

UNIVERSIDAD AUTONOMA DE MADRID

ESCUELA POLITECNICA SUPERIOR



**Grado en Ingeniería de Tecnologías y Servicios de
Telecomunicación**

TRABAJO FIN DE GRADO

**RECONOCIMIENTO BIOMÉTRICO EGOCÉNTRICO PARA
ENTORNOS DE REALIDAD VIRTUAL**

**Daniel de Alcalá Valcárcel
Tutor: Aythami Morales Moreno
Ponente: Julián Fierrez Aguilar**

JULIO 2020

RECONOCIMIENTO BIOMÉTRICO EGOCÉNTRICO PARA ENTORNOS DE REALIDAD VIRTUAL

AUTOR: Daniel de Alcalá Valcárcel

TUTOR: Aythami Morales Moreno

**Dpto. Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
JULIO 2020**

RESUMEN

En este trabajo se presenta el primer entorno experimental para desarrollar sistemas biométricos de reconocimiento palmar en entornos virtuales. El entorno propuesto consta de una base de datos y de un sistema de reconocimiento inicial que sirva como base para futuros desarrollos. El sistema se divide en tres bloques principales: detección de pose de la mano, extracción de la palma y comparación entre palmas. Se crea uno automático que no necesita de supervisión humana, y otro donde la detección de pose se hace manualmente.

El objetivo es crear un entorno que sirva como punto de partida para estudios futuros y que proponga distintas alternativas válidas para la implementación de estos sistemas. También intentar acompañar el auge de los entornos virtuales con un reconocimiento biométrico necesario en ciertas aplicaciones.

Para lograr esto se ha hecho, en primer lugar, un estudio del estado del arte de la biometría y en concreto del reconocimiento palmar. A continuación, se han planteado los retos de este tipo de sistema y a partir de ellos se ha desarrollado el sistema completo. Por último, se han analizado los resultados y se han planteados posibles mejoras.

ABSTRACT

In this Bachelor Thesis the first experimental environment to develop a palmprint biometric systems for virtual environments is displayed. The proposed environment consists of a data base and an initial recognition system for future research. The system is divided into 3 principal blocks: hand pose estimation, palmprint extraction and palmprint matching. Two alternatives were created, an automatic one which does not require human supervision, and another where pose estimation is handmade.

The goal is to create a domain that serves as a starting point for future studies and that comes up with several useful alternatives for implementation of these systems. Also try to go along with the rise in virtual environments with a necessary biometric recognition in some applications.

To achieve that a study of the biometric state of the art has first been done and specifically palmprint recognition study. Next the challenges of this type of system have been raised and from them the complete system has been developed. Finally, the results have been analyzed and possible improvements have been proposed.

PALABRAS CLAVE

Reconocimiento, biometría, palma, reconocimiento palmar, pose de la mano, detección de pose automática, detección de pose manual, extracción de la palma, comparación entre palmas, base de datos, entorno virtual, sistema egocéntrico, realidad virtual, gafas realidad virtual, imagen RGB, imagen de color.

KEYWORDS

Authentication, recognition, biometric, hand, palmprint, hand pose, automatic hand pose detection, manual hand pose detection, palm extraction, palm matching, data base, virtual environment, egocentric system, virtual reality, virtual-reality goggles, RGB image, color image.

AGRADECIMIENTOS

Quiero agradecer este trabajo en primer lugar a mi tutor Aythami Morales, por la ayuda, esfuerzo y comprensión dados. También a BiDA Lab por su apoyo y en especial a Roberto e Iván por ofrecerme su ayuda.

Agradezco también a mi familia, mis padres, mi hermana, mis abuelos, mis tíos y mis primos. A mi novia que me ha ayudado en las cosas que peor se me dan y siempre está apoyándome y ayudándome. Y a todos mis amigos que os quiero.

En último lugar agradezco a Charlie de HB por el arte que das al mundo y lo que disfruto con él.

INDICE

1. INTRODUCCIÓN.....	1
1.1 Objetivo.....	1
1.2 Motivación.....	1
1.3 Organización de la memoria	2
2. ESTADO DEL ARTE.....	3
2.1 Historia del reconocimiento.....	3
2.2 La biometría.....	3
2.3 Reconocimiento palmar	5
2.3.1 Partes e información útil para el reconocimiento	5
2.3.2 Métodos según el origen de las imágenes y la resolución.....	7
2.3.3 Características de entornos y entornos favorables para reconocimiento palmar.....	8
2.3.4 Proceso y módulos de reconocimiento palmar.	9
2.3.4.1 Módulo de extracción de características: con contacto.....	10
2.3.4.2 Módulo de extracción de características: sin contacto	11
2.3.4.3 Módulo de extracción de características: 3D	12
2.3.4.4 Módulo de extracción de características: Deep Learning.....	13
3. RECONOCIMIENTO PALMAR EN ENTORNOS VIRTUALES. RETOS Y HERRAMIENTAS.....	15
3.1 Algoritmos de detección de pose.....	16
3.2 Base de datos.....	18
3.3 Sistema completo diagrama de bloques.....	20
4. DESARROLLO.....	21
4.1 Bloque de detección de pose.....	21
4.2 Bloque de selección de palma.....	24
4.2.1 Transformación geométrica	24
4.2.2 Recorte de la palma.....	29
4.3 Bloque de comparación.....	33
5. INTEGRACIÓN PRUEBAS Y RESULTADOS.....	34
6. CONCLUSIONES Y TRABAJO FUTURO.....	39
6.1 Conclusiones.....	39
6.2 Trabajo futuro y posibles mejoras.....	40
BIBLIOGRAFIA	41
ANEXOS.....	43
Anexo 3.1.....	43
Anexo 4.1.....	46
Anexo 4.2.....	54

INDICE DE FIGURAS

Fig. 2.1: Relación de rasgos biométricos y propiedades [9].....	4
Fig. 2.2: Impresión a carbón de una palma.....	6
Fig. 2.3: Partes de la distribución de crestas. [9].....	6
Fig. 3.1: Posibles ángulos de giro de la mano en el plano imagen.....	15
Fig. 3.2: Diagrama de bloques del código de detección de pose del artículo <i>Learning to Estimate 3D Hand Pose from Single RGB Images</i> . [23].....	16
Fig. 3.3: Diagrama de bloques del código de detección de pose del <i>3D Hand Shape and Pose Estimation from a Single RGB Image</i> [24].....	17
Fig. 3.4: Instrucciones emitidas para la adquisición de la base de datos.....	18
Fig. 3.5: Algunas imágenes de la base de datos.	19
Fig. 3.6: Diagrama de bloques del sistema a desarrollar.....	20
Fig. 4.1: Ejemplo de los puntos de la detección con el algoritmo de detección automática de pose.....	21
Fig. 4.2: Resultado del código de detección automática.....	22
Fig. 4.3: Ejemplo de los puntos de la detección con el algoritmo de detección manual de pose.....	23
Fig. 4.4: Posibles transformaciones afines sobre una imagen. [26]	24
Fig. 4.5: Posibles transformaciones de la imagen. [26].....	25
Fig. 4.6: Transformaciones geométricas sobre la palma.....	26
Fig. 4.7: Dos ejemplos de la base de los dedos índice y corazón.....	27
Fig. 4.8: Rotación con el método automático.....	28
Fig. 4.9: Rotación con el método manual.....	29
Fig. 4.10: Ejemplo del código automático de detección de pose sobre la base de datos.....	29
Fig. 4.11: Histograma de la relación de aspecto de los recortes de la base de datos con el método automático.....	31
Fig. 4.12: Recorte de palma del método automático.....	32
Fig. 4.13: Recortes de palma del método manual.....	32
Fig. 5.1: Etiquetado manual de dos imágenes de la base de datos.....	34
Fig. 5.2: Etiquetado automático de dos imágenes de la base de datos.....	35
Fig. 5.3: Recortes de la palma con el método automático y el manual.....	35
Fig. 5.4: Distribución de número de puntos detectados entre palmas para cada uno de los métodos de comparación con los resultados manuales.....	37
Fig. 5.5: Distribución SURF de número de puntos detectados entre palmas para cada uno de los métodos de comparación con los resultados automáticos.....	37

1. INTRODUCCIÓN

1.1 Objetivo

Las tecnologías de entornos virtuales están en constante expansión. En los últimos años se ha mejorado notablemente en sensores y en contenido. El mercado es cada vez mayor y surgen nuevas aplicaciones (p.ej. espacios de conferencia virtual durante la pandemia). Dentro de este desarrollo, no existen alternativas de gestión de la identidad distintas a las ya conocidas (p.ej. passwords). El desarrollo de nuevos métodos de reconocimiento de identidad adaptados a estos entornos virtuales es una oportunidad de investigación y desarrollo industrial con gran potencial presente y futuro.

El objetivo de este trabajo de fin de grado será por tanto crear un entorno de experimentación para reconocimiento biométrico palmar egocéntrico. Este entorno permitirá desarrollar sistemas que logren hacer un reconocimiento biométrico utilizando la palma de la mano, a partir de una imagen RGB adquirida desde una vista egocéntrica con el posible uso de unas gafas de realidad virtual. Este va a ser un entorno semi-controlado y en el que se va a contar con la colaboración del usuario.

Para llevar a cabo esta tarea, el sistema biométrico se dividirá en tres bloques, un primer bloque de adquisición de la pose de la mano, el segundo bloque consistirá en un recorte de la palma y el último será el comparador entre palmas.

Para desarrollar este entorno, se creará una base de datos de imágenes que cumplan las condiciones de este sistema, adquiridas desde un punto de vista egocéntrico y en las que se disponga de una visión adecuada de la palma. Además, se etiquetarán los puntos de interés para así poder comparar el rendimiento de diferentes algoritmos automáticos.

1.2 Motivación

La motivación de este trabajo es facilitar una futura línea de investigación en este tipo de sistemas virtuales con un punto de vista egocéntrico. Con el auge que presentan estos sistemas viene un incremento de las posibles aplicaciones de estos entornos. Este auge debe ir acompañado de cierta seguridad para no verse limitado en este aspecto.

Se van a estudiar las principales complicaciones que pueden tener estos sistemas y a investigar posibles soluciones eficaces para ellas. Desde el punto de vista científico, los retos que se afrontarán incluyen: alta variabilidad de pose, iluminación no controlada, interoperabilidad entre dispositivos y escasez de cualquier base de datos pública.

Adicionalmente, otra motivación que ha impulsado esta línea de investigación es aprender sobre el funcionamiento de los sistemas biométricos y proponer la palma, una alternativa menos utilizada que la huella y el reconocimiento facial, que puede resultar muy útil en este ámbito. Este reconocimiento palmar podría llevarse a otros campos también en los que no es tan utilizado.

1.3 Organización de la memoria

Este trabajo está organizado en los siguientes puntos que se describen a continuación.

- Introducción: Se citan los objetivos de este trabajo y los motivos que llevaron a su desarrollo.
- Estado del arte: Historia de los sistemas biométricos centrándose en los sistemas biométricos basados en la palma de la mano y en su trabajo pasado.
- Reconocimiento palmar en entornos virtuales: Se detallan las particularidades y retos de este tipo de sistemas en estos entornos, y se explica el sistema y partes basándose en estas dificultades.
- Desarrollo: Explica cómo se han realizado los distintos bloques. Se comentan las etapas y los algoritmos utilizados detalladamente.
- Integración, pruebas y resultados: Se especifican los resultados finales del sistema completo desarrollado y su viabilidad.
- Conclusiones y trabajo futuro: Desde un marco crítico se analiza la capacidad de reconocimiento y se describe el camino a seguir para mejorar el proyecto propuesto.

2. ESTADO DEL ARTE.

2.1 Historia del reconocimiento.

La necesidad de poder identificar a las personas ha hecho que el reconocimiento sea una herramienta ampliamente implementada y necesaria en muchos ámbitos de la actualidad. Por ejemplo, desde principios del siglo XX se ha utilizado la huella dactilar para identificar sospechosos de algún delito.

Todo comenzó en 1882 con el policía francés Alphonse Bertillon que presentó el primer sistema de identificación a partir de características físicas. Aunque al principio la idea no fue muy apoyada, 2 años después de esto consiguió identificar a casi 250 criminales. [1] Previamente hubo casos aislados de identificación de personas por características físicas, en China en el siglo VIII se utilizaron las huellas dactilares como firma de los escultores en sus obras o casos aislados más adelante de identificación de un crimen por una huella. [2]

En el sistema de Alphonse Bertillon se tenían en cuenta características como la altura, la anchura de los brazos extendidos, anchura y largura de la cabeza, de las orejas y de algunos dedos, entre otras. [1] Sin embargo, el problema de estas características es que no eran únicas de cada individuo si no que muchas personas podían coincidir en ellas, dando lugar a errores que no se pueden asumir.

A finales del siglo XIX el jefe de policía Juan Vucetich descubrió al asesino de sus dos hijas gracias a una huella dactilar. Y un año después se aceptó el hecho de que dos personas no podían tener dos huellas dactilares iguales por lo que se solucionaba el problema anterior y comenzó a utilizarse en el ámbito forense. [1]

El reconocimiento por tanto se basa en la capacidad de poder identificar a una persona con la mayor fiabilidad posible. Este reconocimiento sigue siendo utilizado, no exclusivamente para el mundo del crimen, sino que, con el desarrollo de la tecnología, ha surgido la necesidad de poder identificarse unívocamente para poder acceder a ciertos contenidos o realizar alguna compra y que nadie más pueda hacerlo. Así se ha desarrollado en otros campos como por ejemplo en el ámbito de seguridad nacional, comercio electrónico o la banca. [3,4] Para esto históricamente se han utilizado los métodos tradicionales que consistían en contraseñas, en alguna clave física o en información personal que te identifique y de esta manera se conseguía un reconocimiento del usuario sin necesidad de implementar un sistema que reconociese alguna característica.

2.2 La biometría.

La biometría hace uso tanto de lo fisiológico como de algunas características de la persona tales como la cara, la huella, la palma, el iris, la voz o la escritura. [3,4] El reconocimiento basado en estos métodos proporciona ventajas sobre métodos no biométricos como contraseñas o PIN debido a motivos de comodidad y seguridad, ya que esta información se puede obtener u olvidar. La exactitud de reconocimiento sumado a dicha comodidad y seguridad ha contribuido al incremento del uso de características biométricas sobre los métodos más tradicionales. [3,5]

Los métodos biométricos pueden dividirse en dos grupos: El primero, basado en fisiología donde entra la huella, la palma, el iris, la cara etc. Un segundo grupo basado en comportamiento donde se encuentra la escritura, la firma, la forma de hablar o incluso la forma de andar.

No cualquier rasgo puede ser un identificador biométrico, para que un rasgo pueda serlo se necesita que cumpla ciertas condiciones:

1. Universalidad: Que todo el mundo posea ese rasgo de manera natural. Si no se tiene que sea por una anomalía. [1]
2. Particularidad: Que sea lo suficientemente distinto en todas las personas. [1]
3. Permanencia: Debe permanecer invariante al tiempo y a otros factores como el momento del día o el lugar donde se encuentre la persona. [1]
4. Medible: que se pueda de alguna manera cuantificar. [1]
5. Rendimiento: Que garantice un correcto funcionamiento para las distintas condiciones de adquisición. [1]
6. Aceptabilidad: Que los usuarios permitan utilizarlo, no se puede pedir partes que puedan incomodar mostrar al usuario. [1]
7. No falsificable: No puede ser fácil de suplantar. [1]

Cualquier rasgo tiene que procurar cumplir estas condiciones, aunque es imposible cumplir todas a la perfección. En la figura 2.1, se muestra una tabla con distintos rasgos biométricos y la medida en la que cumplen cada condición. De las características citadas en la figura 2.1 algunas cumplen mejor unas condiciones u otras, pero funcionan suficientemente bien para poder ser utilizadas.

		Característica						
		Universalidad	Unicidad	Permanencia	Mensurabilidad	Rendimiento	Aceptabilidad	Evitabilidad
RASGO BIOMÉTRICO <i>fisiológico</i>	Cara	Alto	Bajo	Medio	Alto	Bajo	Medio	Alto
	Geometría de la mano	Medio	Medio	Medio	Alto	Medio	Medio	Medio
	Huella dactilar	Medio	Alto	Alto	Medio	Alto	Medio	Medio
	Huella palmar	Medio	Alto	Alto	Medio	Alto	Medio	Medio
	Iris	Alto	Alto	Alto	Medio	Alto	Bajo	Bajo
	Oreja	Medio	Medio	Alto	Medio	Medio	Alto	Medio
<i>de comportamiento</i>	Dinámica de tecleo	Bajo	Bajo	Bajo	Medio	Bajo	Medio	Medio
	Firma	Bajo	Bajo	Bajo	Alto	Bajo	Alto	Alto
	Forma de andar	Medio	Bajo	Bajo	Alto	Bajo	Alto	Medio
	Voz	Medio	Bajo	Bajo	Medio	Bajo	Alto	Alto

■ Alto
■ Medio
■ Bajo

Fig. 2.1: Relación de rasgos biométricos y propiedades [9]

Otros rasgos posibles menos conocidos son:

1. Termograma facial: Obtiene un mapa de calor de la cara del usuario, este tiene el problema de la permanencia ya que, por ejemplo, si el usuario tiene fiebre o si viene de correr no funciona. [1,6]
2. Olor: Cada persona emite un olor que es único compuesto por distintos elementos químicos en distintas proporciones. [7]
3. Forma de escribir: También en cuanto a rasgos de comportamiento se puede usar la forma de escritura para identificar a una persona. [1]

Los rasgos biométricos de comportamiento tienen la ventaja de ser menos intrusivos, sin embargo, presentan mayor variabilidad con el tiempo y son más difíciles de cuantificar y medir. [1]

La huella dactilar ha sido la más implementada en el pasado y sigue teniendo gran importancia, pero personas de avanzada edad o ciertos trabajadores tienen problemas con este por el desgaste de la piel. Otros reconocimientos como el de iris, el de voz o el facial también son muy relevantes y están muy extendidos. El reconocimiento facial es capaz de lograr un reconocimiento similar al que los humanos son capaces de realizar al observar la cara de otra persona, aunque en general todas estas características pueden conseguir un alto rendimiento. [4,8]

2.3 Reconocimiento palmar

Debido a que este trabajo se centrará en el reconocimiento a nivel palmar, se profundizará en dicho contenido. En comparación con otros rasgos biométricos, las ventajas que presenta el uso de la palma son: la disponibilidad de área para la extracción de características diferenciales, la fácil captura y la alta aceptación por parte del usuario. [3]

El primer sistema de reconocimiento palmar fue desarrollado en Hungría por una empresa privada. Unos años más tarde, en 1997, EE. UU. desarrolló otro sistema de reconocimiento de palmas. En 2004 se empezó a utilizar por los cuerpos policiales y en 2006 Australia desarrolló una base de datos de más de 5 millones de palmas utilizadas para identificar a posibles criminales. [9]

Aunque en la mayoría de las ocasiones se utilizaba para aspectos forenses, el desarrollo de la tecnología ha llevado este tipo de reconocimiento a aplicaciones comerciales de usuarios.

2.3.1 Partes e información útil para el reconocimiento

La palma es la parte interna de la mano, localizada entre los dedos y la muñeca. Aquí la presencia de crestas y pliegues forman un relieve ondulado conocido como la huella palmar, muy parecida a la dactilar.

En la figura 2.2 se ve una huella palmar donde hay ciertas discontinuidades a las que se les llama pliegues de conexión. Estas dos características son las que son únicas de cada persona y las que se pueden identificar unívocamente. [10]

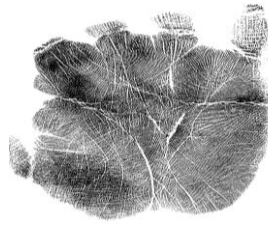


Fig. 2.2: Impresión a carbón de una palma.

Sin embargo, en la palma de la mano hay otras características que también puede utilizarse para reconocer a una persona. Son las siguientes:

- Líneas principales: Estas son las líneas de mayor tamaño de la palma que se pueden ver fácilmente a simple vista. No presentan casi cambios a lo largo de una vida y son una característica de gran ayuda. [3,5,10]
- Forma: Distintos tamaños y áreas de las zonas de la mano. Por ejemplo, anchura de la palma y huecos entre los dedos etc. [3,5,10]
- Distribución de las crestas: Las crestas son como las líneas principales de la mano, pero son más finas y cortas y no tan rectas, son muy variantes. [10]
- Minucias: Las líneas anteriores provocan diferentes puntos. Por ejemplo, cuando terminan, cuando se cruzan o cuando se dividen. [9,10] En la figura 2.3 se pueden ver algunas de ellas.
- Puntos delta: Delta es una distribución concreta de las crestas que muestra la figura 2.3. Y el punto delta es la zona central de esa delta. [9,10]

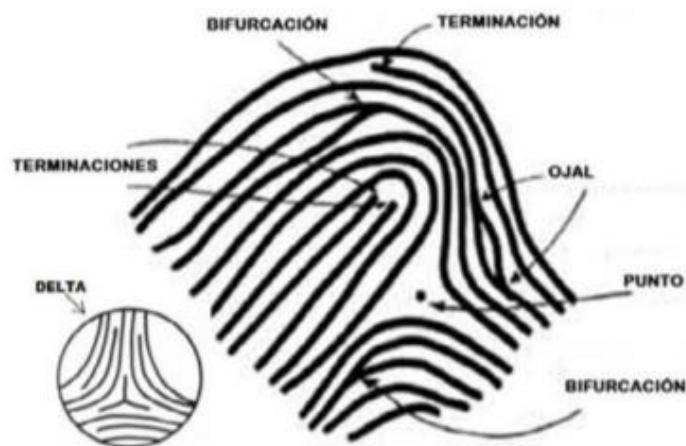


Fig. 2.3: Partes de la distribución de crestas. [9]

Se pueden usar otras características como la textura. Ésta, es muy útil y proporciona información extra cuando la imagen es de baja resolución. Aunque también se pueden obtener diferentes texturas con imágenes de alta resolución. [3,5]. Otra posibilidad podría ser la distribución de las venas. Todo esto ofrece una información parecida a la de la huella dactilar y permite hacer una identificación unívoca de cada persona.

