comparar con ello los resultados obtenidos con los continuos cambios en el vector, buscando la mejor combinación de descriptores.

3.3.1.2. Selección y tratamiento previo de las caracteristicas que conformarán el vector.

Una de las principales limitaciones del sistema inicial es que las coordenadas del vector de caractaristicas son absolutas, esto es, no se tiene en cuenta en ningún momento la posición de la mano o la distancia de la misma a la cámara. Esto no es algo que afecte de manera muy significativa, debido al reducido tamaño de la zona de activación, y a la tendencia de los usuarios a realizar los gestos centrando la imagen. Pese a todo, es algo que debe ser mejorado para dotar de una mayor robustez y preparar al sistema de cara a mejoras en la captura que permitan la realización de gestos a mayor distancia.

Por tanto, se establece como objetivo principal de esta nueva selección de características la dotación del sistema de robustez frente a distancia, giro y posición.

a) Robustez frente a la distancia.

Para ver como afecta la distancia al tamaño devuelto por la cámara de un objeto presente en la escena veamos el dibujo de la figura 3.14.

Observemos en dicha figura, el objeto representado y proyectemos su imagen sobre un plano que se encuentra a una distancia d_2, suponiendo para ello simetria radial y perpendicularidad del objeto frente a la cámara. Su tamaño sería t_2 cuando se encuentra a esa distancia, d_2. Si se acerca a distancia $d_1$su tamaño proyectado en el plano dicho sería t_1. Teniendo esto en cuenta y usando trigonometría básica,

$tg(\alpha) = \frac{t_2}{d_1} = \frac{t_1}{d_2}$, de donde se extrae que $t_2 \cdot d_2 = t_1 \cdot d_1$, o lo que es lo mismo,

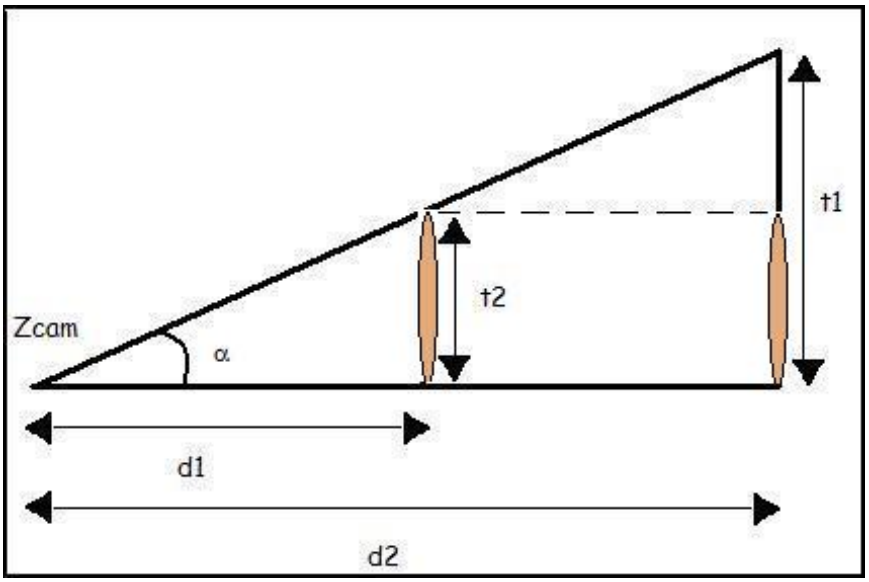

Figure 3.14: Tamaño de un objeto presente en la escena a distintas distancias de la cámara.

$$t{\cdot}d = cte$$

Hemos obtenido con esto una primera forma de dotar de independencia con respecto a la distancia a las características del vector. Multiplicar la dimensión obtenida por su distancia a la cámara.

Otra forma de conseguirlo será seleccionar descriptores que lleven intrínseca la independencia deseada.

Ninguna de las características extraídas en el tratamiento de bajo nivel de la imagen goza de independiencia frente a la distancia, pero si buscamos relaciones de tamaño que se mantengan constantes y las convertimos en componentes de nuestro vector, estaremos consiguiendo de nuevo el objetivo.

Imaginemos una protuberancia de intensidad I y amplitud A. Sus valores cambiarán en función de lo cerca o lejos que nos encontremos del punto donde este situada la cámara. En cambio, lo que no cambiará será la proporción entre ambas, es decir, la relación $\frac{I}{A}$ se mantendrá constante sea cual sea la distancia a la cámara.

Buscando dar una explicación que sirva de base a lo anteriormente planteado, se demostró con anterioridad que el producto *tamaño·distancia* se mantiene

constante. Por lo tanto, si multiplicamos la intensidad de la protuberancia, o su amplitud, por la distancia de la misma a la cámara el resultado será constante, $I \cdot d_p = cte$ y $A \cdot d_p = cte$, entonces la relación de proporción entre ambas también será constante, $\frac{I \cdot d_p}{A \cdot d_p} = cte$, y en consecuencia,

$$\frac{I}{A} = cte$$

Lo mismo sucederá con la relación entre los ejes mayor y menor de la elipse, su cociente también será independiente de la distancia.

b) Robustez frente al giro.

Había que tener cuidado con la forma de afrontar este apartado, pues dotar al sistema de robustez total frente al giro puede dar lugar a confusiones y equivalencias no deseadas entre gestos. La figura 3.15 muestra tres SHPs que se podrían llegar a ver perjudicados por lo anteriormente comentado.

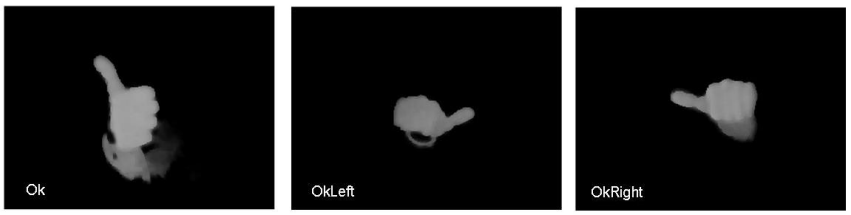

Figure 3.15: SHPs que resultarían equivalentes dotando al sistema de reconocimiento de robustez total frente al giro.

Para dotar al sistema de la robustez mencionada se decide fijar un ángulo para ser cuantificado, y tratar los demás como diferencias con el mismo. Dicho ángulo a cuantificar debe estar presente en la mayoría de los SHPs, pero además debe ser un ángulo que contenga información fiable. El único ángulo siempre presente en la imagen será el correspondiente a la inclinación de la elipse, pero su gran variación dentro de una misma imagen podria provocar fallos al tratar los demás como diferencias. El siguiente ángulo en número de apariciones es el ángulo de la primera protuberancia. Éste será el ángulo cuantificado, pues sus diferencias con el resto de ángulos de inclinación de las otras protuberancias se mantendrá en un rango de valores de poca variación y ayudará con ello a la separación entre posturas tipo.

Notar que la información relativa al ángulo de la primera protuberancia es muy importante para determinar la posición de la mano, pudiendo ser utilizada para determinar su orientación.

Su cuantificación se realiza con 8 niveles posibles de salida, concretamente los múltiplos de $\frac{\pi}{4}$ radianes. Para ello se lleva a cabo la siguiente transformación,

$$Ang'_1 = round(Ang_1/\tfrac{\pi}{4})\cdot\tfrac{\pi}{4}$$

Para el resto de ángulos, no se tiene en cuenta su valor absoluto sino que se toma como coordenada su diferencia con el ángulo de la primera protuberancia, antes de su cuantificación.

$$Ang'_n = Ang_n - Ang_1$$

Resulta coherente pensar que, salvo extrañas realizaciones de los gestos, esta diferencia tendrá valores muy cercanos entre realizaciones del mismo gesto, mientras que su valor absoluto dependerá en todo momento del ángulo de inclinación con que se realice el gesto correspondiente. Así lo muestran la figura 3.16 y la tabla 3.7, donde se pueden ver los valores absolutos y los tomados tras dotar al sistema de cierta independencia frente al giro, de dos realizaciones del mismo SHP pero con distinta inclinación.

Figure 3.16: Realización del SHP EnumFive con distintas inclinaciones.

Ang_n\Ejecución	1	2	Ang'_n\Ejecución	1	
Ang_1	215.94	227.86	Ang'_1	225	225
Ang_2	192.84	202.14	Ang'_2	-23.10	-25.72
Ang_3	173.66	182.60	Ang'_3	-42.28	-45.26
Ang_4	147.99	159.78	Ang'_4	-67.95	-68.08
Ang_5	92.79	98.84	Ang'_5	-123.15	-129.02

Table 3.7: Valores absolutos (Ang_n) de los ángulos y valores tras el tratamiento de los mismos(Ang'_n).

Siguiendo con el mismo ejemplo y puesto que la separación en posturas tipo esta basada en el cálculo de distancias entre distintos vectores de características, a continuación se calcula la distancia que se acumularía entre estas dos realizaciones del mismo gesto, en primer lugar usando los valores absolutos Ang_n, y por otro lado los valores sujetos al tratamiento de independencia frente al giro, Ang'_n.

Empezaremos por calcular la distancia que se acumularía utilizando los ángulos sin tratamiento previo.

$$d_{acum} = \sqrt{(ang_{i1} - ang_{j1})^2 + (ang_{i2} - ang_{j2})^2 + (ang_{i3} - ang_{j3})^2 + (ang_{i4} - ang_{j4})^2 + (ang_{i5} - ang_{j5})^2}$$

$$d_{acum} = \sqrt{(215.94 - 227.86)^2 + (192.84 - 202.14)^2 + (173.66 - 182.60)^2 + (147.99 - 159.78)^2 + (92.79 - 98.84)^2}$$

$$d_{acum} = 22.0024$$

Utilizando la misma fórmula realizamos los cálculos para el segundo caso, en el que se ha cuantificado un ángulo y los demás son tratados como diferencias con éste.

$$d_{acum} = \sqrt{(225 - 225)^2 + (23.10 - 25.72)^2 + (42.28 - 45.26)^2 + (67.95 - 68.08)^2 + (123.15 - 129.02)^2}$$

$$d_{acum} = 5.87$$

Dado que se trataba de dos realizaciones del mismo SHP, cuanto menor sea la distancia acumulada entre ellas, más fácil será llegar a esa conclusión tras el análisis y detección de postura estática.

c) Relleno de coordenadas inexistentes.

La ausencia de protuberancias provoca que haya un número determinado de coordenadas del vector de características se queden sin valor. Si recordamos el apartado 3.2.3.2, todos aquellos valores relacionados con una protuberancia inexistente se rellenaban con '-1', independientemente de que se tratase de un ángulo o una dimensión.

Considerando la posible aparición de protuberancias, insignificantes en cuanto a tamaño, debido a diferentes naturalezas de las manos de los usuarios, en este proyecto se propuso llevar a cabo un relleno coherente de los valores relativos a las protuberancias inexistentes. Entendiendo por coherente que una protuberancia aparecida por error debería parecerse más al relleno de una protuberancia inexistente que a una existente y correctamente extendida.

Por ello, el relleno de aquellas características relacionadas con tamaños o dimensiones de protuberancias no existentes son rellenadas con '0', encontrando su explicación en el pequeño tamaño de una protuberancia aparecida por error, como se puede apreciar en la figura 3.17.

Figure 3.17: Aparición de una protuberancia errónea en la realización del SHP EnumOne.

Siguiendo esta línea de coherencia, el relleno de los ángulos tenía que ser cambiado. La decisión tomada en este aspecto fue la de rellenar su valor con la media de los valores de dicha coordenada, tomando para su cálculo los

vectores usados para el entrenamiento de las SVMs.

d) Vector de características resultante.

Atendiendo a los resultados obtenidos en la separación de SHPs (ver tabla 3.2), se observan tres problemas: El reconocimiento de 'Fist', el reconocimiento de 'OpenHand', y la distinción entre 'EnumOne' y 'Ok', dos gestos con un gran parecido en la realización.

Con esto en mente y siguiendo la línea descrita anteriormente para dotar al sistema de robustez frente a la distancia y frente al giro se crearon varios vectores de características. Añadiendo y eliminando coordenadas y evaluando objetivamente los resultados, en base a los vídeos de prueba utilizados anteriormente. Esta evolución puede ser consultada en la sección 5.3.1.

Una de las medidas tomadas, fue eliminar del vector la coordenada que indicaba el número de máximos, entendiendo que dicha información quedaba ya suficientemente representada con el relleno de protuberancias inexistentes.

Siguiendo lo explicado anteriormente, todas las carácterísticas relativas a los dedos son dotadas de robustez frente a la distancia y frente al giro, dando lugar a tres coordenadas por máximo, que serán $Area_{P_n}/Area_{Elipse}$, I_{P_n}/A_{P_n}, y Ang_{P_n},siendo diferente la forma de tratar está última característica dependiendo de si es la primera protuberancia o cualquier otra, y rellenándose las inexistentes de la manera ya explicada (cuantificación para la primera protuberancia, y el resto como diferencias).

Lo mismo sucede con la información relativa a la elipse, que tendra representación mediante dos coordenadas, la relación entre ejes, Eje_{max}/Eje_{min}, y su ángulo de inclinación.

Una vez tratada la información relativa a los máximos encontrados en la mano y a la elipse, el objetivo era añadir toda la información posible presente en cada gesto, independientemente del número de protuberancias presentes. Dicha información será importante para separar los SHPs sin protuberancias ('Fist' y 'OpenHand') y ayudará a separar aquellos con un solo dedo extendido, en especial los más parecidos y problemáticos, que son 'EnumOne' y 'Ok'.

Estudiando las posibilidades existentes para introducir la información deseada en el vector, contamos con tres puntos significativos, presentes en

cualquiera de los gestos. Estos son el centro de gravedad de la mano (CoG), el centro de la elipse (CoGE) y el punto más cercano a la cámara (Zmin).

La primera característica finalmente seleccionada para formar parte del vector, fue denominada triángulo de profundidad, y hace referencia al formado por los tres puntos mencionados, CoG, CoGE y Zmin. Y se traduce en dos coordenadas del vector: la relación entre las distancias de CoG a Zmin y de CoGE a Zmin ($Rel_{TriPro} = dist_{CoG-Zmin}/dist_{CoGE-Zmin}$), y el ángulo formado por los dos vectores anteriores (Ang_{TriPro}). Se puede ver una representación del denominado triángulo de profundidad en la figura 3.18.

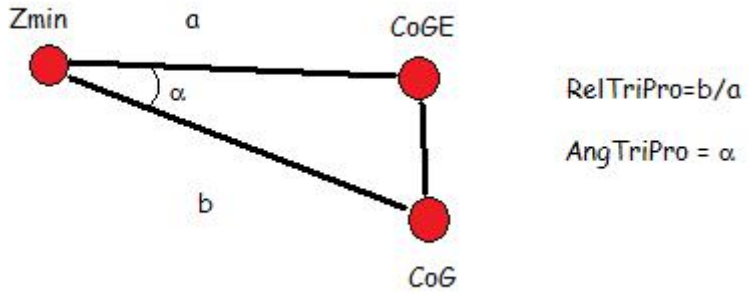

Figure 3.18: Representación triángulo de profundidad.

Reforzando las características representadas por estos puntos, buscando más información relacionada con la diferencia de posición entre el Zmin y la mano, y con una base empírica fundamentada en los resultados obtenidos en distintas pruebas: las tres coordenadas que completan las características seleccionadas son las correspondientes al vector CoGE-Zmin. Para seguir la línea de independencia con la distancia ya descrita, se incluye el vector $\overrightarrow{v} = \overrightarrow{CoGE - Zmin}$ como cocientes de sus coordenadas($\frac{v_x}{v_y}$, $\frac{v_y}{v_z}$, $\frac{v_z}{v_x}$), las cuales se pueden observar en la figura 3.19.

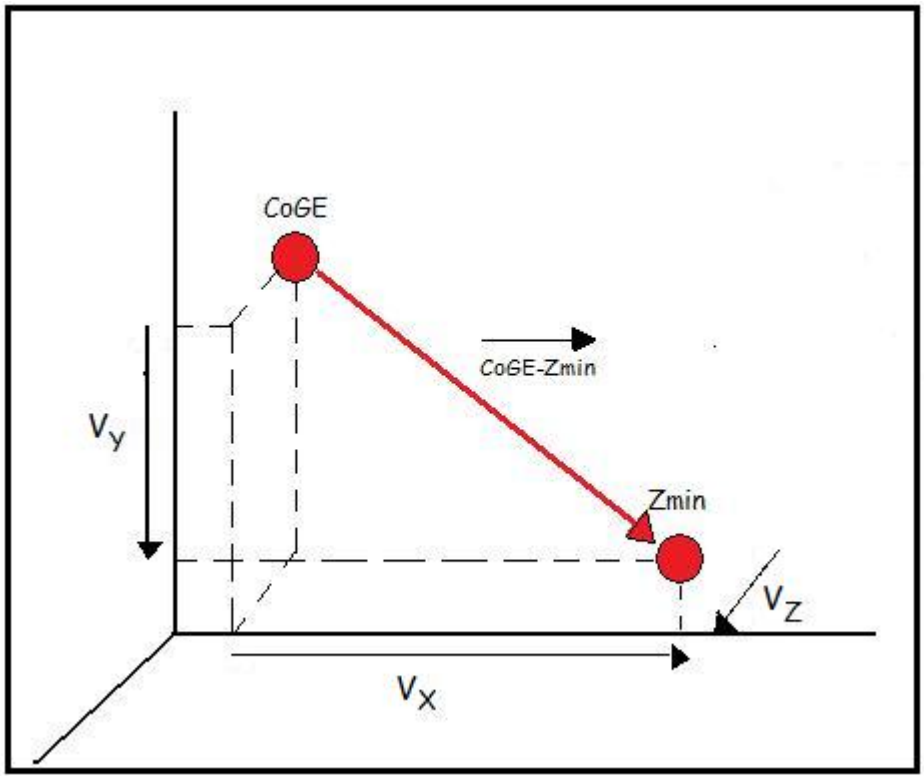

Figure 3.19: Vector CoGE-Zmin.

