



Escola Tècnica Superior d'Enginyeria
de Telecomunicació de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA

PROYECTO FINAL DE CARRERA

RECONSTRUCCIÓN DE LA FORMA DEL ROSTRO A PARTIR DE CONTORNOS.

Estudios: Ingeniería Superior de Telecomunicaciones

Autor: Laia del Pino López

Directores: Verónica Vilaplana Besler / Josep Ramon Morros i Rubió.

Fecha: Barcelona, Julio 2015

“Aquel cuyo rostro no irradia luz nunca será estrella”.

William Blake

Resum del Projecte

La reconstrucció i el modelatge de cares 3D s'han convertit en els últims anys en una línia d'investigació molt activa a causa de la seva utilització en gran nombre d'aplicacions com són el reconeixement facial en 3D, modelatge de cares en videojocs, cirurgia estètica, etc. Durant les últimes dècades s'han desenvolupat múltiples tècniques computacionals de reconstrucció facial 3D. Una de les tecnologies més utilitzades està basada en imatges 2D i mètodes estadístics (*3D Morphable Models*).

L'estimació de la forma de la cara abordada amb models estadístics, té com a objectiu principal trobar un conjunt de paràmetres de la cara que millor s'ajusten a una imatge o un conjunt d'imatges. Per a aquest projecte es disposa d'un model estadístic capaç d'estimar la forma 3D del rostre quan una cara o un conjunt de contorns 2D està disponible des de múltiples punts de vista. Aquest model estima directament l'estructura de la cara 3D mitjançant l'ús d'una matriu de regressió construïda a través de PLS (Partial Least Squares).

Després de la validació del model i els bons resultats obtinguts en la predicció de subjectes sintètics, aquest treball proposa un nou enfoc, entrenar el model amb subjectes reals a partir d'una seqüència de vídeo. D'aquesta forma s'obtenen reconstruccions de la forma de la cara 3D amb dades reals.

Per tant, l'objectiu d'aquest projecte és la definició, la implementació software i l'anàlisi d'un procediment que ens permeti ajustar un model estadístic facial tridimensional genèric a les característiques facials específiques d'un individu a partir dels contorns del seu rostre.

Paraules Claus: Reconstrucció facial 3D, Mètodes estadístics, Contorn

Resumen del Proyecto

La reconstrucción y el modelado de caras 3D se han convertido en los últimos años en una línea de investigación muy activa debido a su utilización en gran número de aplicaciones como son el reconocimiento facial en 3D, modelado de caras en videojuegos, cirugía estética, etc. Durante las últimas décadas se han desarrollado múltiples técnicas computacionales de reconstrucción facial 3D. Una de las tecnologías más utilizadas está basada en imágenes 2D y métodos estadísticos (*3D Morphable Models*).

La estimación de la forma del rostro abordada con modelos estadísticos, tiene como objetivo principal encontrar un conjunto de parámetros del rostro que mejor se ajustan a una imagen o un conjunto de imágenes. Para este proyecto se dispone de un modelo estadístico capaz de estimar la forma 3D del rostro cuando una cara o un conjunto de contornos 2D está disponible desde múltiples puntos de vista. Este modelo estima directamente la estructura de la cara 3D mediante el uso de una matriz de regresión construida a través de PLS (Partial Least Squares).

Tras la validación del modelo y los buenos resultados obtenidos en la predicción de sujetos sintéticos, este trabajo propone un nuevo enfoque, entrenar el modelo con sujetos reales a partir de una secuencia de vídeo. De esta forma se obtiene reconstrucciones de la forma de la cara 3D con datos reales.

Por lo tanto, el objetivo de este proyecto es la definición, la implementación software y el análisis de un procedimiento que nos permita ajustar un modelo estadístico facial tridimensional genérico a las características faciales específicas de un individuo a partir de los contornos de su rostro.

Palabras Claves: Reconstrucción facial 3D, Métodos estadísticos, Contorno

Abstract

The reconstruction and modelling of 3D faces have become, in the last few years, a very active field of research due to their use in a great number of applications such as those involved in 3D facial recognition, face modelling in videogames, surgery, etc. During the last decades, various and many computer techniques of 3D facial reconstruction have been developed. One of the most widely-used technologies is based on 2D images and statistical methods (*3D Morphable Models*).

Estimating the shape of the face by means of statistical models has as its main objective finding a set of facial parameters that best suit and fit an image or set of images. For this project, a statistical model is available which is capable of estimating the 3D facial shape when a face or a set of 2D contours are available from a various range of angles. This model directly estimates the 3D structure of the face by using a regression matrix built by means of PLS (Partial Least Squares).

After the validation of the model and the good results obtained in the prediction of synthetic individuals, this project proposes a new approach training the model with real individuals from a video sequence. In this way, one obtains reconstructions of the shape of the 3D face from real data.

Therefore, this project's objective is the definition, the software implementation and the analysis of a method that allows us to adjust a generic three-dimensional facial statistical model to the specific facial characteristics of an individual, based on the contours of his / her face.

Keywords: 3D Facial Reconstruction, 3D Morphable Models, Contour.

Colaboraciones y agradecimientos

Colaboraciones

Este Proyecto ha sido desarrollado en colaboración con el Departamento de Teoría de Señal y Comunicaciones (Grupo Procesado de Imagen y Video) de la Universitat Politècnica de Catalunya (UPC). Ellos han dirigido el proyecto, me han facilitado información, bibliografía y herramientas para realizar mi Proyecto Fin de Carrera.



Image Processing Group
Signal Theory and Communications Department
Universitat Politècnica de Catalunya. BARCELONATECH

Por otro lado, también se ha contado con la colaboración del Dr. Mario Castelán, investigador en Centro de Investigación y de Estudios Avanzados (Cinvestav) del Instituto Politécnico Nacional de México.

Agradecimientos

Me gustaría agradecer a mis directores de proyecto Verónica Vilaplana y Josep Ramon Morros por haberme guiado y ayudado durante todo este proceso.

También a mi familia, mis amigos y mi pareja, por apoyarme y acompañarme durante toda la carrera.

Índice general

Resum del Projecte.....	v
Resumen del Proyecto.....	vi
Abstract	vii
Colaboraciones y agradecimientos	viii
Índice general.....	ix
Listado de figuras y tablas	xi
Capítulo 1. Introducción	1
1.1. Contexto del proyecto	1
1.2. Objetivos.....	2
1.3. Estructura de la memoria	3
Capítulo 2. Estado del arte y descripción de las técnicas principales	4
2.1. Background.....	4
2.1.1. Aproximación con una sola vista	5
2.1.2. Aproximación con múltiples vistas y considerando contornos.	6
2.1.3. Modelo estadístico de Regresión [15]	8
2.2. Descripción de las técnicas principales.....	12
2.2.1. Partial Least Squares (PLS).....	12
2.2.2. Ultrametric Contour Map (UCM).....	15
2.2.3. Transformada de Hough Generalizada	17
2.2.4. Chehra Face Tracker	20
Capítulo 3. Descripción del método	22
3. 1. Segmentación de los frames del vídeo	23
3. 2. Extracción y procesado de los contornos	24
3. 3. Regresión	25
3.3.1. Contorno principal y primera aproximación del 3D rostro.....	26
3.3.2. Selección del frame con contorno de pitch $90-k\cdot\alpha$ ($k = 1$).....	27
3.3.3. Re-entrenamiento del Regresor con más perfiles	28
3.3.4. Contornos de pitch 0° y 180°	30
3.4. Ejemplo completo.....	32

Capítulo 4. Resultados	38
4. 1. Base de Datos	38
4. 2. Proceso de evaluación	39
4. 3. Resultados.....	39
4. 4. Limitaciones	45
Capítulo 5. Conclusiones	47
Capítulo 6. Referencias	48
Anexo I. Reconstrucciones	51
Anexo II. Inventario de funciones	59

Listado de figuras y tablas

Figura 1. Reconstrucción 3D del rostro	9
Figura 2. Información de entrenamiento	9
Figura 3. Contornos y variaciones de la pose	10
Figura 4. Muestra de frames de una secuencia de vídeo	10
Figura 5. Detección de contornos gPb y Oriented Watershed Transform	16
Figura 6. Segmentación jerárquica de contornos.....	16
Figura 7. Geometría de la detección de la forma para la GHT	18
Figura 8. Chehra Face Tracker. Ejemplos de <i>Landmarks</i> de cara y ojos.	21
Figura 9. Chehra Face Tracker. Numeración de los 49 puntos faciales	21
Figura 10. Reconstrucción de la forma del rostro a partir de los contornos.....	22
Figura 11. Segmentación jerárquica UCM	24
Figura 12. Adquisición y procesamiento de los contornos.....	25
Figura 13. Primera aproximación 3D del rostro	26
Figura 14. Selección del frame con pitch $90-\alpha$	27
Figura 15. Primera reconstrucción 3D del rostro	28
Figura 16. Segunda reconstrucción 3D.....	29
Figura 17. Detección de los puntos faciales mediante Chehra Face Tracker.	30
Figura 18. Distancias para la detección del frame con el rostro más frontal	30
Figura 19. Contornos de diferentes sujetos para pitch $0^\circ/180^\circ$	31
Figura 20. Segmentación UCM para diferentes niveles	31
Figura 21. Localización de la región rostro y extracción de los contornos pitch $0^\circ/180^\circ$..	32
Figura 22. Ejemplo - Muestra de frames segmentados.....	32
Figura 23. Ejemplo - Primera aproximación 3D del rostro	33
Figura 24. Ejemplo - Segunda aproximación 3D del rostro	33
Figura 25. Ejemplo - Tercera aproximación 3D del rostro.....	34
Figura 26. Ejemplo - Cuarta aproximación 3D del rostro	34
Figura 27. Ejemplo - Quinta aproximación 3D del rostro	35
Figura 28. Ejemplo - Sexta aproximación 3D del rostro	35

Figura 29. Ejemplo - Séptima aproximación 3D del rostro.....	36
Figura 30. Ejemplo - Octava aproximación 3D del rostro.....	36
Figura 31. Ejemplo – Reconstrucción final de la forma del rostro	37
Figura 32. Resultados Sujeto 1 - Comparativa de perfiles.....	40
Figura 33. Resultados Sujeto 1 - Reconstrucciones de la forma del rostro.....	40
Figura 34. Resultados Sujeto 1 - Reconstrucción final de la forma del rostro.....	41
Figura 35. Resultados Sujeto 2 - Comparativa de perfiles.....	41
Figura 36. Resultados Sujeto 2 - Reconstrucciones de la forma del rostro.....	42
Figura 37. Resultados Sujeto 2 - Reconstrucción final de la forma del rostro.....	42
Figura 38. Resultados Sujeto 3 - Comparativa de perfiles.....	43
Figura 39. Resultados Sujeto 3 - Reconstrucciones de la forma del rostro	43
Figura 40. Resultados Sujeto 3 - Reconstrucción final de la forma del rostro°	44
Figura 41. Limitaciones del sistema.....	45
Figura 42. Resultados Sujeto 4 - Comparativa de perfiles.....	51
Figura 43. Resultados Sujeto 4 - Reconstrucciones de la forma del rostro.....	51
Figura 44. Resultados Sujeto 4 - Reconstrucción final de la forma del rostro.....	52
Figura 45. Resultados Sujeto 5 - Comparativa de perfiles.....	52
Figura 46. Resultados Sujeto 5 - Reconstrucciones de la forma del rostro	53
Figura 47. Resultados Sujeto 5 - Reconstrucción final de la forma del rostro.....	53
Figura 48. Resultados Sujeto 6 - Comparativa de perfiles.....	54
Figura 49. Resultados Sujeto 6 - Reconstrucciones de la forma del rostro	54
Figura 50. Resultados Sujeto 6 - Reconstrucción final de la forma del rostro.....	55
Figura 51. Resultados Sujeto 7 - Comparativa de perfiles.....	55
Figura 52. Resultados Sujeto 7 - Reconstrucciones de la forma del rostro	56
Figura 53. Resultados Sujeto 7 - Reconstrucción final de la forma del rostro.....	56
Figura 54. Resultados Sujeto 8 - Comparativa de perfiles.....	57
Figura 55. Resultados Sujeto 8 - Reconstrucciones de la forma del rostro	57
Figura 56. Resultados Sujeto 8 - Reconstrucción final de la forma del rostro.....	58
Tabla 1. R-table de la GHT.	18
Tabla 2. Base de datos. Inventario de secuencias de vídeo disponibles	38

Capítulo 1. Introducción

En este primer capítulo se presenta una descripción general del proyecto con el contexto en el que se encuentra actualmente la línea de investigación, la definición de los objetivos de este proyecto y la estructura del documento.

1.1. Contexto del proyecto

La reconstrucción y el modelado de caras 3D se han convertido en los últimos años en una línea de investigación muy activa debido a su utilización en gran número de aplicaciones como son el reconocimiento facial en 3D, modelado de caras en videojuegos, videoconferencias en 3D, cirugía estética, etc. Esta área de investigación abarca múltiples disciplinas, como el reconocimiento de patrones, procesamiento de imágenes, visión por ordenador y redes neuronales.

Durante las últimas décadas se han desarrollado múltiples técnicas computacionales de reconstrucción facial 3D. Actualmente el método más eficaz para construir un modelo facial 3D a partir de un rostro real es a través de un escáner 3D. Sin embargo estos dispositivos son costosos y sólo factibles en entornos controlados y estáticos donde se requiere la cooperación del usuario. La tecnología disponible y más utilizada para el monitoreo de entornos no controlados está basada en imágenes 2D y métodos estadísticos (*3D Morphable Models*).

La estimación de la forma del rostro abordada con modelos estadísticos, tiene como objetivo principal encontrar un conjunto de parámetros del rostro que mejor se ajustan a una imagen o un conjunto de imágenes.

Para este proyecto disponemos de un modelo estadístico capaz de estimar la forma 3D del rostro cuando una cara o un conjunto de contornos 2D está disponible desde múltiples puntos de vista. Este modelo estima directamente la estructura de la cara 3D mediante el uso de una matriz de regresión construida a través de PLS (Partial Least Squares).

Este modelo ha sido entrenado con 124 sujetos sintéticos, de los que se dispone las superficies 3D del rostro y los contornos 2D para diferentes poses. Este método ha obtenido muy buenos resultados en la predicción de un nuevo sujeto, es decir, al estimar la forma 3D del rostro de un sujeto sintético que se había dejado fuera de entrenamiento.

Tras la validación del modelo anterior y los buenos resultados obtenidos se propone un nuevo enfoque, entrenar este modelo para predecir sujetos reales a partir de una secuencia de vídeo. De esta forma se consigue obtener reconstrucciones de la forma de la cara 3D con datos reales.

Para reconstruir la forma del rostro de un sujeto real se requiere extraer los contornos de los frames del vídeo correspondientes a un conjunto de poses determinadas. Para obtener dichos contornos, previamente se debe segmentar los frames del vídeo. Después, mediante PLS y la información de entrenamiento, se obtiene la matriz de regresión \mathbf{B} con la que podemos estimar el rostro del sujeto a partir de los contornos extraídos del vídeo.

Cabe destacar que conforme más contornos del sujeto se procesen, la reconstrucción que obtenemos más se asemeja al sujeto del vídeo, puesto que tenemos más información de éste.

Para validar el método, disponemos de una base de datos de 12 vídeos con sujetos en movimiento (es decir, con diferentes poses del rostro).

Por lo tanto, el objetivo de este proyecto es la definición, la implementación software y el análisis de un procedimiento que nos permita ajustar un modelo estadístico facial tridimensional genérico a las características faciales específicas de un individuo a partir de los contornos de su rostro.

1.2. Objetivos

Este proyecto tiene como objetivo principal la reconstrucción de la forma del rostro a partir de una secuencia de vídeo.

Para ello, se definieron los siguientes objetivos:

- Revisión de conceptos básicos de Procesado de Imagen.
- Revisión del Toolbox de Imagen de Matlab.
- Revisión de la literatura sobre la Reconstrucción facial.
- Evaluación de la factibilidad de utilizar un modelo 3D genérico que se ajuste a los contornos faciales de una imagen 2D.
- Creación de un programa, en lenguaje Matlab, que realice la Reconstrucción de la forma del rostro a partir de contornos.
- Evaluación de los Resultados.

Por último, uno de los objetivos más importantes es el aprendizaje que aporta un Proyecto Final de Carrera; es decir, no solo el conocimiento en una materia sino que también las habilidades que trabaja (organización de ideas, proceso de investigación, resolución de problemas, redactado de una memoria, preparación de una presentación, etc.).

1.3. Estructura de la memoria

La presente memoria está estructurada en seis capítulos.

En este primer capítulo, *Introducción*, se explica el problema que se desea resolver y la solución que se propone. También se contextualiza la línea de investigación y se detallan los objetivos del trabajo.

En el segundo capítulo, *Estado del Arte y descripción de las técnicas principales*, se da una visión general de las líneas de investigación más importantes que se han desarrollado. También se explican los fundamentos en los que se basa este proyecto y la teoría de las técnicas que utilizamos.

En el tercer capítulo, *Descripción del método*, se detalla la metodología de la reconstrucción de la forma del rostro, y se muestra un ejemplo con los pasos más significativos.

En el cuarto capítulo, *Resultados*, se evalúan los resultados obtenidos en las reconstrucciones, se muestra la Base de Datos utilizada y se explican las limitaciones que se han encontrado a lo largo de la investigación.

En el quinto capítulo, *Conclusiones*, se detallan las conclusiones y se presenta un posible enfoque como trabajo futuro.

Y en el sexto capítulo, *Referencias*, se muestra la Bibliografía que se consultado durante la investigación.

Por último, los *Anexos*, en los que se resume las funciones de código más importantes que se han desarrollado en el programa. Y también se añade más ejemplos de reconstrucciones que se han realizado.

Capítulo 2. Estado del arte y descripción de las técnicas principales

Las técnicas de estimación 3D del rostro se han convertido en un tema de especial interés debido a su implicación en aplicaciones de reconocimiento facial, modelado de caras en videojuegos, cirugía estética, realidad virtual, etc.

Durante las últimas décadas se han desarrollado múltiples técnicas computacionales de reconstrucción facial 3D. El método más eficaz logrado hasta ahora para construir un modelo facial 3D a partir de un rostro real es a través de un escáner 3D. Sin embargo estos dispositivos son costosos y sólo factibles en entornos controlados y estáticos donde se requiere la cooperación del usuario. La tecnología disponible y más utilizada para el monitoreo de entornos no controlados está basada en imágenes 2D. La reconstrucción facial 3D a partir de una o varias imágenes 2D es compleja y ha sido muy estudiada con métodos basados en el comportamiento de la luz (*shape-from-shading*), en la visión humana o estereoscópica (*structure-from-motion*) y métodos estadísticos (*3D Morphable Models*) entre otros.

Los métodos basados en modelos estadísticos parten de la idea de modelar la forma, superficie o textura de un rostro en particular, como la combinación lineal de un conjunto de modelos faciales 3D estadísticamente representativos por etnias y sexos. Esta combinación lineal puede ser vista como una interpolación de rostros. Teóricamente, si el conjunto es lo suficientemente representativo y los coeficientes de la combinación son correctamente restringidos, todo elemento resultante presentaría una fisonomía facial correcta y todos los rostros humanos podrían modelarse de esta manera.

2.1. Background

El problema de la estimación de la forma 3D del rostro se ha abordado habitualmente utilizando modelos estadísticos, donde la forma del rostro y la textura están relacionados. El objetivo principal de los métodos estadísticos es encontrar un conjunto de parámetros que mejor se ajusten a una o varias imágenes. Atick et al. fueron pioneros en esta área con su trabajo en [1], donde utilizaron PCA (Principal Component Analysis) para descomponer la estructura 3D de los sujetos de una base de datos. La idea principal era realizar un proceso de optimización de los parámetros del modelo estadístico con el fin de ajustarlos a la imagen de

entrada. Este enfoque se convirtió en el fundamento del *morphable model* de Blanz y Vetter [2], que formularon una optimización para el problema de la estimación 3D de la forma de la cara y la textura de una o más imágenes de entrada.

2.1.1. Aproximación con una sola vista

En [3] se utilizó PCA para construir modelos separados de la textura y de la forma 3D de la cara, considerando únicamente la información contenida en un conjunto de puntos de referencia. La información de la textura se obtuvo a través de una estrategia basada en la subdivisión triangular, mientras que la geometría facial se obtuvo con la proyección de la textura sobre un modelo de la forma 3D. Desafortunadamente, los datos calculados a partir de PCA no bastaban para conseguir una forma 3D precisa, haciendo necesario aplicar un filtro laplaciano para reducir el ruido.

Un trabajo similar se presentó en [4], donde se construyó una malla triangular mediante la interpolación de puntos característicos en la imagen de entrada, utilizando coordenadas *barycentric*. En este trabajo se construyeron dos modelos de PCA separados; uno de ellos contenía puntos sobre la malla triangular y el otro todos los puntos en la imagen. La idea principal de este trabajo fue calcular, a través de un modelo, los valores propios asociados a la malla triangular y luego utilizarlos para calcular la forma 3D de salida. Sin embargo, al construir dos modelos separados, la información compartida entre la forma 3D y la textura fue ignorada.

Una característica común entre los métodos de estimación de forma 3D del rostro mencionados anteriormente es que utilizan PCA. Sin embargo, cuando PCA se realiza por separado para la forma 3D de la cara y la textura, se ignora la información compartida entre ambos espacios. Para aprovechar dicha información aparece como alternativa MLR (Multiple Linear Regression). En [5] se propuso un método para estimar la forma 3D de la cara a partir de una imagen frontal utilizando el análisis de correlación canónica (CCA). CCA intenta encontrar dos conjuntos de variables canónicas de tal manera que la correlación entre las proyecciones sea máxima. Sin embargo, el criterio de correlación tiende a explicar la información con pequeña variación, es decir, la regresión da importancia a las características que son más relevantes para la textura que para la estimación de la forma 3D.

Otro enfoque basado en técnicas de MLR se introdujo en [6], donde se aplicó PLS (Partial Least Squares) en el proceso de regresión. PLS busca para un conjunto de componentes que realizan una descomposición simultánea de los espacios textura y forma 3D de la cara, con la restricción de que estos componentes explican la máxima covarianza entre ambos conjuntos. Aunque los resultados fueron alentadores, el uso de coordenadas cartesianas para la representación de la forma 3D complicó la tarea. Para superar este problema, se presentó un método en [7], donde la estructura 3D de la cara estimada se trató a través de coordenadas cilíndricas y PLS.

2.1.2. Aproximación con múltiples vistas y considerando contornos.

Los contornos de un rostro contienen información geométrica muy significativa de la forma 3D y se han incorporado en la investigación para mejorar el rendimiento de la estimación de la forma 3D de los modelos estadísticos.

En [8] los autores proponen dos métodos para recuperar la forma 3D de la cara a partir de imágenes binarias. El primer método utiliza cámaras calibradas para generar once imágenes de un sujeto. Para ello, se realizó una transformación 3D rígida entre las cámaras y los ejes de la cabeza del sujeto con el fin de alinear un modelo 3D genérico para la estimación de pose. Después se asignaron manualmente puntos de referencia de la cara y se anotaron en al menos tres imágenes del sujeto, mientras que los correspondientes puntos del modelo 3D se calcularon utilizando una optimización de *Levenberg-Marquardt*. Una vez que la alineación estaba solucionada, se extrajeron las imágenes binarias para la entrada y los correspondientes puntos de vista obtenidos de una aproximación inicial de los parámetros de la forma del modelo estadístico. Estos parámetros fueron iterativamente perfeccionados utilizando una optimización probabilística no lineal y una función de coste. Para el segundo método, un algoritmo similar se aplicó a una secuencia de vídeo de una cámara fija y una cara giratoria. Aquí, se necesitaron cuatro puntos de referencia para localizar una máscara 3D que rastrease la zona facial a lo largo de los fotogramas de vídeo. Tras encontrar los parámetros de movimiento para cada frame, se seleccionó un pequeño subconjunto de poses significativas para realizar la recuperación de la forma como en el primer método. Este trabajo demostró ser eficaz para el propósito de la estimación 3D de la forma de la cara. Sin embargo, mientras la aproximación es sensible a errores de alineación, el tiempo consumido en el proceso de optimización se ve afectado por la generación y la alineación de imágenes binarias. Por otra parte, la segmentación de la silueta de la cara de las imágenes de entrada se supone que es sencillo, pero en la práctica esto no es un problema trivial.