2.3.2 Métodos según el origen de las imágenes y la resolución

Cada una de las características previamente mencionadas se pueden conseguir con distintas resoluciones. Por ejemplo, las líneas principales son adquiribles con una imagen de baja resolución, sin embargo, los pequeños detalles necesitan una imagen de alta resolución. Por lo que, dependiendo de la imagen, permite obtener unas características u otras. En función de la aplicación, interesará obtener más características con mayor o menor resolución. Por ejemplo, el trabajo forense de identificación pretende obtener la mayor cantidad de información posible, incluyendo aquella que requiera una mayor resolución para tener más fiabilidad en el reconocimiento. No obstante, en aplicaciones para el usuario las imágenes de menor resolución, aunque proporcionen menos características, ocupan un espacio menor y permiten un funcionamiento a tiempo real. [3,5]

El reconocimiento palmar lleva estudiándose más de 2 décadas, y los estudios sobre el tema deben dividirse según distintos métodos y las características que pueden obtener. El reconocimiento por contacto no obtiene las mismas características que si se realiza sin contacto. Concretamente, las imágenes por contacto tienen mayor resolución y por tanto permiten acceder a las características que requieren de ellas. La suma de los distintos métodos existentes ha conseguido un elevado rendimiento siempre que haya un entorno controlado. [4]

Las líneas principales de la mano dan características muy importantes debido a que, como se ha visto, se pueden obtener tanto en imágenes de baja resolución como de alta resolución, pero no obtienen un buen porcentaje de acierto debido a la similitud que tienen entre distintas personas. Para conseguir mayor acierto con baja resolución se usan otras características que se pueden obtener a partir de estas imágenes como la textura.

Si bien es cierto que la impresión palmar tiene una historia de investigación más corta que otros métodos como la huella digital o el reconocimiento facial, en los últimos años ha surgido como un campo prometedor de cara al futuro. [3]

También existen dos categorías de reconocimiento palmar: offline y online, dependiendo del momento de adquisición de las imágenes. La offline consiste en el reconocimiento a partir de imágenes ya obtenidas anteriormente por lo que no es a tiempo real. La online consiste en capturar la imagen en el momento y se envía a un ordenador que lo procesa en tiempo real. [4]

En esta última, al tener un funcionamiento en tiempo real normalmente se necesita que el sistema dé una respuesta lo más rápido posible y, por tanto, no se podrá centrar en una imagen con una alta resolución y mucha información, sino que se buscará una menor resolución y menos información para conseguir que funcione a mayor velocidad, aunque se tenga menor fiabilidad.

2.3.3 Características de entornos y entornos favorables para reconocimiento palmar.

Los contextos de las aplicaciones afectan directamente al mecanismo necesario para hacer el reconocimiento.

Se puede hacer una clasificación general de entornos hostiles y no hostiles. Los hostiles son aquellos en los que el usuario no ayuda y las condiciones externas no son las favorables para realizar la tarea. Los no hostiles son aquellos en los que ambos factores sí están controlados y por tanto, facilitan la adquisición. Sin embargo, en las distintas aplicaciones, hay muchos factores que pueden provocar un entorno no favorable u hostil, como las siguientes:

1. Cooperación: El usuario no coopera para colocar por ejemplo la palma centrada en la imagen y de forma adecuada. [1]
2. Habitación: No se accede a ellos de forma habitual y por tanto no conoce las condiciones más favorables para el sistema. [1]
3. Supervisión: El sistema no está supervisado tanto de forma automática como por una persona y por lo tanto el usuario no sabe si lo está haciendo correctamente o no. [1]
4. Estándar: No es un método habitual de adquisición como podría ser actualmente la huella del teléfono, sino que es más complejo y no se sabe la forma de actuar. [1]

Para que puedan funcionar bien y tener un buen comportamiento, las condiciones ideales se situarían en un entorno controlado en el que se puedan obtener las características necesarias con facilidad.

No obstante, esto no tiene por qué ser así y, como se ha visto, estas características se pueden encontrar en un entorno hostil donde no es sencillo obtenerlas. En estos entornos sigue siendo un reto y tema de investigación obtener los datos necesarios para realizar el reconocimiento.

Además de los problemas anteriores, que pueden deberse a la propia implementación de las aplicaciones, hay ocasiones en las que se puede presentar un entorno hostil en el que no se pueda acceder a algunas características por utilizar información que no ha sido adquirida para este fin. Como, por ejemplo, usar información de una grabación de un accidente o un crimen para identificar a la persona. Aquí, esta información, por no estar destinada a este propósito, no será en general favorable para esto.

Para estas situaciones la mejor opción, si no llevan las manos cubiertas, puede ser el reconocimiento palmar ya que las caras pueden estar tapadas, lo que impide el reconocimiento facial. Otros métodos como el iris o el de huella necesitan algo más que una grabación de video adquirida en las condiciones, no muy favorables, que suelen presentar estas.

Por otro lado, se puede dar el caso en el que, por la forma de adquisición, solo sea posible el reconocimiento palmar. Por ejemplo, en el caso de unas gafas de realidad virtual donde la cámara está en la cabeza, también se descarta cualquier tipo de reconocimiento facial y otros como el dactilar, que requerirían demasiada resolución.

Parece claro que hay ocasiones donde la palma puede ser la mejor opción para reconocer y esto unido con el auge general de los sistemas de reconocimiento biométricos ha provocado grandes avances en este aspecto.

2.3.4 Proceso y módulos de reconocimiento palmar.

Para hacer el reconocimiento de la palma, los reconocedores cuentan con distintos módulos que dividen la tarea. Estos son: pre-procesado, segmentación, obtención de la zona de interés de la palma, extracción de características y comparación. Cada módulo le pasa al siguiente la información que necesita. [4]

Para que cada módulo obtenga dicha información se han utilizado distintos métodos tradicionales. Posteriormente, fue muy importante el desarrollo de las redes neuronales convolucionales, las cuales sustituyen a los métodos tradicionales y los llegan a mejorar. Éstos, en vez de obtener las características de cada módulo individualmente, hacen un aprendizaje en conjunto dentro de la misma red de las distintas partes. Para el correcto funcionamiento de estas redes se necesita una gran cantidad de datos de entrenamiento y para esto existen grandes bases de datos. [11]

Hay que diferenciar el reconocimiento con imágenes de alta calidad y colaboración del usuario, y el reconocimiento con entornos hostiles e incontrolables y en el que no existe la colaboración del sujeto. Puesto que, como ya se dijo anteriormente, las características que se pueden obtener son distintas.

Esta parte se centra en el reconocimiento sin contacto, que es lo que se va a utilizar en este trabajo de fin de grado. Este tipo de reconocimiento se produce, en general, en un entorno no tan controlado ya que hay muchas posibles variantes al hacerlo a través de este método: la posición de la mano, sus ángulos de rotación, la iluminación, oclusiones o recortes de la mano por estar esta fuera del rango de visión de la cámara. [4,12,13]

Esta entrada poco controlada debe pasar por las etapas mencionadas con anterioridad (pre-procesado, segmentación, zona de interés, extracción de características y comparación).

1. Para las primeras dos etapas, en las cuales hay que obtener el contorno de la mano y sus puntos clave, si se tiene un entorno controlado con un fondo de un solo color hay muchos métodos que dan buenos resultados (suavizado gaussiano, segmentación por umbral, o detección de bordes) aunque sus resultados tampoco son malos en ambientes no controlados.[4]
2. A continuación, utilizando los puntos clave se puede hacer un alineamiento de las manos para así definir posteriormente la región de interés de la que se extraerán las características. [4,5]
3. Por último, se hace una comparación entre las características extraídas y las que se tenían para asociar esta nueva información y llevar a cabo el reconocimiento.

Posteriormente, se explicará el módulo más complejo de realizar, la extracción de características. Obtener las características que identifiquen bien la mano y que se vean afectadas lo menos posible por la manera en la que han sido adquiridas es una tarea complicada que ha pasado por distintas fases.

2.3.4.1 Módulo de extracción de características: con contacto

Para los métodos de adquisición con contacto existen varias alternativas, entre ellas, PalmCode y OrdinalCode.

PalmCode

Para este método se utiliza un entorno controlado, en el cual hay una cantidad de luz y una posición específica para la mano. Es online, es decir, la imagen se captura en el momento y se envía a un ordenador para ser procesada. [5]

Tiene una etapa de pre-procesado donde se realiza el alineado necesario para realizar la comparación, el cual se realiza utilizando los huecos entre los dedos. Después comienza a obtener las características de la mano. Como las imágenes son de baja resolución, se pueden obtener únicamente características disponibles en esta resolución que son las líneas principales de la mano. Éstas no son suficientes para identificar unívocamente a una persona por lo que utiliza también otra característica que se puede obtener en baja resolución que es la textura. La textura se obtiene mediante la aplicación de filtros de gabor. [5]

Con estas dos informaciones, el método es capaz de representar con baja resolución y en tiempo real las características de una mano suficientes para conseguir un rendimiento aceptable.

Por último, en el bloque de comparación, realizará una comparativa de las informaciones utilizando la distancia de hamming entre las informaciones obtenidas. [5]

OrdinalCode

En primer lugar, se recorta la zona de la que se extraerán las características. Las medidas ordinales se obtienen por comparación cualitativa de varias cantidades que dependen de la imagen. [3]

En la práctica, la transformación y la comparación ordinal puede ser combinada en un solo paso a través del filtrado diferencial. El resultado de la comparación ordinal puede ser el signo de una desigualdad, el orden de rango máximo de todas las mediciones involucradas en la comparación o el índice asociado de valor mínimo. [3]

Después de la comparación ordinal, todos los resultados se cuantifican en bits para hacer robusta la información extraída y facilitar el siguiente paso. Se forman códigos binarios que se concatenan para generar una huella característica que es la entrada. Para terminar la comparación se realiza con la distancia de Hamming entre esta huella extraída y la plantilla almacenada de las huellas existentes. [3]

Algunas características del método son las siguientes:

- 1) Esta representación es robusta contra variaciones como ajustes de iluminación, suciedad o sudores en la palma, cambios en la pose y deformaciones. [3]
- 2) Cada código de bits representa una relación entre varias regiones de imágenes. En una palma compuesta por miles de características ordinales los códigos tienen grados de libertad suficientemente altos para diferenciar a todos los individuos en el mundo. [3]

- 3) La información de la palma obtenida es compacta. Miles de los resultados de comparación ordinales solo necesitan una capacidad de memoria inferior de 1K bytes. [3]
- 4) La diferencia entre dos huellas de palmas puede ser medido por el operador con una XOR bit a bit, que puede calcularse muy rápidamente. [3]

Hay que destacar que, aunque tanto OrdinalCode como PalmCode se explican como métodos con contacto, no es porque necesiten algún tipo de contacto si no porque son métodos que, al no emplear características locales, necesitan que las manos estén perfectamente colocadas y alineadas y esto, en un entorno sin contacto, es prácticamente imposible. Para conseguir que todas las manos tengan una misma postura con una separación exacta se necesita una adquisición por contacto para limitar la libertad de posiciones. [12]

2.3.4.2 Módulo de extracción de características: sin contacto

Cuando se empezaron a desarrollar métodos de reconocimiento sin contacto, en un primer momento, las soluciones que hubo asumían que la mano se encontraba en paralelo al plano de la imagen, lo cual no siempre ocurre, y puede producir un bajo rendimiento. No es lo ideal para hacer una aplicación placentera para el usuario. [4,13]

Por esto, surgieron nuevos métodos que empleaban características locales para no depender de la posición de las manos y que pudiesen emplearse también en el reconocimiento sin contacto. Estas características locales son robustas ante translaciones, rotaciones y cambios de escala. Estos nuevos métodos vienen directamente de los métodos con contacto, pero utilizando estas características locales. [4,13]

Entre estos métodos se encuentran, Scale Invariant Feature Transform (SIFT), Local Binary Pattern (LBP), local line directional pattern (LLDP) y histogram of oriented lines (HOL). [4]

SIFT

Es la técnica que se utilizó para obtener estas características locales como un histograma de gradientes orientados. Básicamente este histograma consiste en aplicar el gradiente a la zona a analizar y construir dicho histograma con las fases y módulos encontrados en esta área. Al principio se empleó esta técnica de la manera original sin tener en cuenta que estaba siendo aplicada en una mano, por lo que no se estaba teniendo en cuenta la posición de estas orientaciones al obtener el histograma, que es una información de gran importancia. Otro problema del método es que la técnica original comparaba un punto con todos los posibles de la otra imagen sin tener en cuenta las relaciones de posición que puede haber en una mano al cambiar de posición. [12]

Para solucionar estos problemas se desarrollaron técnicas que mejoraban el SIFT original y lo adaptaban al problema de una mano. Para ello, se realiza primero un pre-procesado con un filtro para mejorar las texturas de la imagen lo que mejorará el funcionamiento del sistema, y obteniendo después las características SIFT sobre

estas imágenes pre-procesadas. Posteriormente, se comparan para encontrar la correspondencia de cada punto y, por último, se hace un refinamiento de estas correspondencias encontradas. [12]

Este método es muy efectivo ya que la palma con sus líneas, arrugas y texturas tiene mucha información de las orientaciones. [12]

LBP LLDP Y HOL

Son técnicas como la anterior, que hacen una representación local de distintas zonas de la imagen obteniendo características locales que permiten mayor libertad de movimientos de la mano durante el reconocimiento y permitieron el reconocimiento por palma sin contacto. [12]

Más recientemente se han propuesto otros métodos basados en las líneas principales. Estos últimos pretenden obtener estas líneas principales de la mano que, en general, suelen ser tres o cuatro. Estas líneas son robustas frente a ruidos o distintas iluminaciones y son más sencillas de obtener incluso con una imagen en baja resolución. A partir de ellas buscan hacer la comparación mirando el número de píxeles totales que se solapan entre las distintas líneas. [14]

Otros sistemas se basan en codificación, como Double Orientation Code (DOC). Los métodos de codificación se basan en extraer las orientaciones de la imagen como características a través de filtros. Posteriormente estas características se convierten en códigos fácilmente comparables. [14]

DOC

Se basa en asumir que cada punto de la palma pertenece a una línea y que, por tanto, el correspondiente filtro obtendrá su orientación obteniendo un máximo. El problema que presenta es que el número de filtros es limitado y no se pueden obtener todas las orientaciones pudiendo provocar que no se encuentre el resultado exactamente. [14]

Además de los ya mencionados, existen otros métodos que obtienen información 3D a partir de una imagen, como el Difference of Normals (DON). Además de otros basados en aprendizaje como el CR-CompCode. No se deben confundir éstos, con el siguiente que se va a explicar. En los anteriores, la información 3D se obtiene a partir de una imagen 2D, mientras que en el que se explicará a continuación se captura directamente información 3D.

2.3.4.3 Módulo de extracción de características: 3D

Es importante hacer mención al reconocimiento palmar utilizando técnicas 3D. Las técnicas 2D vistas hasta el momento utilizan únicamente el plano en 2D de la imagen y la información que en él se encuentra. Sin embargo, estas técnicas tienen ciertas limitaciones por depender de la proyección de la palma en el plano, donde hay información que no se captura y otra que se puede perder.

Para intentar solventar estas limitaciones surgen las técnicas 3D, con las cuales se puede obtener información de profundidad y curvatura que ayuda a mejorar el funcionamiento de los reconocedores. [15]

Con esta información, además de obtener más características de la mano, se identifica cual es la posición de ésta para hacer una corrección, si es necesaria, y poder seleccionar mejor la parte que interesa para el reconocimiento. [15]

El problema de esto es que necesitan algún tipo de sensor que proporcione esta información 3D y, por tanto, ya no es suficiente con una simple imagen, siendo esto menos accesible. No obstante, el rendimiento que se consigue al añadir esta información es mayor que sin ella. Por lo que, en sistemas de alta seguridad o en los que se quiera una alta fiabilidad, si es posible, se tratará de introducir esta capacidad de adquirir la información 3D.

2.3.4.4 Módulo de extracción de características: Deep Learning.

El módulo de extracción de características también puede hacerse mediante el uso de Deep Learning, como ya se ha descrito. La arquitectura de estas redes puede elegirse de manera que optimice el problema para ello. Un ejemplo de una arquitectura de red que ofrece un buen funcionamiento para estos problemas es la de AlexNet.

Svboda [16] utilizó esta arquitectura introduciendo tres imágenes con las zonas de interés al mismo tiempo, con un tamaño de 128x128 y a la salida se obtienen vectores de características de 32 dimensiones. De las tres imágenes de entrada, dos son de la misma persona y otra es de otro usuario. Con esto llegó a obtener unos resultados muy prometedores. [4,16]

Minaee y Wang [17] utilizaron una deep scattering convolutional network (DSCN) y SVM para el reconocimiento, obteniendo un alto acierto, aunque utilizándolo con una base de datos con contacto. Del mismo modo, otros autores como Meraoumia también utilizaron bases de datos con contacto para implementarlo. [4,17]

Fei [18] implementó finalmente varios métodos sin contacto y varias arquitecturas para redes neuronales convolucionales ya implementadas con bases de datos de la palma sin contacto. Llegando a la importante conclusión de que el aprendizaje profundo llegaba a resultados incluso mejores que los métodos convencionales. [4]

Ramachandra [19] posteriormente, utilizó la red pre-entrenada AlexNet y la refinó con regiones de interés de la mano de bebés. Para clasificar se empleó SVM. [4,19]

Genovese [20], propuso una red de tres capas. En la primera capa se encuentran filtros predefinidos, en la segunda un análisis del componente principal (PCA) y en la última una binarización de los resultados. Sus resultados se evaluaron con una base de datos sin contacto y obtuvo un alto rendimiento. [4,20]

Dian and Dongmei [21] usaron también AlexNet para las características y distancia Hausdorff para comparar, pero no ofrecieron datos sobre el funcionamiento de este trabajo en comparación con otros. [4,21]

Jalalie [22] intentó solucionar la deformación de la información con una red muy pequeña, pero se evaluó con una base de datos de muy pocas imágenes y muchas de éstas se usaron para entrenar. [4,22]

En todos estos métodos no se utiliza el aprendizaje profundo para entrenar todo el sistema desde la primera etapa hasta la última si no que solo se utiliza para la extracción de características. Las partes anteriores se extraen con los métodos tradicionales. En este grupo, para la obtención de la región de interés, solo se consideran deformaciones de la imagen lineales. Trabajos más recientes ya sí que hacen el entrenamiento de principio a final y considerando deformaciones no lineales.

3. RECONOCIMIENTO PALMAR EN ENTORNOS VIRTUALES. RETOS Y HERRAMIENTAS.

Este trabajo en reconocimiento palmar está centrado en un entorno virtual en el que solo se consta de una cámara, por lo que solo se obtiene una imagen RGB. Debido a estas circunstancias se ha trabajado más o menos en distintos aspectos.

El protocolo de verificación se puede resumir en los siguientes pasos: 1) el sistema solicita verificación de identidad (p.ej. requerida por otro usuario, servicio o plataforma); 2) la cámara se activa y el usuario debe mirarse a las palmas; 3) El sistema captura una serie de imágenes y compara la información biométrica con la plantilla guardada en la base de datos. Se requiere por tanto de una fase inicial de registro.

En este contexto hay ciertas características de las imágenes adquiridas que lo diferencian de otros entornos. En las imágenes que se obtendrán, la distancia de la mano no tiene demasiada variabilidad. Al depender de la fisiología del cuerpo humano la mano se separa como mucho un metro y medio de la cámara. Sin embargo, sí que presenta una gran variedad de poses. También se obtienen fondos muy distintos que podrían traer dificultades, pero estos fondos serán generalmente de interior, ya que la utilización de estos sistemas virtuales no suele darse en exteriores. La información será únicamente de color, no habrá información de profundidad o de calor.

Las poses de la mano son bastante variadas a pesar de que el entorno sea más o menos controlado. En éste, se parte de la base de pedir al usuario tener la mano extendida, abierta y mirando hacia la cámara situada en su cabeza. Por tanto, no se tendrán poses con oclusiones de dedos y de distintas zonas, sino que se encontrará una palma mirando a la cámara. Se dispone de cierta colaboración del usuario. A pesar de las condiciones de las que se parten, el número de poses es elevado. Esto se debe a que la mano tiene 3 posibles ángulos de giro que mezclados dan una alta cantidad de poses como se puede ver en la figura 3.1.



Fig. 3.1: Posibles ángulos de giro de la mano en el plano imagen.

Por las razones expuestas, se dedicará una parte importante de este proyecto a buscar un método de obtención de la pose de la mano que funcione adecuadamente bajo estas condiciones.

Otra característica que puede suponer un reto es la escasez de bases de datos que cumplan estas condiciones necesarias para un reconocimiento de este estilo. Se necesita una buena cantidad de imágenes en las que se presente la palma de los usuarios cumpliendo las condiciones de control básicas y así ser capaz de realizar las pruebas correctas del sistema. Por esta razón se ha llevado a cabo la obtención de una base de datos adecuada que es posible obtener gracias a que la información es únicamente de color, es decir, una imagen que puede obtener cualquier persona en su casa.

3.1 Algoritmos de detección de pose.

Como ya se ha dicho, es un algoritmo que detecta la posición de la mano mediante puntos en ella que forman un esqueleto o algunos puntos clave y que además funcione bien en las condiciones descritas anteriormente.

Todos los algoritmos que utilicen otra información como la profundidad no son válidos ya que no se dispone de esta información y, aunque quizás se pierda algo de rendimiento en la detección de la pose, se gana en versatilidad al ser un sistema más fácil y de bajo coste, ya que no necesitará sensor de profundidad si no únicamente una cámara.

Se han explorado dos métodos de detección de la pose mediante imágenes RGB:

El primer método que se ha investigado está en el artículo *Learning to Estimate 3D Hand Pose from Single RGB Images* [23], disponible en el siguiente [enlace](#). Consiste en representar la pose de la mano mediante las coordenadas (x,y,z) de 21 puntos clave de ésta. Estas coordenadas 3D son inferidas a partir de una imagen 2D mediante el uso de una red que estima las coordenadas normalizadas. Estas coordenadas se normalizan para que no dependa de la escala la red. Además, dichas coordenadas se hacen independientes a la translación por la imagen restándoles la localización de un punto definido. [23] El diagrama de bloques del funcionamiento total para obtener estos puntos se puede ver en la figura 3.2.