Con todo esto, el vector de características resultante es el siguiente:

1. $Area_{P_1}/Area_{Elipse}$
2. Int_{P_1}/Amp_{P_1}
3. Ang_{P_1}, cuantificado a $n \cdot \frac{\pi}{4}$, con $n \epsilon \mathbb{Z}$.
4. $Area_{P_2}/Area_{Elipse}$
5. Int_{P_2}/Amp_{P_2}
6. $Ang_{P_2} - Ang_{P_1}$
7. $Area_{P_3}/Area_{Elipse}$
8. Int_{P_3}/Amp_{P_3}
9. $Ang_{P_3} - Ang_{P_1}$

10. $Area_{P_4}/Area_{Elipse}$
11. Int_{P_4}/Amp_{P_4}
12. $Ang_{P_4} - Ang_{P_1}$
13. $Area_{P_5}/Area_{Elipse}$
14. Int_{P_5}/Amp_{P_5}
15. $Ang_{P_5} - Ang_{P_1}$
16. Rel_{TriPro}
17. Ang_{TriPro}
18. Eje_{mayor}/Eje_{menor}(Elipse)
19. Ang_{Elipse}
20. X/Y (vector CoGE-Zmin)
21. Y/Z (vector CoGE-Zmin)
22. Z/X (vector CoGE-Zmin)

3.3.1.3. Entrenamiento de SVMs.

El entrenamiento de las SVMs no cambió de forma significativa. Las modificaciones en este apartado se vieron reducidas al rango de valores adoptados por las constantes C y γ, que definen el kernel de la SVM. Los valores anteriores, $C = 2^{-4,-3,...,5}$combinado con $\gamma = 2^{-15,-13,...,3}$, se ven cambiados tras este proyecto a $C = 2^{-1,0,...,8}$combinado con $\gamma = 2^{1,3,...,9}$.

Recordando el apartado 3.3.1.1, estos valores no necesitarán cambios adicionales, pues la división por el número de coordenadas estabiliza las posibles distancias existentes entre vectores.

Los resultados obtenidos para estos valores, además de ser considerados aceptables, lograban un compromiso con el coste computacional, perdiéndose éste si se seguían aumentando dichos valores, siendo muy pequeño el rango de mejora posible.

3.3.2. Detección de DHGs: Ventana temporal, patrones de movimiento y máquina de estados.

Como ya se explicó con anterioridad, la detección de DHGs no sólo depende de la correcta separación entre distintos SHPs, sino que además se ve afectada por la información de movimiento y por la gestión que hace de todo ello la máquina de estados.

En este aspecto, un nuevo enfoque era necesario para mejorar la detección de gestos con información derivada del movimiento. Las limitaciones presentadas por el proyecto en este punto eran las que más afectaban al funcionamiento global del sistema.

Tras estudiar diferentes formas de detección, la elegida fue la de definir un grupo de patrones sintéticos de movimiento con los que comparar el movimiento real de un determinado punto. Con ello se obtiene el movimiento más parecido de entre los posibles, asumiendo que este ha sido el movimiento realizado por el usuario.

Los objetivos en el desarrollo de este apartado estaban claros:

i) Permitir la detección de diferentes patrones de movimiento, no únicamente de los necesarios para los gestos MenuOpen y MenuClose.

ii) Detectar movimientos realizados con diferentes velocidades y curvaturas.

iii) Detección en tiempo real, no olvidando que este es uno de los objetivos principales del sistema global.

3.3.2.1. Ventana temporal.

Recordando el apartado 3.2.4.1, en el cual se explica cómo se obtiene el SHP resultante en cada instante (ventana temporal), vemos como los positivos tienen el mismo valor independientemente del momento en el que han sido obtenidos. Por ejemplo, para una ventana temporal de 8 frames, un positivo obtenido al comienzo de la misma tiene el mismo peso que uno obtenido en el último frame de la ventana. Parece coherente pensar que los frames

más recientes hayan de tener más peso para la detección del SHP que está siendo realizado. Es por ello que esta igualdad de pesos es sustituida por una función que asinga diferentes valores a los distintos frames de la ventana temporal. Tras barajar el uso de una función lineal se optó por una función logarítmica que da más a peso según nos acercamos a los últimos frames de la ventana.

Más concretamente, la función utilizada fue:

$$y = log(x + 0.25)/log(tamañoVentana + 0.25)$$

, donde x va desde 1 hasta el tamaño de la ventana temporal. En la figura 3.20 podemos ver la representación de esta función, donde se aprecia claramente como los últimos frames tienen un peso superior al que tienen los primeros frames de la ventana.

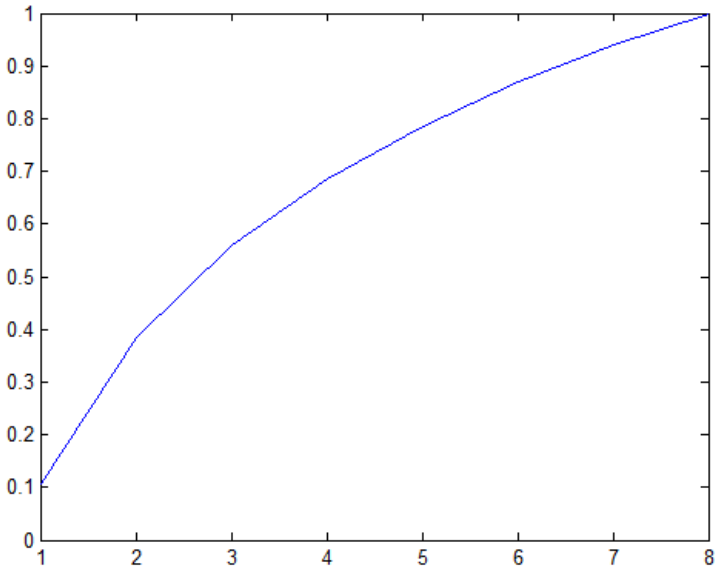

Figure 3.20: Función representativa del peso que tienen los positivos en la ventana temporal.

3.3.2.2. Patrones de Movimiento Detectables.

El objetivo de este proyecto va más allá de mejorar la detección de los gestos MenuOpen y MenuClose. Aumentar el número de movimientos detectables

se antoja necesario para dotar al sistema de una mayor usabilidad. En primer lugar, el interés está centrado en detectar movimientos análogos al vertical, pero en horizontal, con el objetivo de utilizar gestos con esos patrones para pasar página y volver a la página anterior.

Dando un paso más, se incluyen hasta un total de 4 movimientos nuevos diferentes, sumados a los verticales y horizontales ya comentados, que se corresponden con las 4 diagonales posibles.

La figura 3.21, muestra una representación de todos los movimientos objeto de detección.

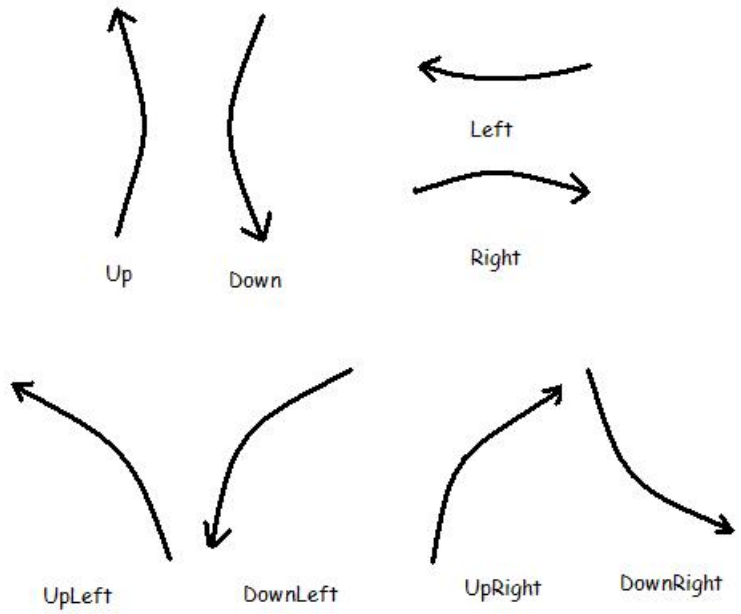

Figure 3.21: Patrones de Movimiento detectados.

3.3.2.3. Definición de los patrones sintéticos.

La variedad en los patrones sintéticos de movimiento definidos establece implicitamente la colección de patrones de movimiento que el sistema será capaz de detectar ante las ejecuciones de usuarios, limitando con ello los gestos detectables y por tanto, el funcionamiento global del sistema.

Partiendo de las grabaciones realizadas por seis usuarios de los distintos DHGs y tomando los gestos MenuOpen y MenuClose como referencia para el

estudio de las trayectorias, se decidió establecer como base para los patrones de movimiento arcos de elipse variando, el plano al que pertenece la elipse, el sentido en el que se recorre la elipse, la longitud del arco y el valor de los semiejes de la elipse (SE1y SE2, ver la Figura 3.22).

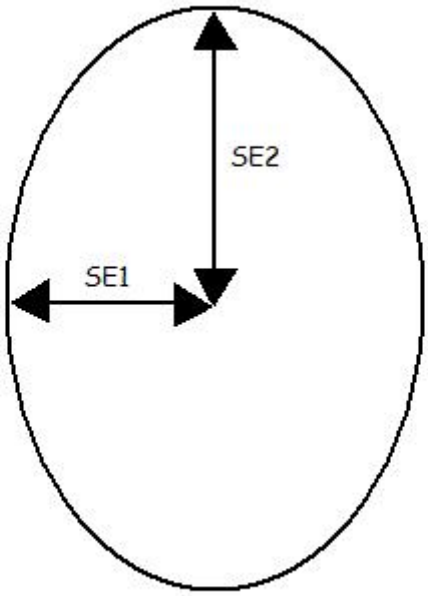

Figure 3.22: Elipse para definir arcos de movimiento.

Variando las diferentes coordenadas y cambiando la ubicación de la elipse, se consigue que los patrones sintéticos se asemejen a arcos correspondientes a movimientos verticales, horizontales y diagonales, situando la elipse en un plano vertical, horizontal o inclinado respectivamente. Con ello se consigue el primero de los objetivos de este apartado, ampliar la colección de movimientos detectables incluyendo, además de los dos de movimiento vertical, dos de movimiento horizontal y cuatro más en las cuatro posiciones diagonales. De esta forma detectamos, finalmente, ocho gestos basados en moviento que se corresponden con "manotazos" en las siguientes direcciones en grados: 0^o, 45^o, 90^o, 125^o, 180^o, 225^o , 270^o y 315^o (ver Figura 3.21).

Para afrontar las complicaciones debidas a las distintas velocidades y recor-

ridos en la realización del movimiento, la decisión tomada al respecto fue mantener constante la longitud de los patrones (número de coordenadas de la evolución). De esta forma se facilita el cálculo de distancias. Tras estudiar las ejecuciones más rápidas, que duraban en torno a 7-8 frames, se decidió establecer 5 frames como tamaño invariable. La rapidez, o amplitud del movimiento, unidas a la posición y distancia a la cámara pueden provocar que en ese espacio de tiempo el punto objeto de estudio complete desde más de media elipse hasta una porción inferior al 25 % de la misma, siendo ambos movimientos válidos para la detección. La figura 3.23 muestra cómo sin variar el tamaño del patrón sintético se pueden tener distintas velocidades de ejecución en consideración. Para ello la elipse se divide en arcos de diferentes longitudes, todos ellos representados mediante 5 puntos (tamaño establecido para todos los patrones).

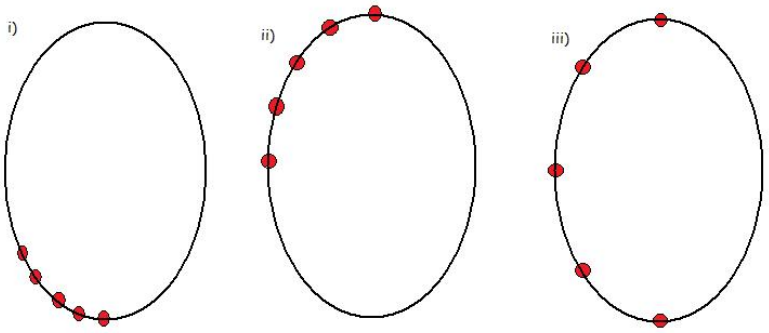

Figure 3.23: Ejemplos de diferentes subarcos de la elipse.

A continuación, en la figura 3.24, se pueden ver todos los subarcos que dan lugar a un patrón sintético de movimiento dentro de una misma elipse, controlando con ello la velocidad de ejecución y en parte, la amplitud del movimiento.

Para terminar de controlar este último punto relativo a la amplitud y además un mayor/menor acercamiento a la cámara al realizar el gesto se combinan 4 valores diferentes para los semiejes de la elipse, dando lugar a 4 elipses distintas. Teniendo en cuenta que a partir de cada elipse generamos 18 patrones diferentes, tendremos un total de 72 patrones para cada patrón de movimiento a detectar (definido por el plano al que pertenece la elipse y el sentido en el que se recorre la misma).Cada uno de estos conjuntos de

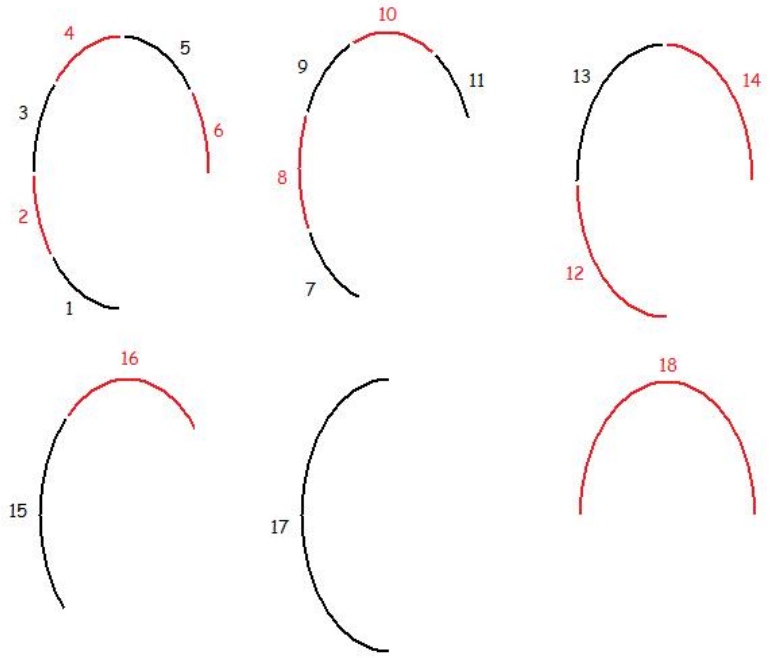

Figure 3.24: Subarcos que dan lugar a un patrón sintético de movimiento dentro de cada elipse.

72 patrones está asociado pues a uno de los ocho patrones de movimiento mencionados.

3.3.2.4. Detección de patrón de movimiento.

La tecnica de detección de patrón de movimiento ante la ejecución de un usuario es sencilla: se obtienen patrones de longitud cinco de la ejecucion del usuario, para cada uno de ellos se calcula el patrón sintético más cercano, asignándole la etiqueta de éste. Al finalizar la ejecución contaremos con un histograma de etiquetas. La salida vendrá definida por la etiqueta con mayor incidencia a lo largo de la ejecución. Esto ocurre con todos los movimientos excepto con el movimiento "estático", cuya presencia en el histograma provocará que sea dicho movimiento la salida.

El algoritmo elegido para calcular la distancia entre patrón real y cada uno de los patrones sintéticos fue Dynamic Time Warping (DTW)[27], tradicionalmente usado para reconocimiento de voz por su capacidad para ajus-

tarse a ejecuciones variables en tiempo y velocidad, lo cuál es un requisito imprescindible para el reconocimiento gestual que contiene información de movimiento. La implementación utilizada es la que sigue a continuación:

La función utilizada para su cálculo parte del siguiente código,

```
int DTWDistance(char s[1..n], char t[1..m]) {
  declare int DTW[0..n, 0..m]
  declare int i, j, cost
  for i := 1 to m
    DTW[0, i] := infinity
  for i := 1 to n
    DTW[i, 0] := infinity
  DTW[0, 0] := 0
  for i := 1 to n
    for j := 1 to m
      cost:= d(s[i], t[j]) DTW[i, j] := cost + minimum(DTW[i-1, j ], DTW[i
, j-1], DTW[i-1, j-1])
  return DTW[n, m]
}
```

devolviendo la mínima distancia entre dos curvas. De ahora en adelante, cuando se hable de distancia entre patrones, siempre será referido a este concepto, y no a la distancia euclídea.

Elegida la función para el cálculo de distancias surgieron dos caminos para su uso. Tratándose de movimiento, su estudio lleva implícito el estudio de la evolución, olvidándonos por tanto de coordenadas absolutas. Una primera opción, finalmente descartada a tenor de los resultados era igualar los comienzos de ambos patrones, el sintético y el real, y relativizar la posición en cada instante a la posición inicial. Esto se conseguía restando a cada posición en cada instante la posición inicial de dicho movimiento, llevando con ello el inicio al orígen de coordenadas y realizando una traslación de todas las posiciones siguientes a su punto correspondiente con respecto al orígen.