Otros esfuerzos se han centrado en la incorporación de los contornos en el sistema de Blanz y Vetter [2]. Por ejemplo, el trabajo de [9] se demostró que la combinación de múltiples características de la imagen tales como intensidad de los píxeles, bordes, reflejos especulares y las limitaciones de la textura, eran útiles para el modelado de la forma 3D del rostro y la textura de una sola imagen. La principal contribución de este trabajo fue un algoritmo de ajuste construido a partir de múltiples clasificadores que era más robusto a los problemas locales. Más tarde, en [10], Keller et. al. exploraron la contribución aislada de los contornos interiores y exteriores en el proceso de ajuste de un modelo *morphable* 3D. Para este fin, se presentó un sistema que encuentra coincidencias entre siluetas / contornos interiores de un modelo 3D *morphable* y una imagen de contorno de entrada. Los autores dividen el problema en dos partes principales: la extracción de características y el ajuste. Para el proceso de ajuste, la idea fue utilizar una función robusta de distancia que ignore bordes y contornos que no coincidan. La extracción de características se llevó a cabo mediante un algoritmo de detección de bordes. La principal conclusión de este trabajo fue que un solo contorno puede garantizar coherencia sin imponer fuertes restricciones sobre la forma plena de la cara 3D. Es decir, la cara 3D reconstruida puede ser muy diferente del *ground truth*, mientras que presenta el mismo contorno 2D que la imagen de entrada.

Como una alternativa a partir de modelos estadísticos de la forma 3D, otros enfoques han explorado el efecto de los contornos cuando está disponible un modelo genérico de referencia. En [11] se introdujo un modelo 3D de la cara basado en los contornos. En una primera etapa, la pose de la cabeza se estimó comparando los mapas de bordes extraídos de los fotogramas del vídeo con los mapas de bordes extraídos de un modelo genérico de cara 3D. Entonces, el modelo genérico de la cara 3D se adapta a los contornos exteriores y a las características internas de las deformaciones globales y locales. Para la deformación global se utilizó un modelo afín, mientras que para la deformación local se calculó una búsqueda estocástica. Al igual que en [8], este algoritmo resultó ser un éxito para la estimación 3D de la cara y más robusto a los cambios de iluminación que muchas estructuras a partir de algoritmos de movimiento. Sin embargo, la velocidad de cálculo también es un problema ya que el tiempo de reconstrucción utilizando 9 fotogramas de vídeo es de aproximadamente 1 hora.

Una solución diferente que combina modelos estadísticos con múltiples vistas se desarrolló en [12]. Como primer paso se realizó un ajuste iterativo para estimar las 5 poses de la cámara y después, se reconstruyó la forma de la cara 3D utilizando curvas de nivel. Desafortunadamente, incluso cuando los contornos añaden limitaciones en el proceso de adaptación, la exactitud de la reconstrucción 3D de la cara se sacrifica, es decir, la superficie recuperada es estable principalmente entorno a las zonas cercanas a la pose de la imagen de entrada. Esto está relacionado con el trabajo desarrollado en [13] y [14], donde se investigó

la flexibilidad en los modelos estadísticos 3D para determinar los modos de deformación generados en una superficie cuando ciertos vértices se ven limitados a permanecer fijos. La conclusión principal de este estudio fue que los modelos estadísticos de caras pueden generar sujetos cuyos contornos (perfiles) son prácticamente iguales, mientras que su aspecto general difiere considerablemente del rostro real.

En [15] se presenta una metodología basada en MLR para estimar la información de la forma 3D de la cara cuando una cara o un conjunto de contornos 2D está disponible desde múltiples puntos de vista. Este enfoque tiene como objetivo estimar directamente la estructura de la cara 3D mediante el uso de una matriz de regresión construido a través de PLS, donde la información común entre los contornos 2D y datos 3D se retiene en el proceso de regresión. La modelización de esta matriz se puede lograr utilizando el algoritmo PLS kernel, que es computacionalmente eficiente y sólo requiere que el número de observaciones (muestras de entrenamiento) sea el mismo para las variables dependientes e independientes. Este trabajo fue el primero en investigar el acoplamiento de contornos y superficies con una regresión lineal. Una limitación de este sistema es que la pose de la cabeza de las imágenes de entrada tiene que ser estimada para poder generar los datos de entrenamiento que construyen la matriz de regresión.

2.1.3. Modelo estadístico de Regresión [15]

El modelo que se presenta en [15] tiene como objetivo estimar la forma del rostro mediante el uso de una matriz de regresión construida con Partial Least Square (en la sección 2.2.1 se detallan los fundamentos matemáticos de PLS).

Este modelo dispone de una base de datos con contornos 2D (matriz de predictores \mathbf{X}) y superficies 3D (matriz de respuestas \mathbf{Y}) de diferentes sujetos generados sintéticamente [Figura 1.a y 1.b]. Mediante estas matrices (información de entrenamiento), el modelo es capaz de generar una matriz de regresión \mathbf{B} que dé las respuestas como una combinación lineal de los predictores [Figura 1.c]. Esta matriz de regresión nos permite estimar nuevos rostros 3D a partir de nuevos contornos 2D.

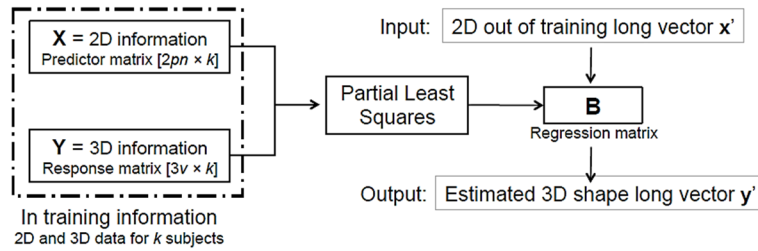
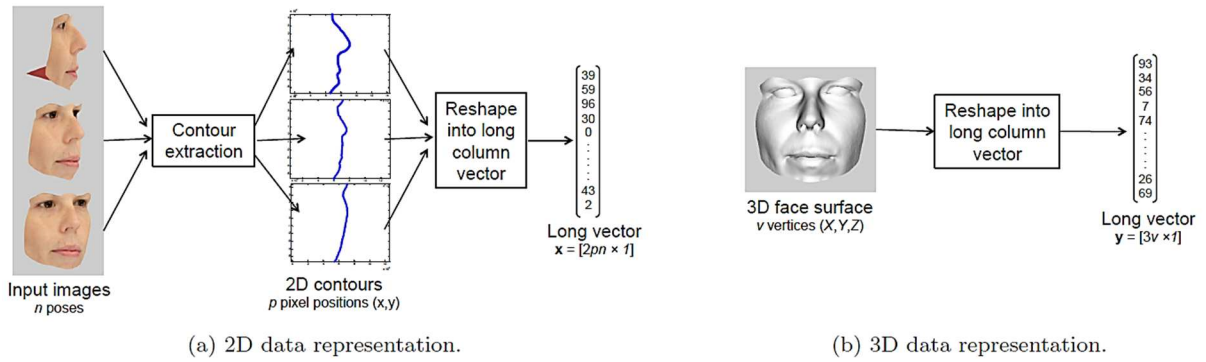


Figura 1. Reconstrucción 3D del rostro. En (a) la información 2D de la cara está representada como un vector columna \mathbf{x} que contiene las posiciones de los píxeles de los contornos. Todos los vectores están alineados con la media. En (b) la información 3D se representa por un vector \mathbf{y} columna que contiene las coordenadas (x,y,z) de cada vértice de la superficie de la cara. En (c) las matrices \mathbf{X} e \mathbf{Y} se utilizan para entrenar al sistema y calcular la matriz de regresión \mathbf{B} . Una vez que la matriz de regresión es conocida, a partir de unos nuevos contornos de entrada \mathbf{x}' puede estimarse el rostro 3D \mathbf{y}' .

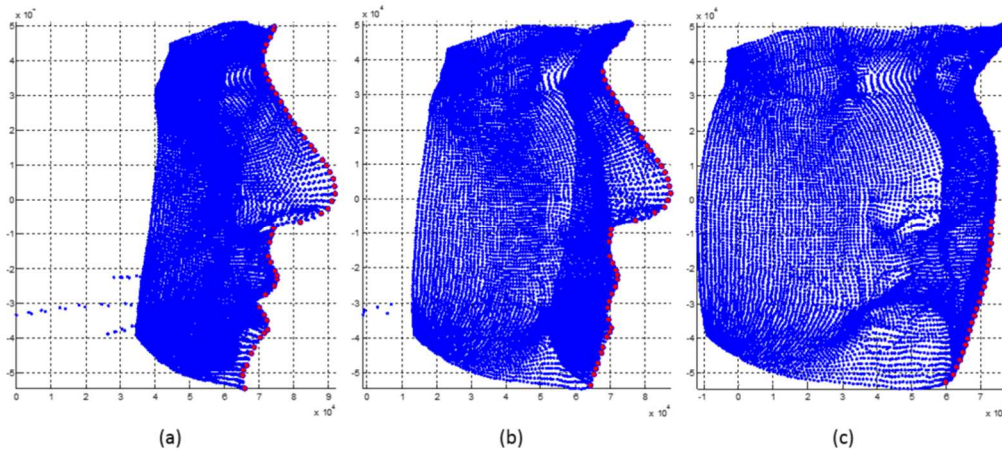


Figura 2. Información de entrenamiento. En esta figura se muestran diferentes poses de un sujeto sintético con los 16.512 vértices en azul y el contorno de pitch marcado en rojo. En (a) la pose de la cabeza se corresponde con el ángulo de pitch 90° . En (b) la pose con 70° . En (c) la pose con 50° .

La información de entrenamiento que se está usando se extrae de 124 sujetos sintéticos diferentes. De cada uno de ellos se dispone:

- 62 poses de la cabeza, correspondientes a las variaciones de los ángulos Roll (31 poses) y Pitch (31 poses) [Figura 3]. Para cada una de estas poses se dispone del contorno 2D extraído como 80 puntos equiespaciados en coordenadas (x,y) [Figura 1.a y 2]
- 16.512 vértices (x,y,z) correspondientes con la superficie 3D de la cara [Figura 1.b y 2].

La pose de la cabeza varía según los ángulos Roll y Pitch. Estos ángulos toman valores comprendidos entre $[-90^\circ, -15^\circ]$ U $[90^\circ, 15^\circ]$ con variaciones de $\Delta=5^\circ$:

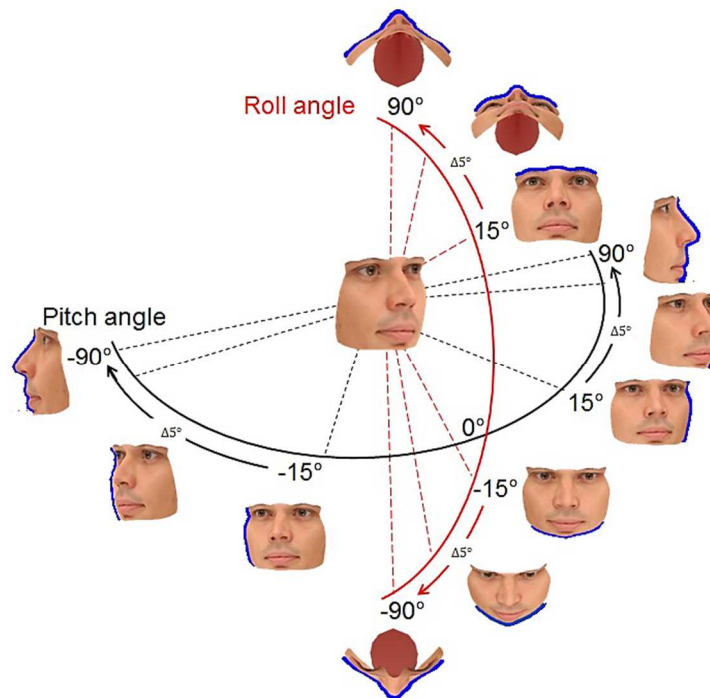


Figura 3. Contornos y variaciones de la pose. Variaciones de la pose según los ángulos Roll y Pitch.

Este proyecto usa el modelo de regresión para estimar la forma del rostro de un sujeto mediante los contornos 2D extraídos de una secuencia de vídeo. La base de datos de vídeos que disponemos contiene sujetos que giran según el ángulo de Pitch [Figura 4]; por ese motivo, para las reconstrucciones, nos hemos centrado únicamente en las variaciones de dicho ángulo, considerando como perfil principal el de Pitch = 90°.



Figura 4. Muestra de frames de una secuencia de vídeo. Ejemplo de frames de una secuencia de vídeo en la que se ve como el sujeto gira según el ángulo pitch $[-90^\circ, 90^\circ]$.

Siguiendo con el modelo de regresión, veamos la metodología con un ejemplo.

Disponemos de n contornos 2D de ángulo pitch $a_1, a_2, a_3, \dots, a_n$ con $a_i \in [-90^\circ, 90^\circ]$ de un sujeto real. Con estos datos podemos predecir la superficie del rostro mediante los siguientes pasos:

1. Preparación de la información de entrenamiento:
 - a. X será la matriz de predictores, que contendrá los contornos 2D de ángulo pitch $a_1, a_2, a_3, \dots, a_n$ de los 124 sujetos sintéticos. Por lo tanto X tendrá dimensiones $[80n \times 2 \times 124]$, donde n es el número de contornos. Recordemos que cada contorno 2D tiene 80 puntos equiespaciados en coordenadas (x,y) , es decir, es una matriz de $[80 \times 2]$.
 - b. Y será la matriz de respuestas, que contendrá las superficies 3D de los 124 sujetos sintéticos. Por lo tanto Y tendrá dimensiones $[16512 \times 3 \times 124]$. Recordemos que las superficies 3D disponen de 16.512 vértices (x,y,z) .
2. Conversión de las matrices X e Y en vectores largos de dimensiones: X $[160n \times 124]$ e Y $[49536 \times 124]$.
3. Obtención de la matriz de regresión B mediante Partial Least Squares y la información de entrenamiento: 2D (X predictores) y 3D (Y respuestas).
4. Una vez conocemos la matriz de regresión, mediante el nuevo vector columna de contornos reales x' podemos predecir el rostro del sujeto real $y' = B \cdot x'$. Cabe destacar que los contornos extraídos del vídeo se guardan concatenados uno tras otro en una matriz que es de dimensiones $[80n \times 2]$. Esta matriz posteriormente se transforma en un vector columna x' de dimensiones $[160n]$.

2.2. Descripción de las técnicas principales

2.2.1. Partial Least Squares (PLS)

Uno de los principales estudios en el análisis estadístico consiste en modelar problemas de gran dimensión, es decir, problemas que involucran un gran número de variables basados en pocas observaciones. Para estos tipos de problemas la meta es encontrar una estructura que exprese las características de los datos y los represente en un menor número de variables.

El método de Partial Least Squares (PLS) tiene la característica principal que puede ser aplicado para modelar cualquier tipo de problema aproximadamente lineal, aun cuando las variables que estén involucradas sean linealmente independientes o no. El estudio del PLS fue desarrollado inicialmente para aplicaciones quimiométricas; sin embargo, se extendió a otras ramas de la ciencia por ser una técnica útil en el modelado de problemas de regresión y clasificación, en especial en situaciones donde los datos se caracterizan por tener una alta dimensionalidad, elevada multicolinealidad (fuerte correlación entre variables) y pocas observaciones.

Partial Least Squares (PLS) tiene como objetivo obtener un modelo en un espacio de menor dimensión que maximice la covarianza entre una matriz independiente X (matriz de predictores) y otra matriz dependiente de X , Y (Matriz de respuestas). PLS captura la máxima variación en X con el número mínimo de variables y maximizando la correlación entre X e Y .

Partial Least Squares (PLS) relaciona la matriz de las variables de entrada $x \in X \subset R^m$ y la matriz de las variables de salida $y \in Y \subset R^p$ a través de sus variables latentes (r), las cuales se pueden definir como un conjunto de variables artificiales ortogonales que permiten hacer una reducción del espacio:

$$X = T \cdot P^T + E$$

$$Y = U \cdot C^T + F$$

$$Y = B \cdot X$$

B es la matriz de regresión. T y U son las matrices compuestas por los vectores latentes. P y C son las matrices compuestas por los vectores de peso. E y F son las matrices residuales.

La matriz de regresión B se puede calcular con el algoritmo Kernel PLS (K-PLS) [16], que es computacionalmente eficiente y sólo requiere que el número de observaciones sea el mismo para las variables dependientes e independientes.

Algoritmo Kernel Partial Least Squares

El algoritmo K-PLS se resume con el siguiente pseudocódigo:

X _ matriz de predictores \rightarrow \mathbf{X} vector columna de predictores
 Y _ matriz de respuestas \rightarrow \mathbf{Y} vector columna de respuestas

$$X_i = X_i - \text{media}(X)$$

$$Y_i = Y_i - \text{media}(X)$$

```
for (i = 1 ; i ≤ r ; i ++)  
    kernel =  $XX^T YY^T$   
     $t_i = Y(:,1)$   
     $t_i^T t_i = 1$   
     $u_i = YY^T t_i$   
     $G_i = I - t_i t_i^T$   
     $XX^T = XX^T - G_i \cdot \hat{X}_{i-1} \cdot \hat{X}_{i-1}^T \cdot G_i$   
     $YY^T = YY^T - G_i \cdot \hat{Y}_{i-1} \cdot \hat{Y}_{i-1}^T \cdot G_i$   
     $T = [T, t_i]$   
     $U = [U, t_i]$   
end
```

$$W = X^T \cdot U$$

$$P = (T^T \cdot X) \cdot (T^T \cdot T)^{-1}$$

$$C = (T^T \cdot Y) \cdot (T^T \cdot T)^{-1}$$

$$B = W \cdot (P^T \cdot W)^{-1} \cdot C^T$$

Previamente, se preparan las matrices de predictores y respuestas, \mathbf{X} e \mathbf{Y} , como vectores columnas centrados en media y escalados apropiadamente:

$$\mathbf{X}_i = X_i - \text{media}(X)$$

$$\mathbf{Y}_i = Y_i - \text{media}(X)$$

Después, se utilizan las matrices de asociación XX^T e YY^T para crear la matriz kernel $XX^T YY^T$. El primer vector propio de la matriz kernel se convierte en el primer vector latente del conjunto de predictores t_i ($i=1$), que debe cumplir la condición $t_i^T t_i = 1$ para asegurar el criterio de máxima covarianza. Después, se calcula el primer vector latente del conjunto de respuestas como $u_i = YY^T t_i$.

Tras el cálculo inicial de t_i y u_i se debe actualizar la matriz kernel. En las siguientes ecuaciones se muestra como XX^T e YY^T se actualizan mediante la matriz $G_i = I - t_i t_i^T$ para cada iteración de $i = \{1, 2, \dots, r\}$; donde i es el índice de la variable latente que se está calculando:

$$\begin{aligned}\widehat{X}_i \cdot \widehat{X}_i^T &= G_i \cdot \widehat{X}_{i-1} \cdot \widehat{X}_{i-1}^T \cdot G_i \\ \widehat{Y}_i \cdot \widehat{Y}_i^T &= G_i \cdot \widehat{Y}_{i-1} \cdot Y_{i-1}^T \cdot G_i\end{aligned}$$

Cabe destacar que las matrices obtenidas siguen siendo de dimensiones $k \times k$. El objetivo de la actualización de las matrices \widehat{X}_i e \widehat{Y}_i es calcular los residuos de X e Y .

Una vez que la dimensión calculada se ha restado de la matriz kernel ($X_i X_i^T - \widehat{X}_i \widehat{X}_i^T$ e $Y_i Y_i^T - \widehat{Y}_i \widehat{Y}_i^T$), se calculan los siguientes vectores propios para obtener los nuevos vectores t_i y u_i .

Cuando se alcanza el número de dimensiones r , todos los vectores t y u se almacenan como vectores columna en las matrices T y U respectivamente. Finalmente se calcula la matriz de regresión B como:

$$\begin{aligned}B &= W \cdot (P^T \cdot W)^{-1} \cdot C^T \\ W &= X^T \cdot U \\ P &= (T^T \cdot X) \cdot (T^T \cdot T)^{-1} \\ C &= (T^T \cdot Y) \cdot (T^T \cdot T)^{-1}\end{aligned}$$

Las matrices W y C contienen las variaciones de la información extraída de los predictores y las respuestas; P es la matriz de carga de los predictores.

Una vez finalizada la regresión, se puede estimar un nuevo vector de respuestas y' usando la matriz B y un nuevo predictor x' :

$$y' = B \cdot x'$$

2.2.2. Ultrametric Contour Map (UCM)

El ser humano es capaz de percibir las imágenes digitales, formadas por píxeles, como un conjunto de formas y objetos. La segmentación de imágenes en regiones permite un análisis más cercano a la percepción humana y una reducción significativa de procesamiento, puesto que pasamos de trabajar con millones de píxeles a decenas/cientas de regiones.

La segmentación es el proceso por el cual somos capaces de dividir una nube de puntos en distintas regiones de interés o extraer las características más importantes de la imagen. La gran mayoría de los métodos de segmentación se pueden clasificar en tres categorías: métodos de detección de bordes, métodos de crecimiento de regiones y métodos híbridos.

Para este proyecto hemos utilizado un método de detección de bordes llamado *Ultrametric Contour Map* (UCM) [17], en el que se construye una segmentación jerárquica de la imagen a partir de sus contornos.

El algoritmo de segmentación Ultrametric Contour Map (UCM) consta de tres módulos principales:

El primero, el método de detección de contornos gPb , combina linealmente dos tipos de técnicas para detectar contornos, una de ellas proporciona un conjunto de imágenes donde los píxeles representan la fuerza del contorno en una orientación dada y la otra utiliza un *clustering* espectral sobre los resultados anteriores. Esta información se combina para formar la componente espectral del detector.

Los resultados obtenidos por el detector de contornos gPb , $E(x, y)$ [Figura 5.b], no son contornos cerrados y por lo tanto no se puede crear una segmentación a partir de ellos.

El siguiente paso de la segmentación, *Oriented Watershed Transform* (OWT), crea un conjunto de regiones (contornos cerrados) a partir de los resultados de gPb [Figura 5.c].

Por último, con la partición obtenida anteriormente, se crea UCM que es un árbol jerárquico de regiones que resulta de la fusión de las regiones más similares [Figura 6]. Esta fusión se crea según una escala finita y la localización de los contornos se preservará en las distintas escalas. Incrementar esta escala equivale a eliminar contornos y fusionar las regiones.

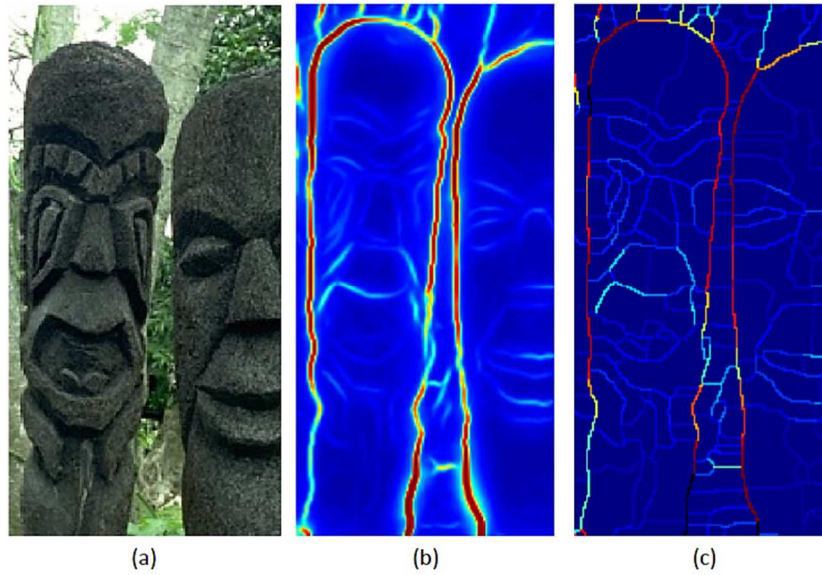


Figura 5. Detección de contornos gPb y Oriented Watershed Transform. En (a) la imagen original a segmentar. En (b) la imagen de contornos $E(x, y)$ obtenida mediante gPb . En (c) los resultados tras aplicar OWT.