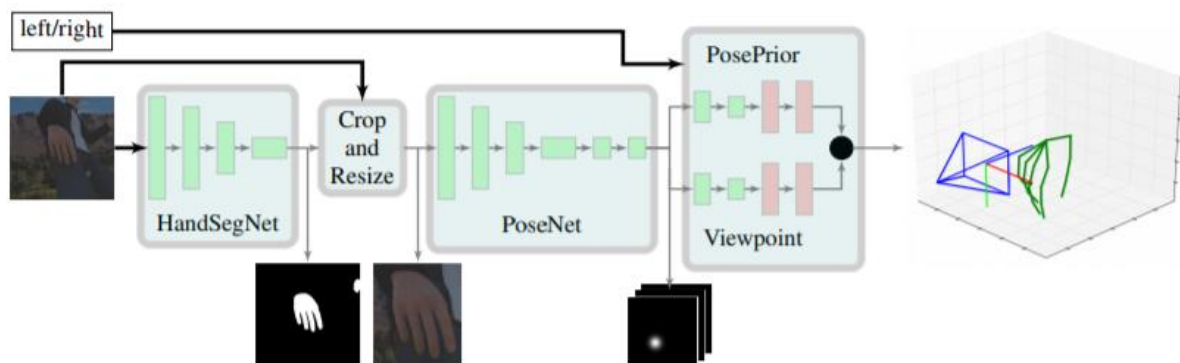


Fig. 3.2: Diagrama de bloques del código de detección de pose del artículo *Learning to Estimate 3D Hand Pose from Single RGB Images*. [23]

La imagen de entrada RGB, se introduce en el primer bloque. Éste localiza la mano mediante la red de segmentación HandSegNet. Con esto obtiene una máscara de la mano en la cual el valor más alto es el más probable de ser el centro de la mano. Seguidamente selecciona un cuadrado en el que se encuentra la mano en la imagen original. Este resultado se introduce en el siguiente bloque, la red PoseNet. El resultado de esta red es un conjunto de mapas de calor de puntos clave en la mano. Para finalizar, el conjunto de puntos clave se introduce en el PosePrior, siendo éste el último bloque. Esta red selecciona la combinación de puntos clave anteriores que es más probable que sea el esqueleto real de la mano. [23]

El segundo método explorado es el del artículo *3D Hand Shape and Pose Estimation from a Single RGB Image* [24] disponible en el siguiente [enlace](#). Este método es algo diferente ya que no contiene solo información de la posición de unos cuantos puntos en la mano, si no que aquí, mediante una red neuronal convolucional, se busca obtener una malla completa en 3 dimensiones de la cara de la mano. Esto, por tanto, tiene más información, ya que, a parte de los puntos que representan la pose, ha obtenido una malla de la cara de la mano. [24] Para ello, proponen su método con el siguiente diagrama de bloques de la figura 3.3.

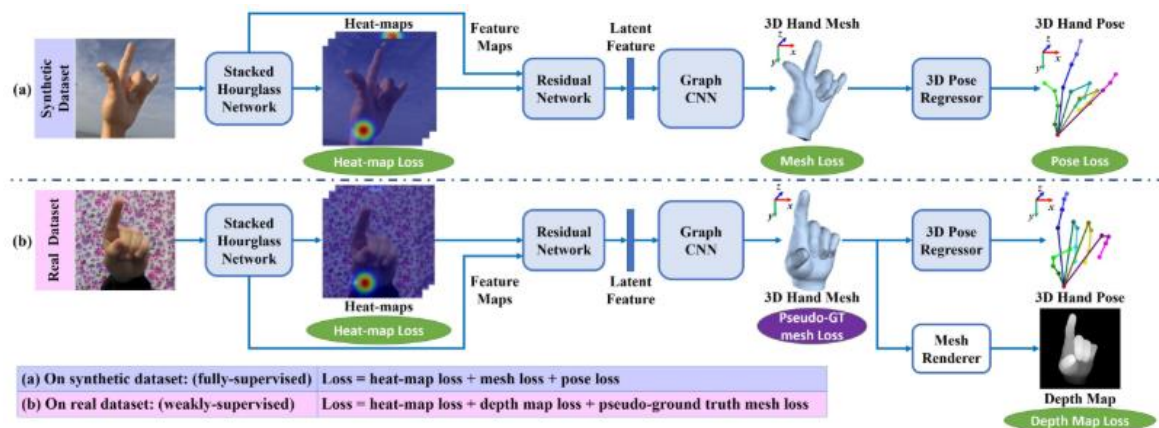


Fig. 3.3: Diagrama de bloques del código de detección de pose del *3D Hand Shape and Pose Estimation from a Single RGB Image* [24].

En la figura 3.3 se ven dos diagramas. Esto es debido a que la red ha sido entrenada de dos formas. En la parte a) se utilizó un entrenamiento más supervisado en el que se introducían los mapas de calor, los puntos 3D de la malla y la pose final. Así se hacía una supervisión completa de toda la red. Esto se realizó con el uso de imágenes sintéticas.

En la siguiente parte de entrenamiento hubo una supervisión más débil, siendo esta vez con imágenes reales. Aquí se introducían otros valores para la supervisión: el mapa de calor de nuevo, un mapa de profundidad y por último los puntos 3D de la malla. [24]

Para ello, en primer lugar, hay una **Stacked Hourglass Network**, en la cual se obtienen los mapas de calor y de características. Después la red residual combina y codifica los mapas de calor y características en un vector. A continuación, esto ya se introduce en la **Graph**

CNN mediante la cual son calculadas las coordenadas de los vértices de la malla. Por último, con una regresión lineal obtiene la pose de la mano. [24]

El primer método fue presentado en la ICCV en 2017, el segundo en la CVPR en 2019. Por tanto, ambos son buenos. Se procedió a probar con los dos, sin embargo, el segundo presentaba una incompatibilidad de librerías la cual no se pudo resolver. Por lo que se optó por utilizar únicamente el primer código.

Por otro lado, también se implementó desde cero un **método manual** de detección de la pose de la mano. En éste, la obtención de los puntos sobre la mano se realiza seleccionando el lugar en la imagen de cada punto.

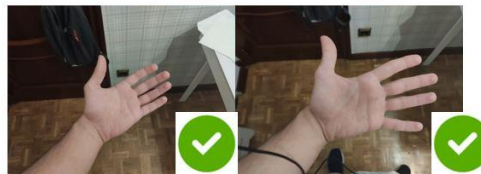
3.2 Base de datos.

Para obtener la base de datos se distribuyeron ciertas instrucciones a los usuarios. Las instrucciones fueron las siguientes: lo primero, tomar las imágenes desde un punto de vista egocéntrico a una de las dos manos, la que el usuario prefiriese, y que se viese la palma entera y sin cortes en los dedos. Con el fin de evitar problemas, se pidió que los dedos no estuviesen totalmente flexionados y así facilitar el método de detección de pose y que no se produjesen oclusiones de la palma por los dedos. También se solicitó variación de las posturas entre las imágenes. Se pidieron cinco imágenes por sesión y dos sesiones. En cuanto a condiciones de fondo de la imagen, no se pusieron restricciones con la intención de hacerlo más realista. Uno de los criterios de inclusión fue hacerlas con la mayor calidad posible y para que no disminuyese, las imágenes se debían enviar por correo electrónico sin disminuir el peso de las mismas. Con el objetivo de facilitar la comprensión de las instrucciones se envió la imagen de la figura 3.4, donde se explican los puntos principales y se dan cinco ejemplos visuales de cómo deben y cómo no deben realizarse.

Instrucciones:

Poner la cámara a la altura de la cabeza, de forma que simule unas gafas con cámara que visualicen las manos. A continuación, realizar 5 fotos de la mano de forma que se vea la palma. La pose que sea libre y natural mientras que se vea la palma de la mano. Importante que los dedos no se encuentren flexionados. **IMPORTANTE:** cambien la posición de la mano después de cada foto.

A continuación, os pongo dos ejemplos de imagen correcta:



Y unos ejemplos de imagen incorrecta.



Es decir, que no haya nada que ocluya la palma, que no falte ningún dedo en la imagen y que no esté demasiado cerca y con mala calidad ni a oscuras.

Enviadlas a este correo: imagenesmanotfg@gmail.com (por whatsapp pierden calidad)

Fig. 3.4: Instrucciones emitidas para la adquisición de la base de datos.

Finalmente, se adquirieron un total de 50 usuarios, con las dos sesiones de 5 imágenes. En total, 500 imágenes. En la figura 3.5 se pueden ver algunos ejemplos de estas imágenes de la base de datos obtenida.

El formato de las imágenes es el siguiente: “*persXX_imagY_sZ*”, siendo los valores posibles de XX desde 01 hasta 50, los de Y desde 1 hasta 5 y los de Z desde 1 hasta 2. De esta forma, todas las imágenes tienen el mismo formato y son fáciles de leer. Además, los números están siempre en las mismas posiciones de la cadena y son por tanto, fáciles de obtener.



Fig. 3.5: Algunas imágenes de la base de datos.

A pesar de los intentos para conseguir una buena resolución en las imágenes, debido a la calidad de la cámara de algunos móviles y a los fallos de enfoque de los usuarios, entre otras razones, la calidad de algunas de ellas no es demasiado elevada. Como se observa en el Anexo 3.1, hay imágenes con alta calidad, otras normal y algunas con mala calidad, aunque la mayoría son de una resolución intermedia.

3.3 Sistema completo diagrama de bloques.

A partir de la base de datos y el código de detección de esqueleto en la mano, se puede desarrollar la siguiente parte del sistema.

La figura 3.6 muestra el sistema que tiene los siguientes bloques:

- Primero las imágenes de la base de datos adquirida se introducen en el código de detección de esqueleto. Como salida de este bloque se obtienen los puntos detectados sobre la mano.
- Después estos puntos junto con la imagen entran en el segundo bloque, en el cual se extrae la palma de la mano de la imagen.
- Por último, estas palmas se dirigen al último bloque en el cual, cada palma se caracteriza con unos puntos y se miden las distancias entre los puntos para tratar de reconocer esa palma como un usuario y distinguirlo del resto.



Fig. 3.6: Diagrama de bloques del sistema a desarrollar.

4. DESARROLLO

Partiendo de las herramientas explicadas anteriormente, se deben desarrollar las distintas partes del diagrama de bloques que componen el sistema. Al tener dos métodos de detección de pose distintos, aunque el resto de los bloques funcionalmente serán iguales, dentro de ellos se pueden encontrar pequeñas diferencias adaptadas a cada método para conseguir un mejor rendimiento.

4.1 Bloque de detección de pose.

Como ya se ha explicado, existe un sistema automático de detección de pose el cual consta de una red de segmentación en primer lugar que centra la imagen en torno a la mano. La siguiente parte es otra red que detecta puntos clave en la imagen centrada de la mano. Para terminar, una capa selecciona el esqueleto más probable combinando los puntos clave localizados de la anterior parte. Con esto se obtienen 21 puntos del esqueleto que se guardan en una matriz, donde cada fila corresponde a las coordenadas de un punto. En la figura 4.1 se destacan los distintos puntos y su numeración correspondiente a la fila de la matriz a la que pertenecen. La figura 4.2 representa un resultado obtenido por este código, donde a parte del esqueleto hay un mapa de calor de la ubicación de la mano en la imagen.

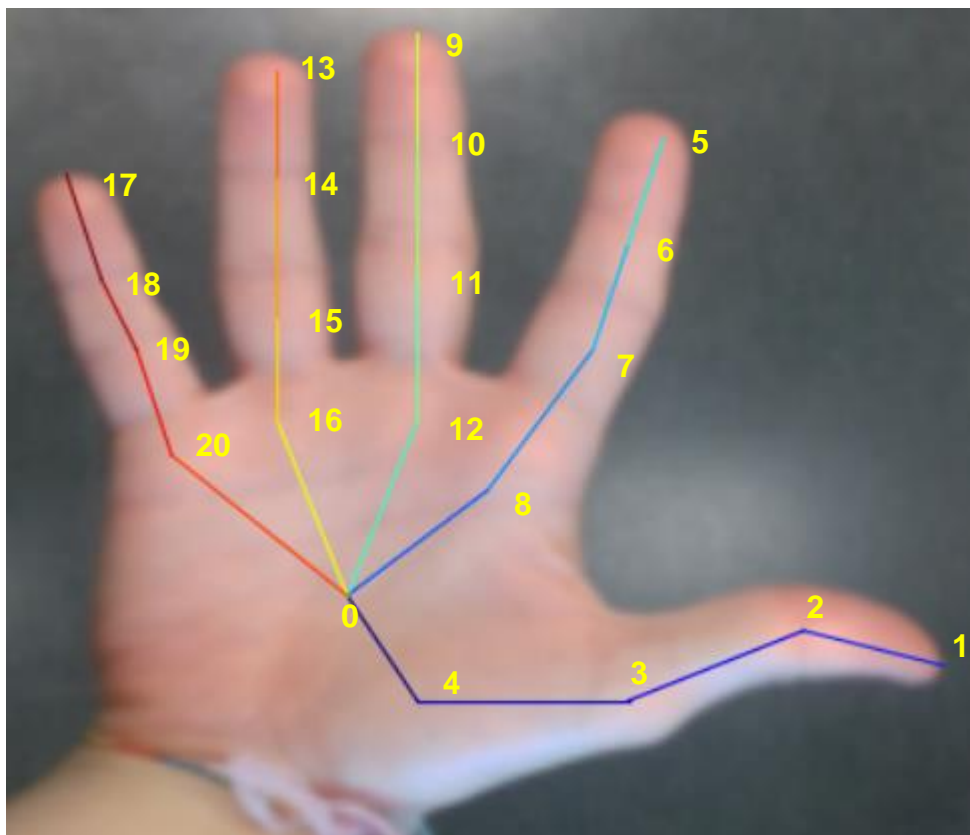


Fig. 4.1: Ejemplo de los puntos de la detección con el algoritmo de detección automática de pose.

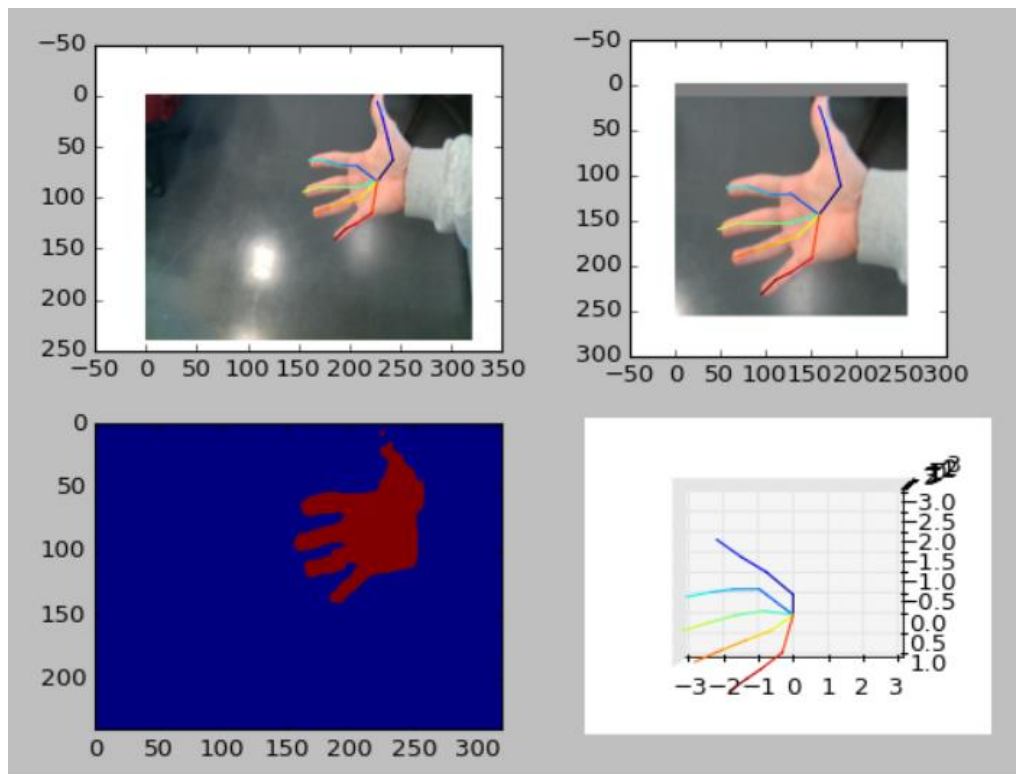


Fig. 4.2: Resultado del código de detección automática.

Antes de probarlo con la base de datos, se llevó a cabo un análisis para comprobar la viabilidad de este código en las condiciones descritas. Para ello se hicieron imágenes de una mano en un entorno controlado con un fondo oscuro y sin objetos que pudiesen dificultar la tarea, variando los posibles grados de libertad que tendría el usuario: los tres ángulos explicados en la figura 3.1 y la distancia a la cámara. Las imágenes del análisis se localizan en el anexo número 4.1.

Los resultados del análisis fueron los siguientes:

- Independientemente de la distancia, limitada en este caso a la longitud del brazo, el código en la primera parte centra la imagen en la palma y los resultados no se ven afectados. El código es capaz de funcionar con independencia de esta distancia. Sin embargo, sí que ocurrieron algunos fallos que, aunque parecen fallos aislados, al realizar un análisis de las imágenes que causaron estos problemas se puede observar que en ellas hay algún dedo o zona de la mano recortada que puede ser la causa.
- Para el ángulo 1, cuando este ángulo es muy grande sí que se produce fallo en la detección, pero este ángulo es relativamente alto y se puede asumir.
- En el ángulo 2, todas las imágenes las reconoce bien exceptuando algunos errores que, al contrario que en el caso de la distancia, son de algún punto aislado sin motivo aparente y ocasionales, lo que no parece suponer ningún problema.
- Con el ángulo 3 ocurre lo mismo que con el ángulo 1. A partir de cierto ángulo de inclinación no detecta correctamente el esqueleto en la mano. Éste no necesita ser tan elevado como el ángulo 1, lo cual sí puede llegar a ser más problemático, pero al contar con la colaboración del usuario, existe la posibilidad de minimizar este error.

Finalmente, este código ofrecía los 21 puntos que formaban el esqueleto sobre una imagen redimensionada y recortada de la original.

En los resultados se detectó un problema: se buscan las imágenes de mayor calidad posible obteniéndose la base de datos con esas intenciones, pero al mirar los resultados se observó que la imagen había perdido resolución, lo cual no interesaba. En la entrada de este sistema la imagen es redimensionada a 240,320 reduciéndose mucho su calidad. Además, la red de segmentación selecciona dentro de esa imagen redimensionada una zona más pequeña centrada en la mano. Al obtener los resultados de esta última imagen se pierde mucha resolución, deteriorando parte del trabajo de la obtención de la base de datos y dando lugar a unos resultados con una calidad demasiado baja para un posible reconocimiento.

Por ello, se trató de trasladar este esqueleto a la imagen original con toda su resolución. La red de segmentación ofrecía el centro y la escala empleada para la selección de la zona que contiene la palma, gracias a lo cual se pudo revertir este paso que realiza la primera parte de la red y se consiguieron las coordenadas sobre la imagen original pero redimensionada. Para solucionar esta redimensión simplemente se multiplicaron las coordenadas por el factor de reducción en cada eje, transformándolas en las correspondientes en la imagen original con la máxima resolución. Éste ahora es el resultado que saca este bloque, ya de mayor utilidad que el anterior.

También se utilizó el método manual, con el cual se obtienen 11 puntos sobre la palma en lugar de 21. Se realizó el etiquetado de éstos sobre todas las imágenes de la base de datos una a una. Se pueden ver estos puntos en la figura 4.3. En este caso, al seleccionar estos puntos directamente sobre la original, no hubo problemas de pérdida de calidad.

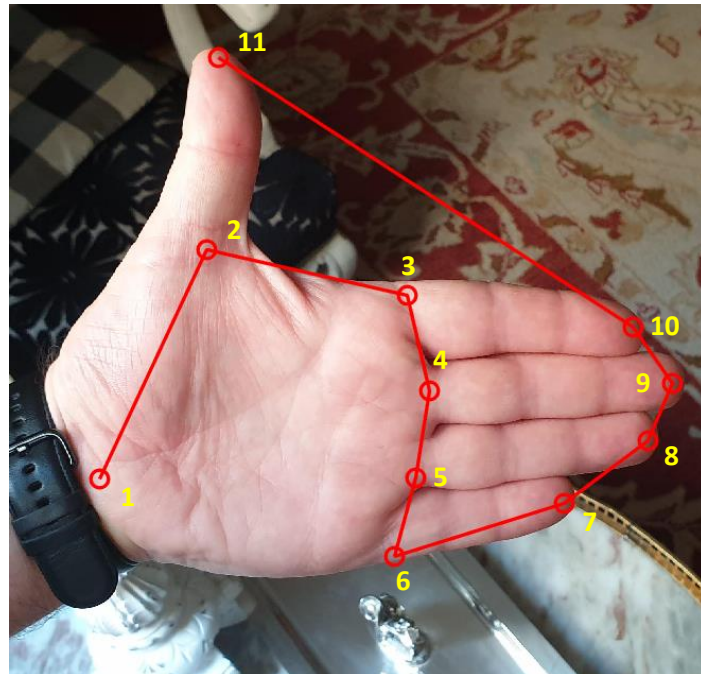


Fig. 4.3: Ejemplo de los puntos de la detección con el algoritmo de detección manual de pose.

4.2 Bloque de selección de palma

En este bloque, se implementaron dos métodos: uno para el método automático del bloque anterior y otro para el manual. Aunque se eligieron los puntos de la detección manual, para que pudiesen implementarse ambos de la misma forma, por ciertas características de los puntos del esqueleto de cada uno se adaptó este bloque para ajustarse mejor.

Esta selección de la palma cuenta con las siguientes partes:

1. Transformación geométrica.
2. Recorte palma.

4.2.1 Transformación geométrica

Lo que se intentó conseguir es eliminar la variación de poses posibles con una transformación geométrica, de tal forma que independientemente de la posición de la mano, el resultado fuera una palma sin rotación en ninguno de los tres ángulos posibles presentados en la figura 3.1. Es decir, una mano en una posición ideal en el plano de la imagen y con la misma orientación que en la figura del 4.1.

En las transformaciones geométricas se explicarán cinco tipos. Si bien es cierto que tres de ellos se podrían agrupar dentro de otro se tratarán por separado por razones que se expondrán más adelante. Los cinco tipos son: escalado, translación, rotación, afín y perspectiva.

La afín contiene, como se puede ver en la figura 4.4, el escalado, la translación, la rotación y además, la inclinación

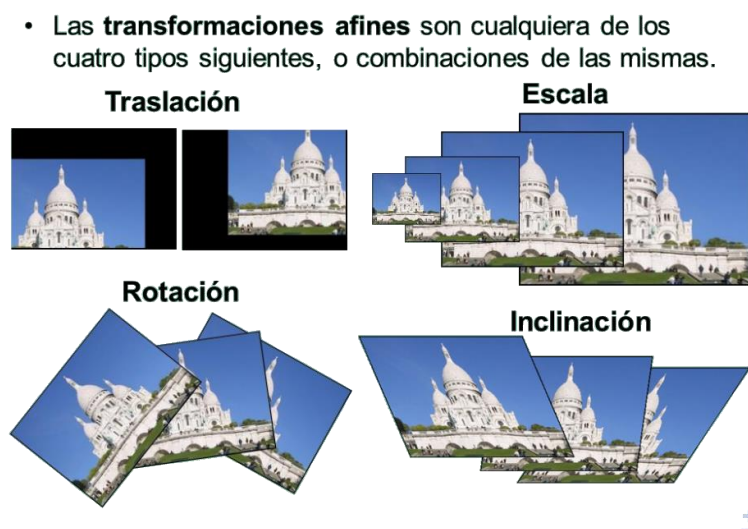


Fig. 4.4: Posibles transformaciones afines sobre una imagen. [26]

Por tanto, la afín puede ser la translación, escala, rotación y la inclinación independientemente o una combinación de ellas. Sea la combinación que sea una transformación afín siempre conserva la rectitud y el paralelismo. [25,26]

Por otro lado, existen las transformaciones perspectivas o bilineales. En ellas, las líneas rectas conservan la rectitud, pero no tienen por qué conservar el paralelismo como se ve en la figura 4.5 [25,26].