La segunda opción, la elegida por su mejor adaptación a las pruebas realizadas, era relativizar la posición en cada instante a la posición en el instante anterior, o dicho de otra forma, convertir la evolución absoluta en una evolución de la pendiente local. Para ello, a cada posición se le restaban las tres coordenadas del mismo punto en el frame inmediatamente anterior, llevando también el inicio al orígen.

A estas dos formas de calcular la distancia se unió la selección del punto cuyo movimiento sería objeto de estudio y determinaría con más fidelidad el movimiento realizado por el usuario. Los puntos elegidos para ser estudiados y comparados fueron el CoG, el CoGE y el Zmin.

Recopilando, tenemos dos métodos, pendiente local y global, y tres puntos posibles para su aplicación, CoG, CoGE y Zmin, lo que da lugar a 6 combinaciones posibles que fueron estudiadas con los vídeos grabados con anterioridad para la evaluación del sistema. Comprobando los resultados obtenidos con los gestos dependientes del movimiento, 'MenuOpen' y 'MenuClose', dando como resultados los mostrados en las tablas 3.8 y 3.9. En estas tablas se ve cómo el mejor método es el estudio de la pendiente local y el punto que mejores resultados presenta es el centro de gravedad de la mano, CoG.

	Up				Down			
	tp	tn	fp	fn	tp	tn	fp	fn
CoG	10	108	2	1	11	108	2	0
CoGE	9	108	2	2	11	108	2	0
Zmin	8	101	9	3	10	103	7	1

Table 3.8: Resultados para los DHGs MenuOpen y MenuClose con estudio de la pendiente absoluta en los distintos puntos.

	Up				Down			
	tp	tn	fp	fn	tp	tn	fp	fn
CoG	10	109	1	1	11	110	0	0
CoGE	8	110	0	3	11	107	3	0
Zmin	9	107	3	2	9	105	5	2

Table 3.9: Resultados para los DHGs MenuOpen y MenuClose con estudio de la pendiente local en los distintos puntos.

3.3.2.5. Máquina de estados.

La máquina de estados fue mejorada por socios del proyecto en el que se engloba este trabajo. Su nuevo diseño permite su configuración mediante un archivo de texto donde se definen los gestos detectables mediante los siguientes parámetros:

- *numStaticPoses*. Número de SHPs que compone el DHG.
- *pose*. La definición de varios *pose* consecutivos define una secuencia de identificadores de SHPs, tantos como *numStaticPoses* indique. La secuencia de SHPs definirá en gran medida el DHG en concreto.
- *gesture*. Su longitud también será igual al valor de *numStaticPoses*, conteniendo los DHGs de salida tras cada uno de los SHPs detectados.
- *outRequirement*. Su valor puede ser *true* o *false*, e indica la necesidad o no de que se produzca desactivación (ver apartado 3.1.1) para dar el DHG por finalizado.
- *movRequirement*. Indica el patrón de movimiento que se ha de dar para el DHG definido.

Clasificaremos los gestos en tres grupos: gestos estáticos, gestos dinámicos y gestos compuestos, ejemplificando a continuación sus definiciones:

a) Gestos simples estáticos.

Se incluyen en este grupo los DHGs que consisten en la ejecución de tan solo un SHP. Para su detección es necesario que la mano este estática durante su realización. Hay que hacer una distinción entre los que pueden dar lugar a gestos compuestos (en la detección se da cierto retardo) y los que no (la detección es prácticamente instantánea), y en función de ello serán definidos de formas diferentes en la máquina de estados.

Empecemos por la definición de aquellos gestos unívocos, que no pueden dar lugar a otros gestos combinándose con alguna secuencia de movimiento o postura estática diferente. La definición básica de un gesto que reúne estas condiciones será:

#<Nombre del gesto>

numStaticPoses=1

pose=<Identificador del SHP>

gesture=<Identificador del DHG>

outRequirement=false

movRequirement=0

Prestar atención a algo ya comentado, para estos gestos no es necesaria la desactivación. No hay que salir de la zona de interacción tras su realización ya que su detección es inmediata, entendiendo por inmediato la salida de una ventana temporal. La única condición que limita su obtención es el requisito de movimiento. Su valor igual a '0' indica que el movimiento ha de ser estático, algo que sucederá cuando el estudio en la ventana temporal del movimiento determine un mayor parecido con el patrón sintético correspondiente a la ausencia de movimiento. Sólo es necesaria una salida de movimiento estática para que el movimiento global sea considerado estático. Esto es algo que no sucede cuando se dan otros patrones de movimiento, como veremos más adelante.

Los gestos que cumplen la definición expuesta y no pertenecen a gestos compuestos se pueden encontrar en la tabla 3.10.

DHG<id>	Gesto	SHP<Id>
2	G_EnumTwo	2
3	G_EnumThree	3
4	G_EnumFour	4
10	G_MenuLeft	11
11	G_MenuRight	12
12	G_Cancel	7

Table 3.10: Gestos simples estáticos unívocos.

Dentro de los gestos estáticos, nos encontramos con algunos que además de ser gestos por sí solos forman parte de gestos compuestos. Su detección no puede ser realizada de igual manera, siendo necesario un tiempo de espera para ver qué es lo que sigue a su realización. Veremos con más detalle su definición en el apartado correspondiente a gestos compuestos.

b) Gestos simples dinámicos.

Forman parte de este grupo todos aquellos gestos definidos por la información de movimiento, siendo irrelevante el SHP detectado. Puesto que el proyecto tenía como uno de sus principales objetivos la detección de patrones de movimiento, los gestos simples dinámicos han sido el grupo más afectado por la ampliación. A los gestos ya existentes, 'G_MenuOpen' y 'G_MenuClose', se ahora se suman otros seis, que podemos ver en la tabla 3.11.

DHG<id>	Gesto	Patrón de movimiento.
7	G_MenuOpen	Ascendente.
8	G_MenuClose	Descendente.
15	G_PageLeft	Horizontal, de derecha a izquierda.
16	G_PageRight	Horizontal, de izquierda a derecha.
17	G_SlapUpRight	Diagonal hacia arriba, de izquierda a derecha.
18	G_SlapDownRight	Diagonal hacia abajo, de izquierda a derecha.
19	G_SlapUpLeft	Diagonal hacia arriba, de derecha a izquierda.
20	G_SlapDownLeft	Diagonal hacia abajo, de derecha a izquierda.

Table 3.11: Gestos simples dinámicos.

La definición de cualquiera de estos gestos en el archivo de configuración de la máquina de estados, es como sigue:

#<Nombre del gesto>

numStaticPoses=1

pose=Cualquiera

gesture=<Identificador del DHG>

outRequirement=true

movRequirement=<Identificador del patrón de movimiento requerido>

Notar, como se dijo anteriormente, que cualquier SHP es válido en uno de estos DHGs. Es imprescindible que estos gestos no se detengan durante su ejecución, pues eso daría lugar a un patrón de movimiento estático, que sale fuera de la definición del DHG. El patrón de movimiento detectado será aquel con mayor representación en las salidas intermedias(salida de cada ventana temporal) durante la realización del gesto.

c) Gestos compuestos.

En el apartado 3.3.2.5 se introdujo el concepto de gesto compuesto, para cuyo reconocimiento es necesaria la detección consecutiva de varios SHPs diferentes.

Son dos los gestos que nos encontramos en este grupo, G_Take&Release y G_Click, indicados en la tabla 3.12.

DHG<id>	Gesto	EvoluciónSHP
9, 14	G_Take, G_Release	EnumFive-Fist-EnumFive
13	G_Click	EnumOne-Fist

Table 3.12: Gestos compuestos.

El primero, ya explicado y detallado con anterioridad (ver apartado 3.3.2.5), responde a la necesidad de coger un objeto presente en la escena, arrastrarlo y soltarlo. Su definición en la máquina de estados será la siguiente:

#Take-Release

numStaticPoses=3

pose=5

pose={8,9,10}

pose={5,4,3,7}

gesture=-1

gesture=9

gesture=14

outRequirement=false

movRequirement=0

Puntualizar en referencia a esta definición que originalmente eran tres los SHPs a detectar: EnumFive-Fist-EnumFive. Para mejorar la robustez en la detección y basándose en problemas aparecidos durante las pruebas realizadas, se decidió añadir SHPs como posibles tanto al 'Fist' como al segundo 'EnumFive'. El primero, era confundido en ocasiones con los SHPs correspondientes a una mano extendida, sobre todo cuando se cogía un objeto

situado en el lateral de la pantalla. En esta situación se da la aparición del antebrazo que es confundido como mano alargando la silueta detectada más de lo conveniente para detectar un 'Fist'. Algo similar sucede con el segundo 'EnumFive', su realización lateral en algunas ocasiones provoca confusiones con un 'EnumFour' o, incluso, un 'EnumThree', por lo que también estos últimos SHPs pasarón a ser considerados válidos en este DHG.

El segundo gesto compuesto mencionado, G_Click, puede asemejarse con el click realizado con un raton de ordenador y estará destinado a la selección de objetos de forma natural. Sus transiciones en la máquina de estados pueden verse en la figura 3.25.

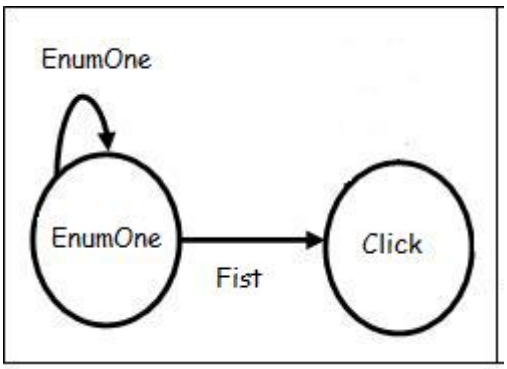

Figure 3.25: Transiciones de la FSM en la ejecución de G_Click

En el archivo de configuración aparece de la siguiente manera,

#Click

numStaticPoses=2

pose={1,6}

pose={8,9,10}

gesture=-1

gesture=13

outRequirement=false

movRequirement=0

Cabe comentar la posibilidad de realizarlo con 'EnumOne' o con 'Ok' como primero de los SHPs a detectar. Esto es debido a que son dos gestos difícilmente diferenciables a partir de la caracterización de mano sobre la que

trabajamos. Con la detección del 'Fist' que cierra el gesto, su tratamiento es el mismo que en el caso del anterior gesto compuesto, añadiendo la posibilidad de detección de un SHP correspondiente a una mano extendida además de la del propio 'Fist'.

Definidos los gestos compuestos es buen momento para ver qué sucede con la definición de gestos simples que pueden dar lugar, dependiendo de las transiciones, a gestos compuestos. Estos gestos son los incluidos en la tabla 3.13.

DHG<id>	Gesto	SHP<id>
1	G_Select	1
5	G_EnumFive	5
6	G_Ok	6

Table 3.13: Gestos simples estáticos que componen parte de algún gesto compuesto.

Su definición de cara a la máquina de estados es la siguiente:

#<Nombre del gesto>

numStaticPoses=2

pose=<Identificador del SHP>

pose=<Identificador del SHP>

gesture=-1

gesture=<Identificador del DHG>

outRequirement=true

movRequirement=0

Traducido a palabras, para detectar uno de estos DHGs es necesario detectar en dos ocasiones consecutivas su SHP asociado. Además estos DHGs requieren de desactivación, algo que no pasaba con el resto de gestos estaticos. De esta forma evitamos la salida de la detección antes de tener la seguridad de que no se trata de algún gesto compuesto.

Capítulo 4

Integración en el sistema

4.1. Introducción.

Este trabajo, como ya se comentó, se lleva a cabo dentro de un proyecto en el que se unen los esfuerzos de distintos grupos de investigación con el objetivo de crear un sistema que engloba desde de la extracción de características de imágenes hasta el nivel de aplicación, pasando por la separación de patrones para la detección de los gestos ejecutados. En consecuencia, la integración de los avances expuestos en este documento requiere de una especial atención.

A lo largo de esta sección se describe la metodología seguida para la integración de los cambios ya descritos.

4.2. Generación de los modelos para la separación de las posturas de mano elegidas *(SHPs)*.

La separación de posturas estáticas de la mano (SHPs) se realiza mediante el entrenamiento de SVMs, capaces de trazar hiperplanos que separan los vectores de características según de los gestos a los que correspondan.

El entrenamiento se realiza mediante scripts matlab que utizan la librería libsvm [35]. Para la realización del mismo se utilizaron inicialmente vídeos grabados por 6 usuarios que durante 10 segundos realizaron en distintos puntos de la zona de interacción cada uno de los SHPs a separar. Esto da lugar

(ver apartado 3.2.3.2) a un total de 14400 muestras, 1200 por cada una de las posturas posibles, que son 12:1200$\frac{muestras}{SHP}$·12$\frac{SHP}{sistema}$ = 14400$\frac{muestras}{sistema}$. Estas cifras pueden verse modificadas cambiando el entrenamiento, pero serán las usadas durante todo este capítulo relativo a la integración.

Las 1200 muestras correspondientes a cada SHP objeto de separación, serán tomadas como muestras positivas, usando el resto, 13200, como negativas en el entrenamiento de la SVM correspondiente. Tras la selección de los valores para los parámetros del núcleo RBF (*Radial Basis Function*), determinados mediante validación cruzada (ver apartado 3.3.1.3), la SVM da como resultado dos archivos, 'model.txt' y 'mv.txt', que contienen la información necesaria para realizar la posterior separación. La figura 4.1 muestra un esquema del funcionamiento de cada SVM.

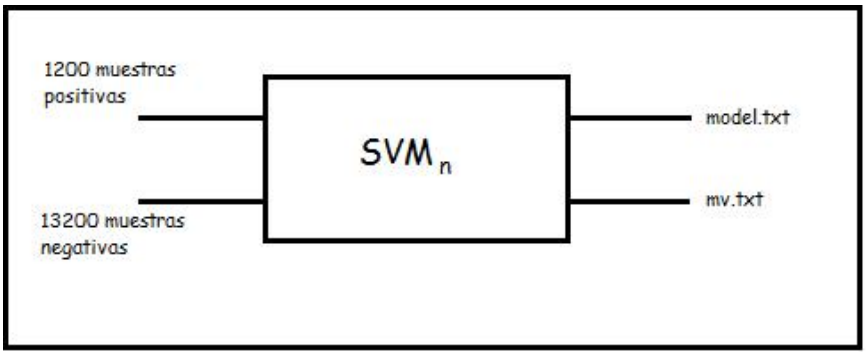

Figure 4.1: Entrada y salida en entrenamiento de cada SVM.

Contamos pues con un modelo de clasificación para cada SHP. El proceso de entrenamiento que da lugar a la generación de modelos queda lejos de poder realizarse en tiempo real, pues tiene un elevado coste computacional. Por otro lado, esto no es necesario, ya que éste se realiza una sola vez previa a la puesta en funcionamiento del sistema. Podemos ver en la figura 4.2, un esquema de como las características extraidas sobre los videos grabados sirven de entrada para las SVM que dan lugar a los modelos usados para la separación en SHPs.

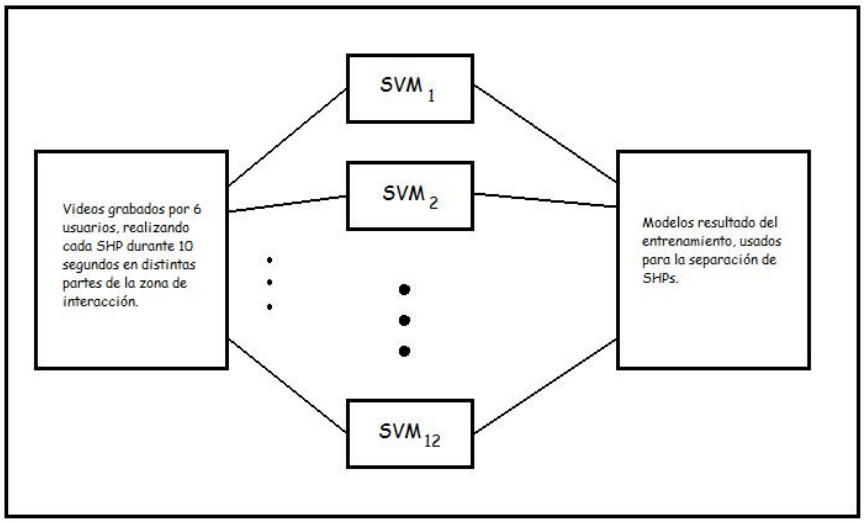

Figure 4.2: Diagrama de entrenamiento.

El número de usuarios, gestos, o muestras por usuario y gesto pueden ser modificados siendo necesario para ello reentrenar el sistema.

4.3. Integración del modelado de la mano.

La nueva selección de características de la mano que componen el vector encargado de caracterizar la postura estática, es un proceso anterior al entrenamiento. Es decir, las nuevas características deben ser seleccionadas antes del entrenamiento, y el vector debe ser normalizado para su utilización. Posteriormente esa misma normalización ha de ser integrada en el sistema en funcionamiento, para que los vectores utilizados en entrenamiento sean comparables con los usados para detección.

Los vectores de características correspondientes a cada frame serán la entrada utilizada por las SVM para el proceso de entrenamiento. En la figura 4.3 se puede observar el tratamiento realizado sobre cada vector de carácterísticas. En dicha figura se aprecia de forma esquemática la evolución desde la captura de las imágenes hasta el entrenamiento de las SVM, pasando por una etapa intermedia en la que tienen lugar la segmentación, la extracción de características, su selección y tratamiento, y finalmente su agrupación en clases y etiquetado manual en función del SHP realizado. La colección de carac-

terísticas y sus correspondientes etiquetas para los frames de la colección de vídeos descrita ya en 4.2 constituyen la entrada para el entrenamiento.