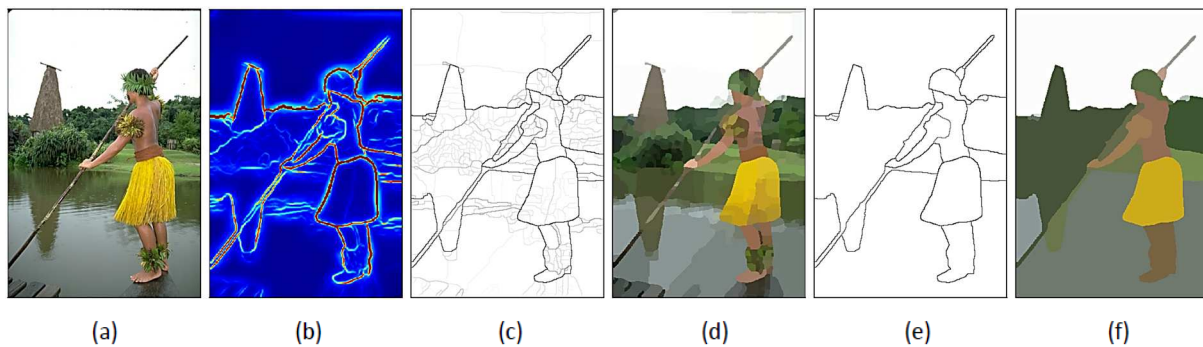


Figura 6. Segmentación jerárquica de contornos. En (a) la imagen original. En (b) la imagen de contornos obtenida mediante la respuesta máxima del detector de contornos gPb . En (c) los resultados tras aplicar OWT-UCM. En (d) la imagen que se corresponde con la partición inicial de UCM. En (e) los contornos correspondientes a la escala 0.5 de UCM y en (f) la imagen segmentada a escala 0.5 de UCM.

Como resultado, el nivel base de esta jerarquía respeta los contornos débiles y tiende a corresponderse con una sobre-segmentación de la imagen, mientras que los niveles superiores solo respeta los contornos fuertes y por lo tanto se corresponde con sub-segmentación.

2.2.3. Transformada de Hough Generalizada

En el análisis automatizado de imágenes, es común encontrar el problema de detectar figuras, tales como rectas o circunferencias. La Transformada de Hough es una técnica muy usada en el campo de Visión por Ordenador, que detecta figuras en imágenes digitales.

La transformada de Hough (HT) fue desarrollada inicialmente para detectar formas definidas analíticamente, por ejemplo: líneas, círculos, elipses, etc. En estos casos, tenemos conocimiento de la forma y el objetivo es conocer su ubicación y orientación en la imagen.

La transformada de Hough generalizada (GHT) [18], es la modificación de la transformada de Hough utilizando el principio de comparación de plantillas. Esta modificación permite utilizar la HT para la detección de objetos descritos con una función analítica y también, objetos arbitrarios descritos en un modelo.

Con la GHT, el problema de encontrar la posición de un objeto (descrito en un modelo) se transforma a un problema de encontrar el parámetro de transformación que mapea el modelo en la imagen. Si determinamos el valor del parámetro de transformación, la posición del modelo en la imagen puede ser determinada.

La implementación original de la GHT utiliza la información de contorno para definir un mapeo desde un punto del contorno al punto de referencia de la imagen. En el caso de una imagen binaria, donde los píxeles pueden ser blancos o negros, cada píxel negro de la imagen puede ser un píxel negro del patrón deseado, creando así un lugar geométrico de puntos de referencia en el espacio de Hough. Cada píxel de la imagen vota por sus puntos de referencia correspondientes. Los puntos máximos del Espacio Hough indican posibles puntos de referencia del patrón en la imagen. Este máximo se puede encontrar mediante el escaneo del Espacio Hough o mediante la resolución de un conjunto de ecuaciones, cada una de ellas correspondiente a un píxel negro.

2.2.3.1. Algoritmo

Para generalizar el algoritmo de Hough a las curvas no analíticas, *Ballard* en [18] define los siguientes parámetros para una forma generalizada: $\mathbf{a} = \{\mathbf{y}, \mathbf{s}, \boldsymbol{\theta}\}$ donde $\mathbf{y} = (x_c, y_c)$ es un origen de referencia para la forma, $\boldsymbol{\theta}$ es su orientación, y $\mathbf{s} = (s_x, s_y)$ describe dos factores de escala ortogonales.

El algoritmo consta de tres etapas principales: Construcción de la *R-table*, Localización del objeto y Generalización del escalado/orientación.

- Construcción de la R-table

Inicialmente se elige un punto de referencia $y = (x_c, y_c)$ para la forma (normalmente se elige dentro de ésta). Después, para cada punto frontera $x = (x, y)$, se calcula la dirección del gradiente $\Phi(x)$, y la distancia $r = y - x$ [Figura 7].

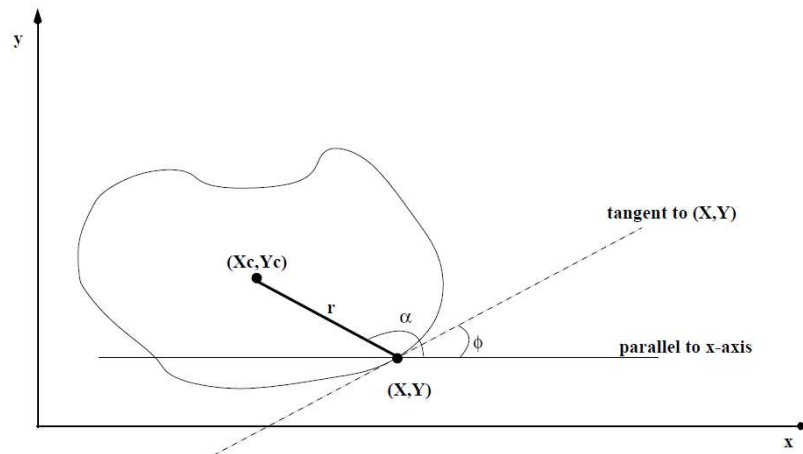


Figura 7. Geometría de la detección de la forma para la GHT. El punto $y = (x_c, y_c)$ es el punto origen de referencia de la forma, $x = (x, y)$ es un punto frontera, Φ se corresponde al ángulo con la perpendicular a la dirección del gradiente y por último, r y α definen la distancia al punto origen de referencia $y = (x_c, y_c) = (x + r \cdot \cos\alpha, y + r \cdot \sin\alpha)$.

La distancia entre el punto de referencia y un punto frontera, r , se puede almacenar como diferencia de coordenadas $(x_c - x, y_c - y)$ o como distancia radial y ángulo (r, α) . Después de haber calculado r y $\Phi(x)$ para cada punto frontera, la *R-table* representará plenamente el objeto del modelo [Tabla 1].

i	Φ_i	R_{Φ_i}
1	0	$(r_{11}, \alpha_{11}) (r_{12}, \alpha_{12}) \dots (r_{1n}, \alpha_{1n})$
2	$\Delta\Phi$	$(r_{21}, \alpha_{21}) (r_{22}, \alpha_{22}) \dots (r_{2m}, \alpha_{2m})$
3	$2\Delta\Phi$	$(r_{31}, \alpha_{31}) (r_{32}, \alpha_{32}) \dots (r_{3k}, \alpha_{3k})$
...

Tabla 1. R-table de la GHT. Todos los puntos frontera del objeto están representados en la R-table a partir de sus parámetros r y $\Phi(x)$.

- Localización del objeto

Se crea una matriz acumulador, A , inicializada con el tamaño máximo de la imagen.

Para cada píxel frontera \mathbf{x} de la imagen, recuperamos el gradiente Φ y se incrementa todos los puntos correspondientes a $\mathbf{x} + \mathbf{r}$ en el acumulador \mathbf{A} (\mathbf{r} es una entrada de la tabla indexada por Φ , es decir, $\mathbf{r}(\Phi)$). Estos puntos de entrada nos dan cada posible posición para el punto de referencia. Aunque se pueden calcular algunos puntos falsos, dado que existe el objeto en la imagen, encontraremos un máximo en el punto de referencia. El máximo de \mathbf{A} se corresponde con el punto de referencia de la forma.

- Generalización de la escala y la orientación

Para una orientación fija de la forma, la matriz acumulador es bidimensional por las coordenadas del punto de referencia. Para buscar formas de arbitraria orientación θ y escalado s , se añaden estos dos parámetros a la descripción de la forma. La matriz acumulador ahora consta de cuatro dimensiones correspondientes a los parámetros (\mathbf{y}, s, θ) .

2.2.3.2. Implementación

Supongamos que el objeto ha sufrido alguna rotación θ y un escalado s uniforme.

Inicializamos el acumulador \mathbf{A} :

$$A[x_c \min \dots x_c \max][y_c \min \dots y_c \max][\theta_{\min} \dots \theta_{\max}][s_{\min} \dots s_{\max}]$$

Para cada punto frontera (\mathbf{x}, \mathbf{y}) , a través de su ángulo gradiente Φ recuperamos todos los valores (\mathbf{r}, α) de la *R-table*. Para cada (\mathbf{r}, α) calculamos los puntos de referencia candidatos:

$$x' = r \cdot \cos\theta$$

$$y' = r \cdot \sin\theta$$

$$\text{for } (\theta = \theta_{\min} ; \theta \leq \theta_{\max} ; \theta ++)$$

$$\text{for } (s = s_{\min} ; s \leq s_{\max} ; s ++)$$

$$x_c = x - s \cdot (x' \cdot \cos\theta - y' \cdot \sin\theta)$$

$$y_c = y - s \cdot (x' \cdot \sin\theta + y' \cdot \cos\theta)$$

$$++ A[x_c][y_c][\theta][s]$$

Las posibles ubicaciones del contorno del objeto son dados por los máximos locales de \mathbf{A} . Si $A[x_c][y_c][\theta][s] > T$, entonces el contorno del objeto se localiza en (x_c, y_c) y ha sufrido una rotación θ y un escalado s .

2.2.4. Chehra Face Tracker

El ser humano es capaz de procesar rostros muy rápidamente; basta con menos de un segundo para reconocer a alguien. En cambio, cuando se trata de un software, la detección de personas en imágenes/vídeos requiere un proceso más complejo.

El rostro humano tiene características únicas como la distancia entre los ojos, anchura de la nariz, forma de la barbilla, pómulos, forma de la boca, etc. El rostro es un objeto dinámico que tiene un alto grado de variabilidad en su apariencia, lo cual hace que su detección sea un problema difícil de tratar en visión por ordenador.

Existe una gran variedad de técnicas de detección facial, desde algoritmos básicos de detección de bordes hasta algoritmos compuestos de alto nivel que utilizan métodos avanzados de reconocimiento de patrones. Estas técnicas se basan en diferentes enfoques:

- *Enfoques basados en rasgos faciales o características locales*, en los que se buscan determinados elementos que componen una cara, como pueden ser los ojos, la nariz, la boca, etc.
- *Enfoques holísticos o basados en la imagen*, en este caso los métodos trabajan con la imagen completa o zonas concretas de la misma, efectuando cálculos que determinan si hay una cara o no, sin buscar rasgos concretos.
- *Enfoques híbridos*, estos métodos usan tanto información local como global para la detección, basándose en el hecho de que el sistema de percepción humano distingue tanto las características locales como globales del rostro.

Para este proyecto requeríamos detectar una imagen en la que el rostro del sujeto apareciese con una pose frontal. Con este propósito hemos utilizado *Chehra Face Tracker* [19], un sistema basado en rasgos faciales que es capaz de detectar puntos faciales.

Este método trata el problema de la actualización de un modelo deformable facial discriminativo. Estos métodos se basan en el uso de la información discriminativa, es decir, un conjunto de clasificadores de puntos de referencia faciales o funciones discriminativas. *Chehra* estudia el entrenamiento gradual de los modelos discriminativos que utilizan una cascada de regresores lineales para el mapeo de la textura a la forma del rostro. En [19] se demuestra que el método incremental propuesto:

- Es capaz de añadir nuevas muestras de entrenamiento y actualizar el modelo sin re-entrenar desde cero, por lo tanto, incrementando la robustez del modelo genérico.
- Puede adaptar de forma automática el sujeto que está rastreando y las condiciones de imagen usando una secuencia de imágenes.

Por lo tanto, *Chehra* (que significa "cara" en indio) es un sistema que rastrea la cara y los ojos de forma completamente automática y en tiempo real. Este software detecta puntos de referencia (*landmarks*) de la cara y los ojos [Figura 8].

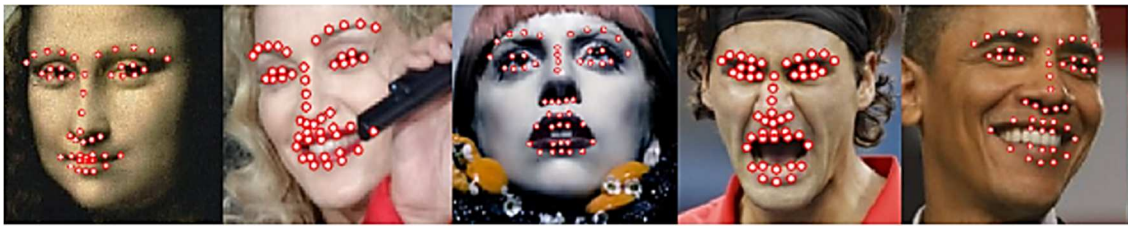


Figura 8. *Chehra Face Tracker*. Ejemplos de *Landmarks* de cara y ojos.

Las características principales que ofrece este detector son las siguientes:

- Sistema completamente automático capaz de rastrear la cara a más de 80 fps.
- Detección de 49 puntos faciales. [Figura 9]
- Detección de 10 puntos de los ojos.
- Estimaciones de la pose de la cabeza 3D.

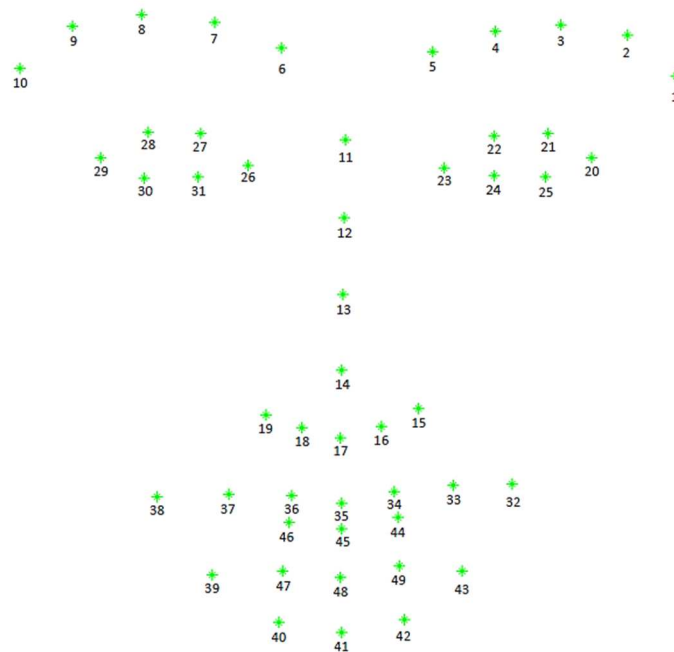


Figura 9. *Chehra Face Tracker*. Numeración de los 49 puntos faciales que detecta. Las coordenadas de dichos puntos son almacenados en un *array*.

Capítulo 3. Descripción del método

En este capítulo se presenta la metodología de la Reconstrucción de la forma del rostro a partir de los contornos extraídos de un vídeo.

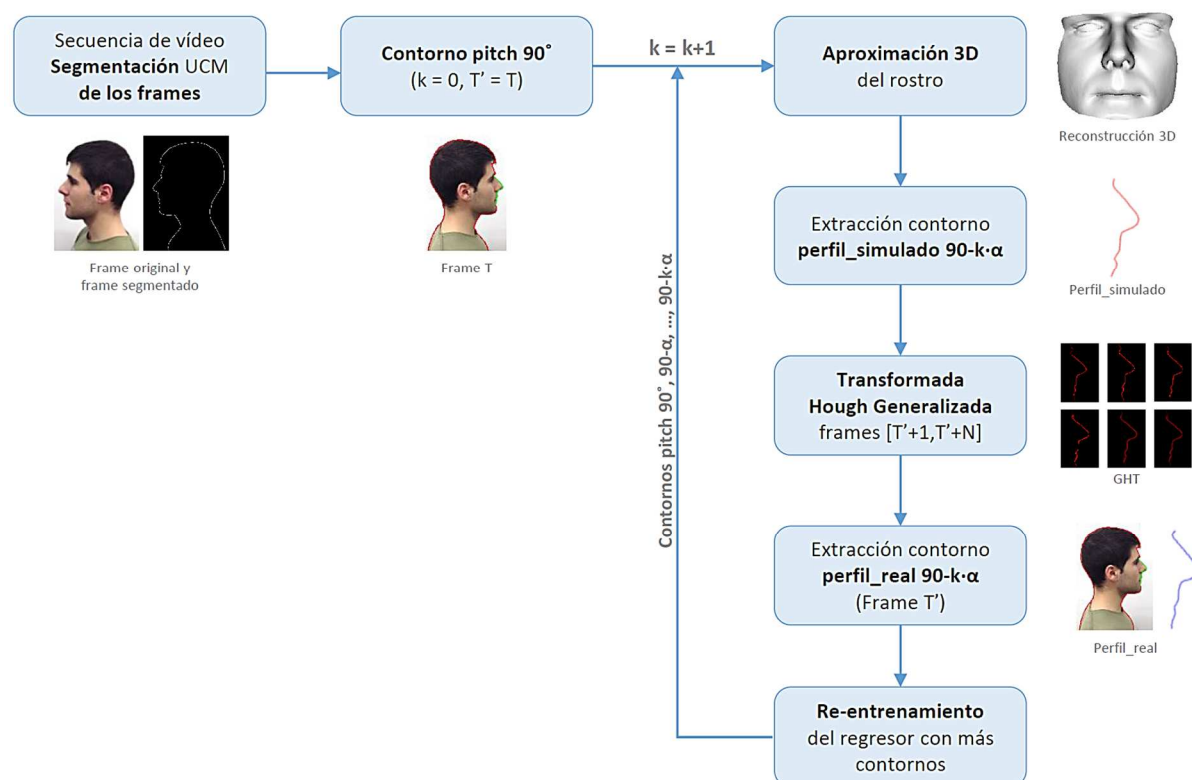


Figura 10. Reconstrucción de la forma del rostro a partir de los contornos de una secuencia de vídeo.

Originalmente se dispone de una secuencia de vídeo en la que un sujeto gira sobre sí mismo (tenemos diferentes poses de la cabeza según variaciones del ángulo pitch [Figura 4]).

Para reconstruir la forma del rostro del sujeto se requiere extraer los contornos de los frames del vídeo correspondientes a un conjunto de ángulos de pitch determinados: $90, 90-\alpha, 90-2\cdot\alpha, \dots, 90-k\cdot\alpha$. Para obtener dichos contornos, previamente se debe segmentar los frames del vídeo.

El problema que tenemos es que la base de datos (vídeos) ha sido grabada con una cámara no calibrada, como puede ser la de un Smartphone, por lo que no sabemos qué frame se corresponde con cada ángulo de pitch. Para abordar este problema se ha usado la Transformada de Hough Generalizada.

Mediante Partial Least Square y la información de entrenamiento, se obtiene la matriz de regresión \mathbf{B} [ver Estado del Arte, sección 2.1.3]. Con dicha matriz podemos estimar el rostro del sujeto \mathbf{y}' a partir de los contornos de entrada \mathbf{x}' .

El proceso de Regresión es iterativo, puesto que se realiza por cada contorno que se añade a la reconstrucción. El regresor debe ser entrenado con los mismos datos que los contornos que procesamos del vídeo; es decir, si se desea reconstruir el rostro con los contornos a° y b° extraídos del vídeo, entonces el modelo debe haber sido entrenado con los datos correspondientes a esos mismos ángulos.

Cabe destacar que conforme más contornos del sujeto se procesen, la reconstrucción que obtenemos más se asemeja al sujeto del vídeo, puesto que tenemos más información de éste.

La reconstrucción de la forma del rostro se compone básicamente de tres pasos:

- Segmentación de los frames del vídeo
- Extracción y procesamiento del contorno del sujeto
- Proceso de regresión

Tras finalizar el Proceso de la regresión, se vuelve al paso previo (extracción y procesamiento del contorno). Por ello que la reconstrucción es un proceso iterativo.

En los siguientes sub-apartados se detalla el proceso de la Reconstrucción de la forma del rostro.

3. 1. Segmentación de los frames del vídeo

Inicialmente partimos de una secuencia de vídeo, en la que tenemos un sujeto en movimiento y del que queremos extraer los contornos del rostro. Para ello tenemos que realizar una segmentación de los frames en la que consigamos separar el contorno del sujeto y el fondo del vídeo. Para obtener estos contornos realizamos la segmentación jerárquica UCM en cada frame de la secuencia [Figura 11.b].

La muestra de vídeos que hemos analizado tiene una característica común: el fondo de todos los frames es blanco. Este detalle nos permite detectar fácilmente la región que se corresponde con el fondo (basta con identificar un punto de esa región, por ejemplo el píxel (1,1), es decir, el píxel superior-izquierdo). Una vez detectada la región del fondo, sabemos que todas las regiones restantes se corresponderán con las del sujeto [Figura 11.c]. Finalmente conseguimos extraer el contorno deseado [Figura 11.d].

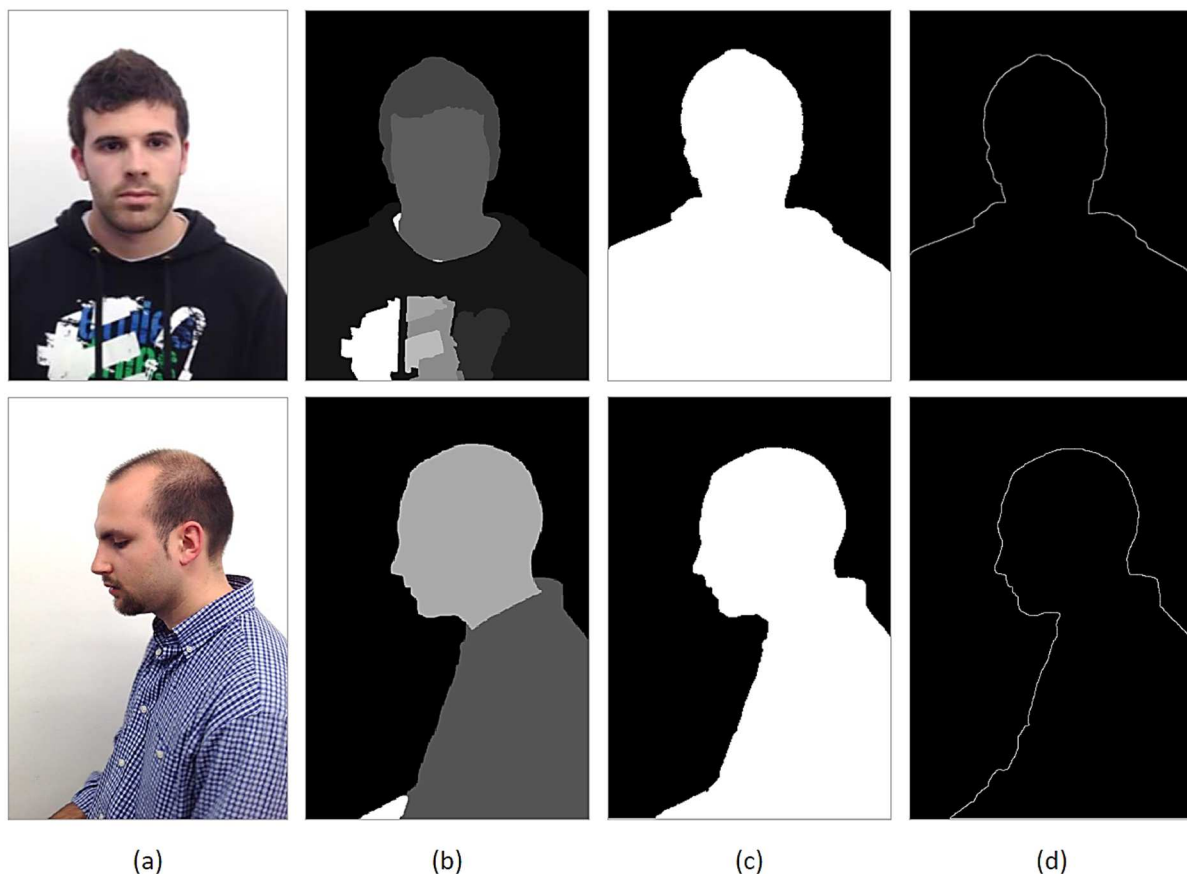


Figura 11. Segmentación jerárquica UCM. En (a) el frame original del vídeo. En (b) la segmentación jerárquica obtenida con UCM de nivel 0.8. En esta imagen se pueden observar las diferentes regiones detectadas. En (c) las regiones sujeto y fondo totales. En (d) el contorno del sujeto extraído.