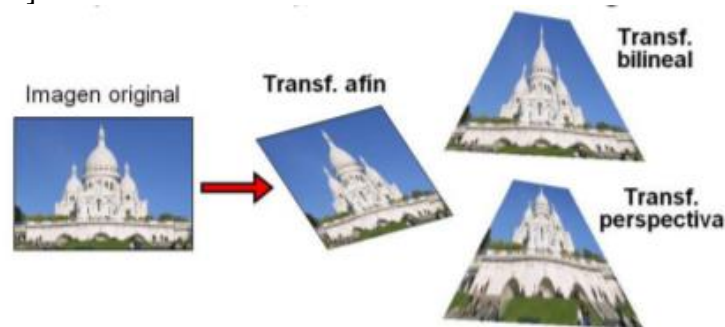


Fig. 4.5: Posibles transformaciones de la imagen. [26]

Para solucionar todas estas posibles transformaciones (tanto la afín como la perspectiva) se utilizó la librería CV2 de Python que tiene funciones para implementar estas transformaciones.

El funcionamiento consiste en escoger unos puntos en la imagen original y donde se quiere que acaben y, a partir de eso, crea la matriz de transformación con una función de la librería utilizada para realizar la transformación geométrica. Concretamente para esta transformación necesita cuatro puntos y sus correspondencias.

Para el método automático se usan los puntos de la mano que se ven en la figura del 4.1 que son los que se encuentran en la misma con este método.

Se exploraron varias opciones:

1. Usar los puntos 10,7,15 y 0 de la figura 4.1. La elección de estos puntos se basa en que son muy estables ya que en general no tienen mucha variabilidad debido a que los dedos corazón y anular son los que menos movilidad poseen. Al contrario que el pulgar que tiene mucha más libertad de movimientos y el índice y meñique que tiene algo menos que el pulgar, pero más que el anular y corazón. Si se cogiesen puntos con más variabilidad en distintas imágenes podrían encontrarse en sitios muy distintos y afectar demasiado a la transformación ya que sería muy difícil asignarles un punto en concreto donde terminar.

Los puntos de destino se han puesto respetando las proporciones de la mano y situándolos en una posición adecuada para conseguir el objetivo. Se observan dos resultados en la figura 4.6.

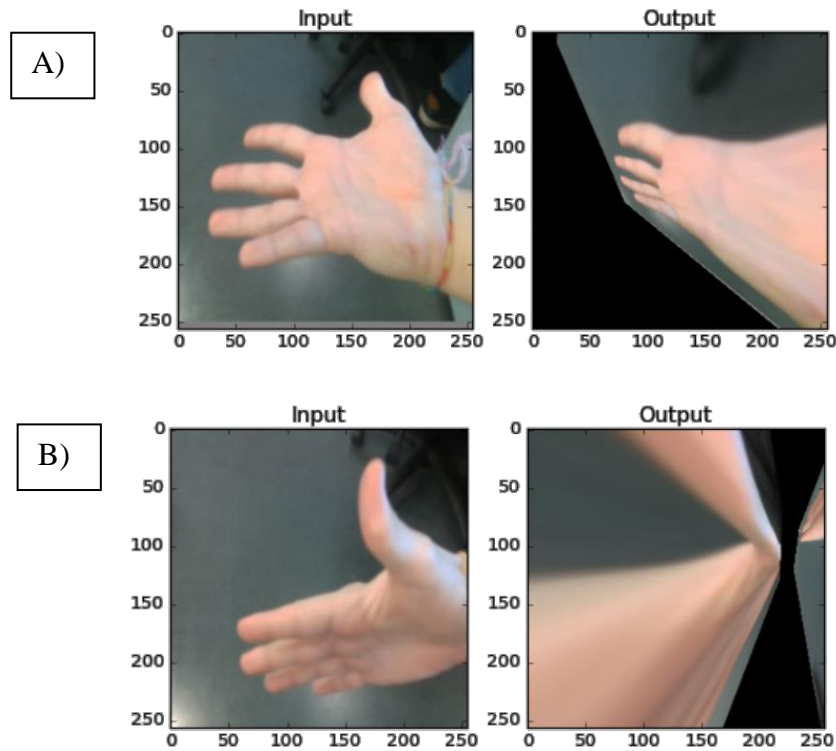


Fig. 4.6: Transformaciones geométricas sobre la palma.

En la parte a) de la figura 4.6 los resultados, aunque no son malos tampoco parece que mejoren la imagen original, no facilitando el acceso a la información de la palma.

Pero en la parte b) de la figura 4.6 se muestra lo que ocurre con una cantidad elevada de las imágenes. La imagen se deforma, no solo no mejorando si no empeorando el acceso a la información de la mano.

Ante estos resultados no muy prometedores decidió descartarse esta idea.

2. Esta segunda opción se implementó teniendo en cuenta que lo importante es la palma y no tanto la información de los dedos. Puesto que los dedos tienen, aunque algunos en mayor medida que otros, bastante variabilidad de posiciones. Por estas razones se optó por usar cuatro puntos situados en la palma de la mano y que cubran la mayor parte de ella.

Los puntos escogidos son el 0,4,12,20 de la figura 4.1 pero los resultados obtenidos tampoco son mejores que con los puntos anteriores y siguen, en algunos casos, dejando una información que no facilita el acceso a la información de la palma como en la parte a) de la figura 4.6. En otros, empeora el acceso a esta información porque deforman la imagen como en la parte b) de la misma figura.

Viendo que intentar aplicar en general las transformaciones geométricas no tiene resultados positivos y que conseguir unos resultados mejores podría suponer una solución demasiado compleja, se opta por otra idea: atender aquí únicamente el escalado, la translación y la

rotación para facilitar obtener la palma, e implementar la última etapa teniendo en cuenta las transformaciones que no se han podido solucionar. Por ello, anteriormente, estas tres se han tratado por separado de la afín a pesar de que forman parte de ella.

Tanto la escala como la translación son revertidas en el primer bloque, ya que los puntos de la mano se detectan igual de bien independientemente de si la imagen de la mano es más pequeña o más grande o de si está más en un lado de la imagen o en otro dentro de los límites impuestos por las condiciones del entorno virtual.

Para solventar únicamente la rotación hay que saber la dirección en la que se encuentra la mano. Aquí también se ven dos posibilidades:

1. El ángulo que se forma al unir los puntos de la base de los dedos anular y corazón y rotar la mano de manera que este ángulo sea 0° respecto a la horizontal.
2. Obtener el ángulo de la unión de dos puntos del dedo corazón y rotarlo de forma que este ángulo sea 90° respecto a la horizontal.

Estas dos opciones parecen adecuadas ya que ambas representan adecuadamente la dirección de la mano.

En la primera se ha escogido la base los dedos corazón y anular y no otros dedos. Esta decisión se debe a que parece que es la unión de dedos que mejor representa la dirección debido a que ambos empiezan a una altura casi igual y los otros tienen una altitud algo distinta.

En la segunda se ha elegido el dedo corazón en lugar de otro porque es el que menos libertad de movimiento tiene hacia los lados y mejor sigue la dirección de la mano.

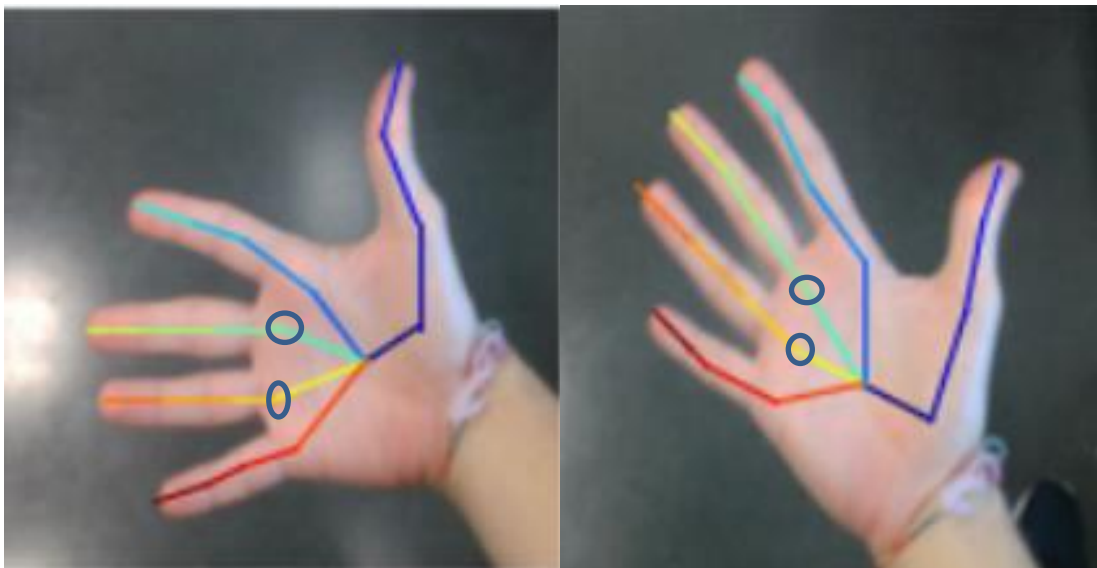


Fig. 4.7: Dos ejemplos de la base de los dedos índice y corazón

Aquí es donde se encuentra la primera diferencia entre el método manual y el automático. En el automático se decidió utilizar la segunda opción por cómo el código obtiene los puntos de la mano. En la figura 4.7 está el esqueleto del método automático en dos imágenes diferentes y se han rodeado los puntos que se utilizarían con la primera opción.

En la imagen de la izquierda sí que se representa bien el ángulo, pero en la de la derecha el ángulo es mayor al que realmente tiene la mano. Esto se debe a que el código no es demasiado preciso situando estos puntos justo en la base del dedo, si no que estos puntos pueden estar más abajo o más pegados de la base por lo que pueden falsear el ángulo como ocurre en esta situación.

Sin embargo, se puede apreciar que los puntos del dedo corazón los representa adecuadamente rectos y siguiendo una dirección representativa de la mano. En general no tiene problemas de desviación de estos puntos hacia los laterales que pueda afectar a este ángulo. Viendo más cantidad de imágenes se ha podido apreciar que la unión de los puntos 12 y 11 de la figura 4.1 son los que menor variación tienen y menos pueden falsear este ángulo. Por lo tanto, para este método automático la decisión tomada ha sido la segunda.

Para el método manual se ha decidido usar la primera opción, al seleccionar los puntos manualmente estos sí que se sitúan en la base del dedo con gran exactitud por lo que no dan un ángulo erróneo. Además, al solo haber seleccionado los 11 puntos de la figura 4.3 no hay un punto intermedio en el dedo corazón por lo que para el segundo método habría que unir la punta con la base y este ángulo que representa si puede llevar a errores más fácilmente si el dedo está un poco doblado o se confunde y se selecciona el punto un poco más a un lateral de lo debido. Por lo que para el bloque anterior de selección manual la forma de obtener el ángulo empleando la base de los dedos es mejor opción.

Una vez adquirido este ángulo con uno de estos dos métodos, simplemente se calcula la matriz de transformación, y con esta matriz y la imagen se obtiene la imagen rotada.

Tras la realización de este proceso, faltaría transformar las coordenadas del esqueleto, ya que ahora la imagen está rotada en base a estas coordenadas, pero el esqueleto continúa perteneciendo a la imagen anterior sin rotar. Para rotar estas coordenadas, se multiplican por la matriz de transformación calculada anteriormente y el resultado son las coordenadas rotadas.

Con esto se logran las transformaciones propuestas solucionadas sobre la imagen y coordenadas con lo que va a ayudar a seleccionar la palma más fácilmente y a introducir en el siguiente bloque información algo más estandarizada. En la figura 4.8 se puede ver esta rotación aplicada sobre el método automático, y en la figura 4.9 sobre el método manual.

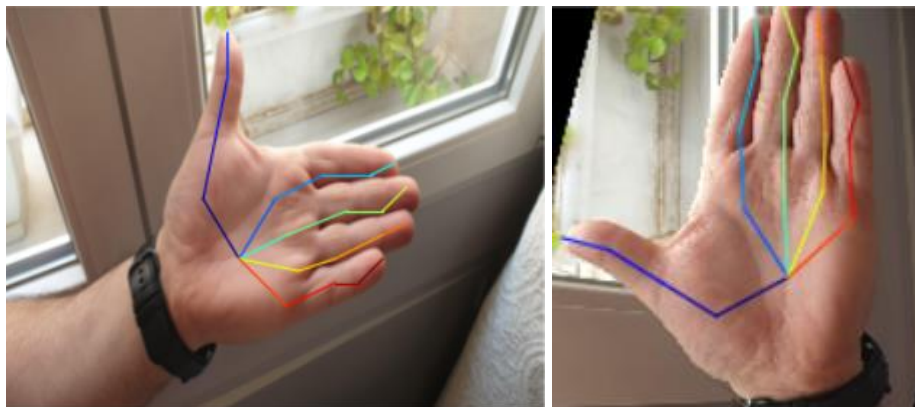


Fig. 4.8: Rotación con el método automático.

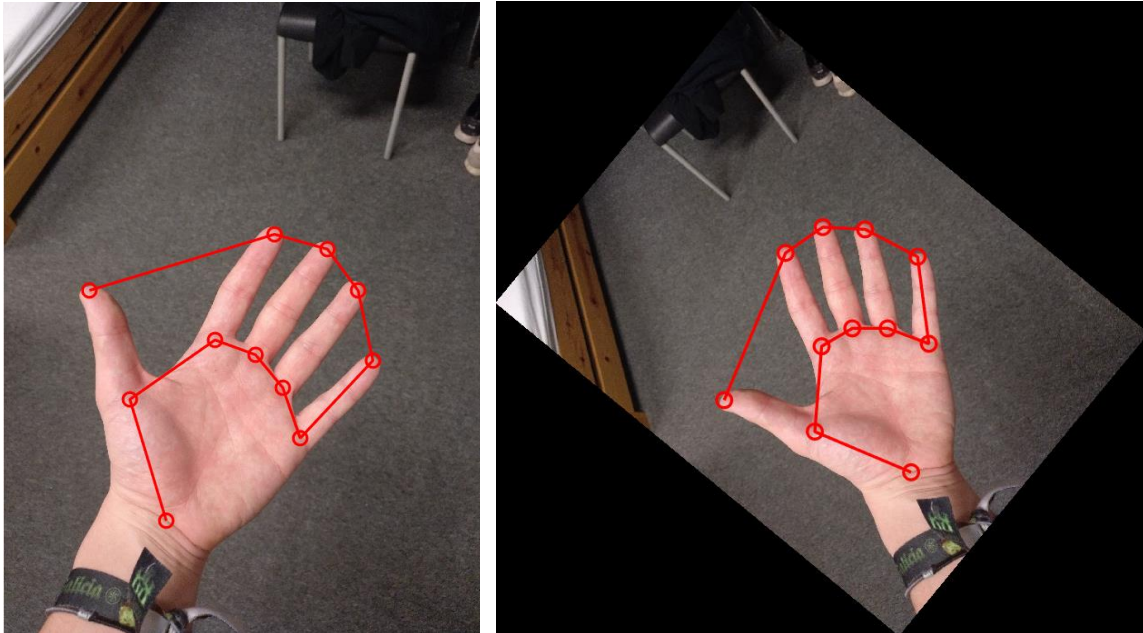


Fig. 4.9: Rotación con el método manual.

4.2.2 Recorte de la palma

Respecto al recorte de la palma también existen diferencias entre el método manual y el automático.

En el automático se probó un primer método que no dio los resultados esperados. Este método consistió en usar la distancia entre la base de los dedos corazón y anular debido a que, aunque para el ángulo la base pudiese estar un poco distorsionada, para esta distancia esa distorsión no afecta y va a permitir obtener una idea de las dimensiones de la mano para poder seleccionar correctamente la palma. Esta distancia se multiplica por tres ya que es aproximadamente el tamaño que va a tener la palma. Y posteriormente, se selecciona un cuadrado con el lado igual a la distancia anterior alrededor del punto 0 de la figura 4.1.

Tras probar con la base de datos de usuarios, se puede apreciar que el punto 0 de la mano tiene una gran variabilidad. Como se puede ver en la figura 4.10, en ocasiones se encuentra más abajo y otras más arriba, pero no siempre se localiza en el centro de la mano, y por tanto el método anterior no puede utilizarse ya que en algunos casos no selecciona bien la palma.

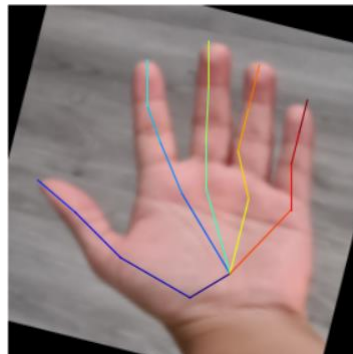


Fig. 4.10: Ejemplo del código automático de detección de pose sobre la base de datos.

Otra limitación que se ha encontrado es la gran variabilidad que presentan las manos, puesto que existen manos que proporcionalmente tienen una palma muy larga en comparación con su anchura o al contrario, tanto por anatomía como quizás por la proyección de la mano en el plano imagen provocada por la pose. Esta proyección no fue solucionada con la transformación geométrica, únicamente se implementó la translación, escala y rotación. Fijándose solo en la anchura de la mano y seleccionando un cuadrado igual en los dos ejes, que es en lo que consiste este método, no es lo más adecuado ya que la palma puede tener dimensiones rectangulares.

Debido a estos problemas se ha decidido explorar una opción que solucione lo anterior, y que no se base en el punto 0 de la mano (figura 4.1) y en la anchura de la mano únicamente.

Esta opción consiste en coger como límites la posición x de los puntos 8 y 20 y la posición y de los puntos 16 y 4 (figura 4.1) que son, en general, los puntos que se encuentran más en el extremo de la palma y generar un rectángulo con ellos. Posteriormente, se les suma o resta una pequeña distancia proporcional a la distancia entre ellos ya que estos puntos no se encuentran justo en el extremo y eliminan en el recorte parte de la palma. Estas distancias se han ajustado con la observación de las imágenes de la base de datos. Se han obtenido las siguientes:

- A la coordenada del punto 8 se le suma o resta, en función de si es la mano derecha o la izquierda, la distancia entre la coordenada del punto 8 y del 20 dividida por 4. A la coordenada del punto 20 se le suma o resta también, dependiendo si es mano izquierda o derecha, la distancia entre los dos puntos multiplicada por $7/16$.
- Para el par 16 y 4 se resta y suma respectivamente la distancia entre ambos dividida por 3.

Con estas posiciones x e y selecciona el rectángulo en la imagen que vaya desde la coordenada " x " más pequeña a la más grande y de la coordenada " y " más pequeña a la más grande. En general este recorte suele coger información del fondo, pero si las medidas que se suman fuesen más pequeñas en muchas ocasiones se perdería parte de la palma por un lado o por el otro.

Al observar los resultados se comprobó que había algunas imágenes de las que no se pudo obtener un recorte de la palma adecuado. Esto se debe a un error del código automático, el cual detecta mal los puntos en esas imágenes. Dependiendo del error que haya en la detección de los puntos se obtienen unos recortes mejores o peores. En el anexo 4.2 se recogen los distintos casos.

Se debe destacar que en aquellas manos en las que se detectan bien los puntos de la palma, el recorte se realiza correctamente y también hay ocasiones en las que, a pesar de que haya puntos que detecte mal, el recorte de la palma es correcto siempre y cuando los puntos utilizados para el recorte estén bien localizados, como se puede ver en el caso 1 y en el caso 2 del anexo 4.2.

Cuando alguno de los puntos utilizados para el recorte se encuentra mal situado pero el resto son correctos, como es el caso 3 (anexo 4.2), se produce un recorte que, aunque se sitúa en la palma, no es adecuado.

Otro caso es aquel en el que los puntos del dedo corazón que se empleaban para la rotación en este método no se hayan ubicado correctamente y por tanto la rotación no se haya podido solucionar. Esto se puede ver en el caso 4 (anexo 4.2). Aquí el recorte de la palma también se encuentra desviado, igual que el caso anterior, y no es conveniente utilizarlo.

El caso número 5 (anexo 4.2) representa aquel en el que el método de detección de pose no ha ubicado bien ningún punto de la mano. El recorte en este caso suele ser el peor ya que ni siquiera incluye la palma.

En el caso 3, 4 o 5 se toma la decisión de eliminar la imagen ya que no es apta para el siguiente bloque, sobre todo el caso número 5. Para eliminar estas imágenes se empleó la relación de aspecto de las imágenes dividiendo la altura de la misma por su anchura. Tras analizar gran parte de las imágenes de la base de datos, se observó que una relación de aspecto por debajo de 0.6 o por encima de 2.9 en la mayoría de los casos corresponde con una imagen de los casos que se deben eliminar y se procedió a calcular en todos los recortes este valor y a eliminar aquellas imágenes que cumpliesen dicha condición. En la figura 4.11 se ve un histograma con la relación de aspecto de las imágenes de la base de datos. El porcentaje de palmas de toda la base de datos con este método que no cumplen las condiciones y que por tanto no pasan al bloque de reconocimiento es de un 22.2%. A pesar de esta medida hay imágenes correctas que se eliminan y algunas incorrectas que no se consiguen excluir con este método. Sin embargo, son una minoría de ellas y se pueden despreciar. En la figura 4.12 se pueden ver algunos de los recortes obtenidos.

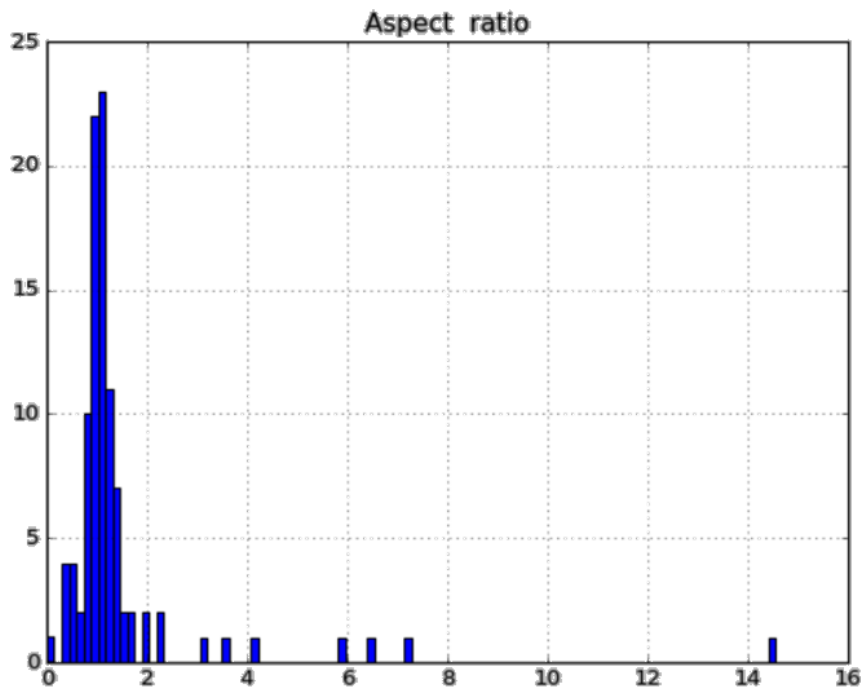


Fig. 4.11: Histograma de la relación de aspecto de los recortes de la base de datos con el método automático.