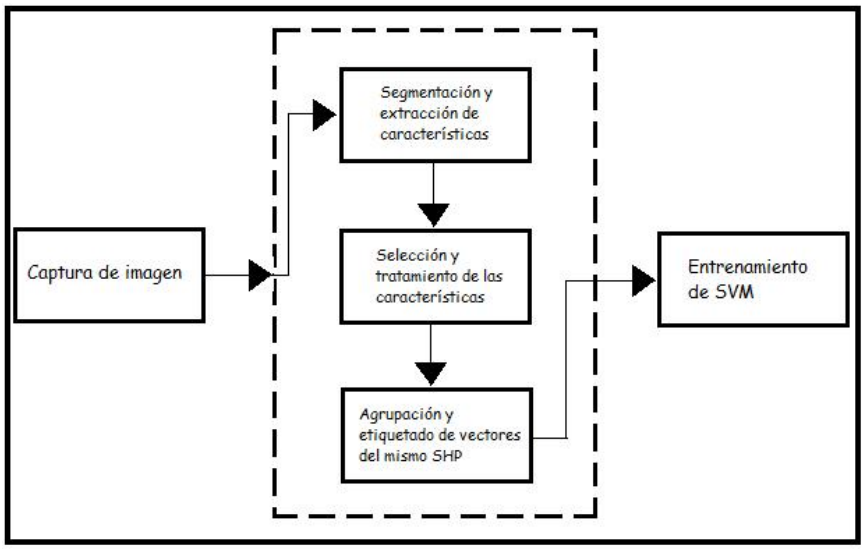

Figure 4.3: Selección y tratamiento previo de las características que compondrán el vector.

El esquema en la fase de predicción (i.e. sistema en funcionamiento) es análogo en términos de extracción y normalización de características. Se recibe una imagen y el sistema tiene que decidir acerca de su parecido con las posturas establecidas. En la figura 4.4 vemos lo que sucede de forma esquemática. Tal y como sucedía en el entrenamiento, tras la extracción de características éstas son seleccionadas y tratadas antes de proceder a la separación de posturas.

Para la separación, son necesarias tantas SVMs como SHPs se contemplan en la detección. En la configuración básica e inicial, se trata de 12 (ver apartado 3.2.3.1). La salida de cada SVM es binaria , esto es, '1' si el vector de características produce una predicción positiva y '0' en caso contrario. Esta información es tratada posteriormente (ver sección 3.2.4.1) para decidir acerca de la postura estática. Básicamente, se tiene en cuenta los resultados dentro de una ventana temporal, el número de positivos y negativos, así como la fiabilidad de los mismos. La figura 4.5 muestra un esquema de cómo tiene lugar este proceso, que concluye con la obtención del SHP detectado.

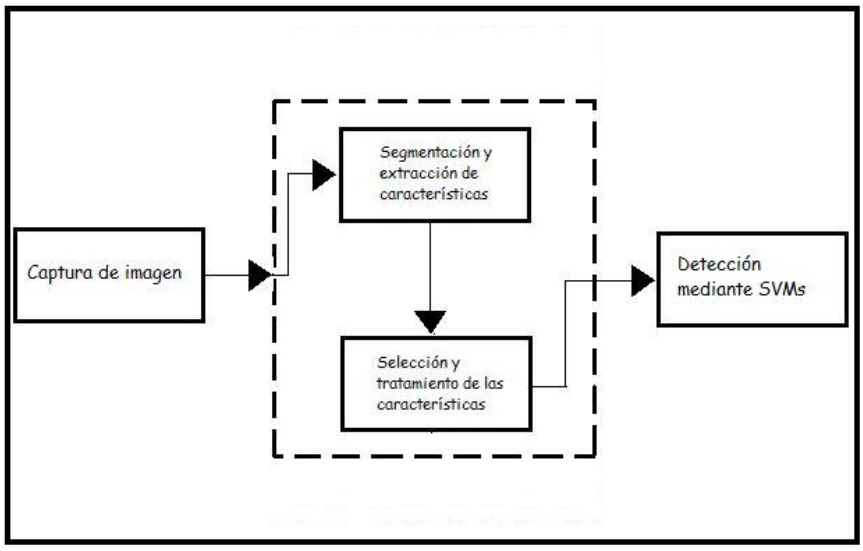

Figure 4.4: Selección y tratamiento de las características para la detección de SHP.

4.4. Integración de la detección de movimiento.

Ya se ha comentado con anterioridad que el sistema de detección de gestos cuenta con dos fuentes de información básicas. Éstas son la predicción de postura estática de la mano y la estimación del movimiento de la misma. El apartado anterior describía el proceso para la estimación de la postura estática de la mano. En paralelo, se realiza la detección de patrones de movimiento, detallándose a continuación todo lo necesario para su integración en el sistema.

Recordando apartados anteriores, el movimiento se detectaba comparando la evolución del centro de gravedad de la mano con un grupo de patrones de movimiento sintéticos, definidos en base a los movimientos objeto de detección. Dicha comparación nos dice cual de los patrones sintéticos es el más parecido, dando como resultado una salida correspondiente al movimiento en cada frame, que después es tratada para concluir cuál es el movimiento global descrito.

En primer lugar, es necesaria la definición de los patrones sintéticos de movimiento que posteriormente se comparan con el movimiento seguido por la mano. Dichos patrones se incluyen en un archivo de texto que es uno de los ficheros de configuración del sistema, siendo necesario para de-

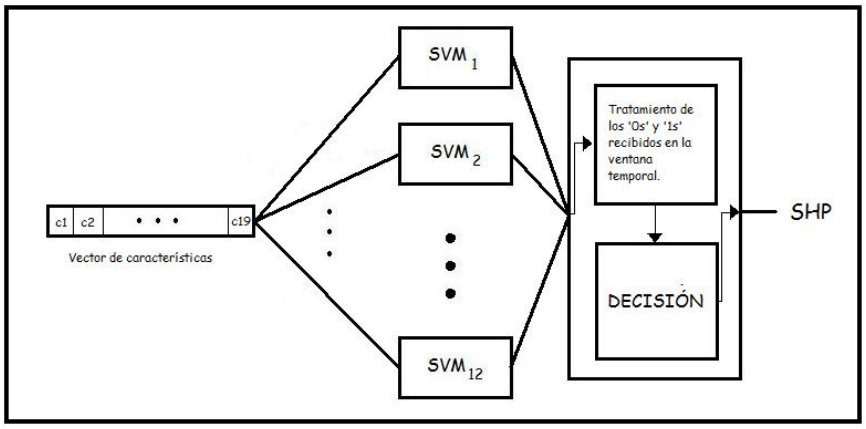

Figure 4.5: Detección mediante SVMs.

terminar el movimiento. En dicho fichero cada línea se corresponde con un patrón de movimiento sintético, y en ella se puede ver la evolución de las tres coordenadas durante la duración del patrón, esto es, 5 frames $(x_1y_1z_1, x_2y_2z_2, x_3y_3z_3, x_4y_4z_4, x_5y_5z_5)$.

En ejecución, contaremos con la información del movimiento realizado por la mano, entendiendo ésta como la evolución temporal de una de las características del vector, concretamente a las tres coordenadas del centro de masas de la mano(CoG), la cual es comparada con los patrones sintéticos ya descritos. La figura 4.6 muestra las entradas al módulo de detección de movimiento.

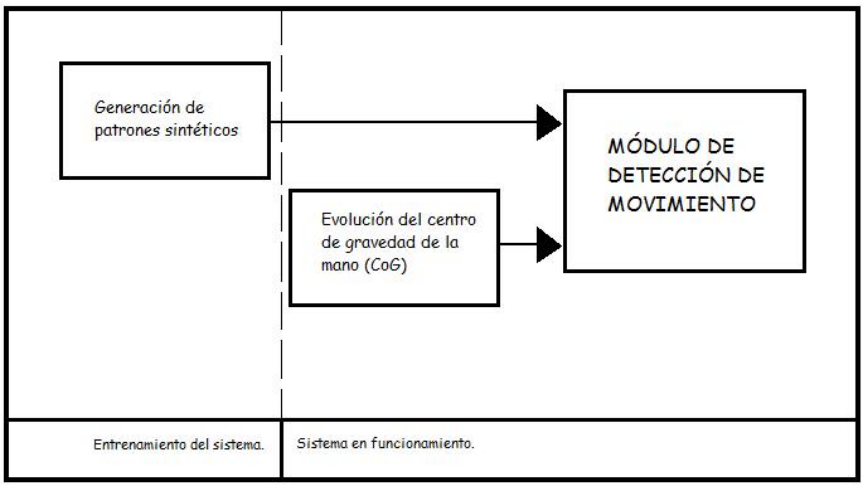

Figure 4.6: Entradas del sistema para la detección de movimiento.

Los patrones sintéticos de movimiento no sufren alteración alguna durante el funcionamiento del sistema, respetando lo definición inicial. Para la comparación se define una ventana temporal destinada a almacenar la evolución del CoG, cuyo tamaño es de 5 frames, siendo así del mismo tamaño que los patrones sintéticos de movimiento. Se corresponde con el tamaño de los patrones sintéticos de movimiento. Cada vez que se llena la ventana (lo cual sucede en cada frame, excepto en los 4 iniciales), su contenido es comparado con el diccionario de patrones sintéticos, dando como resultado el identificador del patrón sintético que presenta la menor distancia al patrón real que esta teniendo lugar en la ejecución (ver apartado 3.3.2.4). Considerando una imagen capturada de duración N frames, esta dará lugar a N-4 movimientos locales extraídos (ver figura 4.7), una por cada frame, excepto los 4 iniciales en los que se llena la ventana temporal la primera vez.

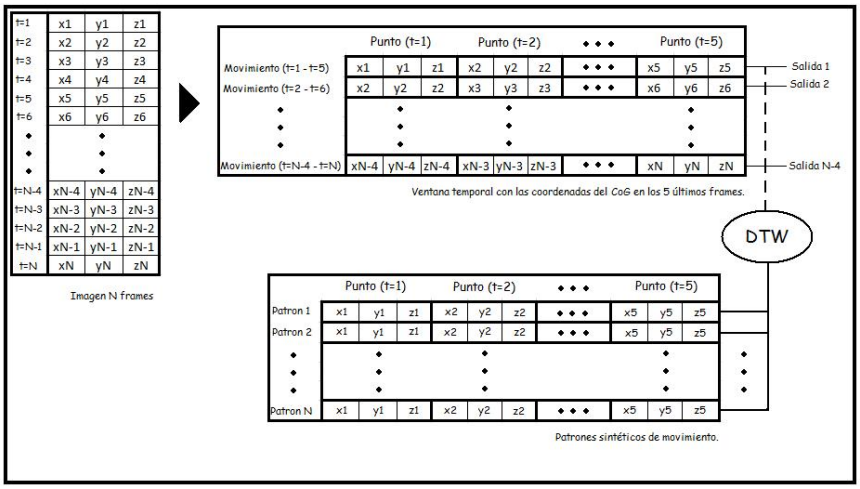

Figure 4.7: Movimiento local. Salida cada frame, excepto los 4 primeros.

Tenemos por tanto una salida correspondiente al movimiento en cada frame, a la que denominaremos movimiento local. El movimiento global, que representa el movimiento a lo largo de la realización de un gesto, se obtiene a partir de los movimientos locales y es el resultado final del módulo de detección de movimiento.

El valor del movimiento global vendrá dado por la etiqueta de patrones de movimiento sintéticos que presente un mayor número de apariciones en las salidas locales hasta ese instante, salvo que alguno de los movimientos locales de como resultato un movimiento “estático” , patrón correspondiente a

la ausencia de movimiento. En este caso el movimiento global será estático hasta que se produzca una desactivación, y nada podrá cambiarlo. Empíricamente se constató que los gestos dependientes del movimiento, manotazos en diferentes direcciones y con distintos sentidos, no realizaban ninguna parada en su recorrido. Por ello se tomó esta decisión de fijar el movimiento a estático con una sola salida que así lo determinase.

El movimiendo global (ver figura 4.8) constituye una entrada más a la máquina de estados, complementando la información derivada de la discriminación entre posturas estáticas. Dicha máquina de estados será la encargada de gestionar la detección de DHGs, como veremos en el siguiente apartado.

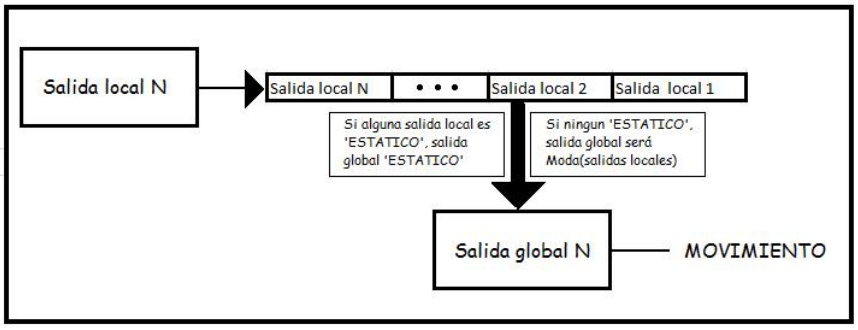

Figure 4.8: Movimiento global.

4.5. Integración de la máquina de estados.

El diseño de la máquina de estados corresponde a trabajo previo realizado fuera del ámbito del trabajo descrito en este documento (ver 3.3.2.5). Partiendo de esta solución, se ha trabajado para exprimir al máximo sus capacidades, intentando alcanzar la configuración óptima en terminos de detección y usabilidad.

Esquemáticamente, el funcionamiento de la máquina de estados se ve resumido en la figura 4.9 . Ésta se puede ver como un módulo diferente que combina la información de SHP detectado y la de movimiento, asi como el propio estado de la máquina dependiendo de lo ocurrido con anterioridad. El objetivo, predecir el DHG correcto, esto es, la salida definitiva del sistema para cada ejecución de un gesto.

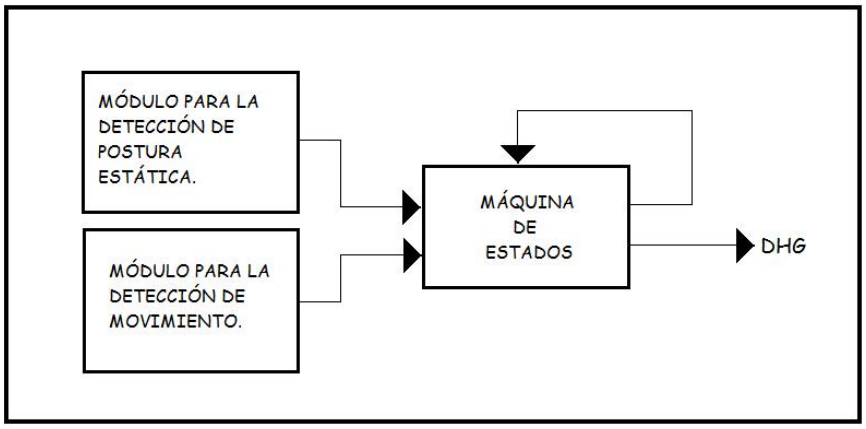

Figure 4.9: Esquema de funcionamiento de la máquina de estados.

Los gestos son definidos en un archivo de configuración, tal como se vio en el apartado 3.3.2.5. Dicho archivo marcará el funcionamiento global del sistema pues en él se definen todos los gestos a detectar, así como las condiciones necesarias para su detección: la postura (o posturas) estática y el patrón de movimiento a cumplir , así como si necesita o no desactivación para la obtención del gesto.

Hay que tener cuidado con la definición de gestos, eliminando la posibilidad de detectar gestos de forma simultánea, lo que podría dar lugar a errores y confusiones.

4.6. Escalabilidad del sistema.

Entendemos por escalabilidad del sistema su capacidad para ampliar el margen de operaciones sin perder calidad. Dicho de otra forma, cuando hablamos de escalibidad del sistema, hablamos de la posibilidad de añadir nuevos gestos al diccionario, así como del grado de complejidad asociado a este proceso. Se plantean tres ámbitos de extensión de capacidades: inclusión de nuevos SHPs, DHGs y/o patrones de movimiento.

La mejora de la escalabilidad del sistema goza de gran importancia. La amplia variedad de gestos que el ser humano es capaz de realizar con sus manos, así como las distintas aplicaciones a controlar hacen necesaria la posibilidad de incluir nuevos gestos sin que sea esta una ardua tarea. Tener

un sistema fácilmente escalable no sólo implica la capacidad de aumentar el número de gestos detectados. También implica la posibilidad de modificar la realización de alguno de los gestos ya incluídos, adaptándose con ello a las necesidades de los usuarios, que son quienes deben verse beneficiados de la usabilidad del sistema.

Hablaremos de tres ambitos de escalabilidad, relacionados con los tres grandes bloques en los que se ve dividida la detección de DHGs: nuevos SHPs, patrones de moviento sintéticos y nuevos DHGs.

4.6.1. Nuevos SHPs.

Un nuevo gesto puede implicar una nueva postura estática que difiera de las ya existentes. En este caso, el primer paso para su detección es la grabación de secuencias con la realización por parte de diferentes usuarios de esa nueva postura. Estas secuencias serán usadas para el entrenamiento de nuevas SVM, y deben ser ejecutadas por diferentes usuarios y en distintas regiones de la zona de activación.

El número de usuarios podría variar, aunque lo recomendable para facilitar el proceso de ampliación del diccionario es que sean los mismos usuarios que han grabado los videos de entrenamiento para el resto de los gestos. Tal y como se explicó en 4.2, donde se describía el proceso de entrenamiento, establecemos el mínimo número de imágenes recomendable en 1200 imágenes del nuevo gesto: 6 usuarios, 10 segundos de grabación por usuario con un frame rate de 20fps.