3. 2. Extracción y procesado de los contornos

Una vez tenemos todos los frames de la secuencia de vídeo segmentados, debemos extraer y tratar los contornos hallados. La finalidad de este proyecto es la aproximación / reconstrucción 3D facial de un sujeto, por lo que únicamente necesitamos procesar los contornos que aportan información del rostro.

El modelo estadístico de reconstrucción que utilizamos (detallado en el Estado del Arte) requiere contornos de 80 puntos equiespaciados en coordenadas (x, y) , normalizados según la media del modelo, y estrictamente delimitados en una región muy concreta del rostro [Figura 12].

En el procesamiento de los contornos intervienen principalmente dos funciones: una que interpola 80 puntos (equiespaciados) a lo largo de la curva 2D del contorno [Figura 12.b]; y otra que determina la transformación lineal (translación, reflexión, rotación ortogonal y escalado) del contorno hallado para adaptarlo mejor a los puntos de la media de contorno del modelo [Figura 12.c].

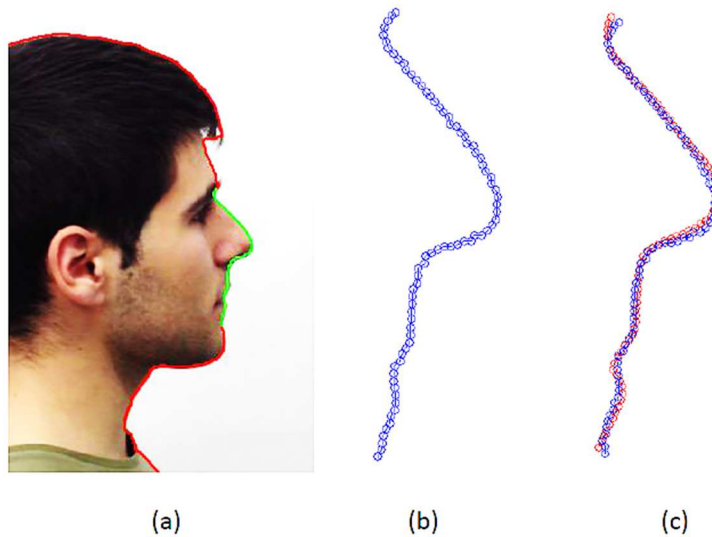


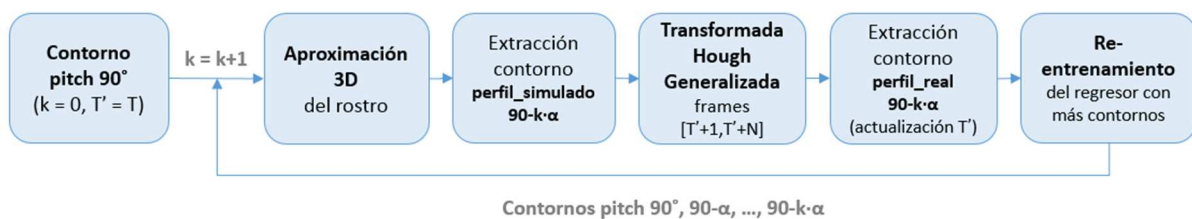
Figura 12. Adquisición y procesamiento de los contornos. En (a) un frame original del vídeo con el contorno del sujeto marcado en rojo y el contorno útil para la reconstrucción en color verde. En (b) el contorno útil extraído es tratado como 80 puntos equiespaciados. En (c) el contorno anterior, en azul, normalizado según la media del modelo, en color rojo.

Los contornos procesados anteriormente son utilizados en conjunto con un algoritmo de reconstrucción de la forma 3D del rostro a partir de contornos y modelos estadísticos. Este algoritmo funciona bajo la condición de que las poses de la cabeza, a partir de las cuales se extraen los contornos del rostro, son conocidas.

El objetivo es extraer del vídeo los contornos correspondientes a un conjunto de ángulos de pitch determinados: $90, 90-\alpha, 90-2\cdot\alpha, \dots, 90-k\cdot\alpha$. Estos contornos son utilizados en la reconstrucción de la forma del rostro.

3. 3. Regresión

El proceso iterativo de la regresión consta básicamente de los siguientes pasos:



3.3.1. Contorno principal y primera aproximación del 3D rostro



Para iniciar el método de Reconstrucción se determina manualmente el frame correspondiente al ángulo pitch 90° (al que llamaremos frame T). Este contorno principal se procesa para que tenga las condiciones requeridas por el modelo, es decir, para que sea un contorno de 80 puntos equiespaciados y normalizado según la media del modelo a pitch 90° .

Ahora, debemos calcular la matriz de regresión B que nos permite obtener la reconstrucción del rostro con el contorno de pitch 90° . Para calcular B se entrena el modelo estadístico con la siguiente información de entrenamiento: X contiene los contornos 2D de pitch 90° de los sujetos sintéticos e Y contiene las superficies 3D de los sujetos sintéticos.

Una vez la matriz B es conocida y el contorno principal se ha preparado como un vector columna x' , podemos obtener la primera aproximación 3D del rostro mediante $y' = B \cdot x'$ [Figura 13].



Figura 13. Primera aproximación 3D del rostro (entrenamiento con contorno principal, pitch 90°)

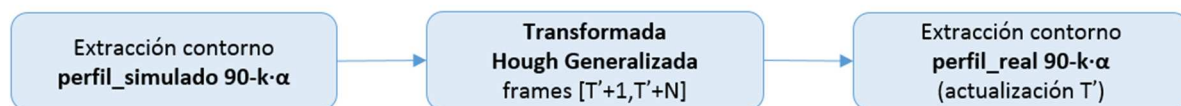
Para obtener reconstrucciones 3D más semejantes al Sujeto bajo estudio, se requiere re-entrenar el Regresor con más perfiles. Como se ha comentado anteriormente, contra más contornos se introducen, más características/datos del sujeto se están procesando y por lo tanto, más semejante deberá ser la reconstrucción al sujeto del vídeo.

Por lo tanto, necesitamos extraer el contorno del frame que contiene la pose de ángulo pitch $90-\alpha$.

3.3.2. Selección del frame con contorno de pitch $90-k\cdot\alpha$ ($k = 1$)

Encontrar el frame del vídeo que se corresponde con el ángulo pitch $90-\alpha$ se convierte en un problema no trivial. La muestra de vídeos que tenemos ha sido grabada con una cámara sencilla, como puede ser la de un Smartphone, por lo que no sabemos cuál frame es el que se corresponde con el ángulo pitch que buscamos. Para abordar este problema se propone la siguiente estrategia:

En el paso previo hemos obtenido la reconstrucción 3D con el contorno principal (pitch 90°). Puesto que el modelo estadístico crea rostros uniformes, podemos intuir que los contornos de la aproximación 3D muy cercanos al contorno principal (es decir, los de pitch $90-\alpha$ con α pequeña) deberían ser muy similares a los contornos reales (es decir, los que extraeríamos del vídeo). Por esa razón, se extrae el contorno de pitch $90-\alpha$ de la superficie 3D generada y se compara con los contornos de un rango de frames del vídeo. El frame que contenga el contorno más parecido al extraído de la aproximación 3D, se asumirá como el de pitch $90-\alpha$.



Por lo tanto, siguiendo el enfoque anterior, debemos extraer el contorno de $90-\alpha$ de la superficie 3D generada (al que llamamos *perfil_simulado*), para compararlo con un rango de frames $[T+1\dots T+N]$ del vídeo. El frame que contenga el contorno más semejante al *perfil_simulado*, que es el que obtiene un *score* mayor en la Transformada de Hough Generalizada, se asume como el frame T' correspondiente al ángulo pitch $90-\alpha$. [Figura 14].

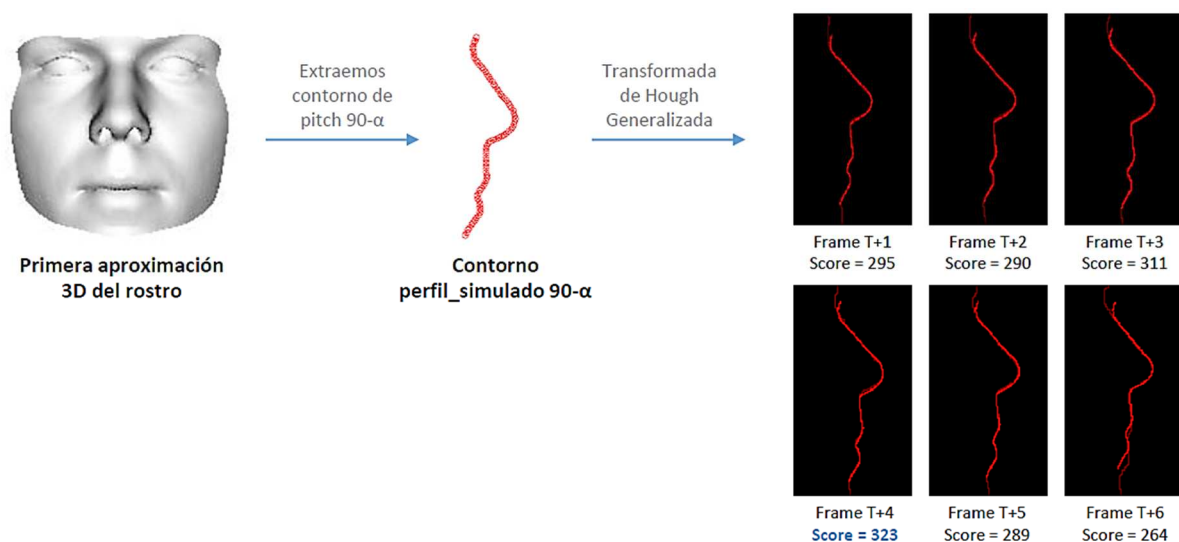


Figura 14. Selección del frame con pitch $90-\alpha$ (el de mayor score).

Recordemos que el *score* de la GHT es la votación que reciben los posibles puntos de referencia del patrón en la imagen. El que obtiene el máximo *score* es el punto de referencia.

Una vez encontrado el frame T' de pitch $90-\alpha$, se extrae el contorno obteniendo el perfil_real de $90-\alpha$ (contorno de 80 puntos equiespaciados y normalizado según la media del modelo a pitch $90-\alpha$) [Figura 15].

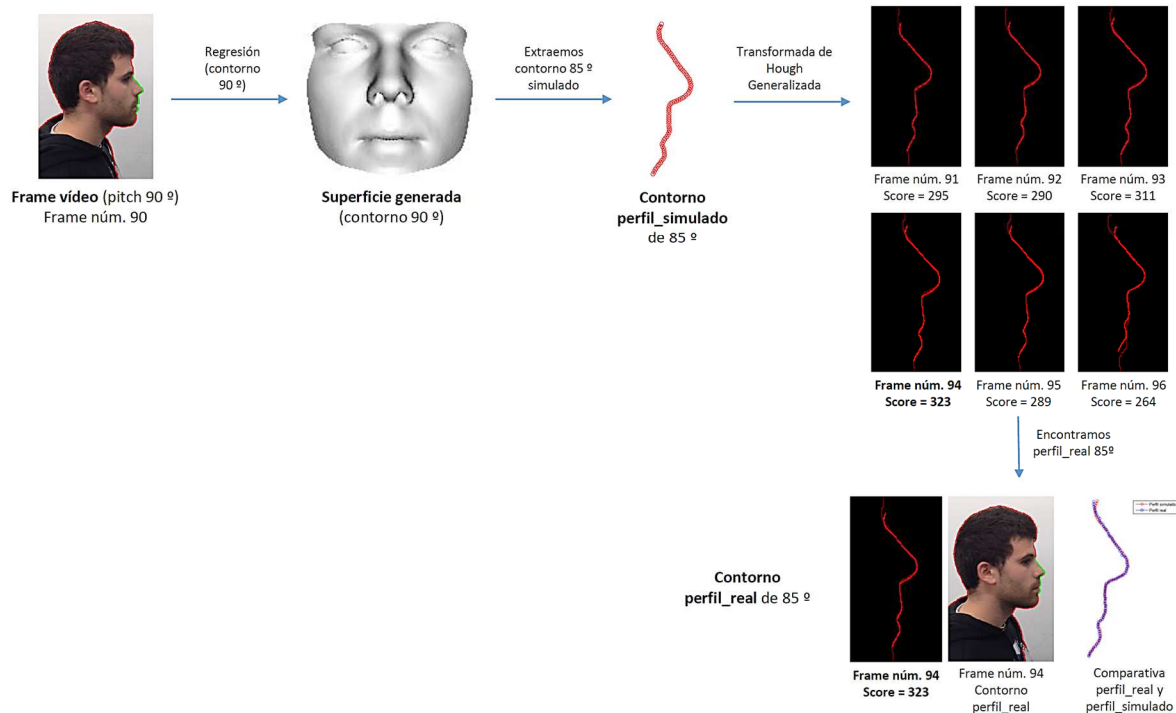
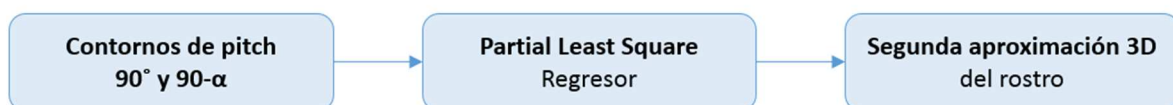


Figura 15. Primera reconstrucción 3D del rostro y obtención del contorno perfil_real $90-\alpha$ (ej. $\alpha=5^\circ$).

3.3.3. Re-entrenamiento del Regresor con más perfiles

A continuación, se quiere re-entrenar el modelo estadístico para obtener una nueva reconstrucción con los contornos de pitch 90° y $90-\alpha$.



Al igual que se hizo en la primera aproximación, debemos calcular la matriz de regresión B que nos permita obtener la reconstrucción del rostro con los contornos de pitch 90° y $90-\alpha$ (perfil_real de $90-\alpha$).

Para calcular B se re-entrena el modelo estadístico con la siguiente información de entrenamiento: X contiene los contornos 2D de pitch 90° y $90-\alpha$ de los sujetos sintéticos e Y contiene las superficies 3D de los sujetos sintéticos.

Una vez la matriz B es conocida y los contornos reales se han preparado como un vector columna x' , podemos obtener la segunda aproximación 3D del rostro mediante $y' = B \cdot x'$

A continuación, se extrae el perfil_simulado de $90-2\cdot\alpha$, que será nuevamente comparado con el rango de frames $[T'+1\dots T'+N]$, para encontrar el frame del vídeo que contenga el contorno que más se asemeje al perfil_simulado (frame de pitch $90-2\cdot\alpha$) [Figura 16].

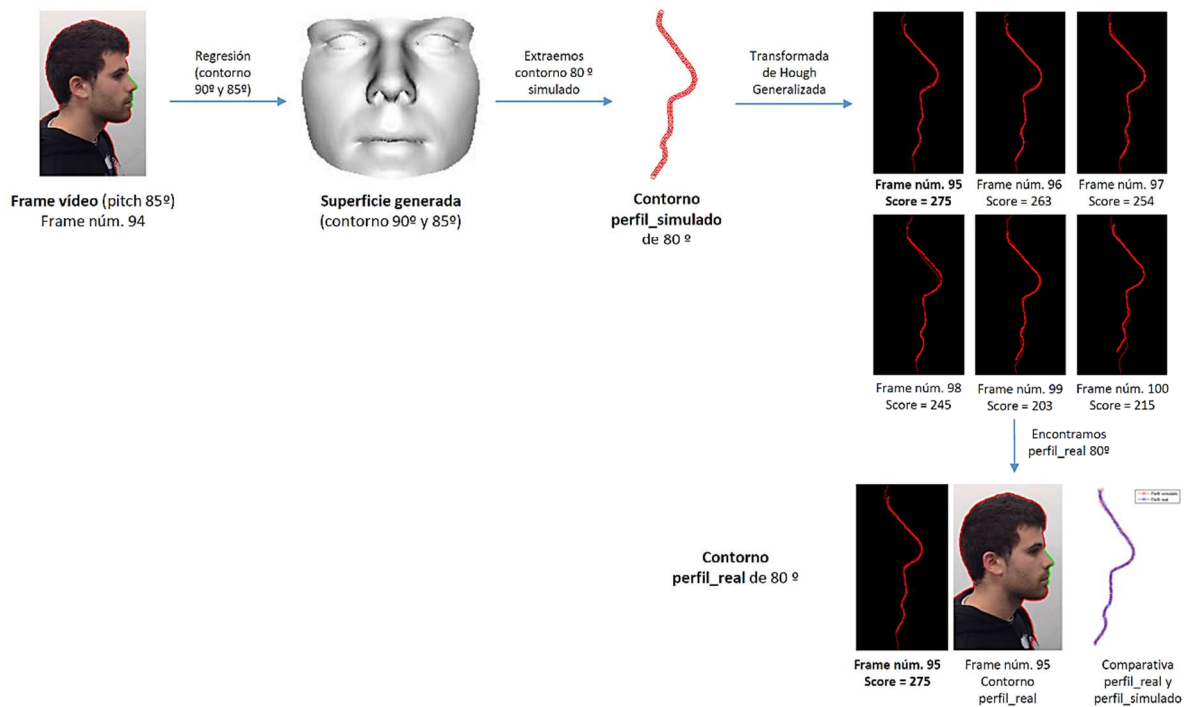


Figura 16. Segunda reconstrucción 3D, entrenamiento con contornos de pitch 90° y 90- α (ej. $\alpha=5^\circ$).

Una vez encontrado el frame de pitch $90-2\cdot\alpha$, se extrae el perfil_real de $90-2\cdot\alpha$, y volvemos a realizar la regresión con los contornos de 90°, $90-\alpha$ y $90-2\cdot\alpha$. Esta vez extraeremos el perfil_simulado de $90-3\cdot\alpha$ que será comparado con un rango de frames del vídeo y hallaremos el perfil_real de $90-3\cdot\alpha$. Con este perfil_real volvemos a realizar la regresión con los contornos de 90°, $90-\alpha$, $90-2\cdot\alpha$, $90-3\cdot\alpha$, y seguiríamos el método.

Por último, para obtener una reconstrucción más real del sujeto bajo estudio, debemos hallar los contornos correspondientes a los ángulos de pitch 0° y 180°. Estos contornos pueden aportar mucha información de la forma o anchura del rostro.

Los contornos de pitch 0° y 180° se extraen y procesan de forma independiente al resto de contornos.

3.3.4. Contornos de pitch 0° y 180°

Al igual que nos pasaba anteriormente, no sabemos qué frame es el que se corresponde con el ángulo pitch 0°/180°, es decir, el que contiene el rostro más frontal del sujeto. Para encontrarlo, se propone analizar los puntos faciales que se obtienen mediante *Chehra Face Tracker* (sistema capaz de rastrear la cara y los ojos de forma completamente automática y en tiempo real).

Siguiendo dicho propósito, analizamos todos los frames del vídeo mediante la detección de los puntos faciales de *Chehra Face Tracker* [Figura 17].

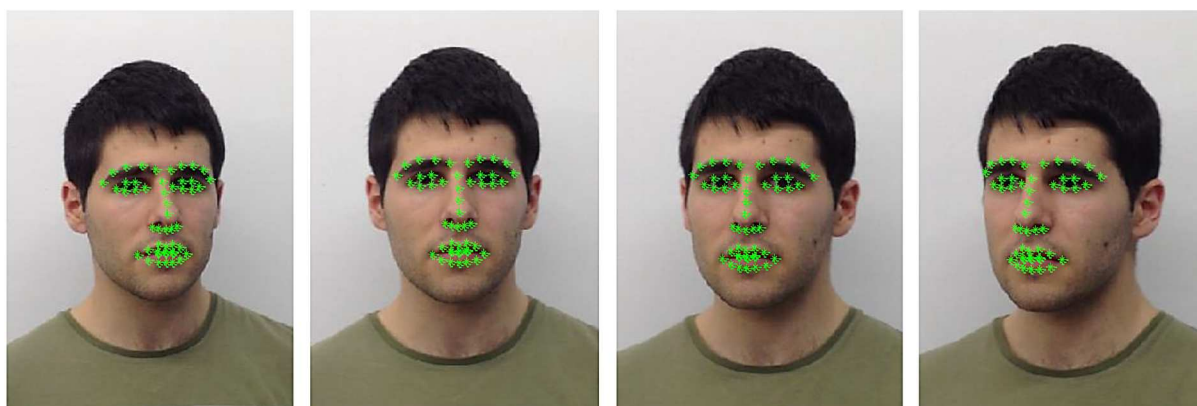


Figura 17. Detección de los puntos faciales mediante *Chehra Face Tracker* en diferentes frames.

Con los puntos obtenidos, podemos encontrar el “frame frontal” a partir del cálculo de distancias simétricas (nariz-ojos) y de la variable medida_fronal [Figura 18].

$$Distancia_1 = |x_{11} - x_{26}|$$

$$Distancia_2 = |x_{23} - x_{11}|$$

$$Medida_fronal = |Distancia_2 - Distancia_1|$$

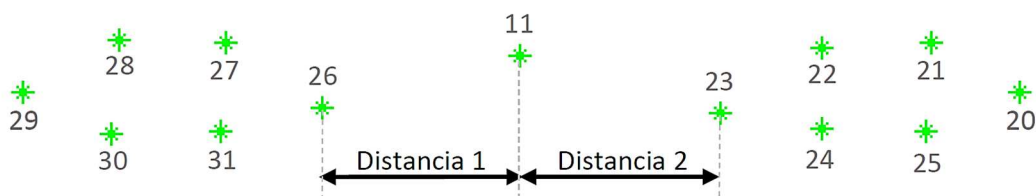


Figura 18. Distancias para la detección del frame con el rostro más frontal del sujeto. Estas distancias se corresponden con la diferencia entre los puntos más internos del ojo (puntos 23 y 26) y el punto más alto de la nariz (punto 11).

El frame que contiene el rostro más frontal es aquel que tenga las distancias nariz-ojos igual, es decir, la variable medida_fronal menor (idealmente 0, ya que los rostros son simétricos).

Una vez detectado el “frame frontal”, podremos extraer los contornos de pitch $0^{\circ}/180^{\circ}$.

Anteriormente habíamos segmentado todos los frames de la secuencia de vídeo, y habíamos obtenido los contornos del sujeto. Si analizamos dichos frames [Figura 19], vemos que los contornos contienen información que no nos interesa (orejas, cabello, etc.). Debemos realizar una segmentación de nivel más bajo para localizar la región de la cara y así extraer el contorno “puro” del rostro.

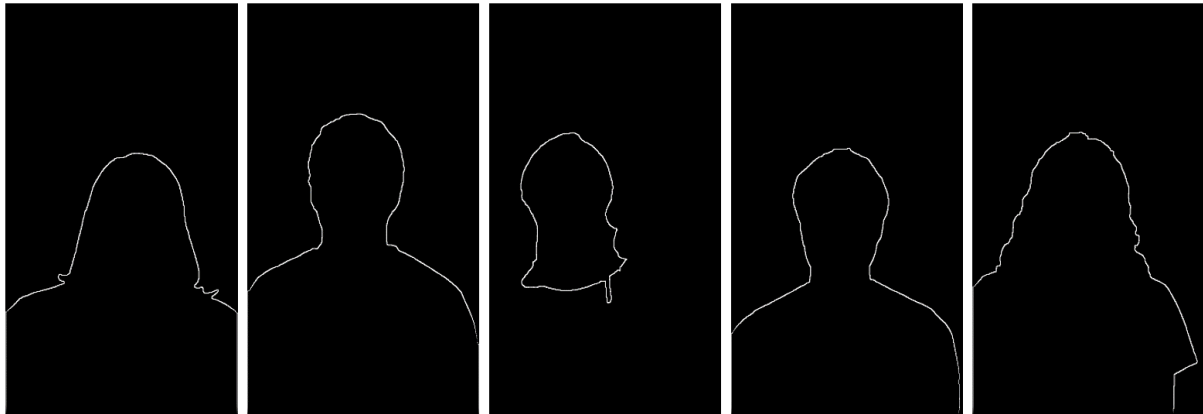


Figura 19. Contornos de diferentes sujetos para pitch $0^{\circ}/180^{\circ}$. Frames frontales segmentados con nivel 0.8 de UCM.