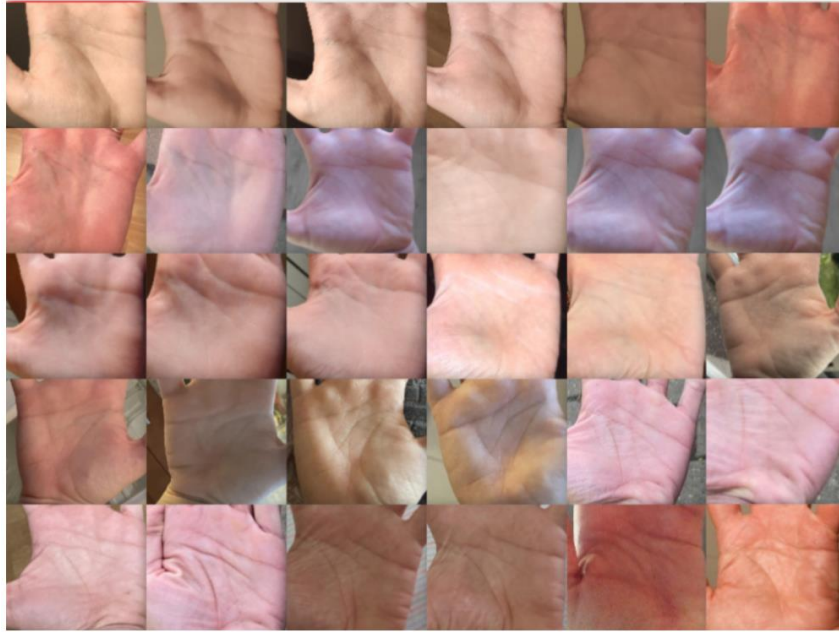


Fig. 4.12: Recorte de palma del método automático.

Para el método manual el método de selección que se ha utilizado ha sido diferente. Por las características que tienen los puntos, se han seleccionado las coordenadas “ x ” de los puntos 3 y 6 y las coordenadas “ y ” de los puntos 4 y 1. A continuación, se selecciona el rectángulo de la imagen igual que antes basado en estas coordenadas sin necesidad de sumar o restar nada ya que la selección de estos puntos está bien situada en el extremo. Se muestran los recortes obtenidos con este método en la figura 4.13.



Fig. 4.13: Recortes de palma del método manual.

En este caso al ser la selección de los puntos manual no hay problemas con el recorte de la palma, resultando todos los recortes válidos y no siendo necesario eliminar ninguna.

4.3 Bloque de comparación.

El bloque de comparación es el último bloque del sistema en el cual, ya una vez se tienen las palmas, se trata de identificarlas. A este punto del sistema, dependiendo de si se ha hecho la detección de pose con el método manual o con el método automático, llegan unos datos u otros, pero la implementación de este bloque es exactamente igual para ambos. Si viene del método manual, llegan el 100% de las imágenes. Por otro lado, si se parte del automático, aunque este no requiere la ayuda de un usuario, también baja la eficiencia y hay algunas palmas que no detecta bien y que, por tanto, descarta el bloque anterior, concretamente llegan un 77,8% de las palmas como se mencionó previamente.

Lo primero que se realiza es redimensionar la imagen y recortar los extremos de las mismas, para así quedarse con una parte central homogénea y que no afecte tanto la parte previa a este bloque. Si las imágenes vienen del método automático suelen incluir más información de fondo y de esta forma se mitiga dicha diferencia. Para la comparación, en primer lugar, se extraen regiones de interés sobre la palma. Existen distintos métodos para llevarlo a cabo que pueden resultar más o menos apropiadas en función de las condiciones. A continuación, se explican brevemente los utilizados en este proyecto.

- SURF: Este método se obtuvo mezclando información de localización de características de distribución de gradientes. El primer paso es fijar información de orientación en una zona circular alrededor del punto. Y después se extrae una zona cuadrada de acuerdo con la orientación antes obtenida. [27]
- Harris: Lo que se propuso en este método fue extraer la información de las esquinas. Pero esta información no era suficiente para la tarea de caracterizar correctamente una imagen porque no tenían información de conectividad entre ellas por lo que también incluyó el obtener la información de los bordes. [28]
- MSER: Este método se enfoca en poder encontrar correspondencias entre una misma imagen tomada desde dos puntos de vista diferentes. Su funcionamiento se basa en ordenar los píxeles por intensidad y luego seleccionar regiones las cuales estén conectadas por la intensidad que tienen. [29]
- FAST: Este detector de puntos lo que hace es, en primer lugar, seleccionar las esquinas de la imagen a partir de un punto y la intensidad de los puntos que le rodean. Estas esquinas son los puntos clave de los que va a obtener la región, que serán un conjunto de píxeles alrededor de esta esquina. [30]

Estos métodos extraen regiones muy robustas alrededor de ciertos puntos clave, y permiten su funcionamiento a pesar de que haya transformaciones geométricas en las imágenes, absorbiendo de esta manera aquí el problema de las transformaciones geométricas que se viene arrastrando desde la primera parte del segundo bloque del sistema. Para implementarlos, se han utilizado las funciones correspondientes en MatLab.

Las características de estas regiones son las que se emplean para el reconocimiento. En primer lugar, se obtienen los puntos clave y las características de la mano que se quiere reconocer. Se recorren el resto de las imágenes de la base de datos y se calculan también sus puntos con sus características. Para cada par de manos se hace la distancia euclídea de todos los puntos con todos. Seguidamente, se calcula la media de estas distancias. Los puntos cuya distancia este por encima de esta media son los que se consideran como puntos reconocidos entre las manos. Idealmente a partir de un cierto número de puntos en la mano es donde se podría decir que las dos manos han sido reconocidas.

5. INTEGRACIÓN PRUEBAS Y RESULTADOS.

Mencionar, en primer lugar, que la base de datos, que es la entrada del sistema, es muy adecuada para este entorno y cumple las condiciones requeridas para poder probarlo de manera correcta.

Los resultados conseguidos con los distintos métodos no son iguales, y para cada bloque se tendrá una salida u otra dependiendo del método seleccionado. A continuación, se realiza un análisis de cada fase y de los resultados que obtiene, y se muestra el proceso que experimenta una entrada de la base de datos a lo largo del sistema.

Primer bloque:

Los resultados de la detección de pose de la mano son bastante prometedores en ambos casos. En el procedimiento manual todas las palmas han sido etiquetadas una por una marcando los 11 puntos indicados en la figura 4.3 y los resultados son correctos e indican bien la posición de la mano para todas las imágenes de la base de datos. En la figura 5.1 se pueden ver dos ejemplos del etiquetado para este método manual. El etiquetado automático en cambio detecta 21 puntos, indicados en la figura 4.1. En la figura 5.2 se encuentran también los dos etiquetados con este método de la misma palma.

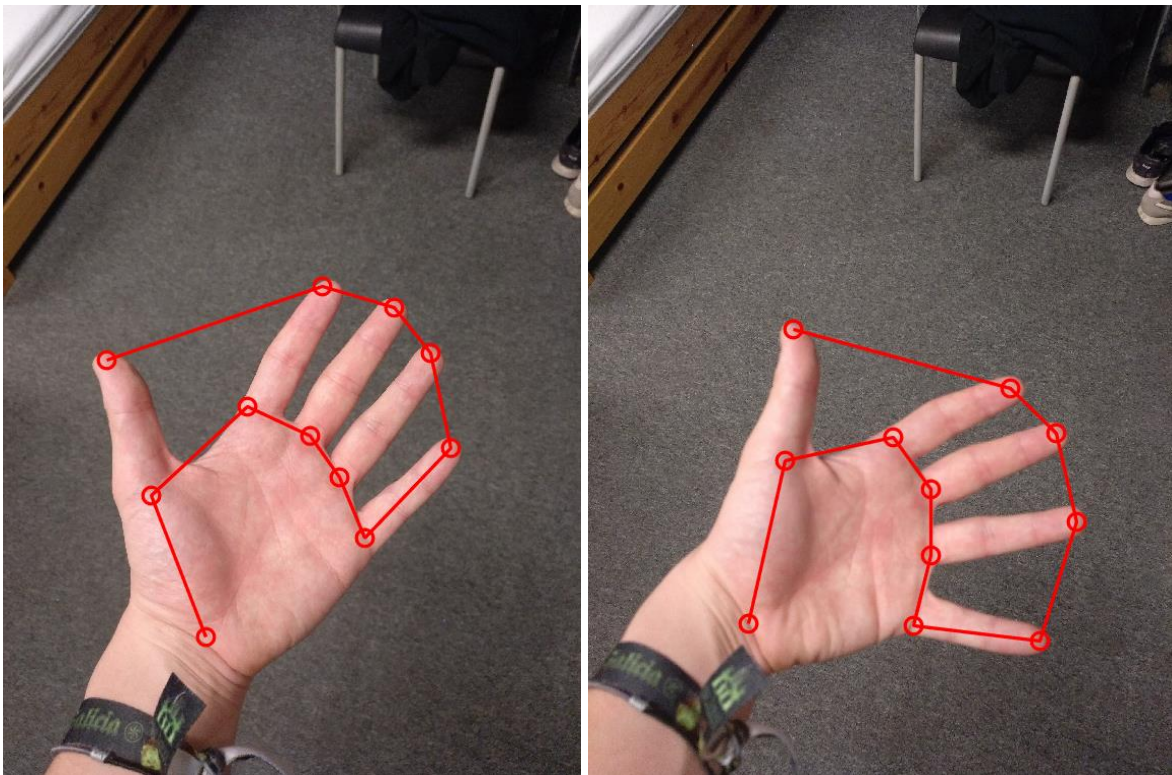


Fig. 5.1: Etiquetado manual de dos imágenes de la base de datos.

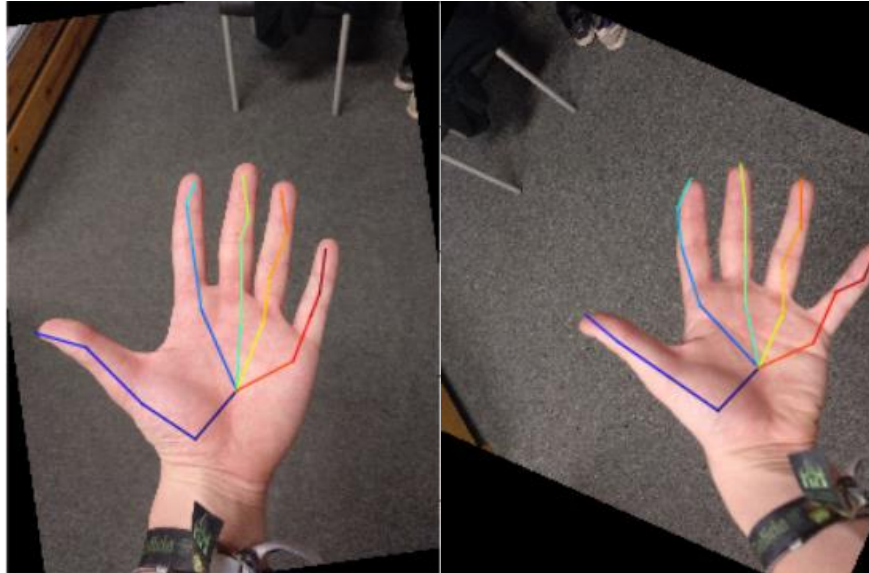


Fig. 5.2: Etiquetado automático de dos imágenes de la base de datos.

Con el método automático se obtienen más puntos que pueden ayudar a conocer mejor la pose de la mano. Además tiene información del centro de la misma que puede resultar útil en algunas aplicaciones, sin embargo, estos puntos tienen algo de desviación, es decir, se encuentran en un área relativamente grande alrededor de su posición ideal y esto conlleva ciertos problemas.

Por otro lado, el automático elimina la necesidad de etiquetar cada imagen de la mano que entra, lo que puede limitar mucho el sistema, ya que impide que sea autónomo. Pero tiene el problema del error en la detección. Hay en torno a un 20% de las imágenes que no se consiguen etiquetar de manera óptima.

Segundo bloque:

El recorte de la palma en base a la pose detectada en la parte anterior funciona de forma óptima en ambos métodos seleccionando de forma muy precisa la palma de la mano, que es lo que interesa para el siguiente bloque. Se puede ver en la figura 5.3 la selección de la primera palma de las figuras 5.1 y 5.2 a la que se ha llegado con los diferentes métodos. A la izquierda está el método automático y a la derecha el manual. Como se puede ver, el recorte de la palma no es idéntico, pero sí muy parecido.



Fig. 5.3: Recortes de la palma con el método automático y el manual.

Para ambos métodos funciona adecuadamente el recorte. Sin embargo, para el automático como los puntos no se encuentran tan establemente colocados, si alguno de los puntos se desvía de su posición ideal, lo cual es bastante habitual, éste puede recortar un poco más de fondo o perder un poco de mano. En general, suelen coger cierta información del fondo debido a que la desviación en los puntos si no ocurre hacia un lado ocurre hacia el otro y, para evitar pérdida de información de la mano que pudiera ser de utilidad, se tomaron las medidas de recorte más grandes, como ya se explicó con anterioridad. Este efecto se ve fácilmente comparando las figuras 4.12 y 4.13. El método manual presenta resultados muy estables teniendo, en general, mayor proporción de palma y menor de fondo, englobando mejor la palma.

Tercer bloque:

Los resultados de comparación obtenidos no van a ser exactamente iguales, ya que las imágenes que llegan al último bloque son diferentes. Aun así, a pesar de que ambos presentan la misma distribución en los resultados, se va a poder analizar algo más con el método manual ya que obtiene mayor número de palmas. Independientemente de las diferencias generales ya nombradas entre las palmas de los distintos métodos, los resultados no se han visto afectados gracias al recorte que se realizó al entrar en este tercer bloque que eliminaba los bordes quedándose con el centro de la imagen.

A continuación, en la figura 5.4 se muestran los resultados que se obtienen a partir de la detección de pose manual. Estos resultados consisten en un histograma del número de puntos, pero representado como una gráfica continua en lugar de como un diagrama de barras. Se muestra un histograma para las muestras genuinas, representado en azul, que son aquellas que deberían resultar correctas puesto que son comparaciones de las manos de un mismo usuario. Se muestra otro histograma para las impostoras, representado en rojo, obtenido de comparaciones con manos que no corresponden a un mismo usuario. Estos dos valores han sido normalizados ya que los resultados impostores tienen muchos más valores, y para poder representar las dos distribuciones juntas y realizar una mejor comparación es necesario normalizar. Los resultados para la detección de pose automática son muy parecidos conservando la distribución, aunque se diferencian en los valores concretos debido a que el número de palmas que entran y las palmas en sí son distintas.

Se muestra que con los métodos SURF, MSERF Y FAST el resultado es muy parecido, presentando una distribución prácticamente igual en los tres. El método Harris, sin embargo, presenta otra distribución diferente que no parece que aporte ninguna información útil para tratar de separar ambas curvas y así poder realizar el reconocimiento.

Se va a analizar el método SURF, aunque los tres ofrecen conclusiones similares. Aquí el valor más alto de los genuinos es 9474 y el de los impostores 8243. Si se calcula la media, el dato para los impostores es 356 mientras que para los genuinos es de 587. Por lo que, efectivamente parece que tienen valores algo más elevados.

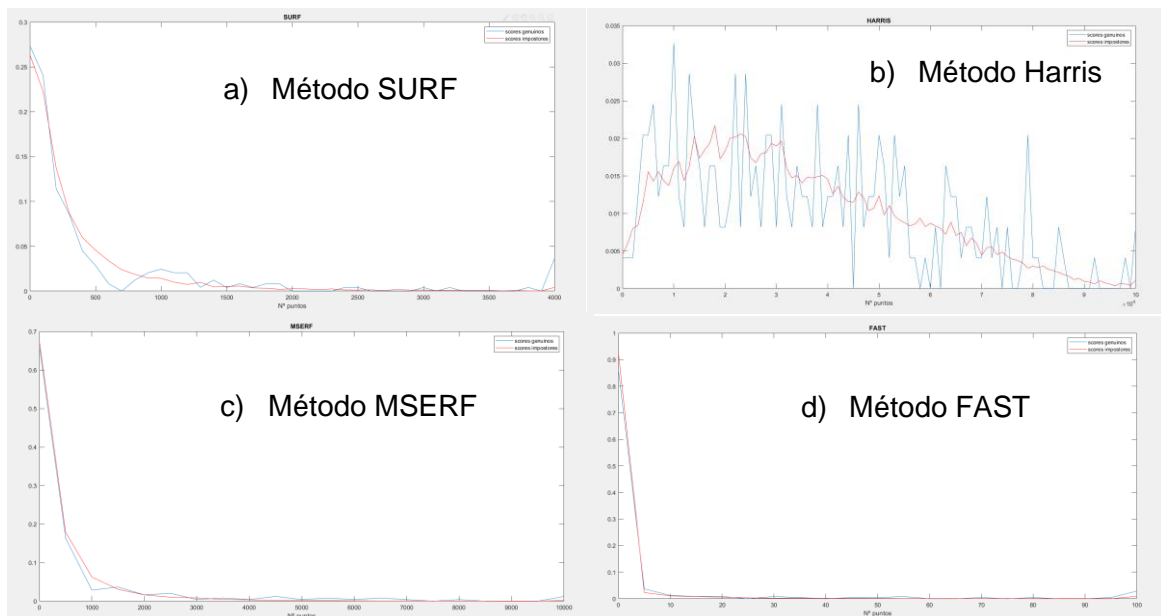


Fig. 5.4: Distribución de número de puntos detectados entre palmas para cada uno de los métodos de comparación con los resultados manuales.

Además, el segundo valor de la gráfica en los genuinos es, en relación al primero, más elevado que en los impostores. En valores de mayor magnitud los impostores no presentan prácticamente ningún valor mientras que los genuinos sí tienen un mayor número.

Todo apunta a que, sí que existe una pequeña diferencia entre ambas distribuciones, pero que no es suficiente para poder separarlas y acertar con una probabilidad razonable en un reconocimiento.

Respecto al método automático, en la figura 5.5 se observa el resultado que se obtiene para el método SURF, que es en el que se centra el método manual, y tiene una distribución que, como se mencionó con anterioridad, es similar a la obtenida con el mismo. Aunque la gráfica no sea exactamente igual, las características y el análisis son parecidos. Para el resto de los métodos también las gráficas presentan la misma distribución.

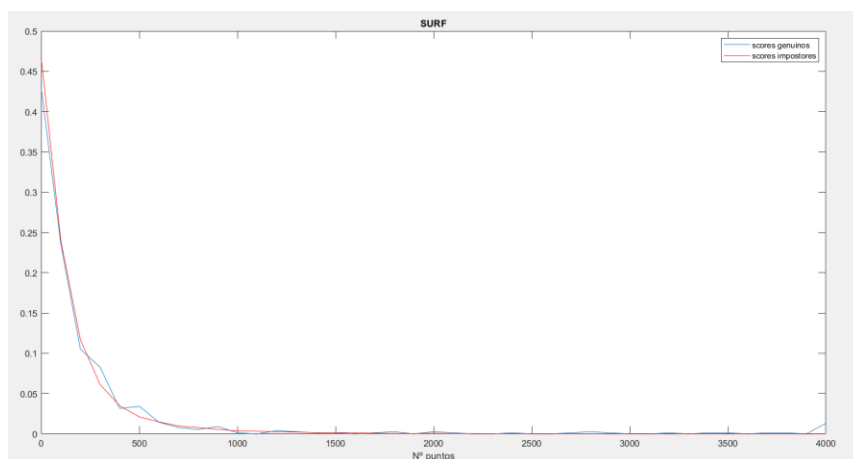


Fig. 5.5: Distribución SURF de número de puntos detectados entre palmas para cada uno de los métodos de comparación con los resultados automáticos.

A pesar de que durante el sistema se ha mantenido la mayor calidad posible en las imágenes, el método de reconocimiento empleado no es capaz, para las imágenes obtenidas, de alcanzar unos resultados que permitan un reconocimiento fiable. Empleando otro método o con unas imágenes de mayor calidad, los resultados podrían verse mejorados.

6. CONCLUSIONES Y TRABAJO FUTURO.

6.1 Conclusiones.

Hoy en día el reconocimiento a través de internet es cada vez más necesario, ya que hay un aumento de tareas que se pueden realizar a través de este medio y que necesitan de identificación por parte del usuario. Los métodos biométricos han sustituido a los tradicionales, basados en contraseñas o algún dato previamente conocido por el usuario, por su comodidad y seguridad.

En concreto, este trabajo se ha centrado en esta identificación en entornos virtuales generalmente basados en una cámara, en este caso únicamente capaz de obtener una imagen RGB. Este ámbito, en gran desarrollo, tiene un gran potencial y en base a las particularidades y a las posibles aplicaciones que puede tener se ha decidido realizar un sistema biométrico basado en el rasgo al que más fácil se puede acceder, la palma.

El reconocimiento palmar en este entorno tiene ciertas características que lo hacen particular y lo diferencian de otros reconocimientos de este tipo. Presenta una distancia, una perspectiva y un ángulo limitados. Además, se puede contar con la colaboración del usuario que tiene que facilitar que se vea la palma dentro de sus posibilidades.

El trabajo desarrollado cuenta con tres bloques principales y con la adquisición de una base de datos que cumple estas condiciones, creando un sistema completo. El primer bloque es para obtener la posición de la mano mediante un esqueleto sobre ella. Este bloque se ha realizado con dos métodos distintos, uno manual y uno automático. El automático es más útil para crear un entorno independiente sin necesidad de ser supervisado, pero también, menos eficaz debido a que hay algunas palmas en las que el esqueleto no está bien ubicado y por tanto el sistema no funciona. A continuación, un segundo bloque de extracción de la palma de la mano. Y para terminar el último bloque centrado en identificación.

La base de datos recogida es adecuada y cumple las condiciones necesarias impuestas, simulando correctamente el entorno requerido.