Una vez grabados los videos, el siguiente paso es el entrenamiento de una SVM por cada SHP que deseemos incluir, para su posterior inclusión en el sistema. Dicho entrenamiento puede realizarse de forma independiente, entrenando sólo la nueva SVM con sus muestras como positivas a la entrada, y el resto como negativas. También pueden reentrenarse todas las SVMs del sistema, de hecho sería lo ideal, incluyendo las muestras del nuevo gesto como entradas negativas para las SVM de los gestos ya existentes. Si no tiene lugar este reentrenamiento completo, sí sería aconsejable al menos reentrenar aquellas SVM que están encargadas de la detección de gestos fácilmente confundibles con el nuevo, con el objetivo de lograr la mayor separación posible entre ambos y evitar errores en el futuro.

Finalmente sólo nos queda incluir la SVM entrenada para detectar y separar la nueva postura de otras, o lo que es lo mismo, la inclusión en el sistema de los nuevos modelos generados. Esto permitirá la detección del SHP con una cierta tasa de acierto siempre que las características de bajo nivel extraídas así lo permitan (ver figura 4.10).

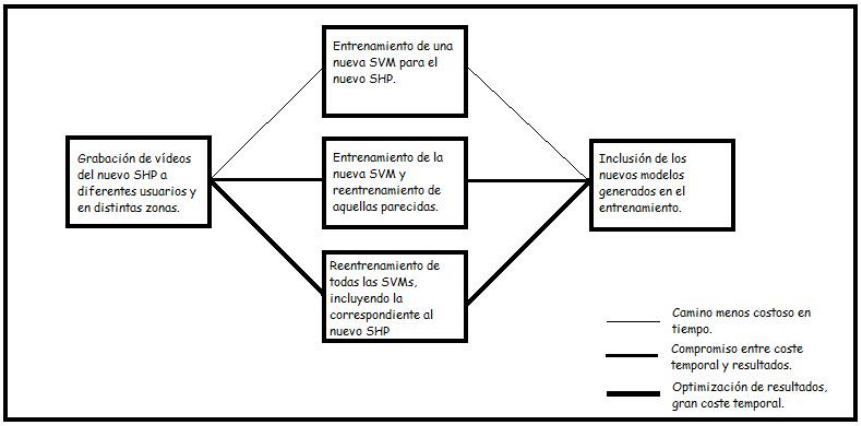

Figure 4.10: Esquema para la aprendizaje de un nuevo SHP.

En última instancia, si con los pasos anteriores no fuese posible detectar el nuevo SHP y éste fuese de importante inclusión en el sistema, se podría modificar la selección y tratamiento de las características que componen el vector de características, asumiendo el riesgo de empeorar los resultados en la detección de otros gestos que esto podría conllevar.

4.6.2. Nuevos patrones de movimiento.

No es la primera vez que se lee en este trabajo que hay gestos que necesitan de la información de movimiento para ser detectados. Muchos de ellos resultan incluso más intuitivos que los que se basan en posturas estáticas, especialmente, y como era de esperar, aquellos que pretenden el desplazamiento de algo en la pantalla: pasar página, desplegar o regoger un menú, avanzar en una determinada dirección, todos ellos son ejemplos de aplicación cuyo equivalente en posturas estaticas gozaría de menor usabilidad.

El estado actual de la detección de patrones de movimiento permite aumentar el número de patrones detectados de una forma muy simple. Basta con añadir al archivo que contiene los patrones sintéticos de movimiento

el nuevo patrón que se quiere detectar, esto es, la evolución numérica de las tres coordenadas de un punto definidas matemáticamente (ver apartado 3.3.2.3). Para que realmente sea detectable lo ideal sería declarar varios patrones para un mismo movimiento cambiando la distancia recorrida por el mismo sin variar el número de frames, siguiendo la línea de detección del movimiento marcada. Esto permitiría detectar movimientos realizados a diferentes velocidades o con diferentes amplitudes.

Dentro del sistema, bastará con asignar un identificador numérico no utilizado con anterioridad al nuevo movimiento para su posterior tratamiento en la máquina de estados. La figura 4.11 incluye un esquema para la detección de un nuevo movimiento.

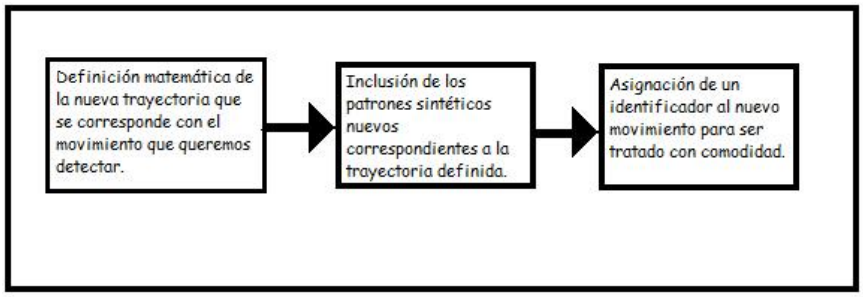

Figure 4.11: Esquema para la detección de un nuevo patrón de movimiento.

4.6.3. Nuevos DHGs.

La definición de un nuevo DHG puede llevar consigo un nuevo movimiento a detectar, un nuevo SHP, o simplemente ser una combinación de lo ya existente. Sea como sea, un paso imprescindible para la detección de un nuevo DHG es su definición en el archivo de configuración de la máquina de estados. Los pasos anteriores descritos en esta sección, la detección de un nuevo SHP, y la detección de un nuevo movimiento, sólo seran necesarias si el nuevo DHG así lo requiere.

Lo mismo sucede si queremos modificar un gesto ya existente, basta con modificar el archivo de configuración adaptándolo a las nuevas necesidades para que el nuevo gesto sea detectable. Esto permite realizar modificaciones adaptando el sistema para solucionar errores que estén teniendo lugar, sin un elevado coste temporal.

Los pasos para la definición de un nuevo DHG se pueden ver en la figura 4.12. Para ello, como ya se ha dicho, hay que tener en cuenta si el nuevo DHG implica una nueva postura y/o un nuevo movimiento, o simplemente una nueva definición basada en posturas y/o movimientos ya existentes.

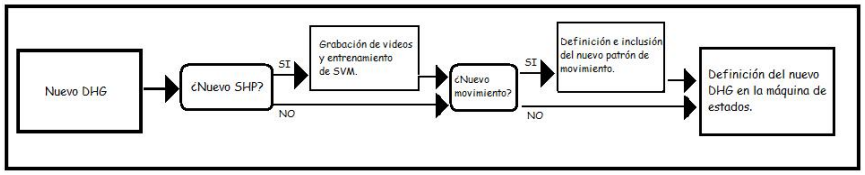

Figure 4.12: Pasos para la inclusión de un nuevo DHG.

4.7. Ampliación del diccionario de gestos.

La evolución y mejoras, especialmente en la detección de patrones de movimiento permite un aumento considerable en el número de gestos detectables, llevando a un sistema mucho más intuitivo para el usuario. Más concretamente, se abre un abanico de nuevos gestos cuya principal característica es su patrón de movimiento. Desplazamientos en la pantalla, abrir o cerrar un menú, pasar página o volver atrás, son algunas de las órdenes que podrán verse representadas de una manera más intuitiva por estos nuevos gestos.

La colección de gestos inicial, que puede verse en el apartado 2.6, se ve ampliada con los gestos que se detallan a continuación.

En lo relativo al movimiento, gracias a la evolución en la detección de patrones se han incluído en el sistema los siguientes gestos:

En primer lugar, aprovechando los 6 nuevos movimientos detectables, se define un gesto para cada uno de ellos. Estos son: PageRight (figura 4.13), PageLeft (figura 4.14), SlapUpRight (figura 4.15), SlapDownRight (figura 4.16), SlapUpLeft (figura 4.17) y SlapDownLeft (figura 4.18).

Figure 4.13: DHG PageRight

Figure 4.14: DHG PageLeft

Figure 4.15: DHG SlapUpRight

Figure 4.16: DHG SlapDownRight

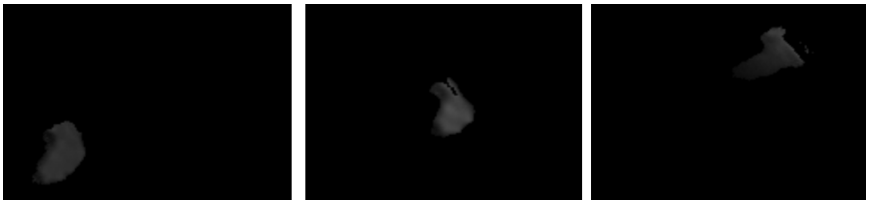

Figure 4.17: DHG SlapUpLeft

Figure 4.18: DHG SlapDownLeft

Además, se incluye un nuevo gesto complejo que se asemeja al click de un ratón. Consiste en la realización del SHP EnumOne seguido del SHP Fist. Su ejecución llevada a cabo se puede ver en la figura 4.19.

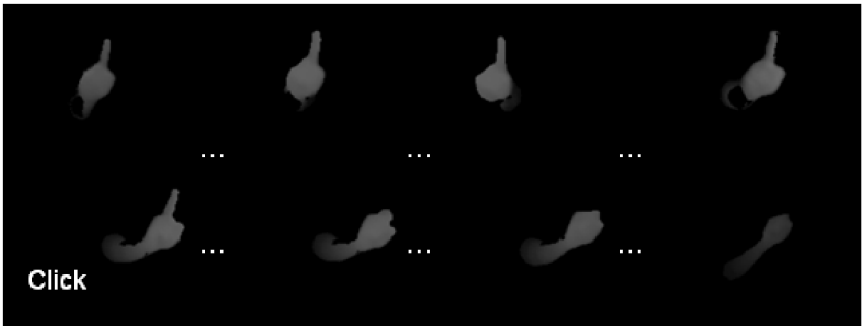

Figure 4.19: DHG Click

4.7.1. Diccionario de gestos.

Finalmente, el diccionario de gestos queda como vemos a continuación, La tabla 4.1 muestra todos los gestos detectables por el sistema, incluyendo la información necesaria para su detección.

DHG<id>	Gesto	SHP<id>	Movimiento	Transiciones
1	G_Select	1	Estático	EnumOne-EnumOne
2	G_EnumTwo	2	Estático	
3	G_EnumThree	3	Estático	
4	G_EnumFour	4	Estático	
5	G_EnumFive	5	Estático	EnumFive-EnumFive
6	G_Ok	6	Estático	Ok-Ok
7	G_MenuOpen	Cualquiera	Up	
8	G_MenuClose	Cualquiera	Down	
9-14	G_Take&Release	5-8-5	Irrelevante	EnumFive-Fist-EnumFive
10	G_MenuLeft	11	Estático	
11	G_MenuRight	12	Estático	
12	G_Cancel	7	Estático	
13	G_Click	1-8	Irrelevante	EnumOne-Fist
15	G_PageLeft	Cualquiera	Left	
16	G_PageRight	Cualquiera	Right	
17	G_SlapUpRight	Cualquiera	Up-Right	
18	G_SlapDownRight	Cualquiera	Down-Right	
19	G_SlapUpLeft	Cualquiera	Up-Left	
20	G_SlapDownLeft	Cualquiera	Down-Left	

Table 4.1: Diccionario DHGs

Capítulo 5

Pruebas y resultados

5.1. Colecciones de vídeos.

Para la evaluación y obtención de resultados se dispone de distintas colecciones de videos que se detallan a continuación. Primero es necesaria una diferenciación entre vídeos grabados para entrenamiento y vídeos grabados para evaluación. En los primeros, cada usuario realiza un SHP durante un determinado tiempo, sin cambiar la postura estática de la mano, pero moviéndola a través de la zona de interacción. Los vídeos grabados para evaluación, por el contrario, son vídeos que recogen ejecuciones completas que un usuario podria realizar para el control de un sistema, incluyendo esto entrada y salida de la zona de interacción (i.e. activación y desactivación).

Hay tres colecciones diferentes de vídeos. La primera de ellas, grabada en las oficinas de TID-Barcelona en Marzo de 2009 con la cámara 3DV. La segunda, grabada en la UAM en Octubre de 2009, también con la 3DV. Finalmente, la última grabación hasta el momento, tuvo lugar en Marzo de 2010, también en la UAM pero en esta ocasión con la cámara SR4000 de Mesa Imaging.

Vayamos al detalle en cuanto a los vídeos grabados, empezando por la primera colección de vídeos. Esta colección fue la utilizada en el trabajo previo descrito en el apartado 3.2 . En ella participaron 6 usuarios, los cuales realizaron cada postura estática durante 10 segundos lo que, teniendo en cuenta la tasa de imágenes de la cámara 3DV (20 fps), da lugar a 1.200 frames por cada SHP. Dado que había un total de 12 SHPs diferentes, esto hace un total de 14.400 muestras para entrenamiento. Por otro lado, estos

mismos usuarios realizaron un video con la grabación de cada DHG, lo cuál sería utilizado para la evaluación del sistema.

En la segunda colección se buscaba tener un mayor número de usuarios, asi como un mayor número de vídeos para la evaluación del sistema con el fin de obtener valores estadísticamente más significativos. En esta grabación tomaron parte 13 usuarios. El proceso de grabación de los vídeos para entrenamiento no sufrió alteración, y los usuarios realizaron cada SHP durante 10 segundos. Nuevamente, 200 frames por usuario y SHP. La grabación de DHGs se vió alterada y en lugar de grabar un único vídeo por usuario y DHG, se capturaron 7 secuencias por cada gesto y usuario. De estas 7 secuencias no todas eran válidas, porque en la grabación el usuario a veces entraba y salía accidentalmente de la zona de interacción dando lugar a vídeos erróneos. Se fijó que como mínimo cada usuario tuviese 2 vídeos por cada DHG, siendo 7 el número máximo, ya comentado.

La tercera, y por el momento última colección de vídeos, consta del mismo número de grabaciones por cada usuario y DHG. En ella participaron 14 usuarios. Las capturas para entrenamiento no sufrieron cambios notables, lo único que varió es la duración de los vídeos. Dado que estas grabaciones se produjeron con la cámara SR4000, que tiene un frame-rate superior a la 3DV, la duración de los vídeos se redujo con el objetivo de mantener constante el número de frames (200 frames por cada usuario y SHP). Para la grabación de DHGs cada usuario realizó 5 veces cada gesto, dando lugar a un total de 70 vídeos de evaluación por cada gesto, y haciendo con ello la evaluación más significativa, al menos en cuanto a número de pruebas. Esta última colección se encuentra accesible en:

http://dymas.ii.uam.es/~vision/intranet/PaginaVideos/pruebaJavi2/GESTUAL_V1.html

5.2. Obtención de resultados.

En este apartado se explican los resultados obtenidos, y es básico para entender de dónde vienen todas las cifras que pueblan las tablas recopiladas.

Hablaremos de resultados de entrenamiento y de evaluación del sistema. Los primeros, se corresponden con una estimación de separabilidad entre los distintos SHPs, mientras que la evaluación del sistema nos da una idea de la capacidad para detectar DHGs, lo que ilustrará la calidad del sistema.

5.2.1. Entrenamiento.

Los resultados de entrenamiento tienen su base en los *true-positives, true-negatives, false-positives & flase-negatives* que se consiguen tras entrenar cada SVM evaluando mediante validación cruzada la F-Score descrita en la sección 3.2.3.2. La validación cruzada nos permite averiguar los valores óptimos de los parámetros del kernel , C y γ , tal y como se describe en la sección 3.2.3.2.

5.2.2. Evaluación del sistema.

Para evaluar el funcionamiento del sistema, este es probado con los vídeos grabados por distintos usuarios realizando los distintos DHGs, los resultados obtenidos son recopilados en una matriz de confusión que nos da una idea de los errores que se estan produciendo, permitiendo apuntar posibles futuras líneas de trabajo.

5.3. Cambios sobre el sistema inicial.

En este apartado veremos los resultados obtenidos para los gestos comentados en el trabajo previo a este proyecto y para los nuevos gestos, cuantificando y analizando los resultados obtenidos. Estas mejoras se analizarán en dos niveles: por un lado las ocasionadas por la selección y tratamiento de las características de la mano, esto es, la composición del vector de características; por otro, las mejoras derivadas de la detección de patrones de movimiento. Éstas últimas sólo pueden cuantificarse en base a los resultados en la detección de G_MenuOpen y G_MenuClose, pues son los únicos gestos presentes en el sistema inicial con información de movimiento, luego los únicos con los que la actual versión es comparable a este nível.

5.3.1. Selección y tratamiento de las características de la mano.

A continuación se muestran los resultados según la evolución seguida hasta alcanzar la composición final del vector de características, explicando en cada uno de los pasos cuáles eran los objetivos, evaluando después su consecución.

Este proceso es desarrollado utilizando la colección de vídeos inicial, grabada por 6 usuarios en Telefónica con la cámara 3DV (ver sección 5.1).

En las tablas de resultados, el valor que aparece es el de la F-score, medida utilizada como objeto de mejora (ver apartado 3.2.3.2), así como los porcentajes de mejora para facilitar la interpretación de las cifras obtenidas.

El punto de partida, refiriéndonos al vector de características, es el definido en el trabajo previo, sección 3.2. Recordar que dicho vector estaba compuesto por 19 coordenadas: número de máximos o protuberancias (Nmax); Intensidad(In), amplitud(An) e inclinación(Angn) de cada una de ellas; Dimensiones de la elipse(DimMa y DimMe) y ángulo de inclinación(AngE). Notar también, que en caso de ausencia de alguno de los máximos el relleno de los parámetros derivados de él se realizaba con valor '-1'.