Si valoramos la segmentación del “frame frontal” a diferentes niveles de UCM [Figura 20], vemos que para niveles más bajos obtenemos más regiones que para niveles superiores. En los niveles más bajos, entre las regiones que localizamos, tenemos la de la cara. Tras analizar varios sujetos, concluimos que para extraer el contorno de la cara debemos segmentar el “frame frontal” con nivel 0.2 de UCM o inferior.

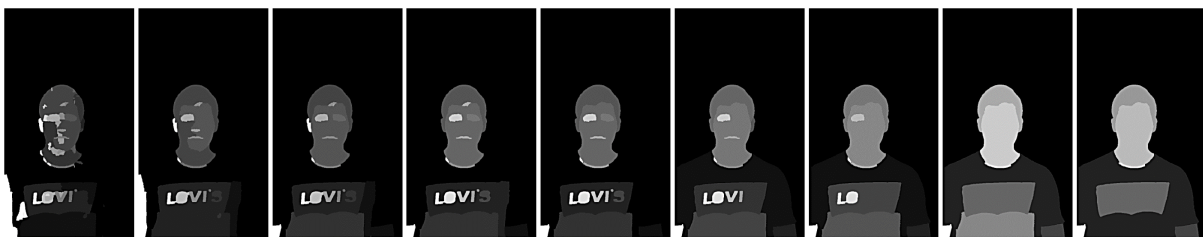


Figura 20. Segmentación UCM para diferentes niveles. En esta figura podemos ver las regiones que se obtienen para diferentes niveles de UCM. La imagen de la izquierda pertenece a un nivel 0.1 de UCM, mientras que la de la derecha es de nivel 0.9; las intermedias se corresponden con saltos de 0.1.

Finalmente, para localizar la región de la cara, nos ayudamos de los puntos faciales obtenidos con *Chehra Face Tracker*. La región de la cara es la unión de todas aquellas regiones que contengan puntos faciales [Figura 21.b y 21.c].

Por último, una vez encontrada la segmentación de la cara, ya podemos extraer los contornos deseados [Figura 21.e y 21.f]

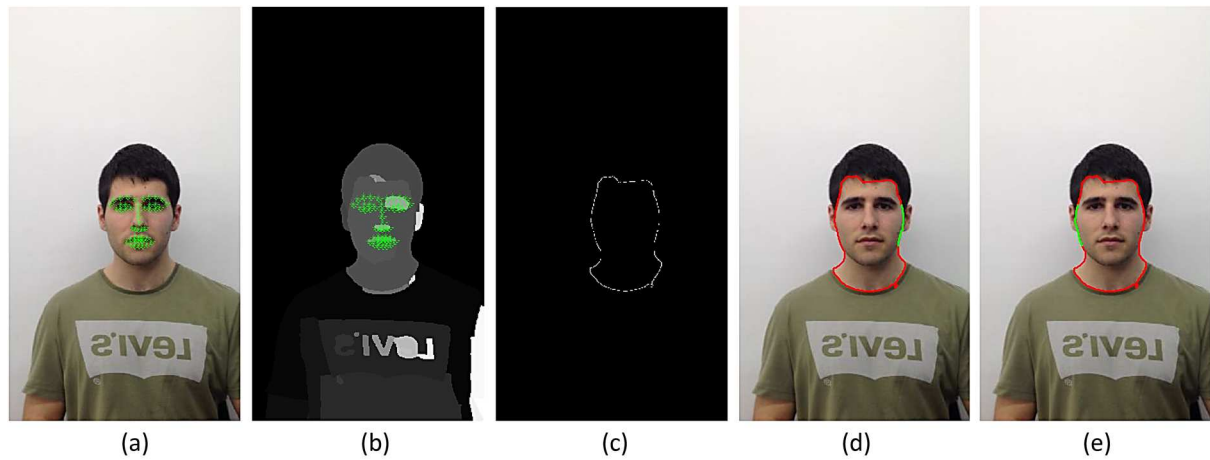


Figura 21. Localización de la región rostro y extracción de los contornos pitch $0^{\circ}/180^{\circ}$. En (a) tenemos el frame frontal con los puntos faciales detectados mediante Chehra Face Tracker. En (b) el frame segmentado a nivel 0.2 de UCM con la superposición de los puntos faciales. En (c) la región de la cara. En (d) y (e) los contornos que extraemos para la reconstrucción.

Tras revisar la metodología de la Reconstrucción de la forma del rostro, a continuación se muestra un ejemplo con los pasos más significativos del proceso. Para este ejemplo tomamos $N=6$ y $\alpha=5^{\circ}$.

3.4. Ejemplo completo

Previamente al proceso de Reconstrucción de la forma del rostro, se procesa la secuencia de vídeo, es decir, se segmentan los frames mediante UCM y se extraen los contornos del sujeto [Figura 22].

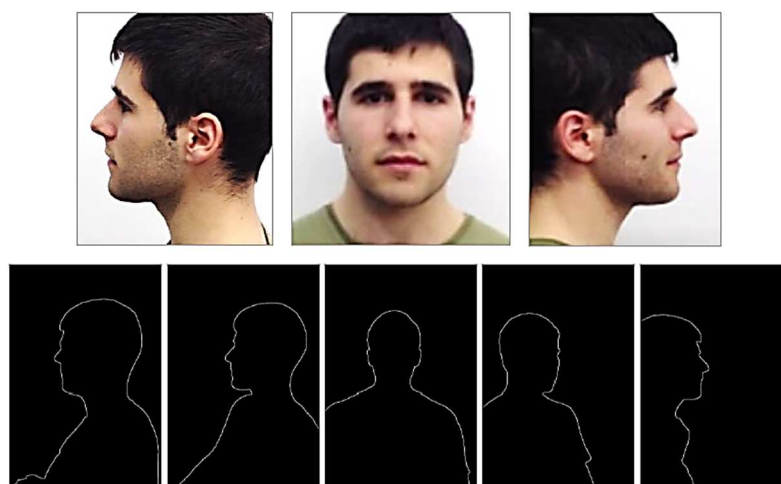


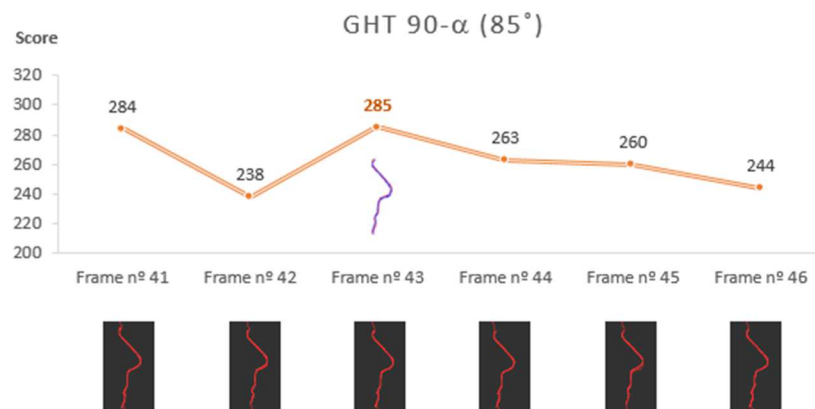
Figura 22. Ejemplo - Muestra de frames segmentados y con los contornos del sujeto extraídos.

Seguidamente, mediante un análisis de los frames, se selecciona el frame de ángulo pitch 90° (para este ejemplo es el frame núm. 40). Con el conocimiento de este frame y tras el procesado del contorno ya nos encontraremos en condiciones para realizar la primera aproximación 3D del rostro [Figura 23].

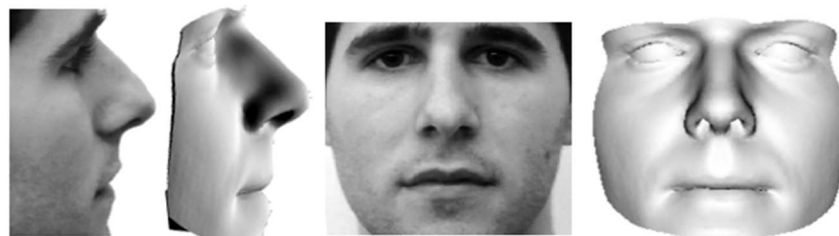


Figura 23. Ejemplo - Primera aproximación 3D del rostro. Previamente, selección del frame pitch 90° y procesado del contorno para que pueda ser introducido en el Regresor (contorno de 80 puntos (x,y) equiespaciados y normalizados según la media del modelo a 90°).

Una vez tenemos la primera aproximación 3D del rostro, podemos empezar el proceso iterativo de la reconstrucción. Comenzamos buscando el frame del vídeo de pitch $90-\alpha$ (para este ejemplo, 85°) mediante la comparación con el perfil_simulado de 85° (método de la GHT) [Figura 24].



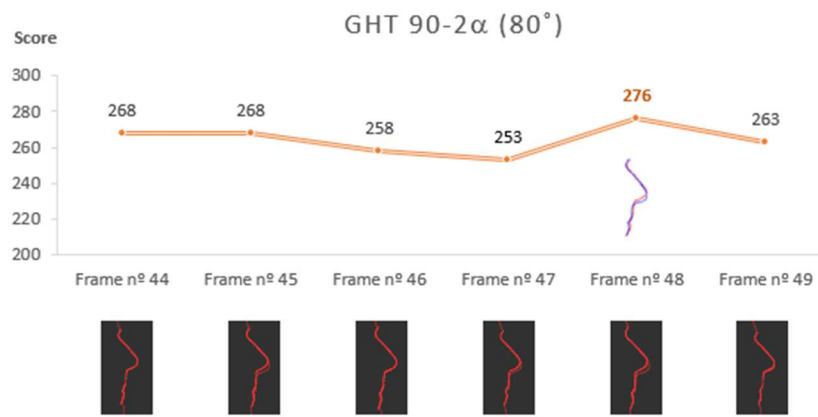
(a) Selección del frame de ángulo pitch $90-\alpha$ mediante la GHT.



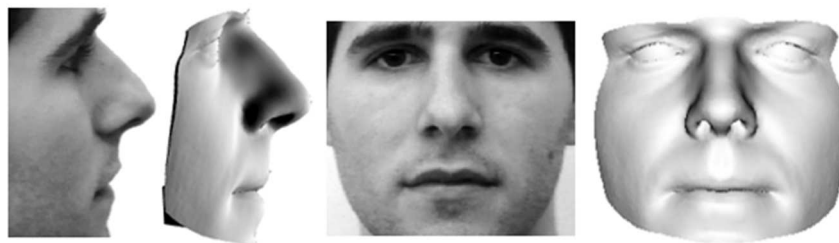
(b) Segunda aproximación 3D de la forma del rostro

Figura 24. Ejemplo - Segunda aproximación 3D del rostro. Previamente, selección del frame pitch $90-\alpha$ (en este caso, frame núm. 43) mediante GHT y reconstrucción 3D con los contornos de 90° y $\pm 85^\circ$.

Continuando con el proceso de reconstrucción obtenemos los siguientes resultados:

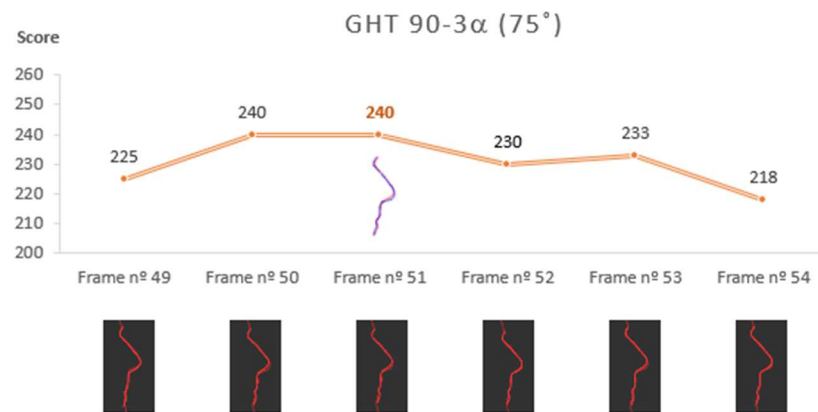


(a) Selección del frame de ángulo pitch $90-2\alpha$ mediante la GHT.

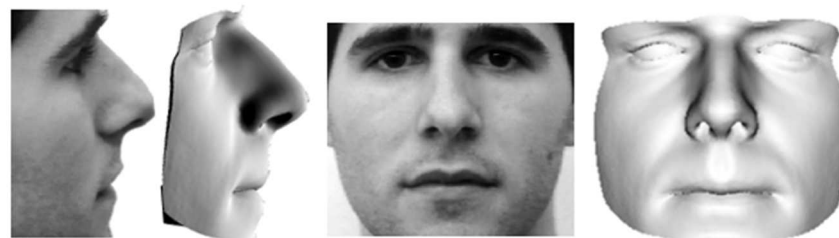


(b) Tercera aproximación 3D de la forma del rostro

Figura 25. Ejemplo - Tercera aproximación 3D del rostro. Previamente, selección del frame pitch $90-2\alpha$ (en este caso, frame núm. 44) mediante GHT y reconstrucción 3D con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$.

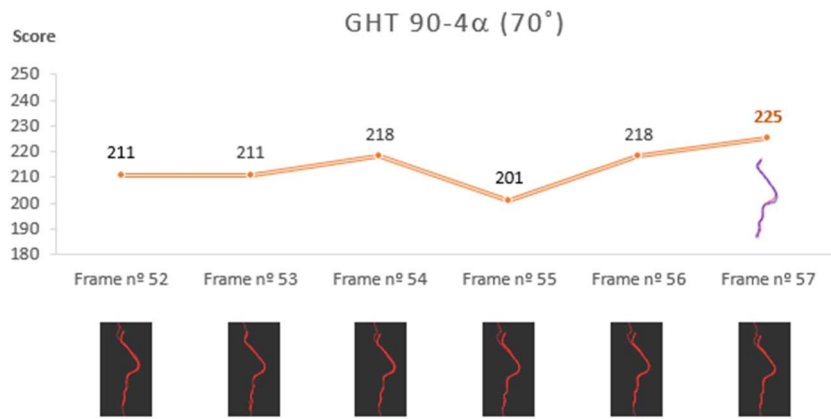


(a) Selección del frame de ángulo pitch $90-3\alpha$ mediante la GHT.

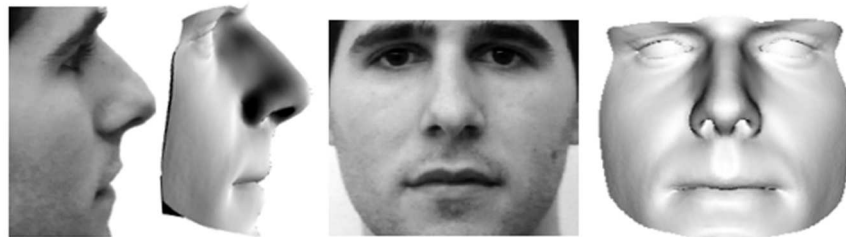


(b) Cuarta aproximación 3D de la forma del rostro

Figura 26. Ejemplo - Cuarta aproximación 3D del rostro. Previamente, selección del frame pitch $90-3\alpha$ (en este caso, frame núm. 45) mediante GHT y reconstrucción 3D con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$.

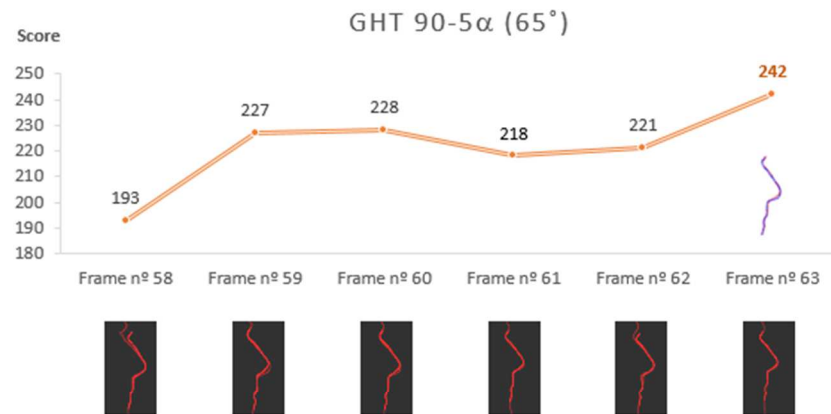


(a) Selección del frame de ángulo pitch 90-4α mediante la GHT.

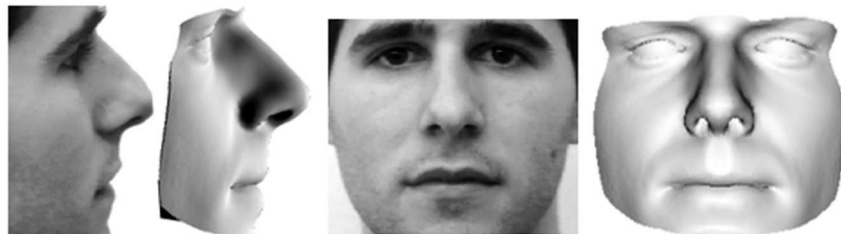


(b) Quinta aproximación 3D de la forma del rostro

Figura 27. Ejemplo - Quinta aproximación 3D del rostro. Previamente, selección del frame pitch 90-4α (en este caso, frame núm. 50) mediante GHT y reconstrucción 3D con los contornos de 90°, ± 85°, ± 80°, ± 75° y ± 70°.

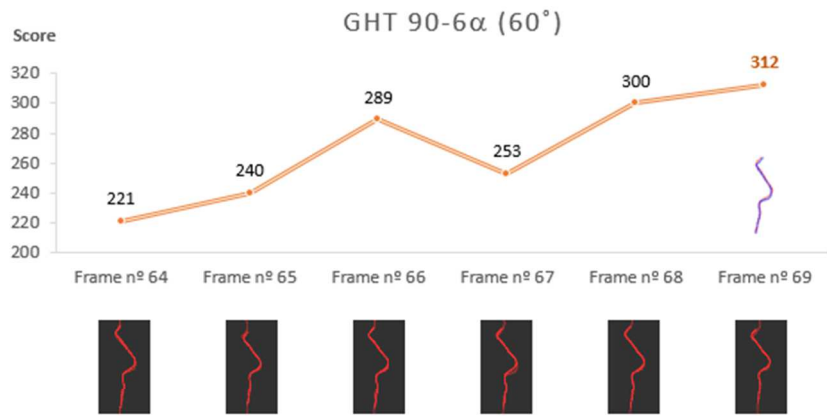


(a) Selección del frame de ángulo pitch 90-5α mediante la GHT.

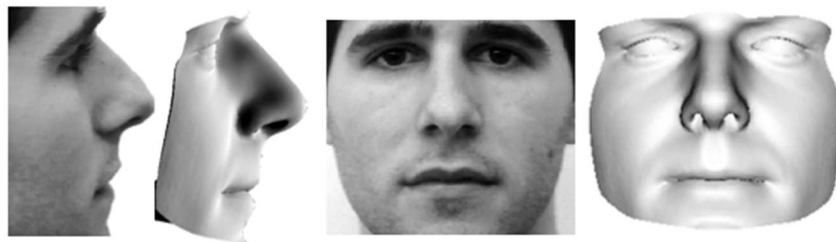


(b) Sexta aproximación 3D de la forma del rostro

Figura 28. Ejemplo - Sexta aproximación 3D del rostro. Previamente, selección del frame pitch 90-5α (en este caso, frame núm. 51) mediante GHT y reconstrucción 3D con los contornos de 90°, ± 85°, ± 80°, ± 75°, ± 70° y ± 65°.

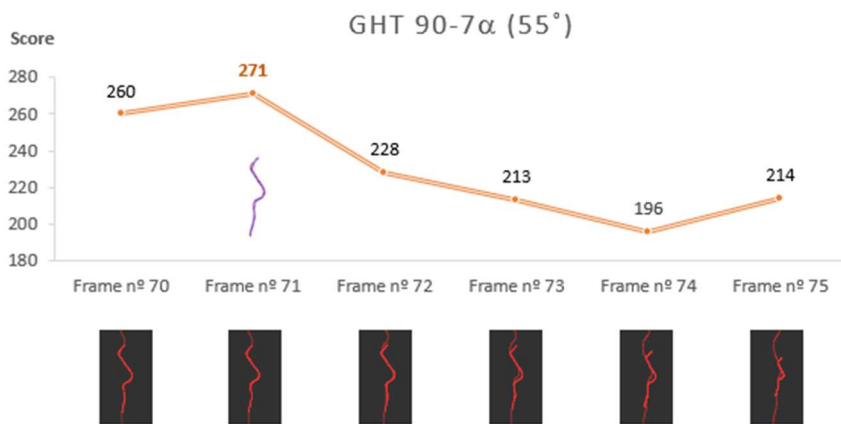


(a) Selección del frame de ángulo pitch 90-6α mediante la GHT.



(b) Séptima aproximación 3D de la forma del rostro

Figura 29. Ejemplo - Séptima aproximación 3D del rostro. Previamente, selección del frame pitch 90-6α (en este caso, frame núm. 57) mediante GHT y reconstrucción 3D con los contornos de 90°, ± 85°, ± 80°, ± 75°, ± 70°, ± 65° y ± 60°.



(a) Selección del frame de ángulo pitch 90-7α mediante la GHT.



(b) Octava aproximación 3D de la forma del rostro

Figura 30. Ejemplo - Octava aproximación 3D del rostro. Previamente, selección del frame pitch 90-7α (en este caso, frame núm. 62) mediante GHT y reconstrucción 3D con los contornos de 90°, ± 85°, ± 80°, ± 75°, ± 70°, ± 65°, ± 60° y ± 55°.

Finalmente nos queda encontrar los contornos frontales para conseguir la Reconstrucción final del rostro. Mediante el detector de puntos faciales Chehra Face Tracker hallamos el frame que contiene el rostro más frontal (en este caso es el núm. 108) y procesamos los contornos [Figura 21]. Una vez introducimos los contornos de pitch $0^{\circ}/180^{\circ}$ en el Regresor, obtenemos la Reconstrucción final del rostro [Figura 31].

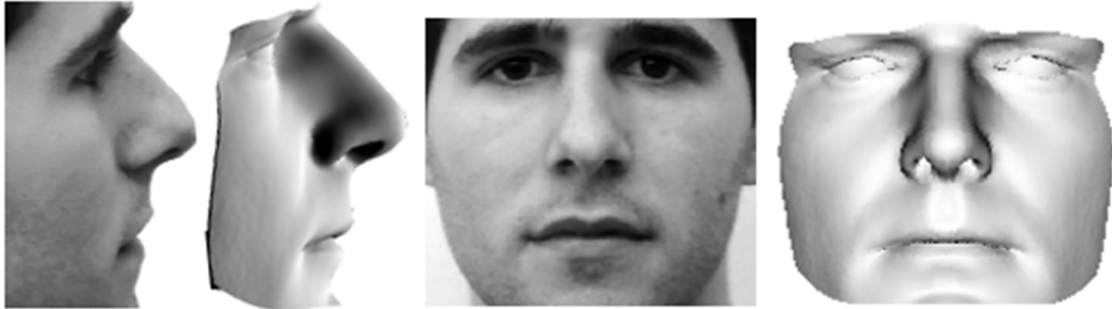


Figura 31. Ejemplo – Reconstrucción final de la forma del rostro. Reconstrucción 3D con los contornos $90^{\circ}, \pm 85^{\circ}, \pm 80^{\circ}, \pm 75^{\circ}, \pm 70^{\circ}, \pm 65^{\circ}, \pm 60^{\circ}, \pm 55^{\circ}, 0^{\circ}$ y 180° .