La salida del primer y segundo bloque tiene un funcionamiento muy prometedor para los dos métodos, aunque cada uno tiene sus puntos fuertes: el manual tiene un funcionamiento perfecto, pero requiere de supervisión humana; y, el automático tienen una eficiencia de entorno a un 80%, que es menos que para el manual, pero no requiere de supervisión humana por lo que aumenta su utilidad.

El tercer bloque, sin embargo, no ha conseguido llegar a un reconocimiento de usuarios dando unos resultados muy parecidos para manos de distintos usuarios y de los mismos usuarios.

El sistema se considera como una primera aproximación para este tipo de reconocimiento y como una alternativa viable, de bajo coste, muy versátil y con gran proyección.

6.2 Trabajo futuro y posibles mejoras.

En el futuro, el trabajo en reconocimiento biométrico va a seguir progresando, incluyendo el reconocimiento palmar. En cuanto a los entornos virtuales la necesidad de reconocimiento va a hacer necesario el avance en este ámbito. Este sistema es un buen punto de partida a un análisis del problema y a una solución mejorable en diversos aspectos.

El primer punto para mejorar es el código de detección automática de pose, aunque tiene un funcionamiento aceptable, no es perfecto y tiene un porcentaje de entorno a un 20% de fallo en la detección de la pose, el cual se podría disminuir. Aunque se propone también un método manual en la detección de la pose que implica no tener fallos si se seleccionan bien los puntos, para conseguir un sistema automático se necesita una detección de pose autónoma.

Sin embargo, el punto importante a mejorar es el bloque de comparación. La palma tiene muchas características para llevar a cabo el reconocimiento. En este sistema solo se adquiere una imagen RGB y por tanto solo se puede adquirir la información presente en ella. En estas imágenes hay gran cantidad de características, pero debido a la calidad de algunas imágenes no son adquiribles en nuestra implementación. A pesar de esto con los datos que se tienen se podría implementar algún método que mejorase el funcionamiento de este bloque y por lo tanto del sistema. Existiendo la posibilidad de utilizar otras características como por ejemplo la textura.

En cuanto al segundo bloque, el recorte de la mano, no hay prácticamente margen de mejora, ya que cuando este falla es por un mal etiquetado del primer bloque.

Se abre camino a mejorar estos puntos para conseguir un sistema funcional completo.

BIBLIOGRAFIA

- [1] SERRATOSA, Francesc. La biometría para la identificación de las personas. *Universitat Oberta de Catalunya*, 2008, p. 8-20.
- [2] WAYMAN, James, et al. An introduction to biometric authentication systems. En *Biometric Systems*. Springer, London, 2005. p. 1-20.
- [3] Z. Sun, T. Tan, Y. Wang, y S. Z. Li, “Ordinal palmprint representation for personal identification,” en Proc. CVPR, Jun. 2005, pp. 279–284.
- [4] MATKOWSKI, Wojciech Michal; CHAI, Tingting; KONG, Adams Wai Kin. Palmprint recognition in uncontrolled and uncooperative environment. *IEEE Transactions on Information Forensics and Security*, 2019, vol. 15, p. 1601-1615.
- [5] D. Zhang, W.-K. Kong, J. You, and M. Wong, “Online palmprint identification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1041–1050, Sep. 2003.
- [6] FIGUEROA, Raúl Aguilar; MONTERO, Raúl Santiago; AZUELA, Juan Humberto Sossa. Estudio comparativo del reconocimiento de rostros térmicos basado en características invariantes. *Research in Computing Science*, 2018, vol. 147, p. 215-228.
- [7] PERCY VERGARA, Lucas Alberto; ONIEL RAMOS, Eduardo Felipe. Evolución de los métodos para el reconocimiento de olores utilizando redes neuronales artificiales. 2003.
- [8] PAULINO, Alessandra A.; FENG, Jianjiang; JAIN, Anil K. Latent fingerprint matching using descriptor-based hough transform. *IEEE Transactions on Information Forensics and Security*, 2012, vol. 8, no 1, p. 31-45.
- [9] QUISILEMA, Stalin Wladimir Yajamín; PÉREZ, Francisco Díaz. Dispositivo LowCost de Seguridad Palmar Biométrico. UNIVERSIDAD POLITÉCNICA DE MADRD. 2017.
- [10] KONG, Adams; ZHANG, David; KAMEL, Mohamed. A survey of palmprint recognition. *pattern recognition*, 2009, vol. 42, no 7, p. 1408-1418.
- [11] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. En *Advances in neural information processing systems*. 2012. p. 1097-1105.
- [12] WU, Xiangqian; ZHAO, Qiushi; BU, Wei. A SIFT-based contactless palmprint verification approach using iterative RANSAC and local palmprint descriptors. *Pattern Recognition*, 2014, vol. 47, no 10, p. 3314-3326.
- [13] KANHANGAD, Vivek; KUMAR, Ajay; ZHANG, David. Contactless and pose invariant biometric identification using hand surface. *IEEE transactions on image processing*, 2010, vol. 20, no 5, p. 1415-1424.
- [14] ZHANG, Chenghao, et al. Double Half-Orientation Code and Nonlinear Matching Scheme for Palmprint Recognition. En *International Conference on Mechatronics and Intelligent Robotics*. Springer, Cham, 2017. p. 36-42.
- [15] ZHANG, David, et al. Robust palmprint verification using 2D and 3D features. *Pattern Recognition*, 2010, vol. 43, no 1, p. 358-368.
- [16] J SVOBODA, Jan; MASCI, Jonathan; BRONSTEIN, Michael M. Palmprint recognition via discriminative index learning. En *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016. p. 4232-4237.
- [17] MINAEE, Shrevin; WANG, Yao. Palmprint recognition using deep scattering network. En *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2017. p. 1-4.
- [18] FEI, Lunke, et al. Feature extraction methods for palmprint recognition: A survey and evaluation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, vol. 49, no 2, p. 346-363.

- [19] RAMACHANDRA, Raghavendra, et al. Verifying the newborns without infection risks using contactless palmprints. En *2018 International Conference on Biometrics (ICB)*. IEEE, 2018. p. 209-216.
- [20] GENOVESE, Angelo, et al. PalmNet: Gabor-PCA convolutional networks for touchless palmprint recognition. *IEEE Transactions on Information Forensics and Security*, 2019, vol. 14, no 12, p. 3160-3174.
- [21] DIAN, Liu; DONGMEI, Sun. Contactless palmprint recognition based on convolutional neural network. En *2016 IEEE 13th International Conference on Signal Processing (ICSP)*. IEEE, 2016. p. 1363-1367.
- [22] JALALI, Amin; MALLIPEDDI, Rommohan; LEE, Minho. Deformation invariant and contactless palmprint recognition using convolutional neural network. En *Proceedings of the 3rd International Conference on Human-Agent Interaction*. 2015. p. 209-212.
- [23] ZIMMERMANN, Christian; BROX, Thomas. Learning to estimate 3d hand pose from single rgb images. En *Proceedings of the IEEE international conference on computer vision*. 2017. p. 4903-4911.
- [24] GE, Lihao, et al. 3d hand shape and pose estimation from a single rgb image. En *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2019. p. 10833-10842.
- [25] VILLAMARÍN, Diego. Estado del Arte, Herramientas y Aplicaciones para Transformaciones geométricas 3D. En *Congreso de Ciencia y Tecnología ESPE*. 2015. p. 226-231.
- [26] Pablo Turmero, M., 2020. *Procesamiento De Imágenes (Traslación, Escala, Rotación, Inclinación) - Monografias.Com*. [online] Monografias.com. Disponible en: <<https://www.monografias.com/trabajos108/procesamiento-imagenes-traslacion-escala-rotacion-inclinacion/procesamiento-imagenes-traslacion-escala-rotacion-inclinacion.shtml>> [consulta:11 Junio 2020].
- [27] BAY, Herbert; TUYTELAARS, Tinne; VAN GOOL, Luc. Surf: Speeded up robust features. En *European conference on computer vision*. Springer, Berlin, Heidelberg, 2006. p. 404-417.
- [28] HARRIS, Christopher G., et al. A combined corner and edge detector. En *Alvey vision conference*. 1988. p. 10-5244.
- [29] MATAS, Jiri, et al. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 2004, vol. 22, no 10, p. 761-767.
- [30] ROSTEN, Edward; DRUMMOND, Tom. Machine learning for high-speed corner detection. En *European conference on computer vision*. Springer, Berlin, Heidelberg, 2006. p. 430-443.

ANEXOS

Anexo 3.1.

Tal y como se explicaba en la base del trabajo, a continuación, se muestran algunos ejemplos de la resolución de las imágenes obtenidas de la muestra poblacional utilizada.

La diferencia de resolución tiene origen en diferentes motivos, entre ellos, la calidad de la cámara del dispositivo móvil correspondiente, la luz natural/artificial o los fallos de enfoque de los usuarios. Por estos motivos, la calidad de la mayoría de las imágenes se vio afectada, provocando que la calidad no fuese elevada.

A continuación, se muestran diferentes ejemplos de la muestra, correspondiendo una calidad alta a la imagen 1, una calidad intermedia a la imagen 2 y una calidad baja a la imagen 3, además de las que fueron imposibles de utilizar en el estudio por no cumplir los criterios de inclusión (una oclusión de la palma, por ejemplo).



Imagen 1. Imagen de muestra palmar, calidad alta.

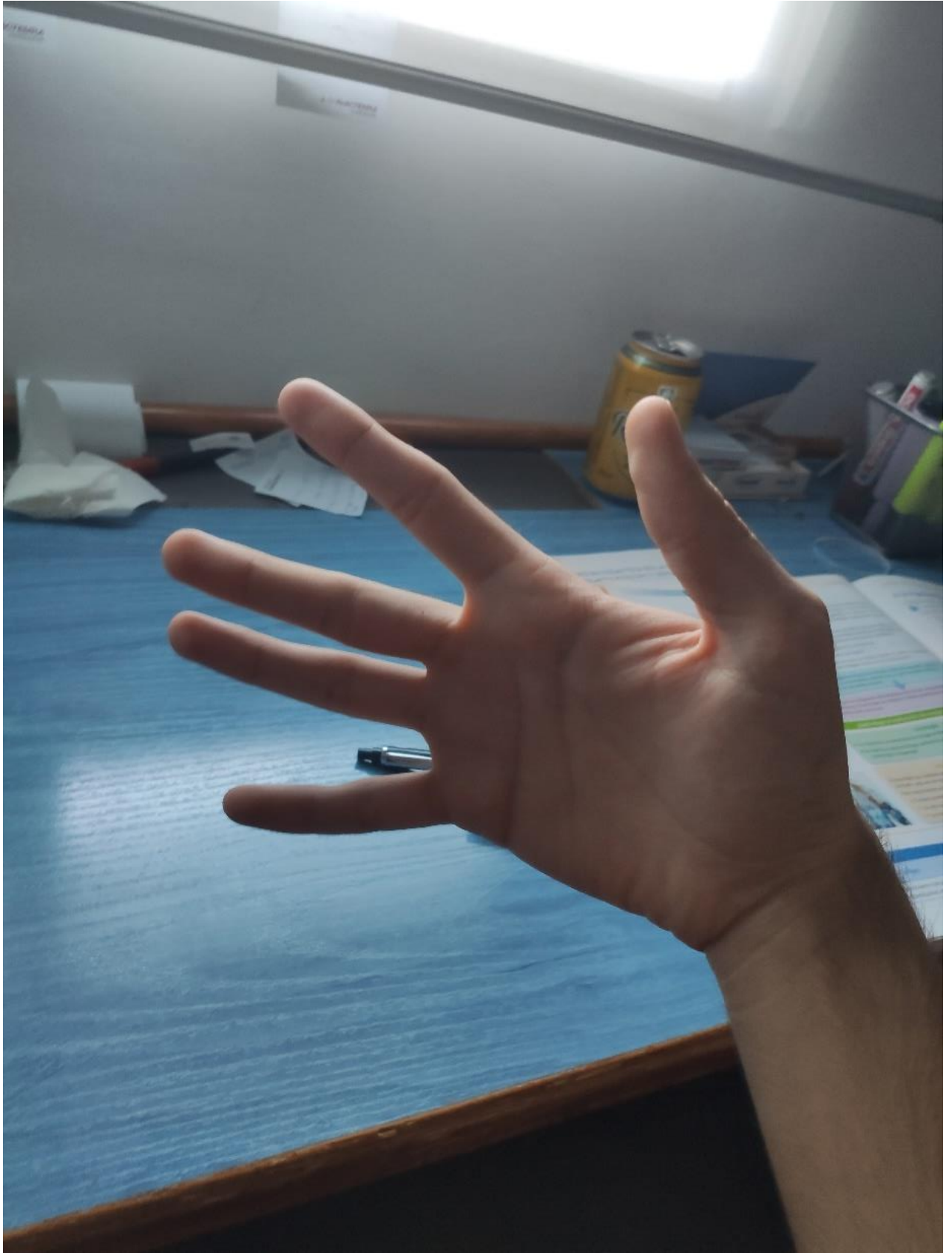


Imagen 2. Imagen de muestra palmar, calidad intermedia.

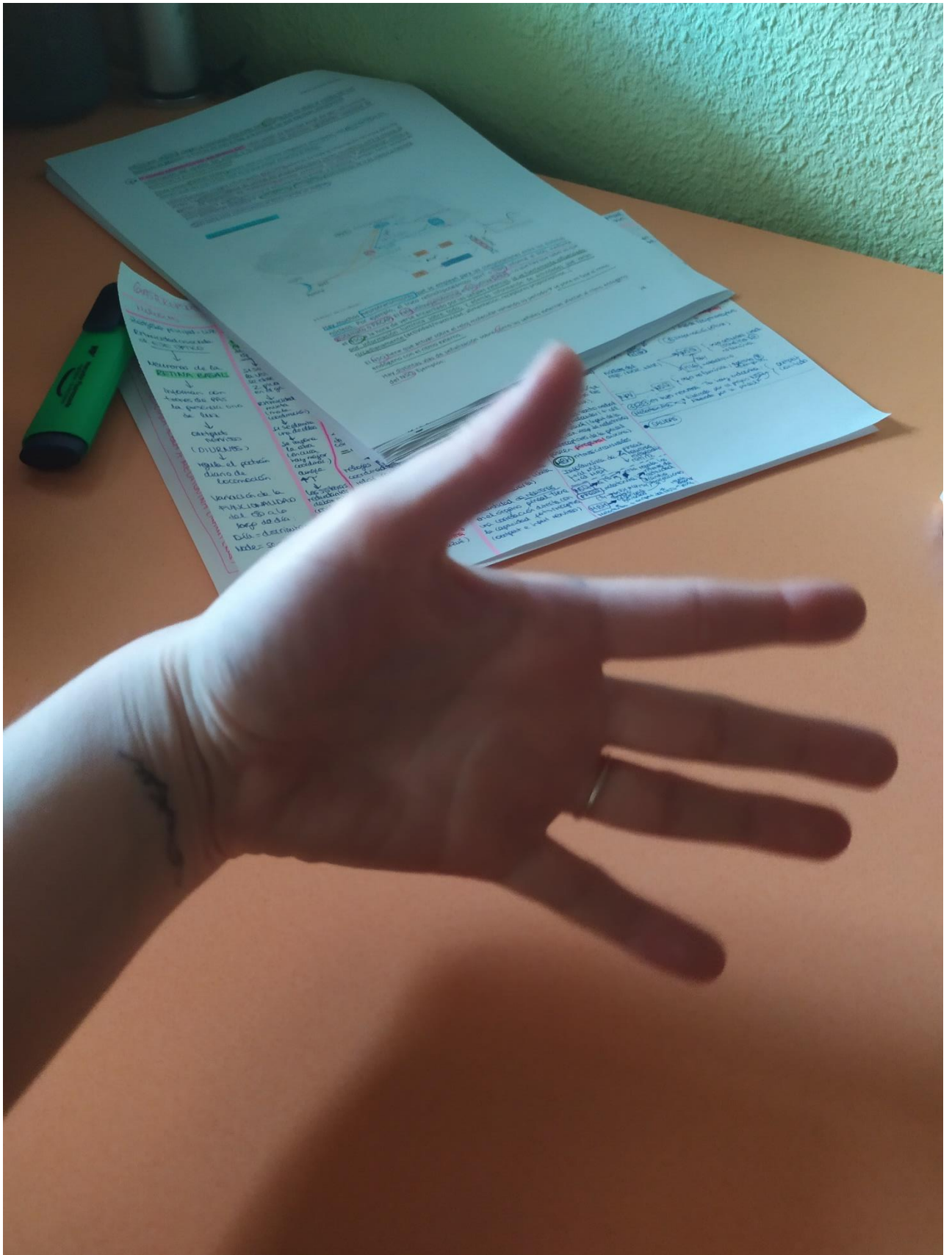
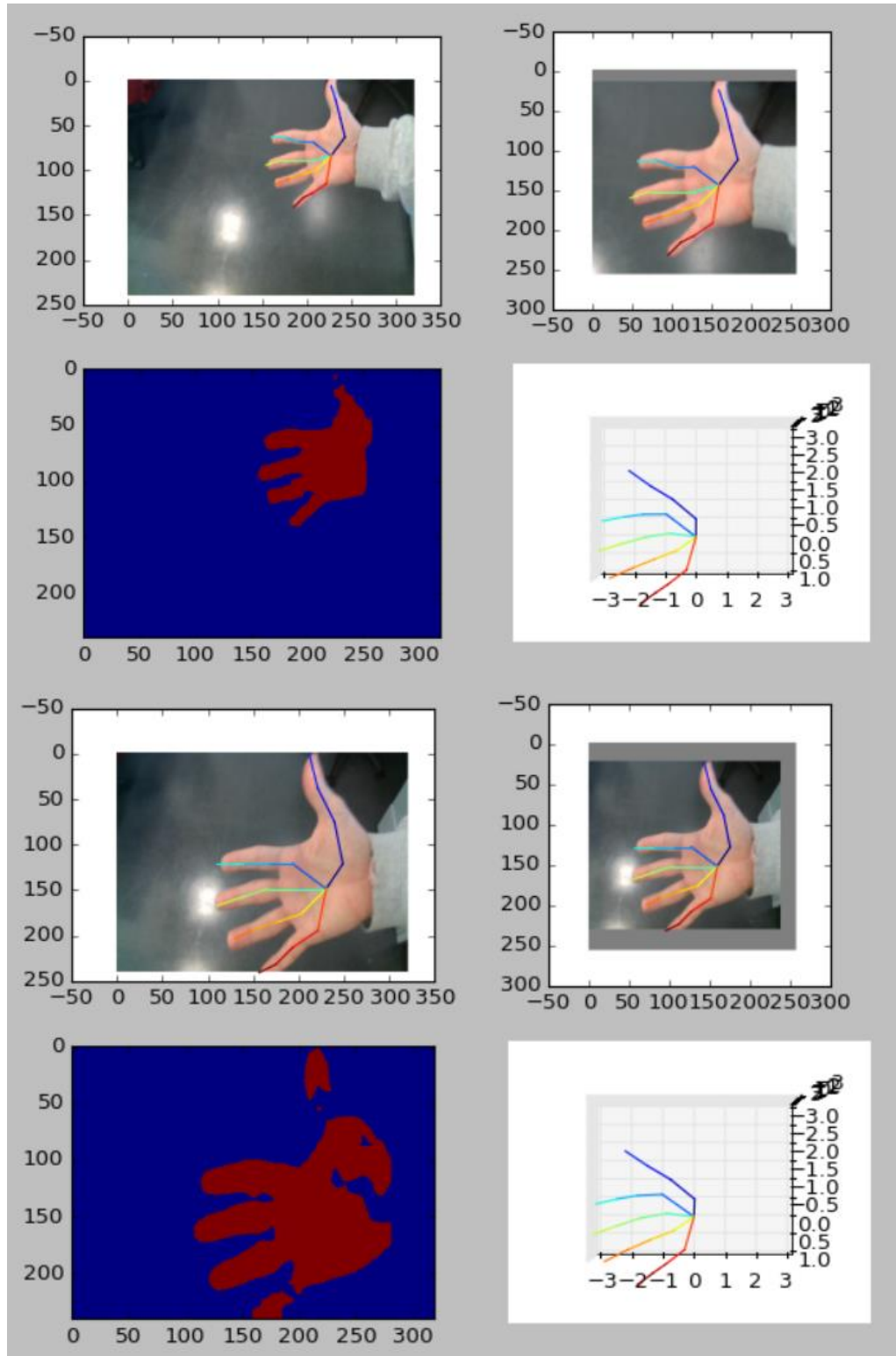