5.3.1.1. Normalización no generalista.

La primera labor afrontada para la realización de este trabajo fue el intento de normalización descrito en el apartado 3.3.1.1. El objetivo era realizar una normalización no generalista (entendemos por generalista, un normalización intracoordenada clásica: varianza 1, media 0) que mantuviese la forma de la mano para una posible reconstrucción. La figura 5.1 permite una comparación de la mano antes de la normalización, la mano tras la normalización generalista y tras la normalización manteniendo las proporciones. En la figura se representa la elipse centrada en el origen y manteniendo su ángulo de inclinación, y los dedos se representan como triángulos con su amplitud por base y su intensidad por altura. Mantienen también el ángulo, sin normalizar, y tienen el centro de la base situado en el orígen.

El objetivo de conservar la proporción y forma de la mano intactas se consigue, pero los resultados tras esta normalización no mejoraron los resultados de una forma significativa, tal y como se muestra en la tabla 5.1.

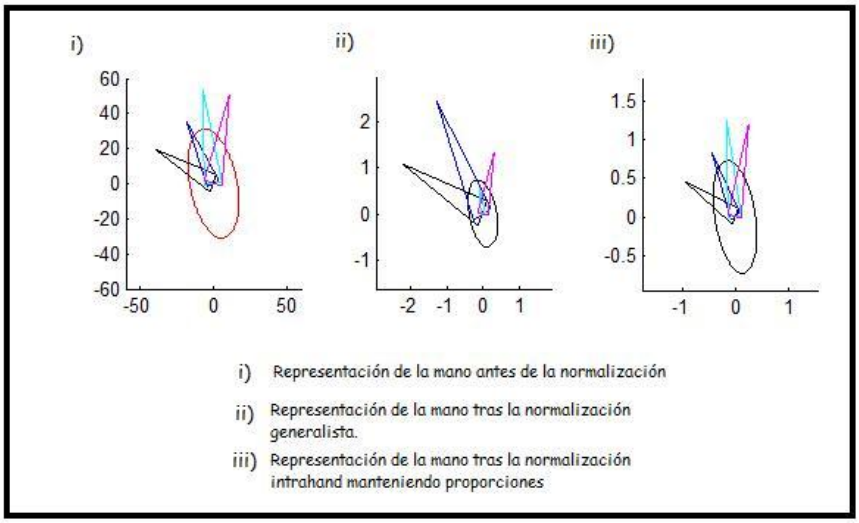

Figure 5.1: Representación de la mano antes y después de la normalización.

Gesture\Version	Marzo2009	Normalizaión intramano	Mejora (%)
EnumOne	0.82101	0.81105	-1.21
EnumTwo	0.97369	0.98037	0.69
EnumThree	0.95871	0.94915	-1.00
EnumFour	0.94677	0.93855	-0.81
EnumFive	0.96699	0.95789	-0.94
Ok	0.63327	0.62758	-0.90
Stop	0.92964	0.94696	1.86
Fist	0.51679	0.52629	1.84
OpenHandUp	0.3074	0.23419	-23.82
OpenHandDown	0.33317	0.3541	6.28
OkLeft	0.8719	0.79324	-9.02
OkRigth	0.84474	0.79115	-6.34

Table 5.1: Separación SHPs tras normalización intrahand.

El siguiente planteamiento, manteniendo la normalización "intrahand", fue cuantificar el ángulo principal, definido como el ángulo de la protuberancia situada en el centro si el número de protuberancias es impar, y el valor medio entre los dos centrales en caso de ser par. Este ángulo se cuantificaba a un múltiplo de 90^{o} dando lugar a cuatro posibles orientaciones. El desplazamiento provocado en este ángulo se le aplicaba al resto de ángulos manteniendo intacta la forma de la mano. Con esto se fija la orientación de la mano pretendiendo una mayor similitud entre gestos iguales con distinto

ángulo de ejecución. Los resultados obtenidos se muestran en la tabla 5.2.

Gesture\Version	Marzo2009	N. intramano + Cuantificación	Mejora (%)
EnumOne	0.82101	0.83138	1.26
EnumTwo	0.97369	0.98283	0.94
EnumThree	0.95871	0.94542	-1.39
EnumFour	0.94677	0.91308	-3.55
EnumFive	0.96699	0.95833	-0.90
Ok	0.63327	0.66159	4.47
Stop	0.92964	0.89504	-3.72
Fist	0.51679	0.51892	0.41
OpenHandUp	0.3074	0.2105	-31.52
OpenHandDown	0.33317	0.34043	2.18
OkLeft	0.8719	0.77445	-11.18
OkRigth	0.84474	0.79339	-6.08

Table 5.2: Normalización intrahand y cuantificación del ángulo principal.

Observando los resultados, aquellos gestos que presentaban peores condiciones para su detección eran los que no tenían protuberancias, por lo que se decidió aumentar la información relativa a la mano en cualquiera de sus posturas. Para ello se añadió una coordenada más al vector de características, la distancia entre el punto más cercano a la cámara, Zmin, y el centro de la elipse, CoGE. Nótese que estos puntos estan presentes en cualquier postura estática, no siendo como los máximos, que pueden o no aparecer. Esto mejoró los valores de F-score en las posturas deseadas, aquellas sin protuberancias, como muestra la tabla 5.3.

Gesture\Version	Marzo2009	Distancia Zmin-CoGE	Mejora (%)
EnumOne	0.82101	0.78659	-4.19
EnumTwo	0.97369	0.97769	0.41
EnumThree	0.95871	0.94629	-1.30
EnumFour	0.94677	0.89792	-5.15
EnumFive	0.96699	0.95753	-0.98
Ok	0.63327	0.62138	-1.88
Stop	0.92964	0.87222	-6.18
Fist	0.51679	0.75147	45.41
OpenHandUp	0.3074	0.37523	22.07
OpenHandDown	0.33317	0.40304	20.97
OkLeft	0.8719	0.81847	-6.13
OkRigth	0.84474	0.82339	-2.53

Table 5.3: Resultado tras añadir la distancia entre Zmin y CoGE.

Los resultados en la separación de posturas sin protuberancias mejoraron notablemente, convirtiendo 'Fist' en una postura más detectable, por lo que se decidió seguir esta línea de aumentar la información relativa a la mano, tal y como veremos en el siguiente apartado.

5.3.1.2. Triángulo de profundidad.

Denominaremos triángulo de pronfundidad al que tiene por vértices tres puntos significativos resultantes de la segmentación y extracción de características de la mano: Zmin y CoGE, cuya distancia ya se incluyó en la anterior prueba, y además, se añade el centro de gravedad de la mano, CoG.

Esta nueva característica se introduce en el vector con dos de sus lados y un ángulo, es decir, un total de 3 coordenadas. Los lados incluídos son las distancias entre Zmin y los otros dos puntos, y el ángulo correspondiente a ese vértice, el de Zmin. Además, para esta prueba se volvió a una normalización generalista, olvidando la anterior que presentaba un mayor coste computacional y no mejoraba los resultados (ver resultados en la tabla 5.4).

Gesture\Version	Marzo2009	Triángulo de profundidad	Mejora(%)
EnumOne	0.82101	0.71102	-13.40
EnumTwo	0.97369	0.97187	-0.19
EnumThree	0.95871	0.96515	0.67
EnumFour	0.94677	0.94635	-0.04
EnumFive	0.96699	0.96013	-0.71
Ok	0.63327	0.52833	-16.57
Stop	0.92964	0.80338	-13.58
Fist	0.51679	0.78102	51.13
OpenHandUp	0.3074	0.47131	53.32
OpenHandDown	0.33317	0.48994	47.05
OkLeft	0.8719	0.86674	-0.59
OkRigth	0.84474	0.80301	-4.94

Table 5.4: Resultado tras añadir el triángulo de profundidad.

Los resultados a la vista están, continúa mejorando la discriminación entre gestos sin protuberancias, pero empeora notablemente, lo que se convierte en un problema, la separación de gestos conflictivos como 'EnumOne' y 'Ok'.

5.3.1.3. Robustez frente a la distancia.

Con la idea de que 'EnumOne' y 'Ok' son posturas difíciles de separar se decide abrir una nueva línea de trabajo con el objetivo de mejorar sus resultados sin empeorar los anteriormente mejorados. El objetivo queda centrado en dotar al sistema de independencia frente a la distancia, entendiendo que las proporciones de la protuberancia en ambos gestos ha de ser lo bastante diferente como para separarlos con cierta seguridad. El modo de conseguir esta robustez está detallado en el apartado 3.3.1.2, y consiste básicamente en sustituir aquellas características del vector correspondientes a distancias en la imagen, por otras que contengan la misma información y además estén dotadas de independencia frente a la distancia a la cámara. Los resultados de este cambio se muestran en la tabla 5.5.

Gesture\Version	Marzo2009	Robustez distancia	Mejora (%)
EnumOne	0.82101	0.82327	0.28
EnumTwo	0.97369	0.99012	1.69
EnumThree	0.95871	0.96644	0.81
EnumFour	0.94677	0.94312	-0.39
EnumFive	0.96699	0.95289	-1.46
Ok	0.63327	0.69921	10.41
Stop	0.92964	0.94918	2.10
Fist	0.51679	0.77957	50.85
OpenHandUp	0.3074	0.50694	64.91
OpenHandDown	0.33317	0.50039	50.19
OkLeft	0.8719	0.82638	-5.22
OkRigth	0.84474	0.83786	-0.81

Table 5.5: Resultado tras dotar al sistema de robustez frente a la distancia.

Los resultados mejoran notablemente, por lo que el cambio fue aceptado y la robustez frente a la distancia se sumó al triángulo de profundidad como mejoras que seguirían presentes en siguientes versiones.

5.3.1.4. Robustez frente al giro.

Con la idea de seguir aumentando la robustez del sistema se decidió cambiar la forma de tratar los ángulos para dotar al sistema de relativa independencia frente al giro (ver apartado 3.3.1.2.b)). Los resultados, presentados en la tabla 5.6, no mejoraron en demasía con respecto a los anteriores , pero el cambio se aceptó ante la idea de que podría mejorar la usabilidad y posibilidades en la realización de los gestos.

Gesture\Version	Marzo2009	Robustez giro	Mejora (%)
EnumOne	0.82101	0.8325	1.40
EnumTwo	0.97369	0.96717	-0.67
EnumThree	0.95871	0.95516	-0.37
EnumFour	0.94677	0.92923	-1.85
EnumFive	0.96699	0.96368	-0.34
Ok	0.63327	0.65909	4.08
Stop	0.92964	0.94167	1.29
Fist	0.51679	0.78961	52.79
OpenHandUp	0.3074	0.51154	66.41
OpenHandDown	0.33317	0.50883	52.72
OkLeft	0.8719	0.79879	-8.39
OkRigth	0.84474	0.81034	-4.07

Table 5.6: Resultado tras dotar al sistema de robustez frente al giro.

El vector utilizado finalmente para la separación de SHPs queda compuesto por las características presentes tras este último cambio. Podemos ver las características finales en el apartado 3.3.1.2.d).

5.3.1.5. Refuerzo de la primera protuberancia.

Observando la tabla de resultados conseguida (5.6), nos encontramos con que han empeorado los resultados en la separación de los SHPs 'OkLeft' y 'OkRight'. Esto, unido a que otro de los problemas sigue siendo diferenciar entre 'EnumOne' y 'Ok', llevó a incluir más información acerca de la primera protuberancia.

En primer lugar, y en base a los buenos resultados tras la inclusión del triángulo de profundidad (ver apartado 5.3.1.2), éste se amplía a tres coordenadas en lugar de dos. La coordenada correspondiente a la relación entre dos de los lados de este triángulo, se divide en las dos distancias, multiplicadas por su distancia a la cámara, para mantener la independencia con la misma.

Además se añaden nuevas coordenadas, que son:

- Pendiente de la base de la primera protuberancia.
- Triángulo de altura. El formado por los puntos Max1(primer máximo), punto más a la derecha de la elipse, y punto más a la izquierda. Se

incluye de forma análoga al triángulo de profundidad, dos de sus lados y un ángulo.

Los resultados tras estas inclusiones se pueden ver en la tabla 5.7, y puede apreciarse como mejora ligeramente la separación de los SHPs con una sola protuberancia.

Gesture\Version	Marzo2009	Refuerzo P1	Mejora (%)
EnumOne	0.82101	0.84514	2.94
EnumTwo	0.97369	0.96515	-0.88
EnumThree	0.95871	0.955	-0.39
EnumFour	0.94677	0.92703	-2.08
EnumFive	0.96699	0.96217	-0.49
Ok	0.63327	0.68844	8.71
Stop	0.92964	0.94183	1.31
Fist	0.51679	0.80559	55.88
OpenHandUp	0.3074	0.55835	81.63
OpenHandDown	0.33317	0.57182	71.63
OkLeft	0.8719	0.82637	-5.22
OkRigth	0.84474	0.84522	0.06

Table 5.7: Resultado tras refuerzo de la primera protuberancia.

Esta versión no llegó a incluirse en el proyecto global.

5.3.1.6. Resultados en la detección de DHGs-Evaluación del Sistema

La detección de DHGs es algo que no sólo depende de la correcta separación de SHPs, sino también de la correcta detección de movimiento ejecutado y de la gestión que de esta información lleva a cabo la máquina de estados. En este apartado sólo se tendrá en cuenta aquellos gestos cuya detección dependa de la postura estática.

Inicialmente, para esta evaluación se contaba con muy pocos vídeos, seis usuarios grabaron una vez cada gesto, lo que da un total de 66 vídeos. Las tablas 5.8 y 5.9 muestran los resultados para la versión 'Marzo2009' y tras los cambios introducidos en la detección de SHPs. Para ello sólo se han tenido en cuenta aquellos DHGs cuya detección depende del SHP ejecutado. La tasa de error en ambos casos es muy baja, pero no debe dársele demasiada relevancia

debido a la escasez de vídeos disponibles. Además, hay que tener en cuenta que los usuarios que realizaron la grabación de los vídeos para evaluación son los mismos que grabaron los vídeos de entrenamiento, haciendo mucho más separables las posturas estáticas.

DHG(SHP)\User	User 1	User 2	User 3	User 4	User 5	User 6
Select (1)	1	1	1	1	1	1
EnumTwo (2)	2	2	2	2	2	2
EnumThree (3)	3	3	3	3	2	3
EnumFour (4)	4	4	4	4	4	4
EnumFive (5)	5	5	5	5	5	5
Accept (6)	6	6	6	1	6	1
MenuLeft (11)	11	11	11	11	11	11
MenuRight (12)	12	12	12	12	12	12

Table 5.8: SHP resultante en la ejecución de DHGs. Marzo2009

DHG(SHP)\User	User 1	User 2	User 3	User 4	User 5	User 6
Select (1)	1	1	1	1	6	1
EnumTwo (2)	2	2	2	2	2	2
EnumThree (3)	3	3	3	3	2	3
EnumFour (4)	4	4	4	4	4	4
EnumFive (5)	5	5	5	5	5	5
Accept (6)	6	6	6	1	6	6
MenuLeft (11)	11	11	11	11	11	11
MenuRight (12)	12	12	12	12	12	12

Table 5.9: SHP resultante en la ejecución de DHGs tras los cambios.

Durante el desarrollo de las pruebas, uno de los problemas constatados, es que los DHGs 'MenuLeft' y 'MenuRight' (ver figura 5.2) no se detectaban correctamente cuando se realizaban en la parte superior de la pantalla (ver figura 5.3). Este problema debería tener solución, en la medida de lo posible, con la nueva selección de características en la que se pretendía dotar al sistema de indepencencia frente a la distancia y la posición.

A continuación podemos ver en la tabla 5.10 una comparación de los resultados de detección de estos gestos, antes y después de los cambios en la detección de posturas estáticas.

Se aprecia una mejora en la detección de estos gestos, que gozan de particular

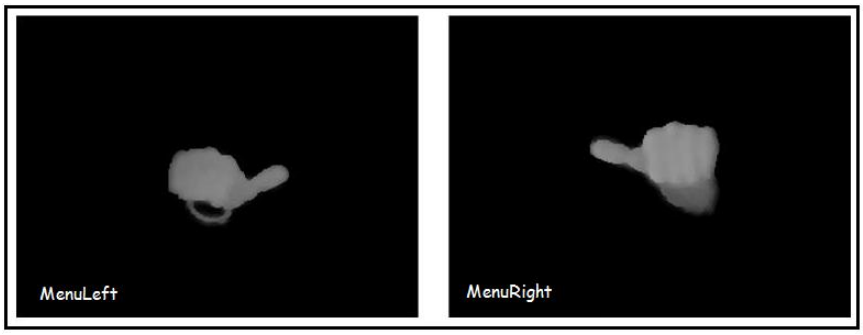

Figure 5.2: DHGs 'MenuLeft' y 'MenuRight'

Figure 5.3: 'MenuRight' realizado en la parte superior de la pantalla.

dificultad dado que su realización provoca que el antebrazo sea interpretado como una parte más de la mano. Esto provoca un aumento en el tamaño de la elipse y dificulta su clasificación. Dado que el sistema solo se ha evaluado hasta el momento con 6 vídeos por cada DHG, no se pueden sacar conclusiones al respecto.

A continuación se exponen los resultados en la detección de DHGs dependientes del SHP detectado, utilizando para ello los videos grabados en Septiembre de 2009 en la UAM. La tabla 5.11 corresponde a la ejecución del sistema inicial con los vídeos ya comentados. Destacar la dificultad para diferenciar entre los gestos "Select" y "Accept" pues con las características obtenidas son dos gestos practicamente iguales. Además, aparecen muchos falsos positivos en el caso de "EnumThree" y la detección del gesto "Cancel" no se termina de conseguir.