Capítulo 4. Resultados

4. 1. Base de Datos

Para esta investigación disponíamos de 20 secuencias de vídeo de diferentes sujetos en movimiento. Esta base de datos ha sido tomada mediante una cámara convencional, no calibrada y sencilla como puede ser la de un Smartphone.

El inventario de los vídeos con los detalles principales se detalla a continuación:

Vídeo	Núm. de frames	Tamaño frames	Velocidad (frames/s)
IMG_6734.MOV	173	RGB24 1920×1080	29.97
IMG_6737.MOV	296	RGB24 1920×1080	29.97
IMG_6738.MOV	461	RGB24 1920×1080	29.97
IMG_6739.MOV	321	RGB24 1920×1080	29.97
IMG_6740.MOV	298	RGB24 1920×1080	29.97
IMG_6741.MOV	282	RGB24 1920×1080	29.97
IMG_6742.MOV	308	RGB24 1920×1080	29.97
IMG_6743.MOV	427	RGB24 1920×1080	29.97
IMG_6744.MOV	477	RGB24 1920×1080	29.97
IMG_6745.MOV	410	RGB24 1920×1080	29.97
IMG_6746.MOV	562	RGB24 1920×1080	29.97
IMG_6747.MOV	411	RGB24 1920×1080	29.97
IMG_6749.MOV	351	RGB24 1920×1080	29.97
IMG_6750.MOV	362	RGB24 1920×1080	29.97
IMG_6752.MOV	390	RGB24 1920×1080	29.97
IMG_6753.MOV	387	RGB24 1920×1080	29.97
IMG_6756.MOV	241	RGB24 1920×1080	29.97
IMG_6757.MOV	357	RGB24 1920×1080	29.97
IMG_6758.MOV	328	RGB24 1920×1080	29.97

Tabla 2. Base de datos. Inventario de secuencias de vídeo disponibles para el proyecto.

De dicha base de datos hemos descartado inicialmente 7 vídeos puesto que no tenían las características indicadas para la reconstrucción 3D (los sujetos llevaban gafas, tenían una barba muy pronunciada, etc.).

4. 2. Proceso de evaluación

Evaluar el modelo de reconstrucción facial 3D a partir de imágenes 2D, utilizando los resultados obtenidos, no da una medida certera de la calidad de la modelización facial. Para una mejor evaluación es necesario comparar la reconstrucción obtenida con el modelo facial del sujeto adquirido a través de un escáner 3D. Para realizar dicha comparación se utiliza la distancia euclidiana entre los vértices del modelo 3D reconstruido y sus correspondientes en el escaneado (Euclidean Based Distance, EBD), de manera que el error puede ser calculado como la media aritmética de estas distancias.

Puesto que no disponemos de un escáner 3D, únicamente podemos hacer una evaluación cualitativa de los resultados.

Para cada sujeto real reconstruido se van a evaluar dos estrategias distintas:

- i. La reconstrucción obtenida con los contornos de ángulos pitch $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ$ y $\pm 55^\circ$, frente al sujeto real.
- ii. La reconstrucción obtenida con los contornos de ángulos pitch $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ, \pm 55^\circ, 0^\circ$ y 180° , frente al sujeto real.

Adicionalmente se van a comparar los contornos 2D extraídos del vídeo (perfiles reales) y los obtenidos en las reconstrucciones (perfiles simulados) mediante la distancia euclidiana:

$$d_j = |\text{perfil}_{\text{simulado}} - \text{perfil}_{\text{real}}| = \sqrt{\sum_{i=1}^{80} |x_i - y_i|^2}$$

4. 3. Resultados

A continuación se presentan los resultados obtenidos para tres sujetos diferentes mediante la comparativa de los contornos 2D, la evolución de las reconstrucciones con los diferentes contornos y la reconstrucción final con las dos estrategias propuestas.

Sujeto 1

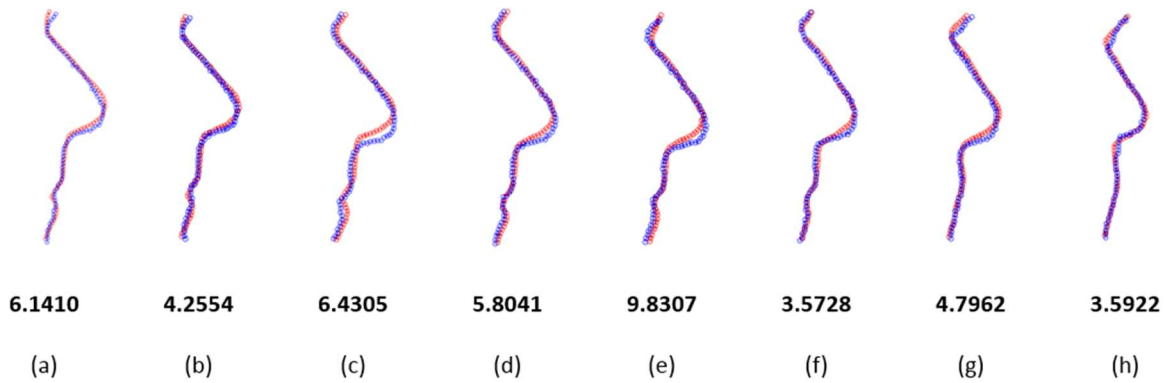


Figura 32. Resultados Sujeto 1 - Comparativa de perfiles. Comparativa de los contornos reales (en azul) con los simulados (en rojo) mediante la distancia euclidiana. En (a) la comparativa con los contornos de pitch 90° . En (b) contornos de pitch 85° . En (c) contornos de pitch 80° . En (d) contornos de pitch 75° . En (e) contornos de pitch 70° . En (f) contornos de pitch 65° . En (g) contornos de pitch 60° . En (h) contornos de pitch 55° .

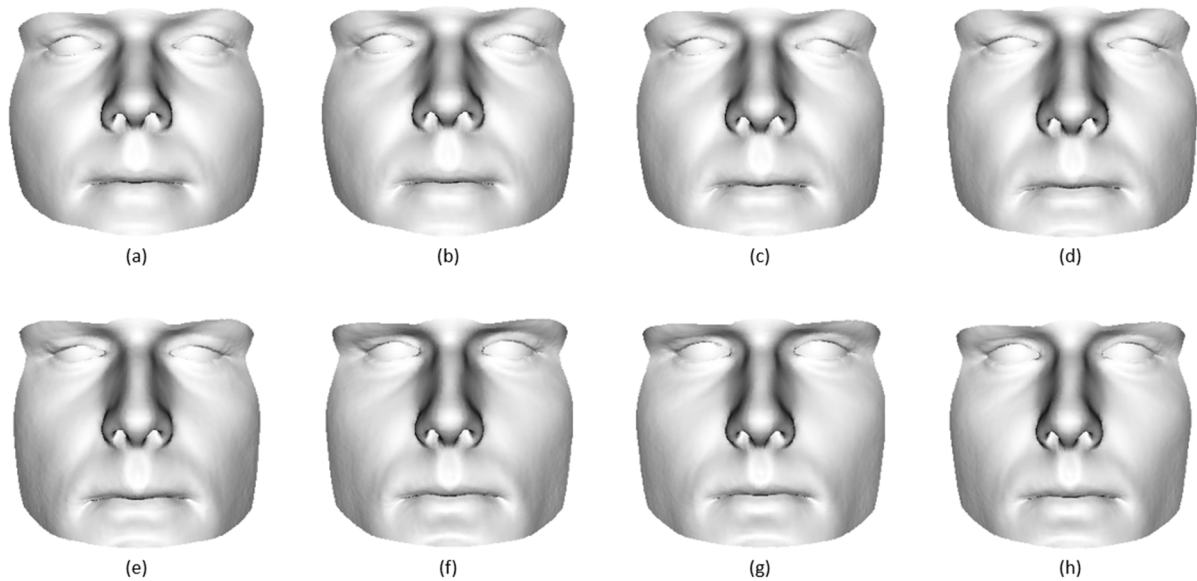


Figura 33. Resultados Sujeto 1 - Reconstrucciones de la forma del rostro. En (a) la reconstrucción con el contorno de 90° . En (b) la reconstrucción con los contornos de 90° y $\pm 85^\circ$. En (c) con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$. En (d) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$. En (e) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 70^\circ$. En (f) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$ y $\pm 65^\circ$. En (g) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$. Y en (h) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$ y $\pm 55^\circ$.

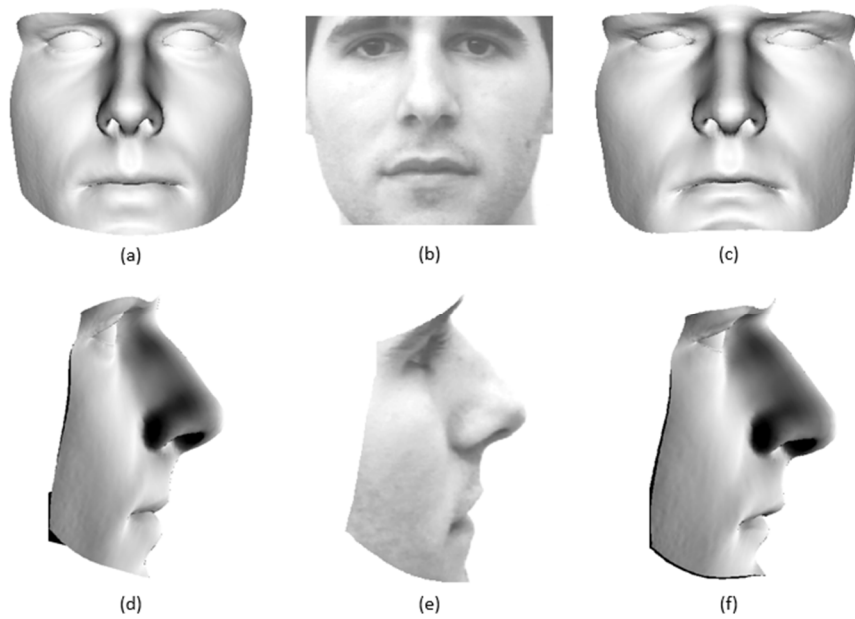


Figura 34. Resultados Sujeto 1 - Reconstrucción final de la forma del rostro. En (a) y (d) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$ y $\pm 55^\circ$. En (b) y (e) el sujeto real. En (c) y (f) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$, $\pm 55^\circ$, 0° y 180° .

Sujeto 2

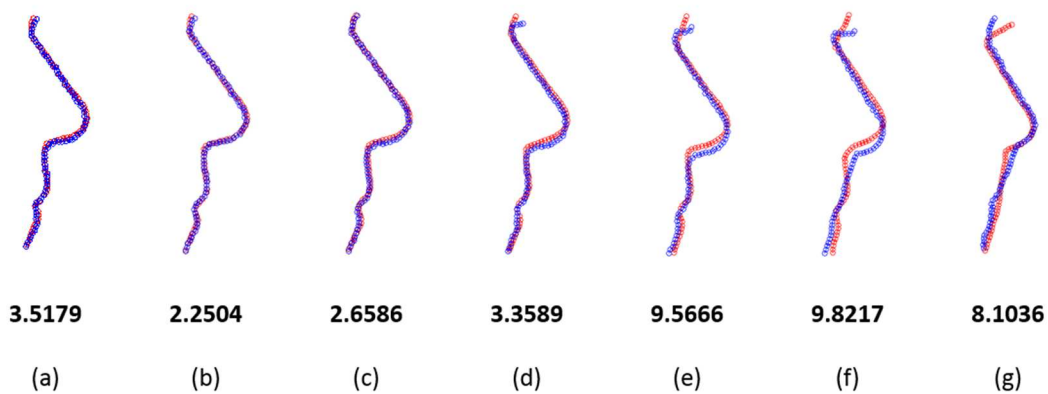


Figura 35. Resultados Sujeto 2 - Comparativa de perfiles. Comparativa de los contornos reales (en azul) con los simulados (en rojo) mediante la distancia euclidiana. En (a) la comparativa con los contornos de pitch 90° . En (b) contornos de pitch 85° . En (c) contornos de pitch 80° . En (d) contornos de pitch 75° . En (e) contornos de pitch 70° . En (f) contornos de pitch 65° . En (g) contornos de pitch 60° .

Para el sujeto número 2, podemos observar que los tres últimos contornos no se ajustan correctamente [Figura 35.e-g]. El modelo no consigue ajustar los datos y además parece haber un error de estimación del ángulo pitch.

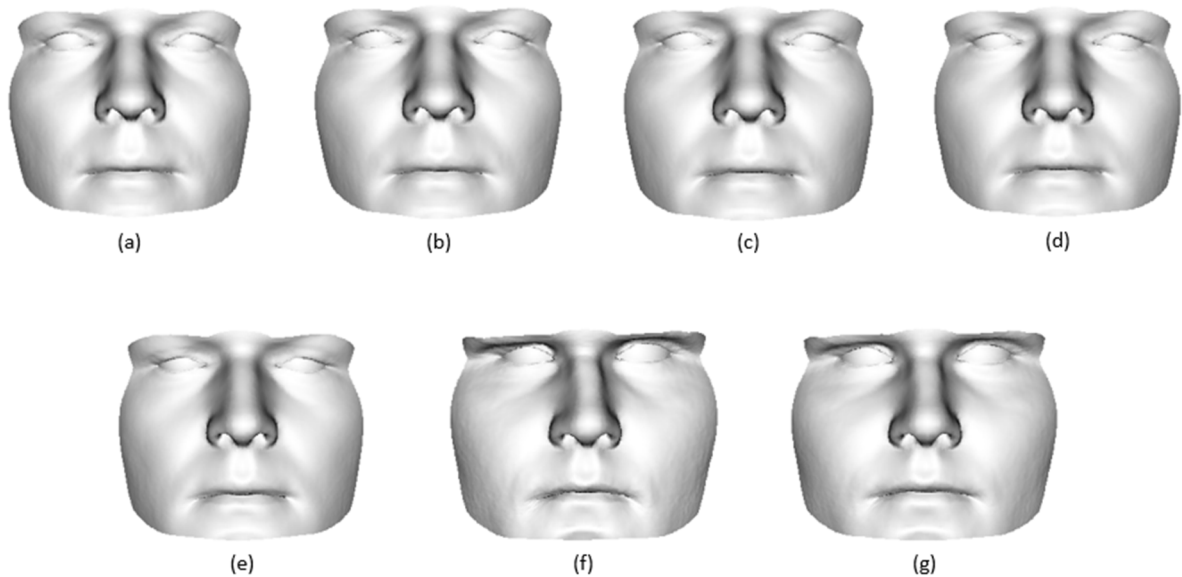


Figura 36. Resultados Sujeto 2 - Reconstrucciones de la forma del rostro. En (a) la reconstrucción con el contorno de 90° . En (b) la reconstrucción con los contornos de 90° y $\pm 85^\circ$. En (c) con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$. En (d) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$. En (e) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 70^\circ$. En (f) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$ y $\pm 65^\circ$. En (g) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$.

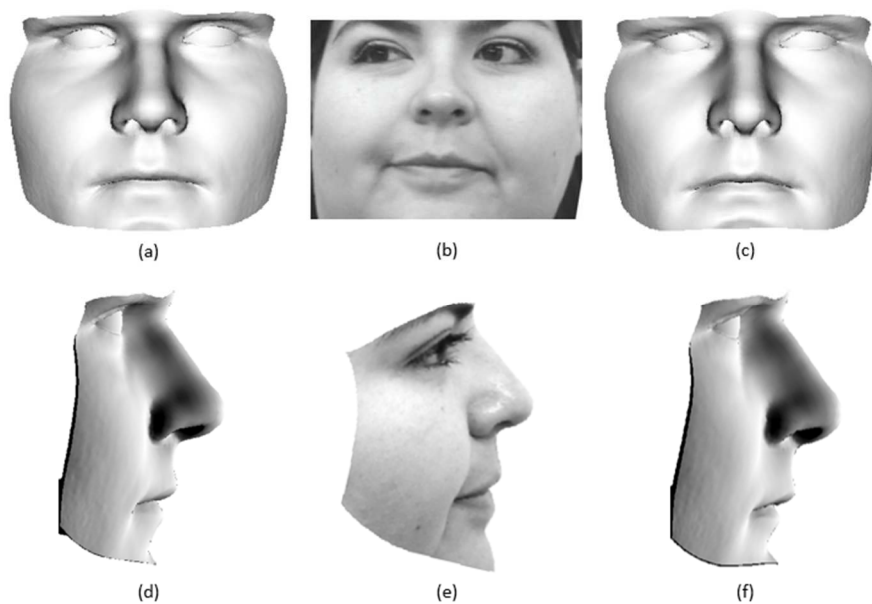


Figura 37. Resultados Sujeto 2 - Reconstrucción final de la forma del rostro. En (a) y (d) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$. En (b) y (e) el sujeto real. En (c) y (f) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$, 0° y 180° .

Sujeto 3

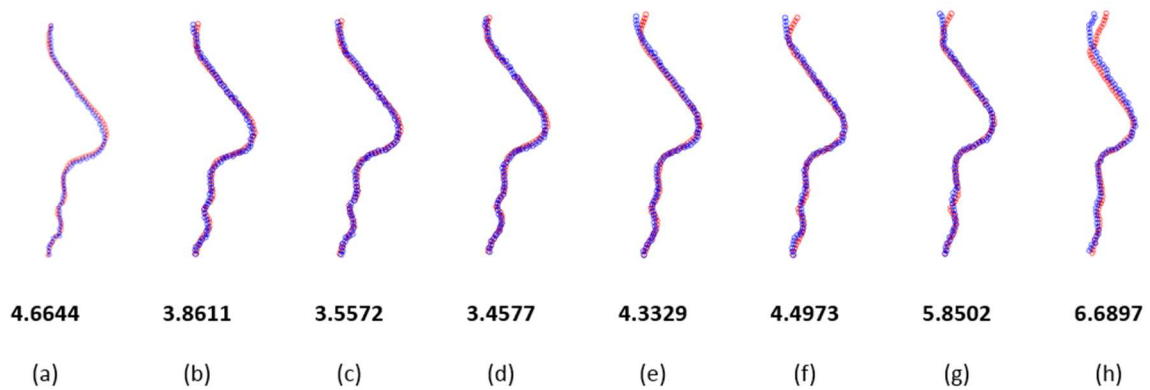


Figura 38. Resultados Sujeto 3 - Comparativa de perfiles. Comparativa de los contornos reales (en azul) con los simulados (en rojo) mediante la distancia euclidiana. En (a) la comparativa con los contornos de pitch 90° . En (b) contornos de pitch 85° . En (c) contornos de pitch 80° . En (d) contornos de pitch 75° . En (e) contornos de pitch 70° . En (f) contornos de pitch 65° . En (g) contornos de pitch 60° . En (h) contornos de pitch 55° .

Para el sujeto número 3, podemos observar que los cuatro últimos contornos no se ajustan correctamente [Figura 38.e-h]. El modelo no consigue ajustar los datos superiores, sin embargo para el resto del contorno parece hacerlo bien.

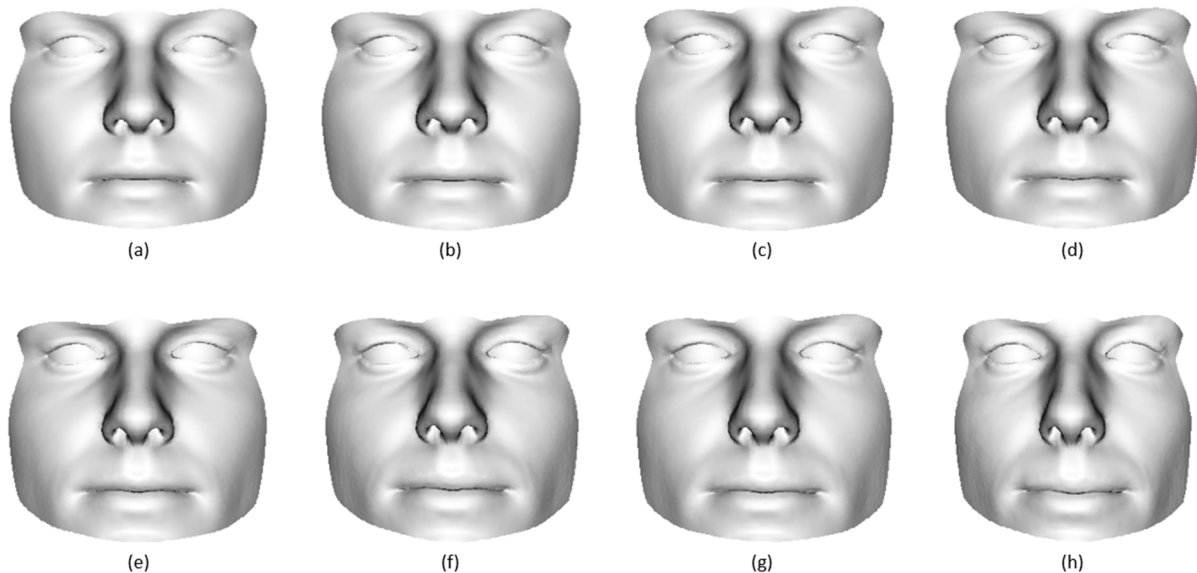


Figura 39. Resultados Sujeto 3 - Reconstrucciones de la forma del rostro. En (a) la reconstrucción con el contorno de 90° . En (b) la reconstrucción con los contornos de 90° y $\pm 85^\circ$. En (c) con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$. En (d) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$. En (e) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 70^\circ$. En (f) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$ y $\pm 65^\circ$. En (g) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$. Y en (h) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$ y $\pm 55^\circ$.

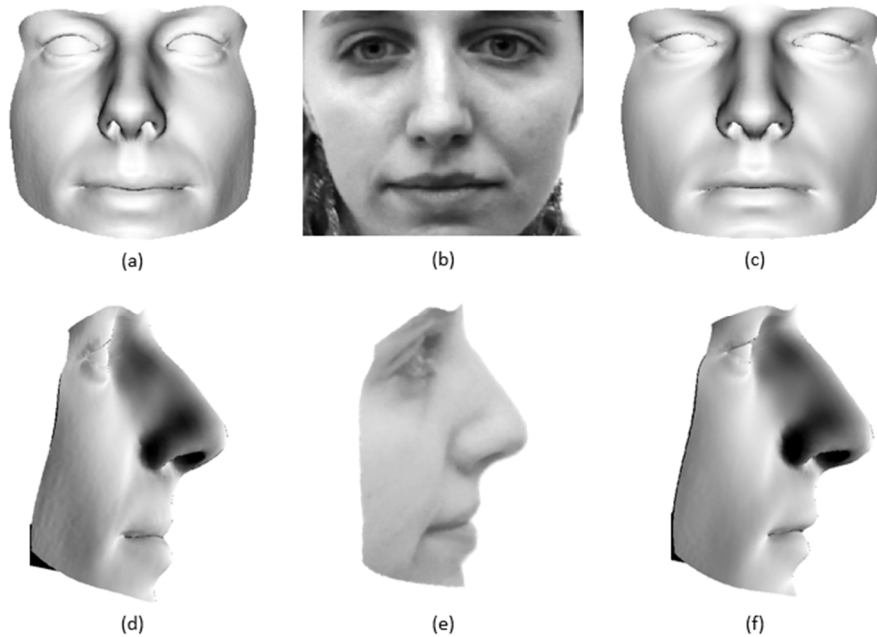


Figura 40. Resultados Sujeto 3 - Reconstrucción final de la forma del rostro. En (a) y (d) la reconstrucción con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ$ y $\pm 55^\circ$. En (b) y (e) el sujeto real. En (c) y (f) la reconstrucción con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ, \pm 55^\circ, 0^\circ$ y 180° .

En el Anexo I se muestran los resultados de las reconstrucciones obtenidas para otros sujetos.

Tras valorar las distintas reconstrucciones, podemos ver que aunque los contornos 2D extraídos del vídeo sean prácticamente iguales a los reconstruidos, el rostro global sigue sin parecerse lo suficiente al sujeto real [Figuras 34, 37 y 40].

Observando la comparativa de los perfiles reales y los extraídos de la reconstrucción [Figuras 32, 35 y 38], podemos ver que el modelo estadístico es capaz de caracterizar los datos pero que la zona de la nariz no la hace tan bien como se esperaba.

Valorando las dos estrategias propuestas:

- i. La reconstrucción obtenida con los contornos de ángulos pitch $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ$ y $\pm 55^\circ$, frente al sujeto real.
- ii. La reconstrucción obtenida con los contornos de ángulos pitch $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ, \pm 55^\circ, 0^\circ$ y 180° , frente al sujeto real.