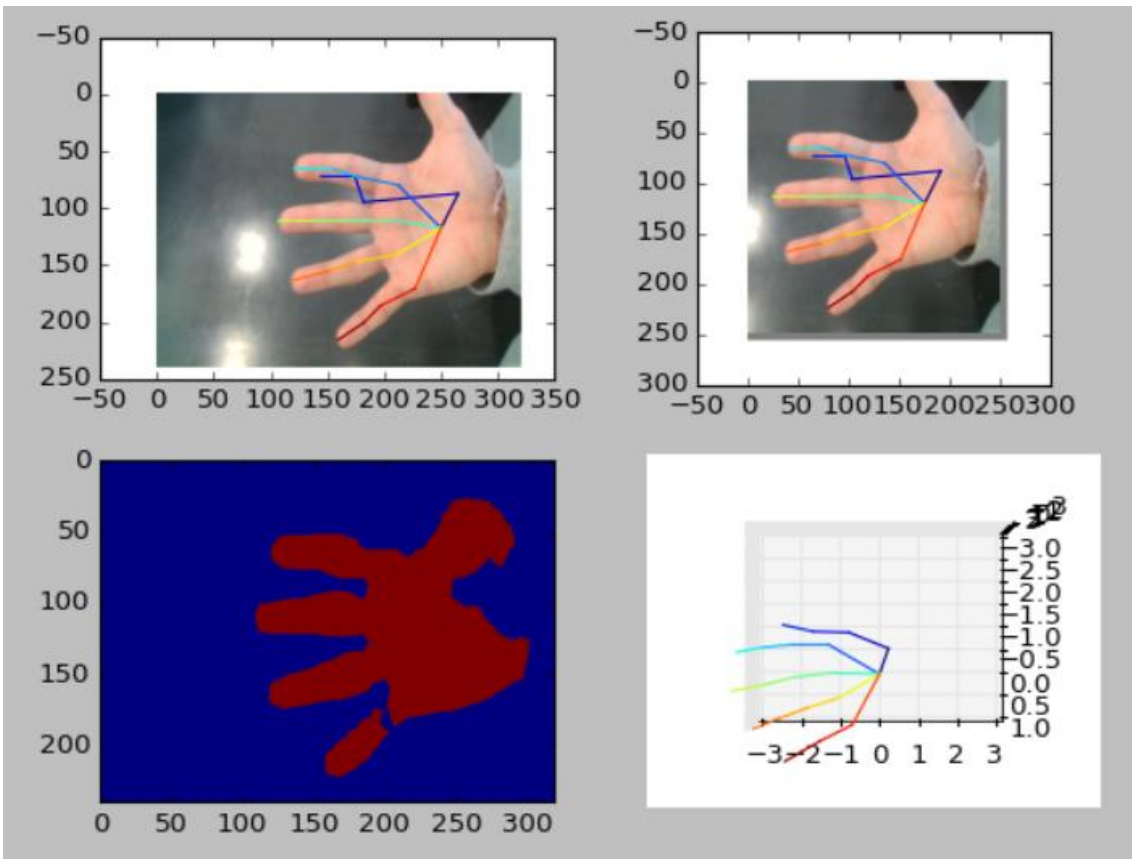
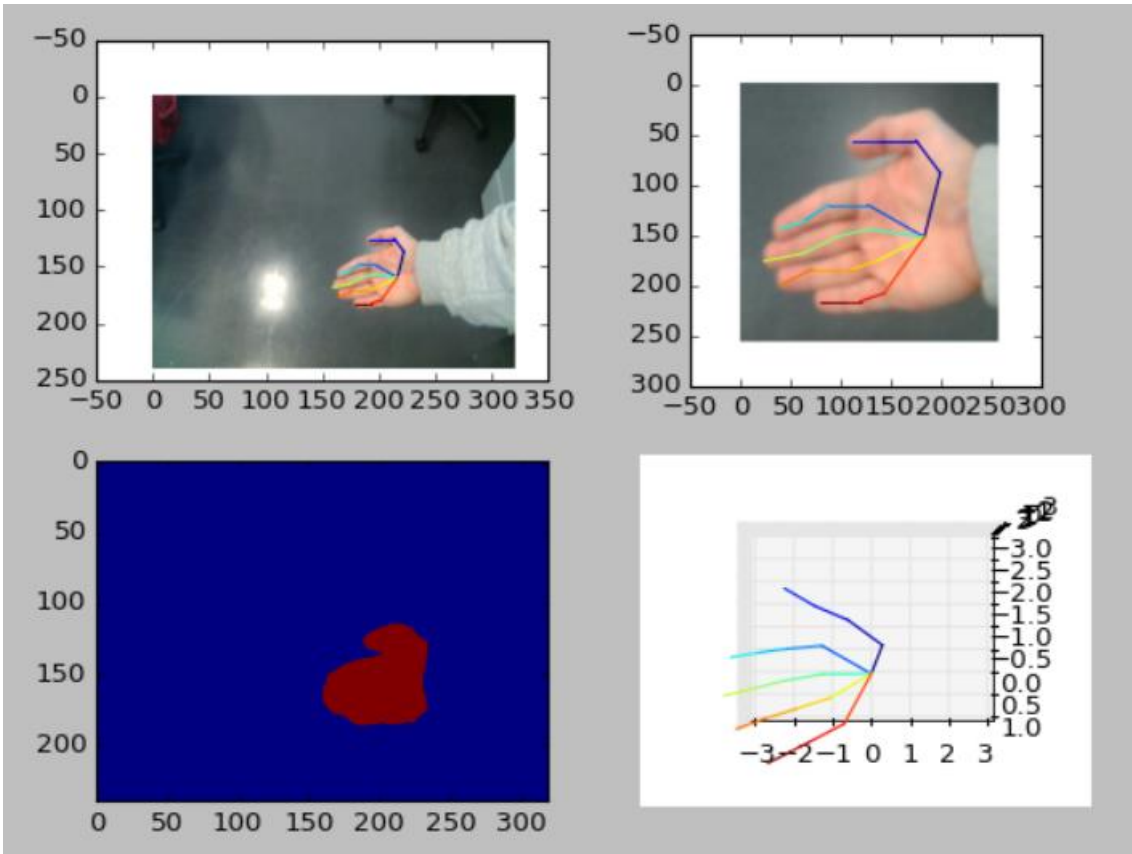
Imagen 3. Imagen de muestra palmar, calidad baja.

Anexo 4.1.

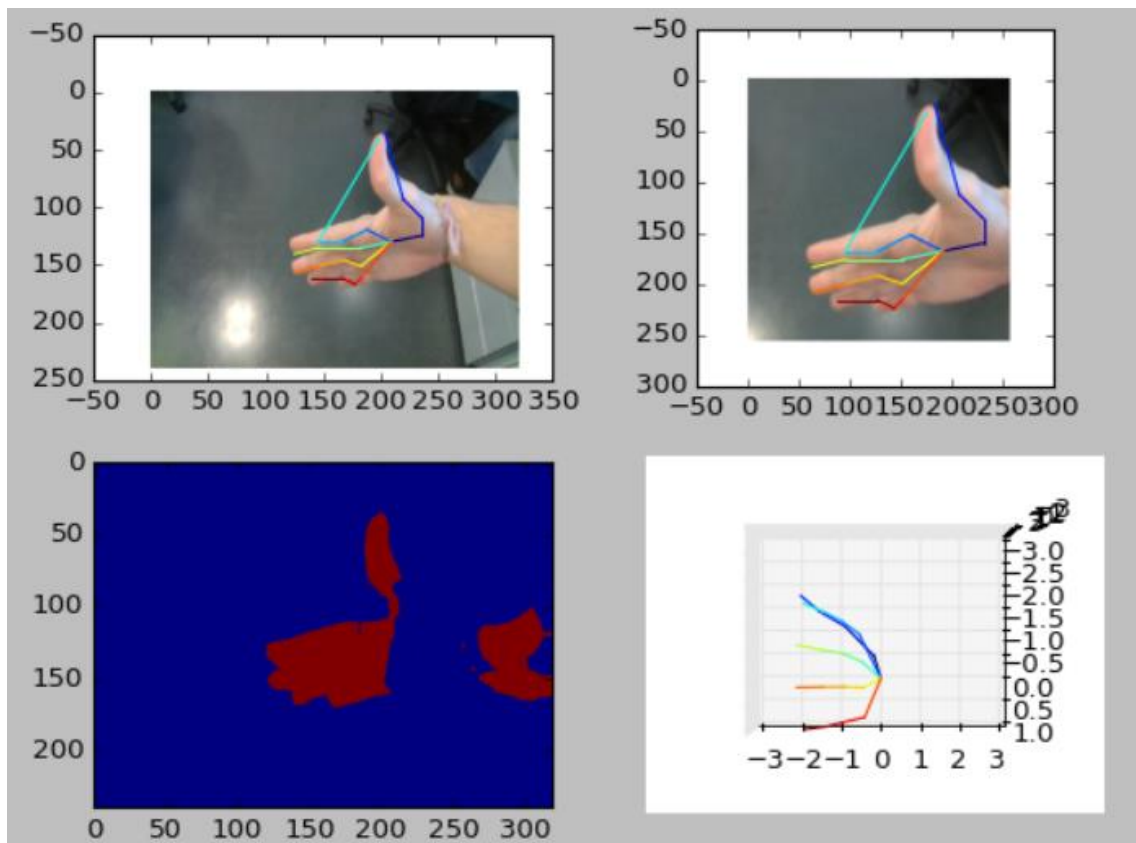
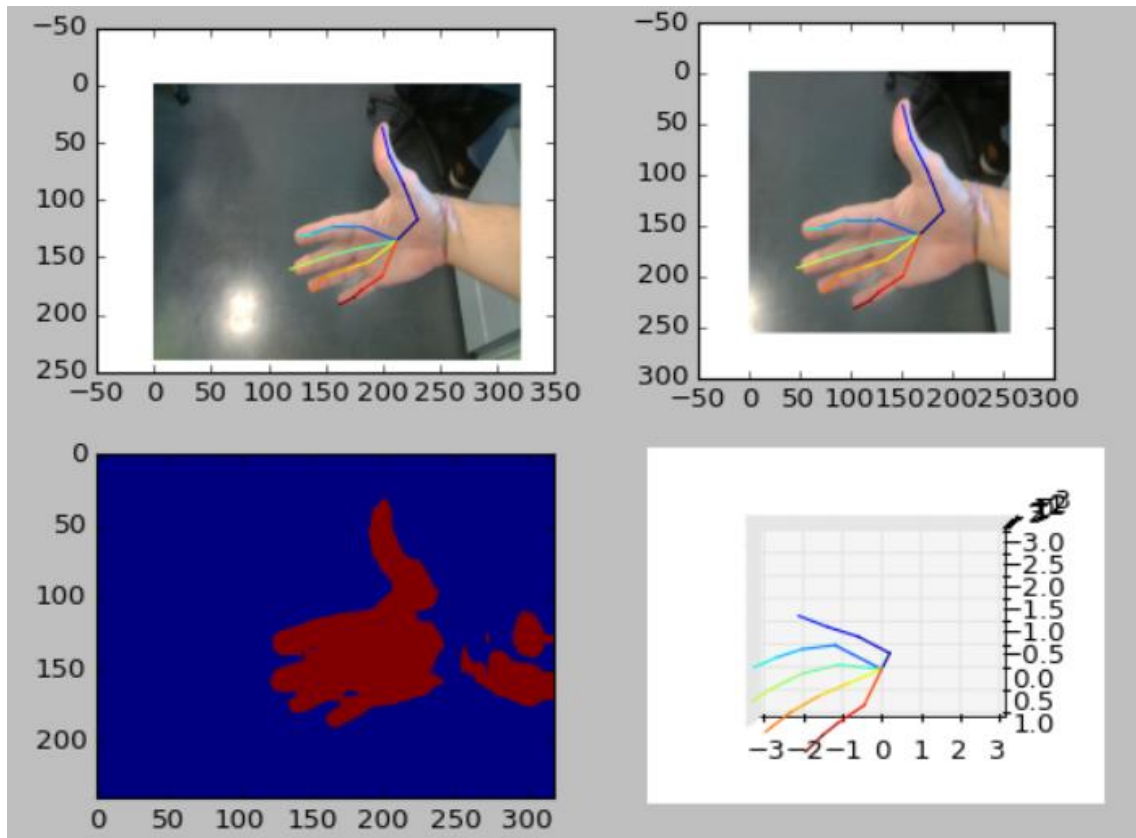
Aquí lo que se puede observar son las imágenes utilizadas para comprobar la viabilidad y resultados del algoritmo, probando los ángulos de giro y la distancia y en un entorno controlado.

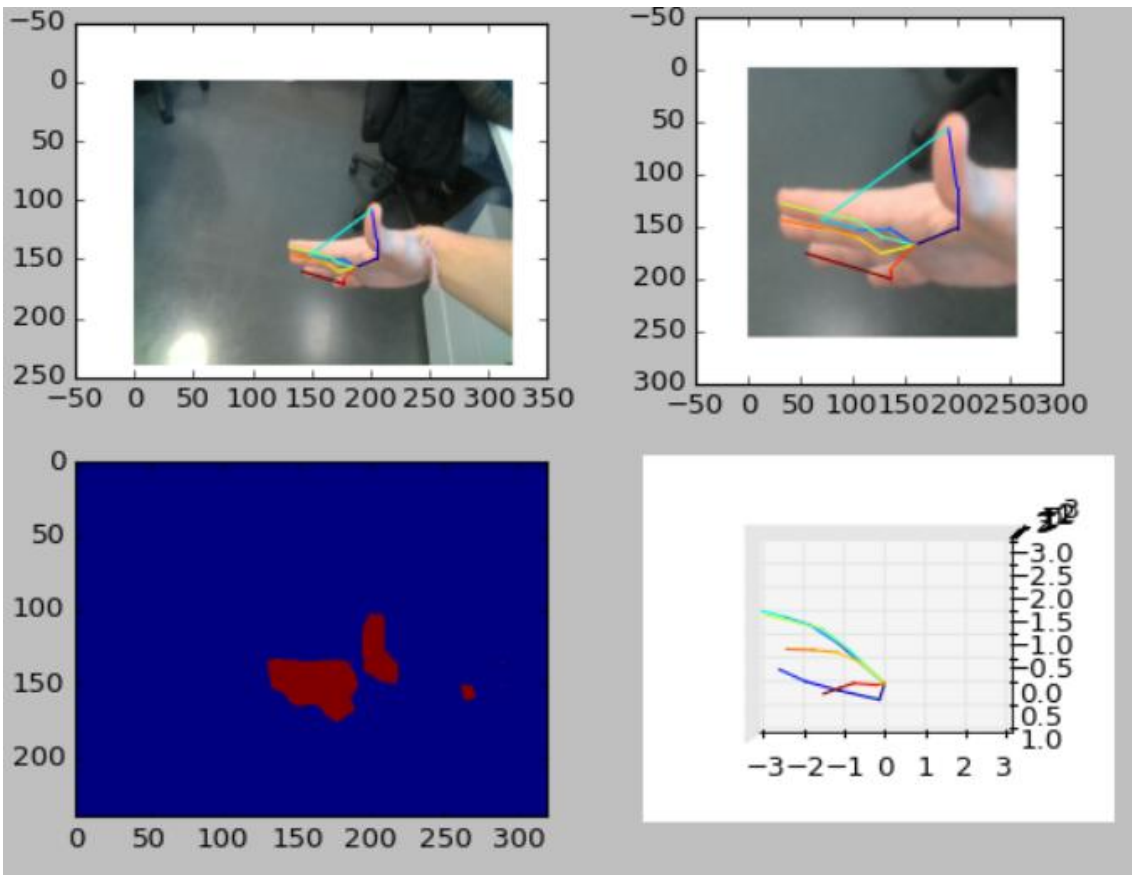
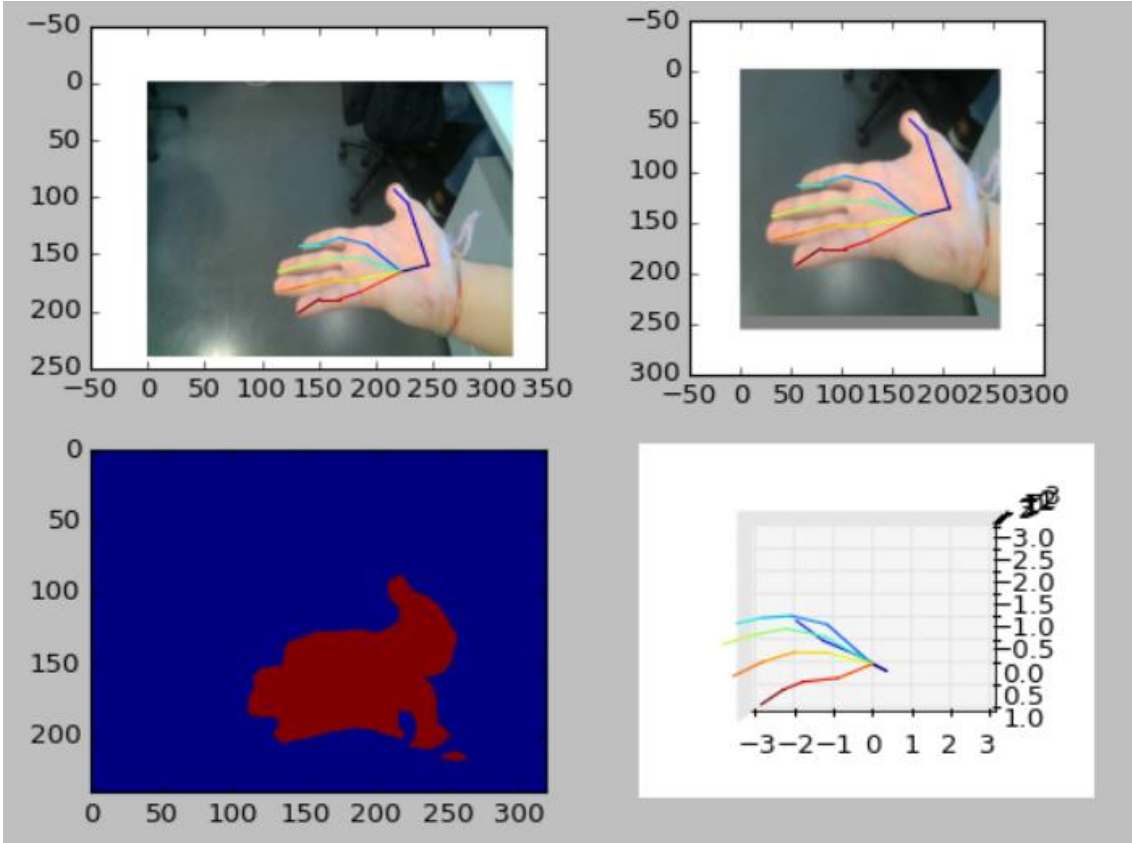
Variando distancia:



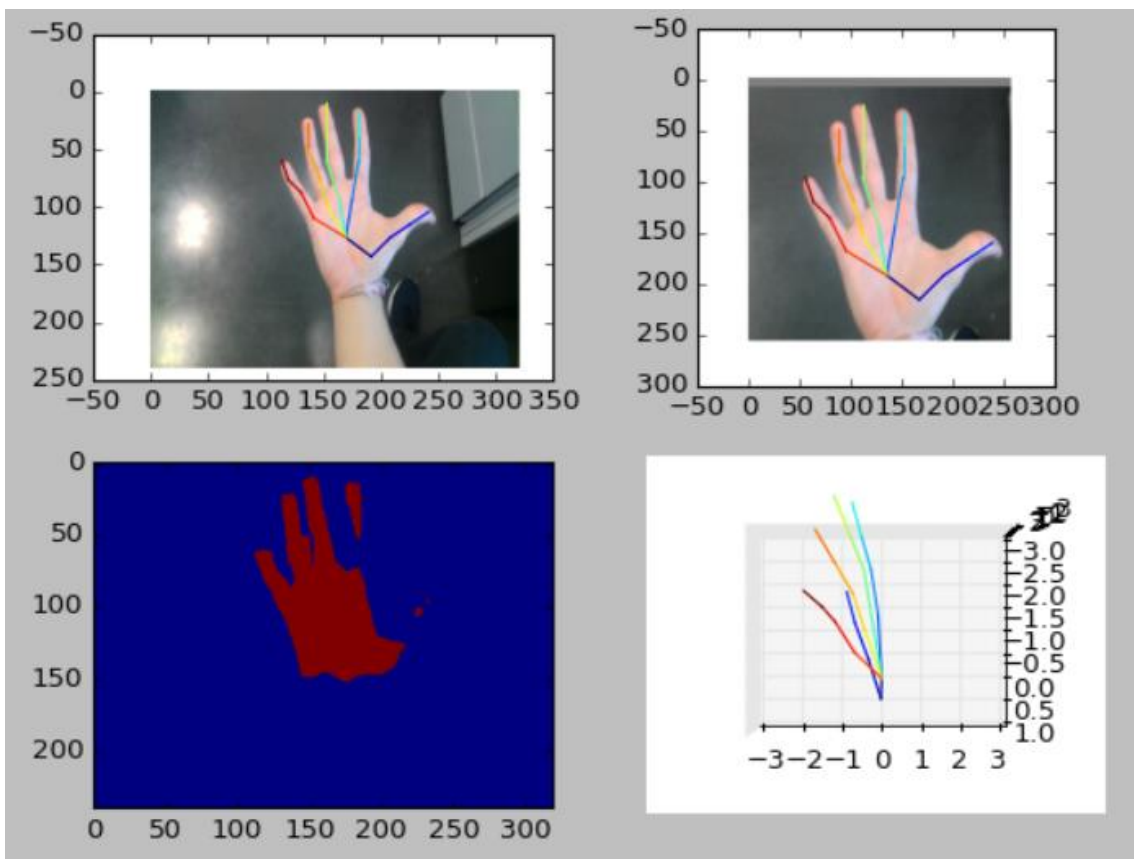
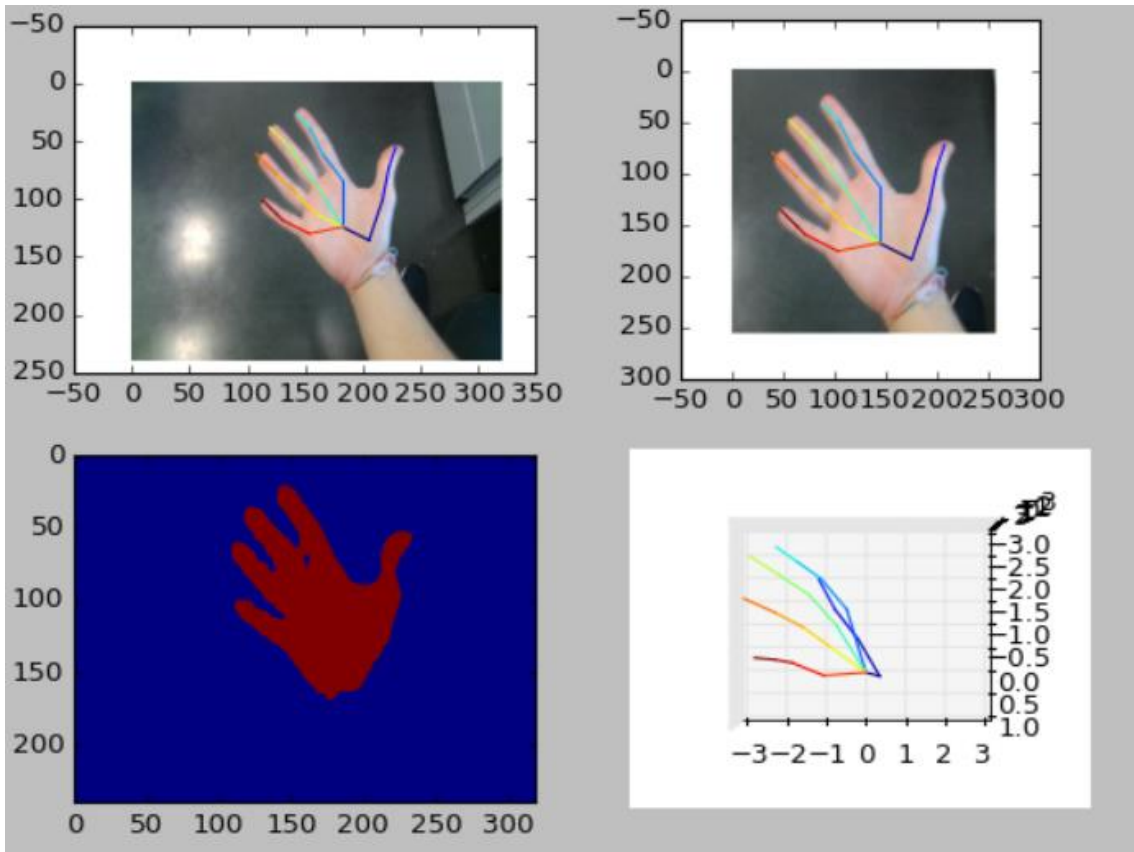