A continuación, en la tabla 5.12 podemos ver reflejado el funcionamiento del sistema tras los cambios introducidos. Diferenciar los gestos "Accept" y "Select" sigue siendo un claro problema. Nótese que la posibilidad de detectar

Gesto en el video	Predicción Marzo 2009	Predicción tras los cambios
OkRight	Cancel	OkRight
OkRight	Cancel	OkRight
OkLeft	MenuOpen	MenuOpen
OkLeft	Cancel	Cancel
OkRight	MenuOpen	Accept
OkRight	MenuOpen	OkRight
OkLeft	OkLeft	OkLeft
OkLeft	Select	OkLeft
OkRight	OkRight	OkRight
OkRight	OkRight	OkRight
OkLeft	MenuOpen	MenuOpen
OkRight	Cancel	Cancel
OkLeft	MenuOpen	MenuOpen
OkRight	MenuOpen	OkRight

Table 5.10: Resultados en la detección de "Oks altos".

nuevos movimientos así como el nuevo gesto compuesto "Click", provoca la aparición de falsos positivos más variados que antes. Este último en concreto, provoca muchos falsos positivos cuando el gesto realizado es "Select" o "Accept", ambos aceptados como primera parte del gesto "Click", seguidos por "Fist", gesto fácilmente detectable de forma equivocada en el abandono de la zona de interacción y en las perdidas accidentales de protuberancia debido a la limitaciones del sistema. Algo similar sucede con el gesto "Take", en el que encontramos falsos positivos cuando la entrada es "EnumFive", o incluso "EnumFour". Además, mejora notablemente la detección de "Cancel", con respecto a la versión inicial.

5.3.2. Mejoras por la detección de patrones de movimiento.

Recordando la sección 3.3.2, la detección de movimiento realizada en el trabajo previo es modificada por completo. Del estudio de la evolución de una coordenada umbralizando después la pendiente de su movimiento y añadiendo alguna condición, se pasa a la comparación con unos patrones sintéticos definidos, determinando con ello el movimiento seguido por la mano.

Esta evolución no solo permite ampliar el número de movimientos que el sistema es capaz de detectar de una forma muy sencilla, sino que además

mejora, como veremos en este apartado, los resultados en los patrones de movimiento que ya se detectaban.

Antes de analizar los resultados obtenidos en la detección de los nuevos patrones de movimiento que constituyen una de las más importantes mejoras de este trabajo, comenzaremos por realizar una comparación entre la detección de los gestos MenuOpen y MenuClose en la versión inicial (Marzo 2009), y en la versión tras la inclusión de los cambios en este aspecto. En las tablas 5.13 y 5.14 podemos ver el gesto detectado por el sistema para los vídeos grabados por 6 usuarios en Telefónica en Marzo de 2009. Se aprecia como mejoran los resultados, pero apenas se pueden sacar conclusiones por la pequeña cantidad de vídeos usados.

Realizando un análisis más exhaustivo, tiene lugar una evaluación con ambas versiones de los videos grabados en la UAM en septiembre de 2009. Los resultados pueden verse en las matrices de confusión mostradas en las tablas 5.15 y 5.16. En este caso es mucho más evidente la mejora en la detección de estos gestos, con casi ausencia de errores en el caso de la nueva estrategia de estimación de movimiento. En cambio, en la versión inicial los gestos son detectados en muy pocas ocasiones.

La mejora en la detección de gestos dependientes de movimiento se planteaba como uno de los objetivos principales de este proyecto y los resultados no dejan duda acerca del aumento en la tasa de acierto al detectar estos gestos.

5.4. Resultados obtenidos en la detección de nuevos gestos.

Hasta el momento se ha realizado ya una comparación entre la detección de gestos previa a este proyecto y posterior a su realización, tanto en gestos dependientes de la postura estática como en los dependientes del movimiento.

A continuación se evalúan y comparan tanto el sistema inicial, como el resultado tras incluir los cambios en la detección de posturas estáticas y en la estimación de movimiento.

En primer lugar, veremos las matrices de confusión correspondientes a la completa evaluación de ambas versiones del sistema, con los videos de Sep-

tiembre de 2009. La tabla 5.17 incluye los resultados de la ejecución de dichos vídeos en el sistema inicial.

La tabla 5.18 contiene la matriz de confusión correspondiente a la evaluación del sistema tras los cambios.

Finalmente, en lo relativo a los vídeos de Septiembre de 2009, una comparativa de ambos sistemas, reflejada en la tabla5.19.

Con la grabación de nuevos vídeos, y la inclusión de una nueva cámara en el proyecto, la SR4000 (ver apartado 3.1.1), surge la posibilidad de realizar nuevas pruebas y ver cómo se adapta el sistema al hecho de tener vídeos para entrenamiento grabados con una cámara, y vídeos para la evaluación grabados con otra diferente. Concretamente, son dos las evaluaciones realizadas. Una primera, en la que los usuarios de entrenamiento siguen siendo los 6 iniciales, evaluando los resultados obtenidos con todos los vídeos grabados en las dos últimas colecciones. Estos resultados pueden verse en la tabla 5.20. Para la otra evaluación, se eligen tres usuarios de entrenamiento y vídeos grabados con la cámara SR4000. Los resultados de esta prueba están plasmados en la tabla 5.21. Los resultados en ambas tablas son aceptables, sobre todo teniendo en cuenta la variedad de los usuarios así como el uso de diferentes cámaras para la grabación de los vídeos.

Input\Output	Select	Enum2	Enum3	Enum4	Enum5	Accept	MenuOpen	MenuClose	OkLeft	OkRight	Cancel
Select	**42**	0	0	0	0	43	0	0	0	0	0
EnumTwo	3	**74**	3	0	0	1	0	5	0	0	0
EnumThree	0	2	**75**	2	0	0	0	0	0	0	0
EnumFour	1	0	9	**76**	0	0	0	0	0	0	1
EnumFive	0	0	5	6	**76**	0	0	0	0	0	1
Accept	6	0	0	0	1	**81**	0	0	2	0	0
OkLeft	0	0	1	0	0	0	2	0	**59**	18	8
OkRight	0	0	3	0	0	0	0	0	0	**78**	9
Cancel	7	0	11	1	0	5	4	14	2	0	**35**

Table 5.11: Detección de gestos estáticos. Versión inicial con vídeos Septiembre 2009.

Input\Output	-	Select	E_2	E_3	E_4	E_5	Accept	M_Open	M_Close	Take	OkLeft	OkRIght	Cancel	Click	P_Left	P_Right	SD_Right	SU_Left
Select	0	**39**	1	1	0	0	7	0	2	0	0	0	0	35	0	0	0	0
EnumTwo	0	0	**81**	0	1	0	0	2	0	0	0	0	0	0	0	2	0	0
EnumThree	0	0	1	**73**	3	1	0	0	0	0	0	0	0	0	0	1	0	0
EnumFour	0	0	0	2	**73**	5	0	0	2	4	0	0	0	1	0	0	0	0
EnumFive	1	0	0	0	8	**54**	0	4	0	19	0	0	1	0	1	0	0	0
Accept	1	26	0	0	0	0	**40**	0	1	0	4	0	0	16	0	0	2	0
MenuLeft	0	0	0	0	0	0	1	5	1	0	**79**	2	0	1	0	0	0	0
MenuRigth	0	0	0	0	0	1	1	2	1	1	0	**79**	0	0	0	4	0	1
Cancel	0	0	0	1	0	1	1	5	1	0	7	0	**58**	2	2	0	0	1

Table 5.12: Detección de gestos estáticos. Nuevo vector de características. Vídeos Septiembre 2009.

DHG\User	User1	User2	User3	User4	User5	User6
MenuOpen	MenuOpen	MenuOpen	MenuOpen	—	OkRight	MenuOpen
MenuClose	—	MenuClose	MenuClose	MenuClose	Cancel	MenuClose

Table 5.13: Detección de gestos MenuOpen y MenuClose versión inicial con vídeos Marzo 2009.

DHG\User	User1	User2	User3	User4	User5	User6
MenuOpen	MenuOpen	MenuOpen	MenuOpen	OkRight	OkRight	MenuOpen
MenuClose	MenuClose	MenuClose	MenuClose	MenuClose	MenuClose	MenuClose

Table 5.14: Detección de gestos MenuOpen y MenuClose versión tras cambios con vídeos Marzo 2009.

Input\Output	Select	EnumThree	EnumFive	Accept	MenuOpen	MenuClose	OkRight
MenuOpen	0	9	0	18	**37**	0	5
MenuClose	1	10	2	14	24	**27**	0

Table 5.15: Detección de gestos MenuOpen y MenuClose versión inicial con vídeos Septiembre 2009.

Input\Output	Unknown	MenuOpen	MenuClose	SlapDownRight	SlapUpLeft
MenuOpen	1	**65**	0	0	3
MenuClose	0	0	**77**	1	0

Table 5.16: Detección de gestos MenuOpen y MenuClose versión tras cambios con vídeos Septiembre 2009.

DHG/id	0	1	2	3	4	5	6	7	8	9	10	11	12	RECALL:
G_Unknown/0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
G_Select/1	0	**42**	0	0	0	0	43	0	0	0	0	0	0	0.49
G_EnumTwo/2	0	3	**74**	3	0	0	1	0	5	0	0	0	0	0.86
G_EnumThree/3	0	0	2	**75**	2	0	0	0	0	0	0	0	0	0.95
G_EnumFour/4	0	1	0	9	**76**	0	0	0	0	0	0	0	1	0.87
G_EnumFive/5	0	0	0	5	6	**76**	0	0	0	0	0	0	1	0.86
G_Accept/6	0	6	0	0	0	1	**81**	0	0	0	2	0	0	0.90
G_MenuOpen/7	0	0	0	9	0	0	18	**37**	0	0	0	5	0	0.54
G_MenuClose/8	0	1	0	10	0	2	14	24	**27**	0	0	0	0	0.35
G_Take/9	0	0	0	7	2	44	0	2	6	**1**	2	1	0	0.02
G_MenuLeft/10	0	0	0	1	0	0	0	2	0	0	**59**	18	8	0.67
G_MenuRigth/11	0	0	0	3	0	0	0	0	0	0	0	**78**	9	0.87
G_Cancel/12	0	7	0	11	1	0	5	4	14	0	2	0	**35**	0.44
PRECISION:	-	0.70	0.97	0.56	0.87	0.62	0.50	0.54	0.52	1.00	0.91	0.76	0.65	

Table 5.17: Evaluación sistema inicial. Videos Septiembre 2009.

DHG/id	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	RECALL:
G_Unknown/0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
G_Select/1	0	**39**	1	1	0	0	7	0	2	0	0	0	0	35	0	0	0	0	0	0	0.46
G_EnumTwo/2	0	0	**81**	0	1	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0.94
G_EnumThree/3	0	0	1	**73**	3	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0.92
G_EnumFour/4	0	0	0	2	**73**	5	0	0	2	4	0	0	0	1	0	0	0	0	0	0	0.84
G_EnumFive/5	1	0	0	0	8	**54**	0	4	0	19	0	0	1	0	1	0	0	0	0	0	0.61
G_Accept/6	1	26	0	0	0	0	**40**	0	1	0	4	0	0	16	0	0	0	2	0	0	0.44
G_MenuOpen/7	1	0	0	0	0	0	0	**65**	0	0	0	0	0	0	0	0	0	0	3	0	0.94
G_MenuClose/8	0	0	0	0	0	0	0	0	**77**	0	0	0	0	0	0	0	0	1	0	0	0.99
G_Take/9	0	0	0	5	4	0	0	0	0	**54**	1	1	0	0	0	0	0	0	0	0	0.83
G_MenuLeft/10	0	0	0	0	0	0	1	5	1	0	**79**	2	0	1	0	0	0	0	0	0	0.90
G_MenuRigth/11	0	0	0	0	0	1	1	2	1	1	0	**79**	0	0	0	4	0	0	1	0	0.88
G_Cancel/12	0	0	0	1	0	1	1	5	1	0	7	0	**58**	2	2	0	0	0	1	0	0.73
G_Click/13	0	0	0	0	0	0	0	1	0	0	0	0	0	**88**	0	0	1	1	0	0	0.96
G_PageLeft/15	3	0	0	0	0	0	0	0	0	0	0	0	0	0	**61**	0	0	0	7	5	0.80
G_PageRight/16	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**59**	12	0	0	0	0.80
G_SlapUpRight/17	15	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	**62**	0	0	0	0.79
G_SlapDownRight/18	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	**63**	0	0	0.84
G_SlapUpLeft/19	13	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	**54**	0	0.78
G_SlapDownLeft/20	9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**68**	0.87
PRECISION:	-	0.60	0.98	0.89	0.82	0.87	0.80	0.75	0.91	0.69	0.87	0.96	0.97	0.62	0.95	0.87	0.83	0.94	0.82	0.93	

Table 5.18: Evaluación sistema tras los cambios. Videos Septiembre 2009

DHG/id	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	RECALL:
Unknown/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0	0	0	0	0	0	0	-
Select/1	0/0	**39/42**	1/0	1/0	0/0	0/0	7/43	0/0	2/0	0/0	0/0	0/0	0/0	35	0	0	0	0	0	0	0.46/0.49
Enum2/2	0/0	0/3	**81/74**	0/3	1/0	0/0	0/1	2/0	0/5	0/0	0/0	0/0	0/0	0	0	2	0	0	0	0	0.94/0.86
Enum3/3	0/0	0/0	1/2	**73/75**	3/2	1/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0	0	1	0	0	0	0	0.92/0.95
Enum4/4	0/0	0/1	0/0	2/9	**73/76**	5/0	0/0	0/0	2/0	4/0	0/0	0/0	0/1	1	0	0	0	0	0	0	0.84/0.87
Enum5/5	1/0	0/0	0/0	0/5	8/6	**54/76**	0/0	4/0	0/0	19/0	0/0	0/0	1/1	0	1	0	0	0	0	0	0.61/0.86
Accept/6	1/0	26/6	0/0	0/0	0/0	0/1	**40/81**	0/0	1/0	0/0	4/2	0/0	0/0	16	0	0	0	2	0	0	0.44/0.90
MOpen/7	1/0	0/0	0/0	0/9	0/0	0/0	0/18	**65/37**	0/0	0/0	0/0	0/5	0/0	0	0	0	0	0	3	0	0.94/0.54
MClose/8	0/0	0/1	0/0	0/10	0/0	0/2	0/14	0/24	**77/27**	0/0	0/0	0/0	0/0	0	0	0	0	1	0	0	0.99/0.35
Take/9	0/0	0/0	0/0	5/7	4/2	0/44	0/0	0/2	0/6	**54/1**	1/2	1/1	0/0	0	0	0	0	0	0	0	0.83/0.02
OkLeft/10	0/0	0/0	0/0	0/1	0/0	0/0	1/0	5/2	1/0	0/0	**79/59**	2/18	0/8	1	0	0	0	0	0	0	0.90/0.67
OkRigth/11	0/0	0/0	0/0	0/3	0/0	1/0	1/0	2/0	1/0	1/0	0/0	**79/78**	0/9	0	0	4	0	0	1	0	0.88/0.87
Cancel/12	0/0	0/7	0/0	1/11	0/1	1/0	1/5	5/4	1/14	0/0	7/2	0/0	**58/35**	2	2	0	0	0	1	0	0.73/0.44
Click/13	0	0	0	0	0	0	0	1	0	0	0	0	0	**88**	0	0	1	1	0	0	0.96
PLeft/15	3	0	0	0	0	0	0	0	0	0	0	0	0	0	**61**	0	0	0	7	5	0.80
PRight/16	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**59**	12	0	0	0	0.80
SURight/17	15	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	**62**	0	0	0	0.79
SDRight/18	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	**63**	0	0	0.84
SULeft/19	13	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	**54**	0	0.78
SDLeft/20	9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**68**	0.87
PREC:	-	0.60/0.70	0.98/0.97	0.89/0.56	0.82/0.87	0.87/0.62	0.80/0.50	0.75/0.54	0.91/0.52	0.69/1.00	0.87/0.91	0.96/0.76	0.97/0.65	0.62	0.95	0.87	0.83	0.94	0.82	0.93	

Table 5.19: Comparativa antes y después de los cambios. Videos Septiembre 2009.

DHG/id	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	RECALL:
G_Unknown/0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
G_Select/1	22	**22**	3	0	0	0	84	3	4	0	0	0	0	13	1	1	0	1	1	0	0.14
G_EnumTwo/2	15	0	**130**	0	0	0	6	1	1	0	0	0	0	1	1	1	0	0	0	0	0.83
G_EnumThree/3	7	0	11	**122**	0	0	0	4	0	0	0	0	0	1	1	1	0	2	0	0	0.82
G_EnumFour/4	9	0	3	1	**122**	9	3	2	3	0	0	0	0	0	0	2	0	3	0	0	0.78
G_EnumFive/5	13	0	1	6	33	**97**	3	1	0	1	0	0	0	1	0	0	1	1	0	0	0.61
G_Accept/6	14	10	0	0	0	0	**105**	5	2	0	0	0	0	16	1	4	1	0	2	0	0.66
G_MenuOpen/7	7	0	0	0	0	1	0	**121**	0	0	1	0	0	0	0	0	2	0	8	0	0.86
G_MenuClose/8	28	0	0	0	0	0	0	0	**105**	0	1	0	0	0	0	1	0	3	2	8	0.71
G_Take/9	4	0	4	8	16	13	0	0	1	**78**	3	0	0	6	0	0	0	1	1	0	0.58
G_MenuLeft/10	10	0	0	0	0	0	0	8	7	0	**118**	5	0	0	0	5	0	0	4	0	0.75
G_MenuRight/11	26	0	0	0	0	0	13	8	4	0	2	**93**	0	1	0	5	1	0	0	1	0.60
G_Cancel/12	36	0	0	1	0	0	73	3	3	0	3	1	**22**	4	0	0	1	0	0	1	0.15
G_Click/13	16	8	0	0	0	0	30	0	1	0	0	0	0	**104**	2	0	1	0	0	0	0.62
G_PageLeft/15	9	0	0	0	0	0	0	0	0	0	1	1	0	0	**98**	1	0	0	27	9	0.67
G_PageRight/16	11	0	0	0	0	0	0	1	0	0	3	0	0	0	0	**102**	23	4	0	0	0.71
G_SlapUpRight/17	43	0	0	0	0	0	0	4	0	0	0	0	0	0	0	2	**99**	0	0	0	0.67
G_SlapDownRight/18	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	**109**	0	0	0.75
G_SlapUpLeft/19	28	0	0	0	0	1	0	6	0	0	0	0	0	0	2	0	0	0	**102**	0	0.73
G_SlapDownLeft/20	30	0	0	0	0	0	0	0	6	0	2	0	0	0	8	1	0	1	0	**100**	0.68
PRECISION:	-	0.55	0.86	0.88	0.71	0.80	0.33	0.72	0.77	0.99	0.88	0.93	1.00	0.74	0.86	0.77	0.77	0.87	0.69	0.84	

Table 5.20: Evaluación del sistema entrenado por 6 usuarios con cámara 3DV.