Vemos que obtenemos mejores resultados en la primera estrategia (es decir, más parecidos con el sujeto real). Esto es debido a que la extracción de los contornos principales no es trivial

ya que es difícil obtener la región del rostro y por consiguiente, los contornos en los límites adecuados.

Los contornos 2D según el ángulo Pitch, únicamente provocan restricciones verticales sobre el rostro. Si en futuras investigaciones se añadiesen al estudio los contornos que varían según el ángulo Roll, el rostro estaría mejor restringido y tendría mejor definidas las características del sujeto (como por ejemplo, anchura de la nariz, pómulos, barbilla, etc).

Después de valorar visualmente los resultados, podemos concluir que las reconstrucciones de la forma del rostro que hemos obtenido no son tan buenas o ajustadas a la forma real como se esperaban.

4. 4. Limitaciones

Tras valorar los resultados obtenidos, encontramos las siguientes limitaciones de nuestro sistema:

- i. El método que hemos presentado es muy sensible a la selección de los puntos inicio (x_1) y fin (x_2) del contorno útil. Dichos puntos se seleccionan manualmente al iniciar la reconstrucción con el frame de ángulo pitch 90° [Figura 41]. Si estos puntos no se eligen correctamente, el resto del proceso será erróneo puesto que la primera aproximación de la forma del rostro no será coherente con el sujeto real.



Figura 41. Limitaciones del sistema. Corrección de la selección de los puntos x_1 y x_2 del contorno útil.

Por otro lado, que la selección de los puntos x_1 y x_2 sea manual, implica que los resultados obtenidos en diferentes pruebas siempre sean diferentes (seleccionar dos veces el mismo punto es muy complicado).

- ii. La Reconstrucción de la forma del rostro tiene un tiempo de procesado bastante elevado por lo que no es útil para aplicaciones en tiempo real.
Tras evaluar todas las funciones, vemos que el paso que requiere más tiempo de procesado es el que realiza la transformada de Hough generalizada (este paso se realiza para encontrar el frame que contiene el contorno de ángulo pitch $90-k\cdot\alpha$).
- iii. El proceso de detección del frame para los distintos ángulos pitch (mediante la transformada de Hough generalizada), está limitado. Solo se puede realizar para los ángulos pitch $\in [55^\circ, 90^\circ]$. Si consideramos ángulos pitch inferiores, el sistema no es capaz de encontrar el frame correctamente.
- iv. El modelo estadístico que utilizamos requiere que la pose de la cabeza sea estimada. La Base de Datos que tenemos ha sido grabada mediante una cámara convencional y sencilla como puede ser la de un Smartphone, por ese motivo no sabemos el ángulo de rotación que ha sufrido el sujeto en cada frame.
Si tuviésemos una Base de Datos tomada con una cámara calibrada, como por ejemplo un sistema *opti-track*, sabríamos con gran certeza la posición de la cabeza del sujeto en cada frame. Con esto se eliminaría el ruido de pose.
- v. El código de Matlab que se ha realizado en este proyecto es compatibles con todos los Sistemas Operativos, excepto las siguientes funciones:
 - La función que realiza la Segmentación UCM únicamente es compatible con Linux/Ubuntu.
 - La función que obtiene los puntos de referencia faciales del Chehra Face Tracker únicamente es compatible con Windows.

Capítulo 5. Conclusiones

Tras finalizar la investigación y después de valorar visualmente los resultados, podemos concluir que las reconstrucciones de la forma del rostro que hemos obtenido no son tan buenas o ajustadas a la forma real como se esperaba.

La evaluación de las reconstrucciones con algún parámetro cuantitativo ha sido muy complicada, por ello únicamente se ha podido hacer una valoración de forma cualitativa. En cambio, en la comparativa de los contornos 2D si se ha podido establecer un parámetro cuantitativo como es la distancia euclidiana.

Creemos que la metodología de trabajo que se ha utilizado es coherente con los objetivos marcados, puesto que al analizar los pasos individualmente obtenemos buenos resultados.

Es posible que el modelo necesite más información de la cara (como la anchura de la barbilla, de la nariz, etc.) para ajustar correctamente los rostros. Estos datos se podrían obtener mediante contornos según variaciones del ángulo Roll. Por ello, un enfoque interesante como trabajo futuro podría ser la estimación del rostro con contornos 2D obtenidos con variaciones de los ángulos Roll y Pitch.

A pesar de las limitaciones que nos hemos encontrado en la investigación, creemos que el modelo estadístico con el que hemos trabajado tiene mucho potencial, y que con otra base de datos se podrían obtener mejores resultados. Por ello, como otro posible trabajo futuro, se propone probar el modelo estadístico con una base de datos tomada con un sistema tipo *opti-track*, puesto que se sabría con gran certeza la pose de la cabeza del sujeto en cada frame y se eliminaría una gran parte del proceso (transformada de Hough, Chehra face tracker, etc).

Con una base de datos así, el proceso requeriría un tiempo de procesado mucho menor pero, en contraposición, perdería la simplicidad que buscaba este proyecto ya que dejaríamos de analizar un vídeo sencillo (grabado con una cámara cualquiera) para analizar un vídeo obtenido con sistemas más complejos, que no están al alcance de todos.

Capítulo 6. Referencias

Bibliografía

- [1] *J. J. Atick, P. A. Griffin, and A. N. Redlich*. "Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images". *Neural Computation*, 8(6):1321-1340, 1996.
- [2] *V. Blanz and T. Vetter*. "A morphable model for the synthesis of 3d faces". In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187-194. ACM Press/Addison-Wesley Publishing Co., 1999.
- [3] *J. P. Mena-Chalco, L. Velho, and R. M. Cesar Junior*. "3d human face reconstruction using principal components spaces". In *WTD SIBGRAPI-Conference on Graphics, Patterns and Images*, pages 1-6, 2011.
- [4] *Y. Ming, Q. Ruan, and X. Li*. "3d face reconstruction using a single 2d face image". In *Educational and Information Technology (ICEIT), 2010 International Conference on*, volume 3, pages V3-32. IEEE, 2010.
- [5] *M. Reiter, R. Donner, G. Langs, and H. Bischof*. "3d and infrared face reconstruction from rgb data using canonical correlation analysis". In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 425-428. IEEE, 2006.
- [6] *M. Castelán and J. Van Horebeek*. "3d face shape approximation from intensities using partial least squares". In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1-8. IEEE, 2008.
- [7] *D. Sánchez-Escobedo and M. Castelán*. "3d face shape prediction from a frontal image using cylindrical coordinates and partial least squares". *Pattern Recognition Letters*, 34(4):389-399, 2013.
- [8] *B. Moghaddam, J. Lee, H. Pfister, and R. Machiraju*. "Model-based 3d face capture with shape-from silhouettes". In *Analysis and Modelling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, pages 20-27. IEEE, 2003.
- [9] *S. Romdhani and T. Vetter*. "Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior". In *Computer Vision and Pattern*

Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 986-993. IEEE, 2005.

[10] *M. Keller, R. Knothe, and T. Vetter.* "3d reconstruction of human faces from occluding contours". In Computer Vision/Computer Graphics Collaboration Techniques, pages 261-273. Springer, 2007.

[11] *H. Gupta, A. K. Roy Chowdhury, and R. Chellappa.* "Contour-based 3d face modelling from a monocular video". In BMVC, pages 1-10, 2004.

[12] *Y. Lin, G. Medioni, and J. Choi.* "Accurate 3d face reconstruction from weakly calibrated wide baseline images with profile contours". In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 1490-1497. IEEE, 2010.

[13] *T. Albrecht, R. Knothe, and T. Vetter.* "Modelling the remaining flexibility of partially fixed statistical shape models". In 2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy, pages 160-169, 2008.

[14] *M. Lüthi, T. Albrecht, and T. Vetter.* "Probabilistic modelling and visualization of the flexibility in morphable models". In 13th IMA International Conference, pages 251-264, 2009.

[15] *D. Sánchez-Escobedo, M. Castelán and W. A. P. Smith.* "Statistical 3D Face Shape Estimation from Occluding Contours". 2014

[16] *S. Rännar, F. Lindgren, P. Geladi, and S. Wold.* "A pls kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm". Journal of Chemometrics, 8(2):111-125, 1994.

[17] *P. Arbelaez, M. Maire, C. Fowlkes and J. Malik.* "Contour Detection and Hierarchical Image Segmentation". IEEE TPAMI, Vol. 33, No. 5, pp. 898-916, May 2011.

[18] *D. H. Ballard,* "Generalizing the Hough transform to detect arbitrary shapes", Pattern Recognition, 111-122 vol. 13(2) (1981).

[19] *A. Asthana, S. Zafeiriou, S. Cheng and M. Pantic,* "Incremental Face Alignment in the Wild". In CVPR 2014.

Adicionalmente se han consultado otras fuentes como: www.wikipedia.org

Anexo I. Reconstrucciones.

En este anexo se muestran las Reconstrucciones de la forma del rostro para diferentes sujetos.

Sujeto 4

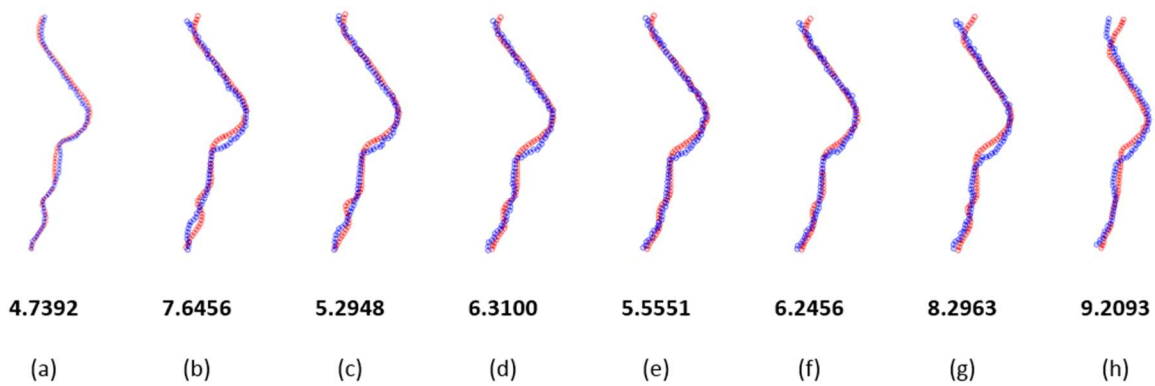


Figura 42. Resultados Sujeto 4 - Comparativa de perfiles. Comparativa de los contornos reales (en azul) con los simulados (en rojo) mediante la distancia euclidiana. En (a) la comparativa con los contornos de pitch 90° . En (b) contornos de pitch 85° . En (c) contornos de pitch 80° . En (d) contornos de pitch 75° . En (e) contornos de pitch 70° . En (f) contornos de pitch 65° . En (g) contornos de pitch 60° . En (h) contornos de pitch 55° .

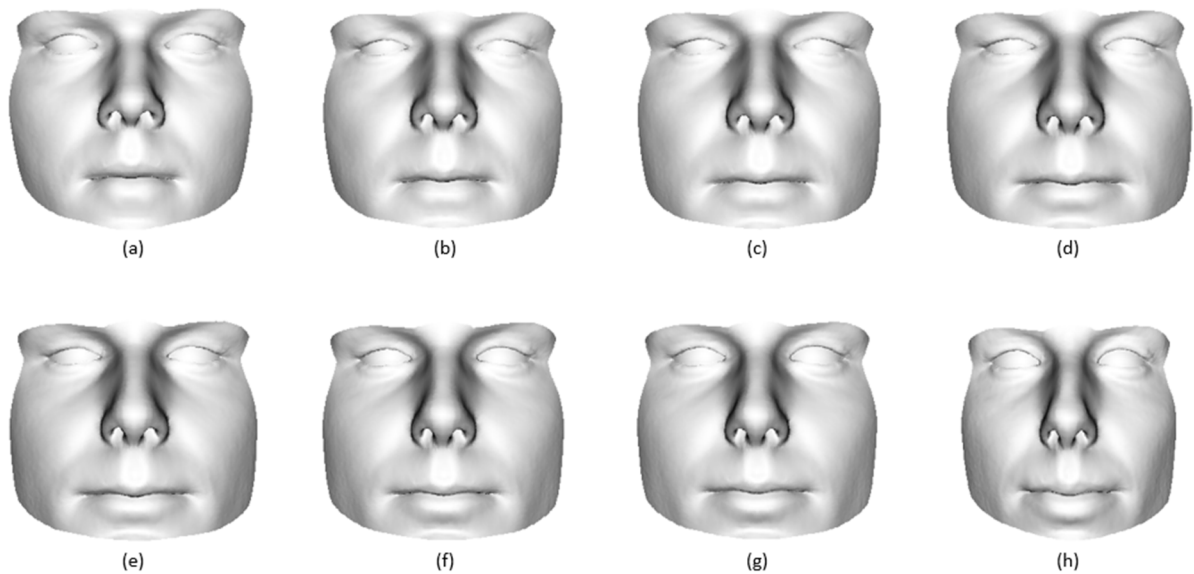


Figura 43. Resultados Sujeto 4 - Reconstrucciones de la forma del rostro. En (a) la reconstrucción con el contorno de 90° . En (b) la reconstrucción con los contornos de 90° y $\pm 85^\circ$. En (c) con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$. En (d) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$. En (e) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 70^\circ$. En (f) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 65^\circ$. En (g) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 60^\circ$. En (h) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 55^\circ$.

$\pm 70^\circ$. En (f) con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ$ y $\pm 65^\circ$. En (g) con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ$ y $\pm 60^\circ$. Y en (h) con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ$ y $\pm 55^\circ$.

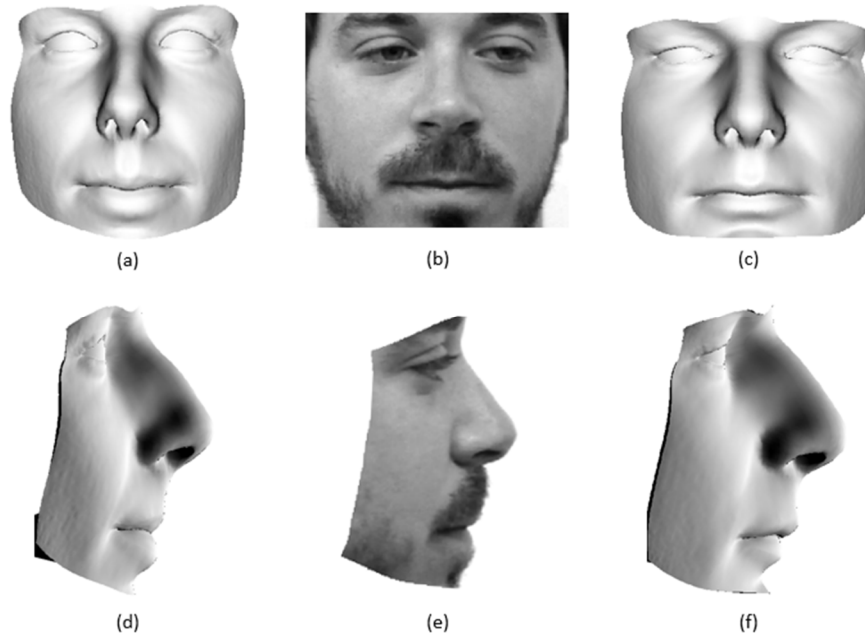


Figura 44. Resultados Sujeto 4 - Reconstrucción final de la forma del rostro. En (a) y (d) la reconstrucción con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ$ y $\pm 55^\circ$. En (b) y (e) el sujeto real. En (c) y (f) la reconstrucción con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ, \pm 55^\circ, 0^\circ$ y 180° .

Sujeto 5

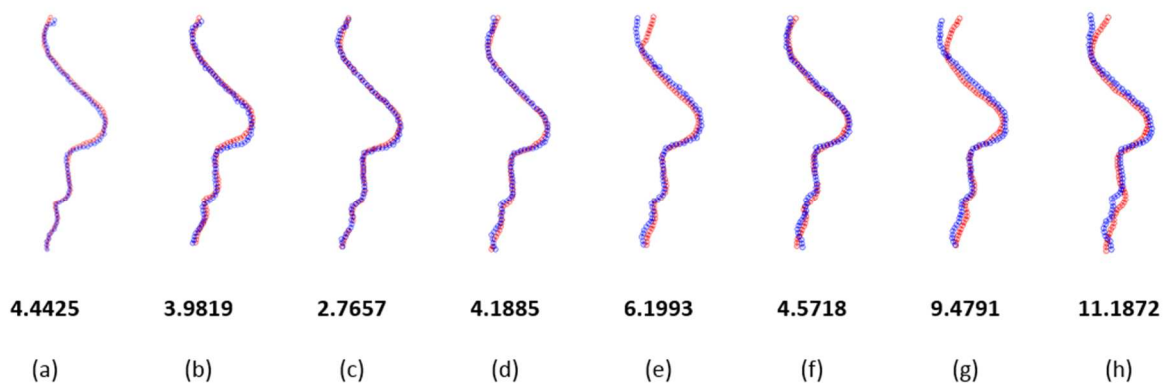


Figura 45. Resultados Sujeto 5 - Comparativa de perfiles. Comparativa de los contornos reales (en azul) con los simulados (en rojo) mediante la distancia euclidiana. En (a) la comparativa con los contornos de pitch 90° . En (b) contornos de pitch 85° . En (c) contornos de pitch 80° . En (d) contornos de pitch 75° . En (e) contornos de pitch 70° . En (f) contornos de pitch 65° . En (g) contornos de pitch 60° . En (h) contornos de pitch 55° .

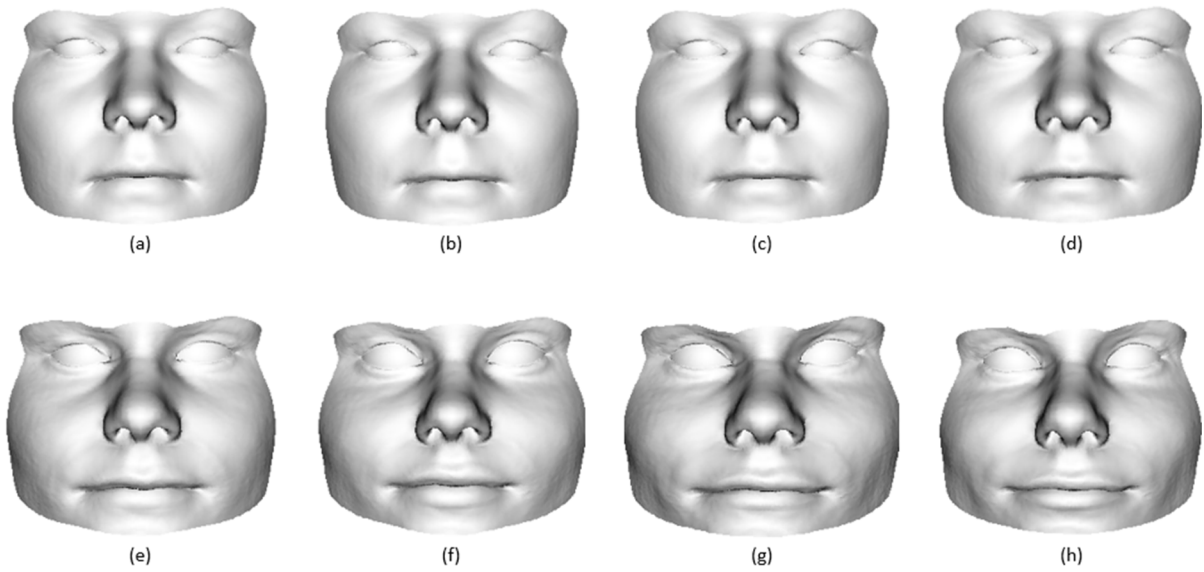


Figura 46. Resultados Sujeto 5 - Reconstrucciones de la forma del rostro. En (a) la reconstrucción con el contorno de 90° . En (b) la reconstrucción con los contornos de 90° y $\pm 85^\circ$. En (c) con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$. En (d) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$. En (e) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 70^\circ$. En (f) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$ y $\pm 65^\circ$. En (g) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$. Y en (h) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$ y $\pm 55^\circ$.

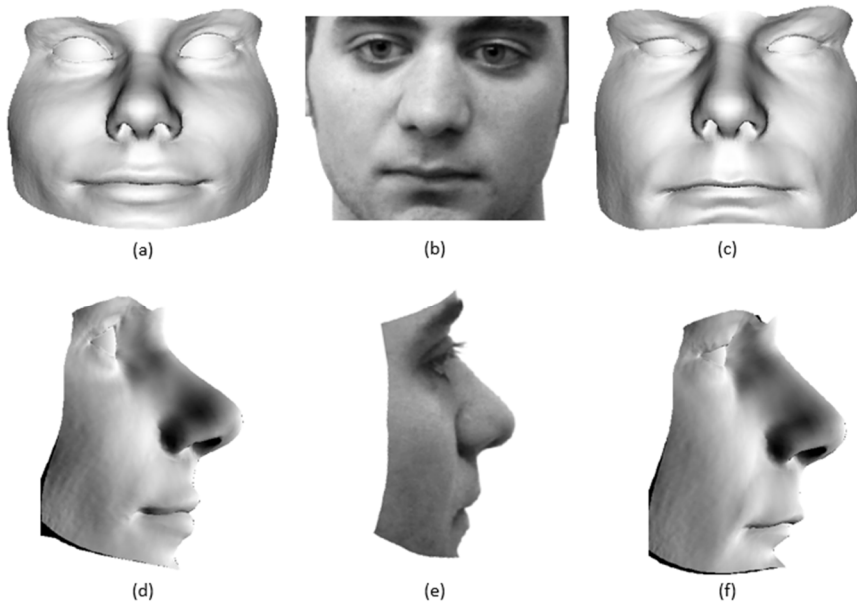


Figura 47. Resultados Sujeto 5 - Reconstrucción final de la forma del rostro. En (a) y (d) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$ y $\pm 55^\circ$. En (b) y (e) el sujeto real. En (c) y (f) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$, $\pm 55^\circ$, 0° y 180° .

Sujeto 6

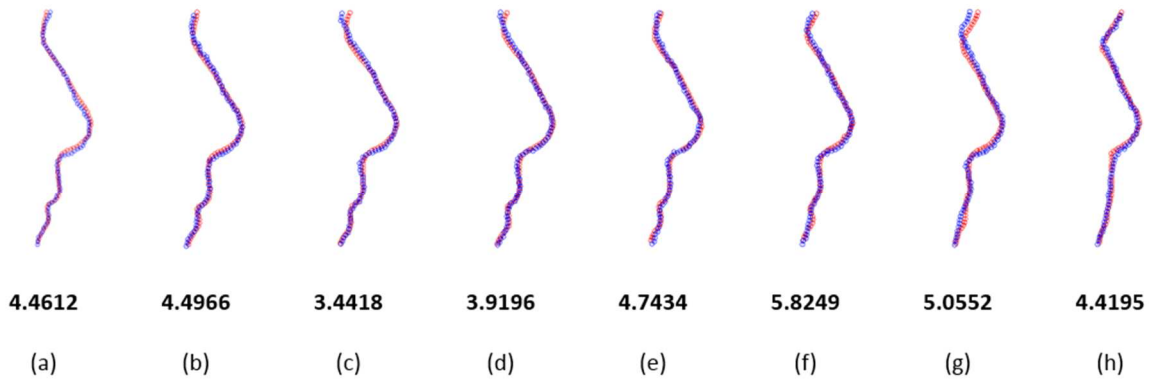


Figura 48. Resultados Sujeto 6 - Comparativa de perfiles. Comparativa de los contornos reales (en azul) con los simulados (en rojo) mediante la distancia euclidiana. En (a) la comparativa con los contornos de pitch 90° . En (b) contornos de pitch 85° . En (c) contornos de pitch 80° . En (d) contornos de pitch 75° . En (e) contornos de pitch 70° . En (f) contornos de pitch 65° . En (g) contornos de pitch 60° . En (h) contornos de pitch 55° .