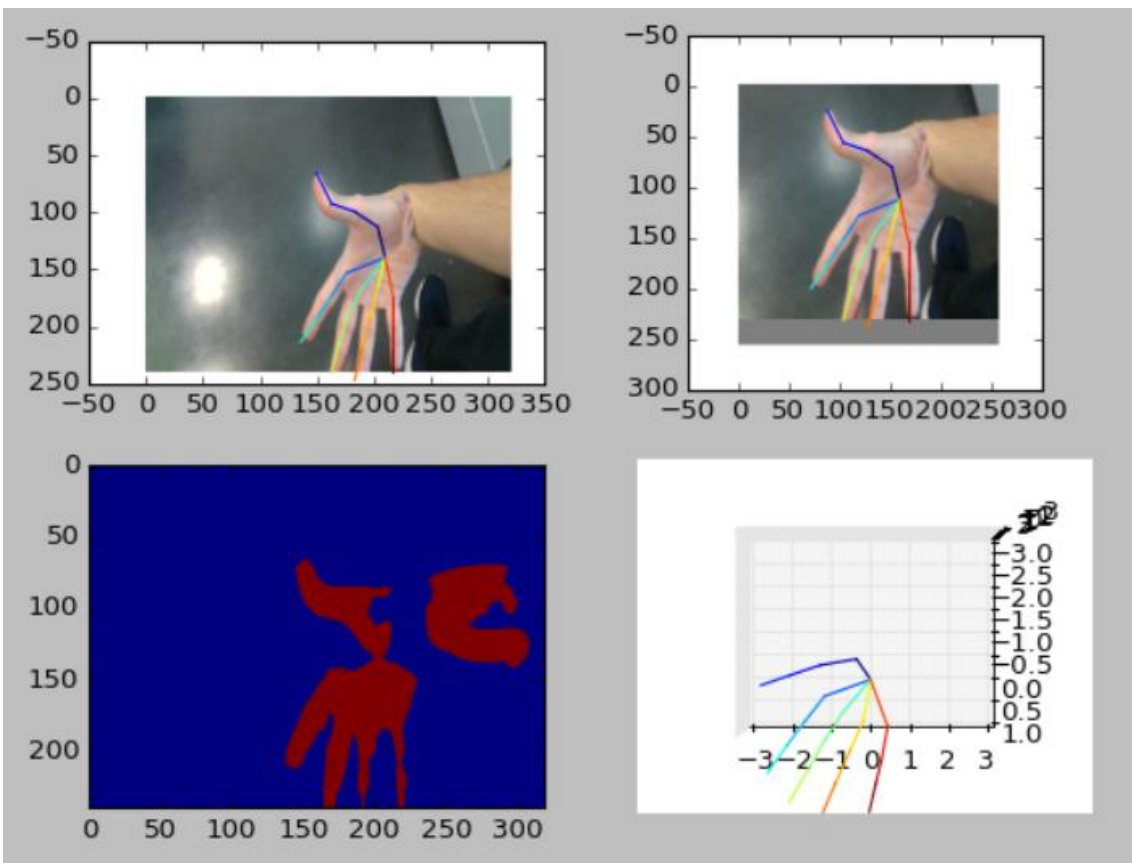
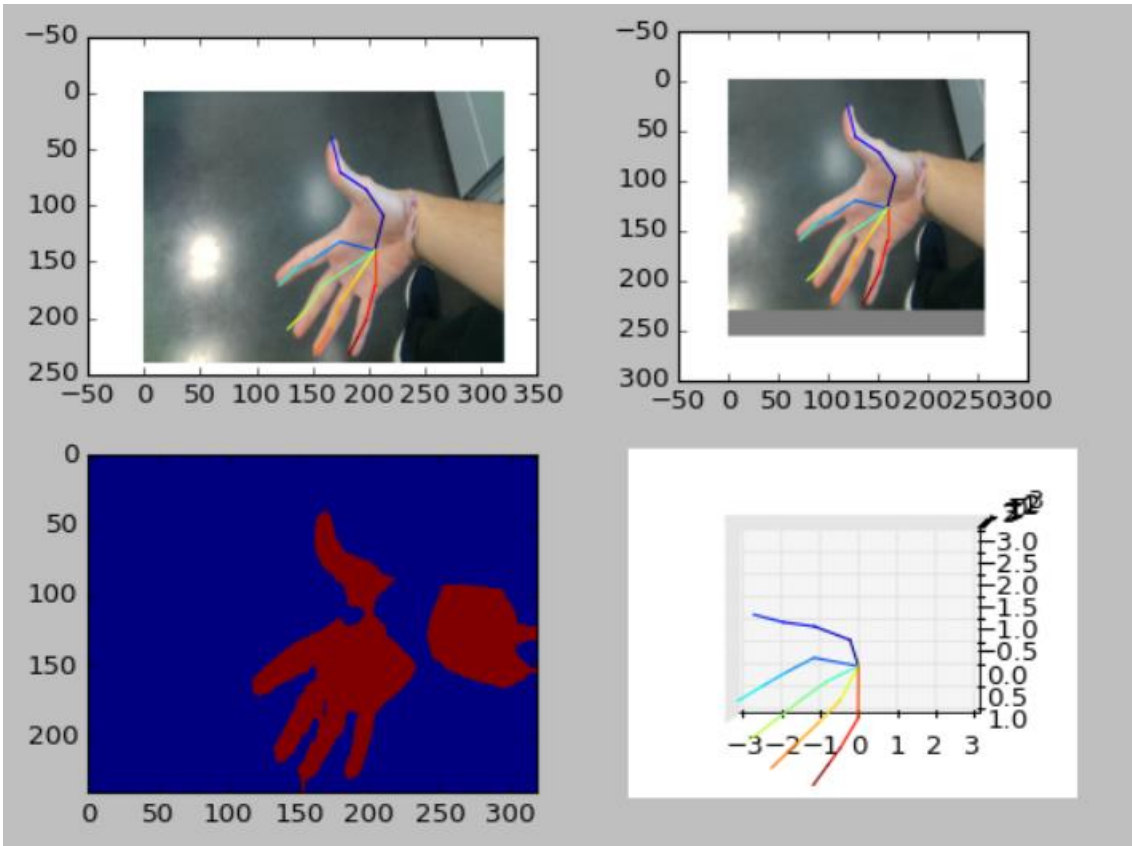
Variando en un primer ángulo:



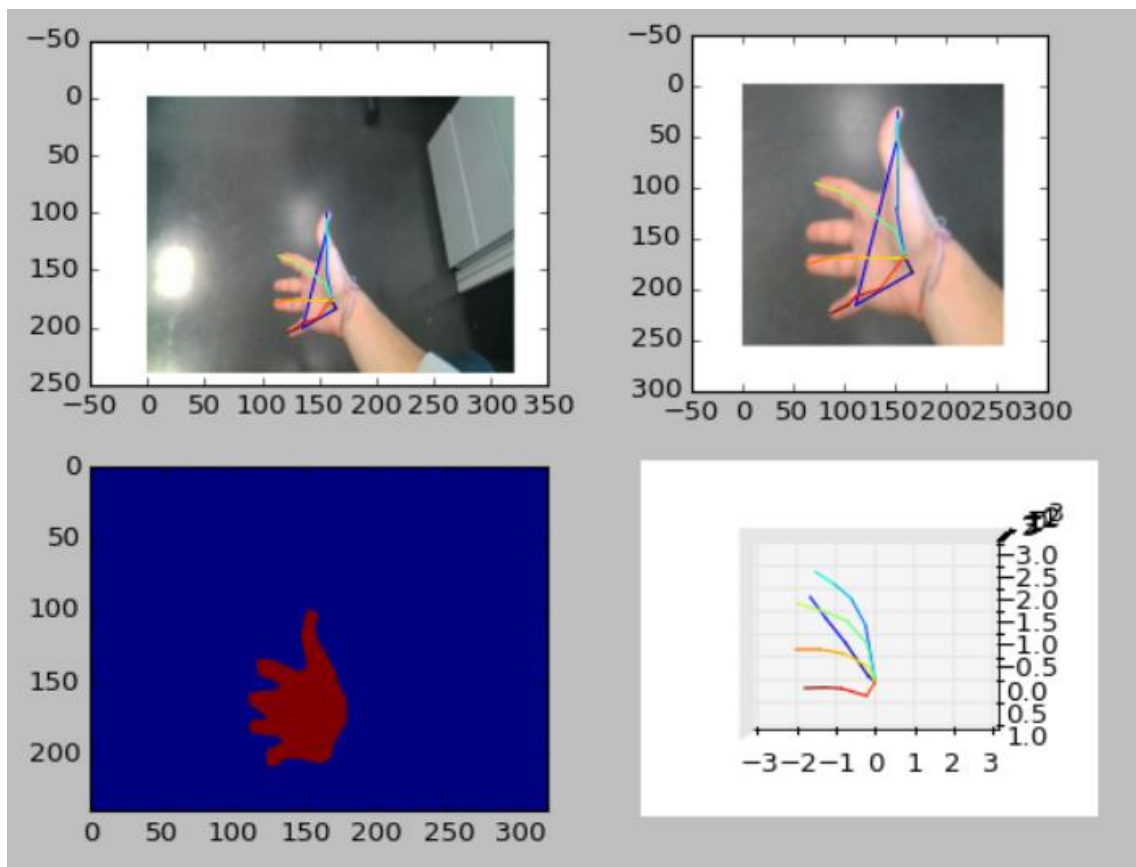
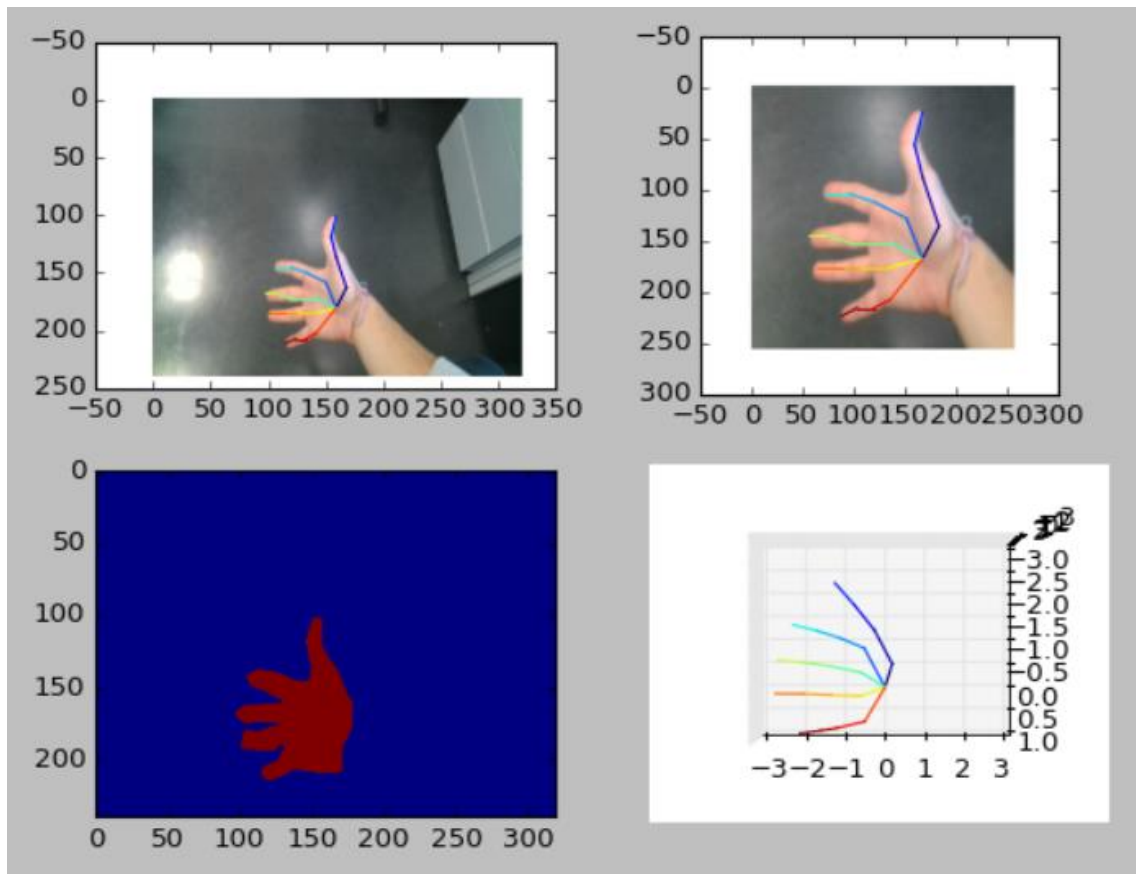


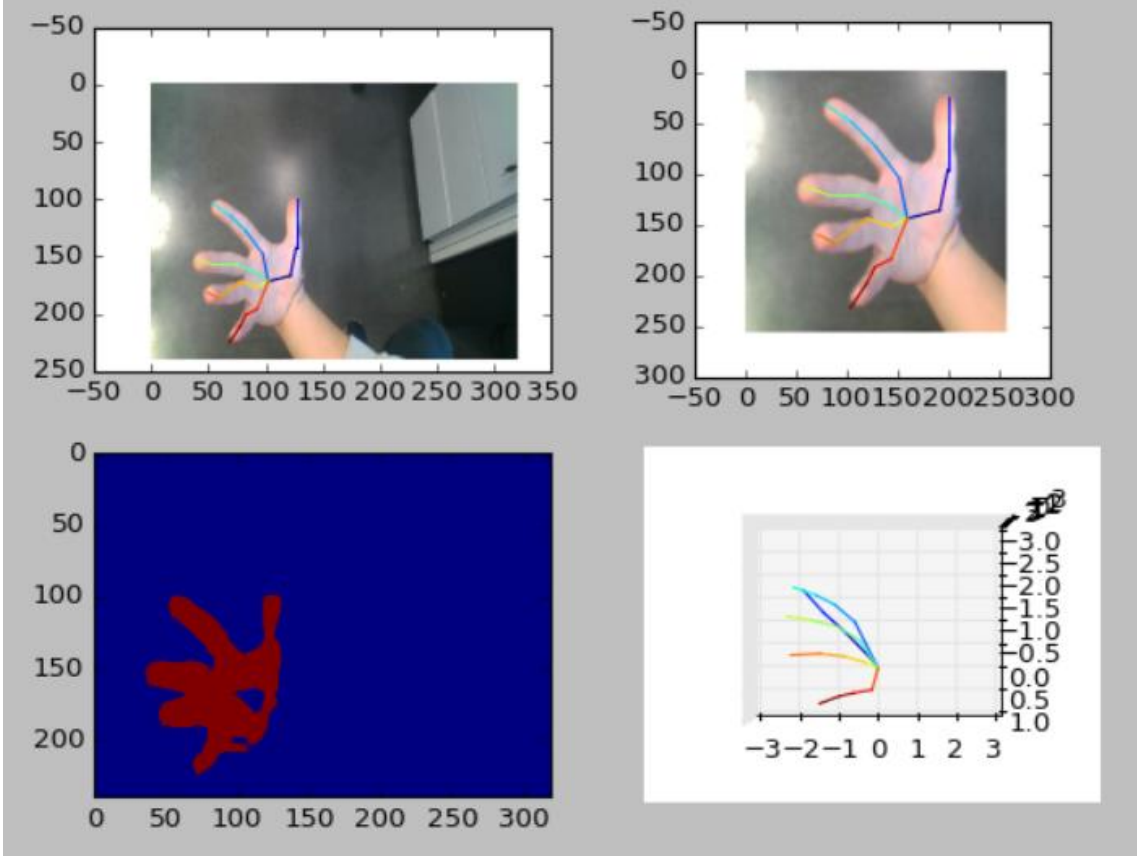
Variando en un segundo ángulo:





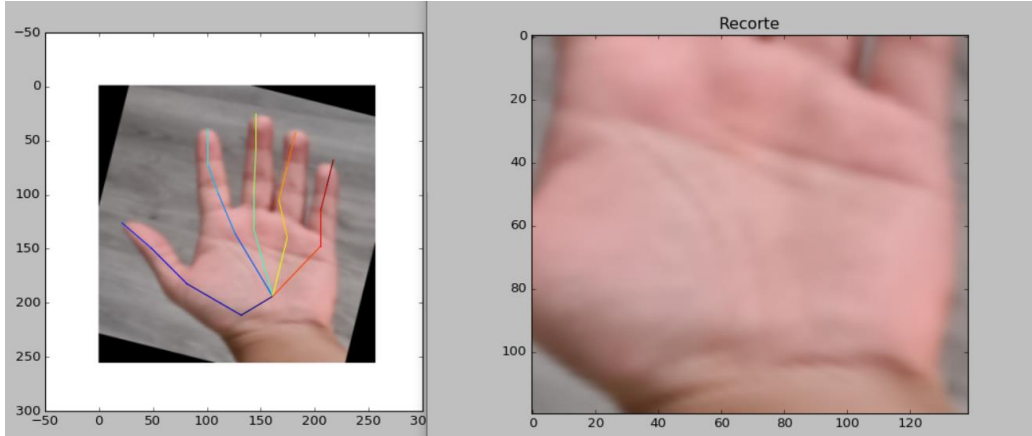
Variando en el tercer ángulo:



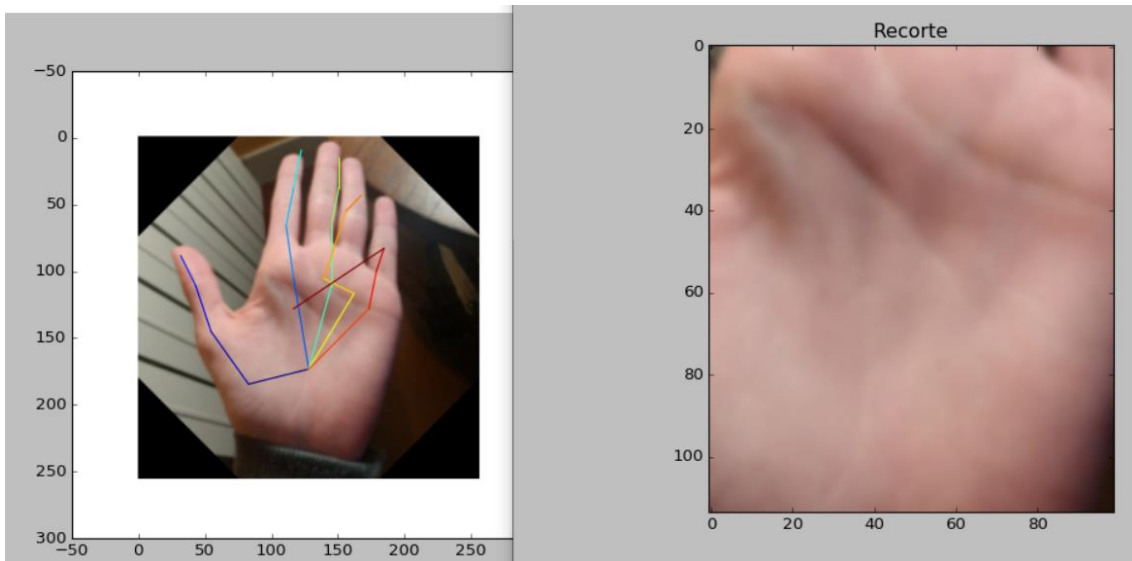


Anexo 4.2.

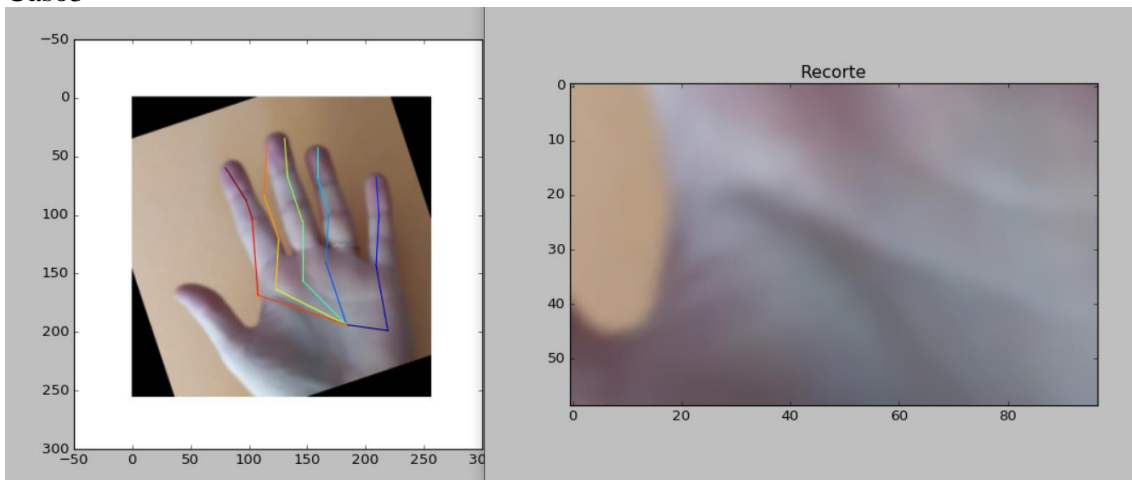
Caso 1



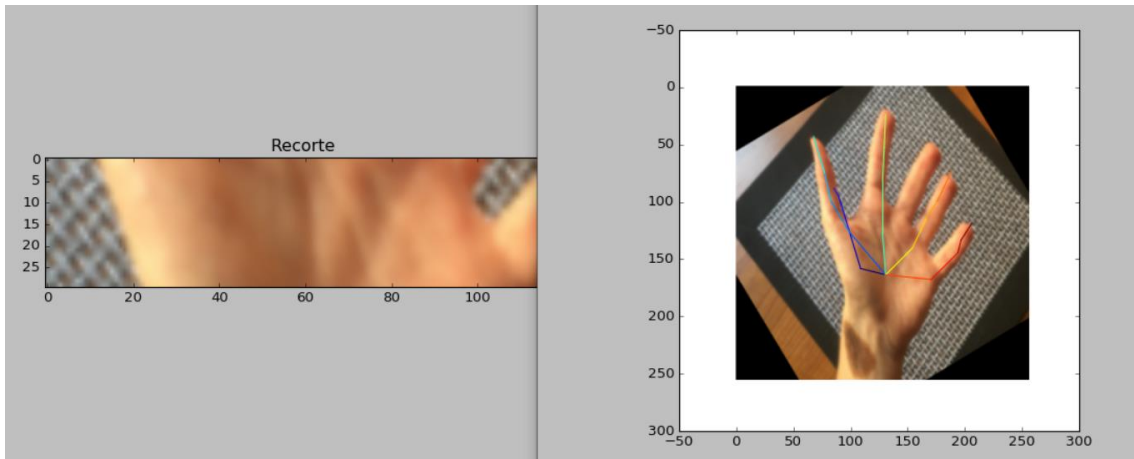
Caso 2



Caso3



Caso 4



Caso 5