DHG/id	0	1	2	3	4	5	6	7	8	9	10	11	12	13	15	16	17	18	19	20	RECALL:
G_Unknown/0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-
G_Select/1	15	**10**	2	0	0	0	104	1	3	0	0	0	0	15	1	1	0	1	2	0	0.06
G_EnumTwo/2	2	0	**142**	0	0	0	8	1	1	0	0	0	0	0	1	1	0	0	0	0	0.91
G_EnumThree/3	2	0	15	**123**	0	0	1	4	0	0	0	0	0	0	1	1	0	2	0	0	0.83
G_EnumFour/4	5	0	4	0	**128**	5	4	2	2	0	0	0	0	0	1	2	0	3	1	0	0.82
G_EnumFive/5	11	0	1	8	37	**92**	4	1	0	2	0	0	0	0	0	0	1	1	0	0	0.58
G_Accept/6	30	0	0	0	0	0	**98**	5	2	0	0	0	0	17	1	4	1	0	2	0	0.61
G_MenuOpen/7	7	0	0	0	0	0	0	**121**	0	0	1	0	0	0	0	0	2	0	8	0	0.87
G_MenuClose/8	28	0	0	0	0	0	0	0	**105**	0	1	0	0	0	0	1	0	3	2	8	0.71
G_Take/9	3	0	3	5	18	13	2	0	1	**75**	4	0	0	9	0	0	0	1	1	0	0.56
G_MenuLeft/10	5	0	0	0	0	0	0	9	9	0	**123**	1	0	0	2	5	0	0	4	0	0.78
G_MenuRight/11	26	0	0	0	0	0	5	11	4	0	3	**95**	0	1	7	5	2	0	0	1	0.59
G_Cancel/12	56	0	0	1	2	1	72	3	3	0	2	2	**2**	3	0	0	1	0	0	1	0.01
G_Click/13	8	1	1	0	0	0	28	0	1	0	0	0	0	**120**	2	0	1	0	0	0	0.74
G_PageLeft/15	9	0	0	0	0	0	0	0	0	0	0	1	0	0	**99**	1	0	0	27	9	0.68
G_PageRight/16	11	0	0	0	0	0	0	1	0	0	3	0	0	0	0	**102**	23	4	0	0	0.71
G_SlapUpRight/17	43	0	0	0	0	0	0	4	0	0	0	0	0	0	0	2	**99**	0	0	0	0.67
G_SlapDownRight/18	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	**109**	0	0	0.75
G_SlapUpLeft/19	28	0	0	0	0	1	0	6	0	0	0	0	0	0	2	0	0	0	**102**	0	0.73
G_SlapDownLeft/20	30	0	0	0	0	0	0	0	6	0	2	0	0	0	8	1	0	1	0	**100**	0.68
PRECISION:	-	0.91	0.85	0.90	0.69	0.82	0.30	0.72	0.77	0.97	0.88	0.95	1.00	0.73	0.79	0.77	0.76	0.87	0.68	0.84	

Table 5.21: Evaluación del sistema entrenado por 3 usuarios con cámara SR4000.

Capítulo 6

Conclusiones y trabajo futuro

Observando las tablas de evaluación del sistema, especialmente las dos últimas (5.20 y 5.21) podemos concluir que casi todos los gestos son separables con una elevada tasa de acierto. Pese a ello, todavía quedan errores que solventar cara al futuro. De ellos, los más preocupantes son: la confusión entre los DHGs Select y Accept; la no detección del DHG Cancel; el alto número de falsos positivos del DHG EnumFour, cuando el gesto realizado es EnumFive.

Para poder diferenciar los gestos Select y Accept, es necesario acudir a los descriptores de la mano. Utilizando la elipse y protuberancias resulta imposible diferenciar un gesto de otro por lo que para su correcta separación habría que cambiar la base de la separación de SHPs, buscando un descriptor capaz de hacer estas dos posturas distinguibles.

En el caso de la no detección del Cancel, es algo que no ocurría en anteriores evaluaciones (ver tabla 5.18). Decir, que durante la realización de este proyecto el sistema global no sólo ha sufrido los cambios introducidos por el mismo(i.e. selección y tratamiento de las características del vector, y detección de movimiento), sino que además otras partes del sistema se han visto modificadas. En primer lugar, la introducción de la nueva máquina de estados, descrita en la sección 3.3.2.5. Esta máquina de estados resulta de gran utilidad para la definición de nuevos DHGs, así como para la modificación de algunos requisitos de los ya definidos. Pero también presenta ciertos problemas de transición entre gestos. Por ejemplo, cuando el sistema detecta SHP=6 (postura estática correspondiente al Ok), el sistema queda pendiente de recibir un nuevo 6, detectando con ello el gesto Accept (a la

espera de desactivación), o un 8 (postura estática correspondiente al Fist, puño cerrado), detectando con ello el gesto Click. El problema es que si ese primer 6, era un falso positivo, y tras el mismo obtenemos durante repetidos cuadros un resultado de SHP=7 (correspondiente al gesto Cancel), este nunca será detectado, pues como ya se ha dicho, el sistema queda a la espera de recibir un nuevo Ok o un Fist para cerrar el gesto. Otro cambio que puede afectar a la detección del gesto Cancel, es la introducción de la nueva segmentación, que utiliza información de profundidad buscando eliminar con mayor precisión el antebrazo. En ocasiones, no sólo elimina el antebrazo, sino que además elimina alguna protuberancia y reduce el tamaño de la mano. Esta reducción puede llevar a que la protuberancia detectada en el Cancel, mucho mayor por definición que la del Accept, se vea reducida siendo con ello mucho más parecidas y dando lugar a falsos positivos, con la consecuencia descrita anteriormente.

La forma de solucionar esto, pasa por sustituir la máquina de estados por otra más tolerante con las transiciones entre estados, así como con el tratamiento de falsos positivos. Una posible línea a seguir es la introducción de HMMs, donde los estados ocultos se corresponderían con los DHGs, estableciendo probabilidades de transición entre los mismos, mediante definición, o mediante entrenamiento. Esto no sólo eliminaría el problema descrito de la no detección del Cancel, sino que además serviría para acabar con la mayor restricción, en cuanto a usabilidad, que presenta el sistema. Ésta es la necesidad de salir de la zona de interacción entre gesto y gesto, pues el sistema solo detecta un DHG por cada activación. Con la utilización de HMMs existiría una cierta probabilidad de pasar de un gesto a otro sin necesidad de desactivar, por lo que el usuario no tendría que estar acercando y alejando el brazo continuamente.

El otro problema comentado, la detección de un elevado número de falsos positivos del gesto EnumFour cuando el usuario realiza EnumFive, tiene que ver con la desaparición de alguna protuberancia en la segmentación. Una posible idea para solucionarlo, sería introducir un rango dinámico de detección, en lugar del actual. Actualmente, el sistema captura un rango de 3m (de 0.3m a 3.3m), siendo este estático y captando todo lo que haya en ese rango. La posible solución pasaria por detectar el punto más cercano a la cámara, asumiendo que este se corresponderá con la mano, y a partir de

ahí y en función de la distancia detectar sólo lo que aparezca entre ese punto y una determinada distancia. Otra solución, tal vez más compleja, sería utilizar además de la información de profundidad, una cámara tradicional combinando ambas informaciones. Tener información de profundidad y de color permitiría separar la mano con una mayor precisión. El problema sería realizar la calibración de ambas cámaras para hacer corresponder cada punto de la imagen de profundidad con los de la imagen 2D.

En cuanto a usabilidad, el sistema se ve muy mejorado con la incorporación de los nuevos gestos correspondientes al movimiento de la mano. El usuario sería capaz de navegar por menús de una forma muy sencilla e intuitiva lo cuál no sólo da mayor sensación de control que la realización de un gesto estático, sino que además se corresponde en mayor medida con el lenguaje natural utilizado por el ser humano. La inclusión de HMMs, propuesta anteriormente, también mejoraría la usabilidad, evitando el problema ya comentado de realizar un único gesto por activación.

Bibliografía

[1] E. Kollorz, J. Penne, J. Hornegger, and A. Barke, "Gesture recognition with a time-of-flight camera," *Int. J. Intell. Syst. Technol. Appl.*, vol. 5, no. 3/4, pp. 334–343, 2008.

[2] M. Flasinski and S. Myslinski, "On the use of graph parsing for recognition of isolated hand postures of polish sign language," *Pattern Recognition*, vol. 43, no. 6, pp. 2249–2264, 2010.

[3] N. Tanibata, N. Shimada, and Y. Shirai, "Extraction of hand features for recognition of sign language words," in *In International Conference on Vision Interface*, 2002, pp. 391–398.

[4] C. Manresa, J. Varona, R. Mas, and F. Perales, "Hand tracking and gesture recognition for human-computer interaction," vol. 5, no. 3, pp. 96–104, 2005.

[5] E. Sánchez-Nielsen, L. Antón-Canalís, and M. Hernández-Tejera, "Hand gesture recognition for human-machine interaction," in *WSCG*, 2004, pp. 395–402.

[6] S. Soutschek, J. Penne, J. Hornegger, and J. Kornhuber, "3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1–6.

[7] A. L. University, A. Licsár, and T. Szirányi, "Hand-gesture based film restoration," in *Proc. of PRIS-02*, 2002, pp. 95–103.

[8] J. J. Laviola, "Bringing vr and spatial 3d interaction to the masses through video games," *IEEE Computer Graphics and*

Applications, vol. 28, no. 5, pp. 10–15, 2008. [Online]. Available: http://dx.doi.org/http://dx.doi.org/10.1109/MCG.2008.92

[9] J. Letessier and F. Bérard, "Visual tracking of bare fingers for interactive surfaces," in *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology.* New York, NY, USA: ACM, 2004, pp. 119–122. [Online]. Available: http://dx.doi.org/10.1145/1029632.1029652

[10] X. Zhu, J. Yang, and A. Waibel, "Segmenting hands of arbitrary color," in *FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000.* Washington, DC, USA: IEEE Computer Society, 2000, p. 446.

[11] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, November 2006. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2006.08.002

[12] R. Grzeszczuk, G. Bradski, M. Chu, and J. Bouguet, "Stereo based gesture recognition invariant to 3d pose and lighting," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. I: 826–833.

[13] K. Nickel and R. Stiefelhagen, "Visual recognition of pointing gestures for human-robot interaction," *Image and Vision Computing*, vol. 25, no. 12, pp. 1875–1884, 2007. [Online]. Available: http://dx.doi.org/http://dx.doi.org/10.1016/j.imavis.2005.12.020

[14] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff, "A unified framework for gesture recognition and spatiotemporal gesture segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1685–1699, 2009. [Online]. Available: http://dx.doi.org/http://dx.doi.org/10.1109/TPAMI.2008.203

[15] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla, "Model-based hand tracking using a hierarchical bayesian filter," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 9, pp. 1372–1384, 2006. [Online]. Available: http://dx.doi.org/http://dx.doi.org/10.1109/TPAMI.2006.189

[16] Guomundsson, R. Larsen, H. Aanaes, M. Pardás, and J. R. Casas, "TOF imaging in smart room environments towards improved people tracking," in *Time of Flight Camera based Computer Vision (TOF-CV)*, jun 2008, pp. 1–6. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php?5665

[17] X. Liu and K. Fujimura, "Hand gesture recognition using depth data," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 529–534.

[18] P. Breuer, C. Eckes, and S. Muller, "Hand gesture recognition with a novel ir time-of-flight range camera: A pilot study," in *Computer Vision/Computer Graphics Collaboration Techniques Third International Conference, MIRAGE*, 2007, pp. 247–260.

[19] S. Malassiotis and M. Strintzis, "Real-time hand posture recognition using range data," *Image and Vision Computing*, vol. 26, no. 7, pp. 1027–1037, July 2008.

[20] Y. T. Chen and K. T. Tseng, "Developing a multiple-angle hand gesture recognition system for human machine interactions," in *Industrial Electronics Society, 2007. IECON 2007. 33rd Annual Conference of the IEEE*, 2007, pp. 489–492.

[21] G. Zheng, C. J. Wang, and T. E. Boult, "Application of projective invariants in hand geometry biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 758–768, 2007.

[22] W. T. Freeman, W. T. Freeman, M. Roth, and M. Roth, "Orientation histograms for hand gesture recognition," in *In International Workshop on Automatic Face and Gesture Recognition*, 1994, pp. 296–301.

[23] O. Aran, I. Ari, L. Akarun, B. Sankur, A. Benoit, A. Caplier, P. Campr, A. H. Carrillo, and F.-X. Fanard, "Signtutor: An interactive system for sign language tutoring," *IEEE Multimedia*, vol. 16, pp. 81–93, 2009.

[24] Y. Liu, Z. Gan, and Y. Sun, "Static hand gesture recognition and its application based on support vector machines," *Proceedings of the 2008 Ninth ACIS International Conference on*

Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, vol. 0, pp. 517–521, 2008. [Online]. Available: http://dx.doi.org/10.1109/SNPD.2008.144

[25] P. Hong, M. Turk, and T. S. Huang, "Constructing finite state machines for fast gesture recognition," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 3, 2000, pp. 691–694 vol.3. [Online]. Available: http://dx.doi.org/10.1109/ICPR.2000.903639

[26] M. K. Bhuyan, D. Ghosh, and P. K. Bora, "Feature extraction from 2d gesture trajectory in dynamic hand gesture recognition," in *Cybernetics and Intelligent Systems, 2006 IEEE Conference on*, June 2006, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1109/ICCIS.2006.252353

[27] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1163055

[28] R. F. Mello and I. Gondra, "Multi-dimensional dynamic time warping for image texture similarity," in *SBIA '08: Proceedings of the 19th Brazilian Symposium on Artificial Intelligence*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 23–32.

[29] M. Wöllmer, M. Al-Hames, F. Eyben, B. Schuller, and G. Rigoll, "A multidimensional dynamic time warping algorithm for efficient multimodal fusion of asynchronous data streams," *Neurocomput.*, vol. 73, no. 1-3, pp. 366–380, 2009.

[30] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. The MIT Press and McGraw-Hill Book Company, 1989.

[31] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, pp. 379 –385, jun. 1992.

[32] T. Starner and A. Pentland, "Visual recognition of american sign language using hidden markov models," in *In International Workshop on Automatic Face and Gesture Recognition*, 1995, pp. 189–194.

[33] T. R. Stephan Hussmann and B. Hagebeuker, *A Performance Review of 3D TOF Vision Systems in Comparison to Stereo Vision Systems.* Stereo Vision, Asim Bhatti (Ed.), 2008.

[34] D. Castilla, I. Miralles, M. Jorquera, C. Botella, R. Baños, J. Montesa, and C. Ferran, "Analysis and testing of metaphors for the definition of a gestual language based on real users interaction: vision project," in *13th International Conference on Human-Computer Interaction*, San Diego, CA, USA, 2009.

[35] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," 2001. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.20.9020

[36] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Advances in Information Retrieval*, ser. Lecture Notes in Computer Science, D. E. Losada and J. M. Fernández-Luna, Eds. Springer Berlin / Heidelberg, 2005, vol. 3408, pp. 345–359.

Parte I

Presupuesto

1) Ejecución Material

- Compra de ordenador personal (Software incluido)..... 2.000 €
- Material de oficina..... 150 €
- Total de ejecución material 2.150 €

2) Gastos generales

- 16 % sobre Ejecución Material 344 €

3) Beneficio Industrial

- 6 % sobre Ejecución Material 129 €

4) Honorarios Proyecto

- 8000 horas a 15 € / hora 12000 €

5) Material fungible

- Gastos de impresión 60 €
- Encuadernación 200 €

6) Subtotal del presupuesto

- Subtotal Presupuesto 14410 €

7) I.V.A. aplicable

- 16 % Subtotal Presupuesto 2305.6 €

8) Total presupuesto • Total Presupuesto 16715 €

Madrid, Septiembre de 2010

El Ingeniero Jefe de Proyecto

Fdo.: José Antonio Pajuelo Martín

Parte II

Pliego de condiciones

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un sistema de reconocimiento de gestos manuales. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego. Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que suponvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no,

se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4 % del presupuesto y la provisional del 2 %.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad “Presupuesto de Ejecución de Contrata” y anteriormente llamado ”Presupuesto de Ejecución Material” que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha

aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.

e este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo

a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4 % del presupuesto y la provisional del 2 %.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la

provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad “Presupuesto de Ejecución de Contrata” y anteriormente llamado ”Presupuesto de Ejecución Material” que hoy designa

otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.