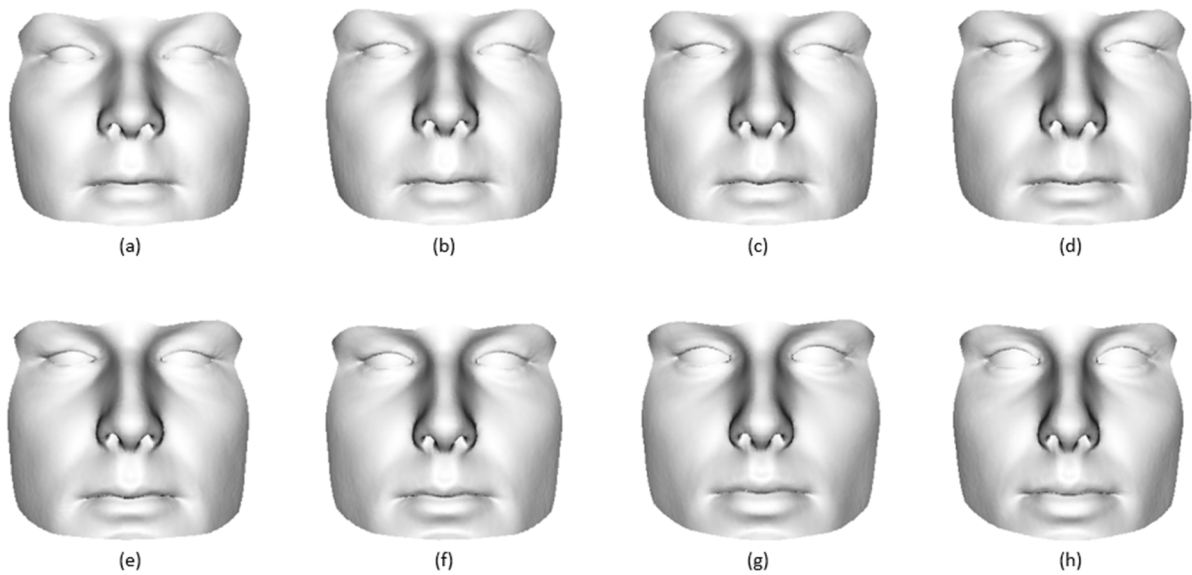


Figura 49. Resultados Sujeto 6 - Reconstrucciones de la forma del rostro. En (a) la reconstrucción con el contorno de 90° . En (b) la reconstrucción con los contornos de 90° y $\pm 85^\circ$. En (c) con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$. En (d) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$. En (e) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 70^\circ$. En (f) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$ y $\pm 65^\circ$. En (g) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$. Y en (h) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$ y $\pm 55^\circ$.

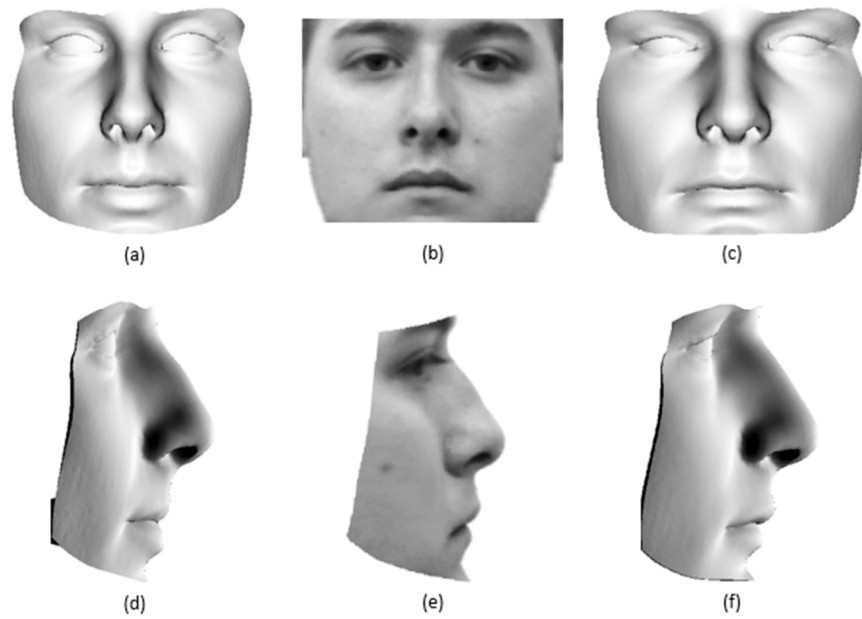


Figura 50. Resultados Sujeto 6 - Reconstrucción final de la forma del rostro. En (a) y (d) la reconstrucción con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ$ y $\pm 55^\circ$. En (b) y (e) el sujeto real. En (c) y (f) la reconstrucción con los contornos de $90^\circ, \pm 85^\circ, \pm 80^\circ, \pm 75^\circ, \pm 70^\circ, \pm 65^\circ, \pm 60^\circ, \pm 55^\circ, 0^\circ$ y 180° .

Sujeto 7

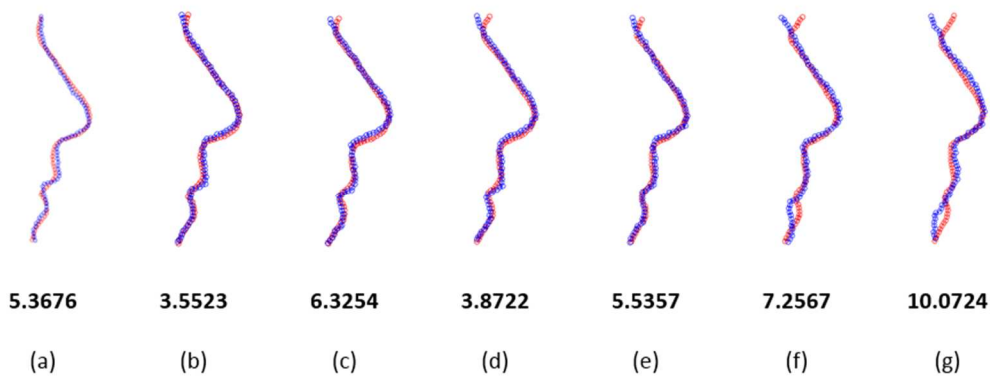


Figura 51. Resultados Sujeto 7 - Comparativa de perfiles. Comparativa de los contornos reales (en azul) con los simulados (en rojo) mediante la distancia euclidiana. En (a) la comparativa con los contornos de pitch 90° . En (b) contornos de pitch 85° . En (c) contornos de pitch 80° . En (d) contornos de pitch 75° . En (e) contornos de pitch 70° . En (f) contornos de pitch 65° . En (g) contornos de pitch 60° .

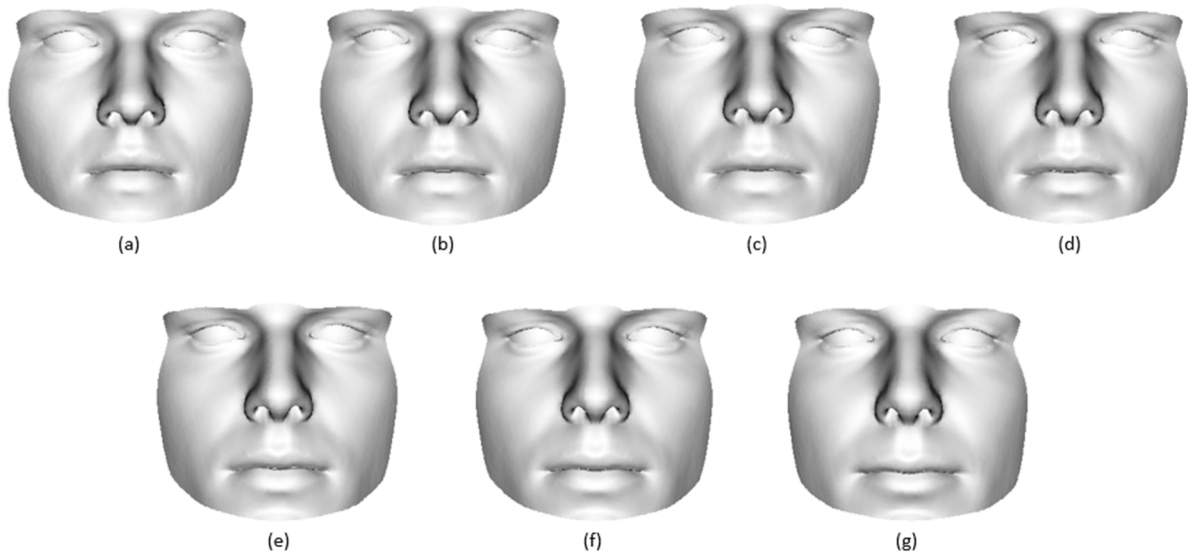


Figura 52. Resultados Sujeto 7 - Reconstrucciones de la forma del rostro. En (a) la reconstrucción con el contorno de 90° . En (b) la reconstrucción con los contornos de 90° y $\pm 85^\circ$. En (c) con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$. En (d) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$. En (e) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 70^\circ$. En (f) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$ y $\pm 65^\circ$. Y en (g) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$.

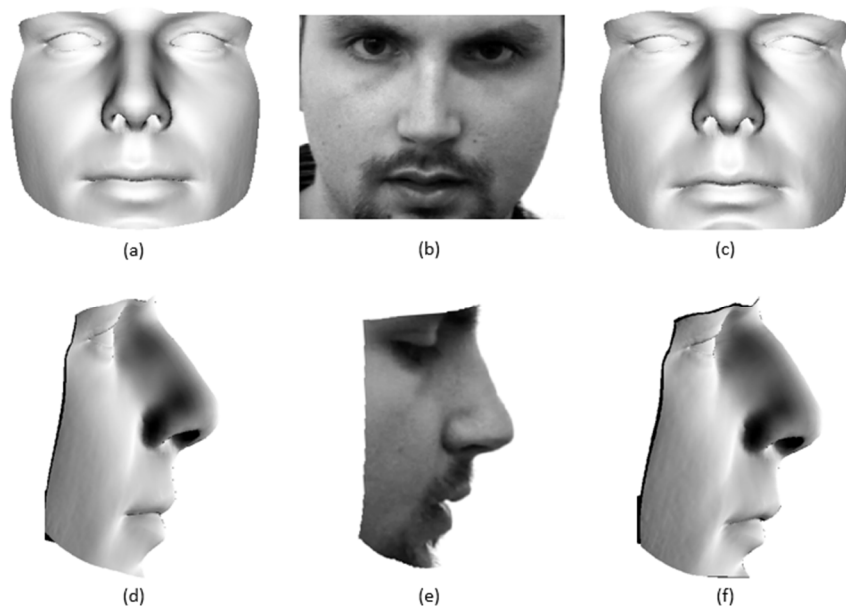


Figura 53. Resultados Sujeto 7 - Reconstrucción final de la forma del rostro. En (a) y (d) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$. En (b) y (e) el sujeto real. En (c) y (f) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$, 0° y 180° .

Sujeto 8

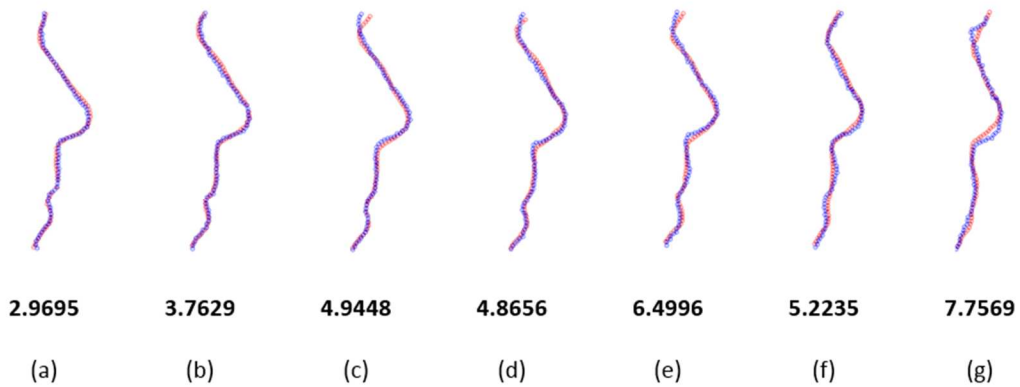


Figura 54. Resultados Sujeto 8 - Comparativa de perfiles. Comparativa de los contornos reales (en azul) con los simulados (en rojo) mediante la distancia euclidiana. En (a) la comparativa con los contornos de pitch 90° . En (b) contornos de pitch 85° . En (c) contornos de pitch 80° . En (d) contornos de pitch 75° . En (e) contornos de pitch 70° . En (f) contornos de pitch 65° . En (g) contornos de pitch 60° .

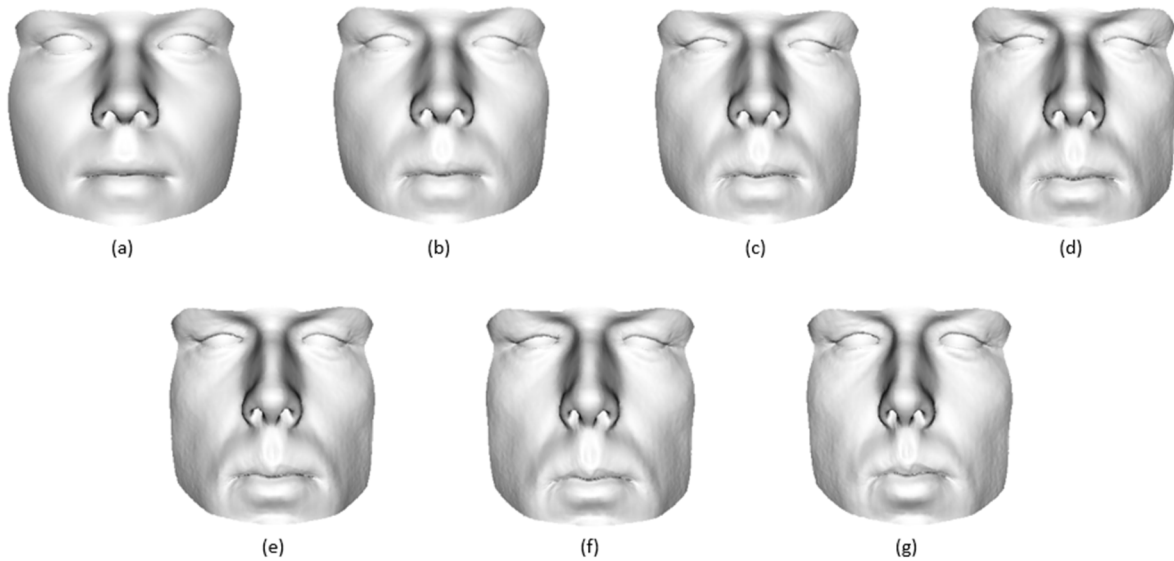


Figura 55. Resultados Sujeto 8 - Reconstrucciones de la forma del rostro. En (a) la reconstrucción con el contorno de 90° . En (b) la reconstrucción con los contornos de 90° y $\pm 85^\circ$. En (c) con los contornos de 90° , $\pm 85^\circ$ y $\pm 80^\circ$. En (d) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$ y $\pm 75^\circ$. En (e) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$ y $\pm 70^\circ$. En (f) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$ y $\pm 65^\circ$. Y en (g) con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$.

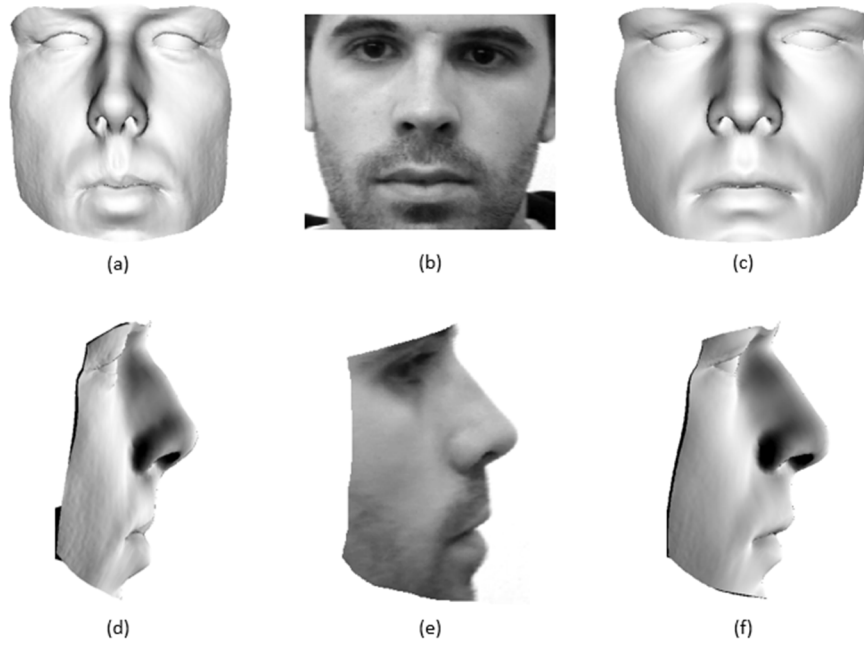


Figura 56. Resultados Sujeto 8 - Reconstrucción final de la forma del rostro. En (a) y (d) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$ y $\pm 60^\circ$. En (b) y (e) el sujeto real. En (c) y (f) la reconstrucción con los contornos de 90° , $\pm 85^\circ$, $\pm 80^\circ$, $\pm 75^\circ$, $\pm 70^\circ$, $\pm 65^\circ$, $\pm 60^\circ$, 0° y 180° .

Anexo II. Inventario de funciones.

En este anexo se presentan las funciones más significativas de este proyecto, con el detalle de los parámetros de entrada/salida de cada una de ellas.

Reconstruccion3D

Función general para la reconstrucción de la forma del rostro a partir de una secuencia de vídeo.

Parámetros de entrada:

- `num_video`: Número del vídeo que se desea procesar. Si se desea reconstruir el sujeto del vídeo IMG_6749.MOV, se debe poner únicamente 6749.

Parámetros de salida:

- `Cont`: Contornos del rostro que se han introducido en el Regresor. Se componen como un vector de $80n \times 2$, donde n es el número de contornos.
- `R`: Superficie recuperada. Nube de puntos de la reconstrucción del rostro.

Comparison_profile

Función que extrae la matriz Profile en la que se encuentran los contornos del vídeo (`perfil_real`) mediante el análisis de los contornos simulados (`perfil_simulado`).

Parámetros de entrada:

- `num_video`: Número del vídeo.
- `range_angle`: Intervalo (ángulo) entre los diferentes contornos.
- `angle_ini`: Ángulo del contorno inicial.
- `angle_end`: Ángulo del contorno final.
- `VL`: Variables latentes.
- `Frame_vid`: Número de frame que contiene el contorno principal (pitch 90°).
- `Frame_frontal`: Número de frame que contiene los contornos frontales (pitch 0°/180°)

Parámetros de salida:

- `Profile`: Contornos del rostro que se han introducido en el Regresor.
- `Rs`: Superficie 3D recuperada.

find_profile

Función que encuentra el contorno del vídeo (perfil_real) que más se asemeja al perfil_simulado mediante la Transformada de Hough Generalizada (GHT).

Parámetros de entrada:

- video: Ubicación del vídeo.
- profile_simulation: Contorno simulado (template para la GHT).
- angle: Ángulo bajo estudio.
- angle_end: Ángulo del contorno final.
- frame_vid: Número de frame que contiene el contorno principal (pitch 90°).
- p1 y p2: Límites superior e inferior del contorno extraído anterior.
- fi_old, S_old: Valores de rotación y escalado del contorno extraído anterior.
- k: Contador para la GHT.
- num_vídeo: Número del vídeo.

Parámetros de salida:

- profile_real: Contorno del vídeo que más se asemeja al perfil_simulado.
- Is_frame_num: Número de frame del vídeo considerado como perfil_real.
- fi_act y S_act: Valores de rotación y escalado del perfil_real.

TGHough

Función que realiza la Transformada de Hough Generalizada (encuentra el template/shape Itm en la imagen grayscale Is).

Parámetros de entrada:

- Is: Imagen en grayscale en la que se encuentra el template Itm.
- Itm: Template.
- p1 y p2: Límites superior e inferior del contorno extraído anterior.
- fi_ant, S_ant: Valores de rotación y escalado del contorno extraído anterior.
- k: Contador auxiliar para la GHT.

Parámetros de salida:

- x, y: Coordenadas del frame que tiene mejor match.
- fi_tf y s_tf: Rotación (grados decimales) y escalado del mejor match.
- score: Score del mejor match.

Repr_TGHough

Función que realiza el escalado, la rotación y la translación al template con los parámetros recuperados en la Transformada de Hough Generalizada

Parámetros de entrada:

- `x_tf`: Coordenada x de translación que se recupera de TGHough.
- `y_tf`: Coordenada y de translación que se recupera de TGHough.
- `fi_tf`: Ángulo de rotación que se recupera de TGHough.
- `s_tf`: Valores de escalado que se recupera de TGHough.
- `profile_x`: Coordenadas x del Template.
- `profile_y`: Coordenadas y del Template.

Parámetros de salida:

- `x_tm` e `y_tm`: Coordenadas x e y tras la transformación (escalado, rotación y translación).

profile_image

Función que selecciona el contorno deseado de un frame y extrae las coordenadas de dicho contorno mediante un vector de 80 puntos equiespaciados y alineado con la media correspondiente.

Parámetros de entrada:

- `IMG`: Frame (imagen) del que se quiere extraer el contorno.
- `Media`: Media de los contornos del modelo en el ángulo que se estudia.
- `video`: Ubicación del vídeo.
- `frame`: Número del frame del que se quiere extraer el contorno.

Parámetros de salida:

- `Z`: Contorno tras normalizarlo con la función Procrustes.
- `D`: Distancia entre los contornos.
- `y1` e `y2`: Coordenadas y del inicio y del final del segmento contorno.

profile_image_2

Función que selecciona el contorno deseado de un frame a partir del matching de la Transformada de Hough Generalizada.

Parámetros de entrada:

- `IMG`: Frame (imagen) del que se quiere extraer el contorno.
- `Media`: Media de los contornos del modelo en el ángulo que se estudia.
- `video`: Ubicación del vídeo.
- `frame`: Número del frame del que se quiere extraer el contorno.
- `x1` y `x2`: Coordenadas x (inicio y fin) del contorno.
- `y1` e `y2`: Coordenadas y (inicio y fin) del contorno.
- `angulo`: Ángulo bajo estudio.
- `directory`: Directorio donde guardar las imágenes del proceso.

Parámetros de salida:

- `Z`: Contorno tras normalizarlo con la función Procrustes.
- `D`: Distancia entre los contornos.

test_images

Función que segmenta los frames de un vídeo y extrae los contornos del sujeto. Esta segmentación se hace con un nivel 0.8 de UCM.

test_segmentation_frontal

Función que segmenta el frame frontal del vídeo que se especifica. Esta segmentación se hace para los valores 0.1, 0.2,..., 1 de UCM.

Parámetros de entrada:

- `video`: Número del vídeo
- `scale`: Escalado de los frames.
- `frame_frontal`: Número de frame que contiene el rostro más frontal.

Esta función no tiene parámetros de salida puesto que se guardan los frames segmentados automáticamente.

face_tracker

Función que encuentra el frame que contiene el rostro más frontal. Este proceso se realiza mediante el detector de puntos faciales Chehra Face Tracker.

Parámetros de entrada:

- `video`: Ubicación del vídeo.
- `scale`: Escalado de los frames (mismo escalado que se ha realizado en la segmentación UCM de los frames).

Parámetros de salida:

- `frame_frontal`: Número de frame que contiene el rostro más frontal.

face_segmentation

Función que extrae los contornos de pitch 0° y 180°.

Parámetros de entrada:

- `frame_frontal`: Número de frame que contiene el rostro más frontal.
- `video`: Número del vídeo
- `i`: Nivel de segmentación UCM que se quiere realizar. Si se desea realizar una segmentación del 0.2, se debe introducir un 2).

Parámetros de salida:

- `profile_real_180`: Contorno extraído de ángulo pitch 180°.
- `profile_real_0`: Contorno extraído de ángulo pitch 0°.

Estimacion3D

Función que extrae los contornos de pitch 0° y 180°.

Parámetros de entrada:

- `NewBD`: Base de datos que contiene la información de entrenamiento (contornos 2D y superficies 3D de los sujetos sintéticos)
- `Angulos_in`: Ángulos que se introducen en el cálculo del regresor.
- `Angulos_out`: Ángulos que se desea extraer el contorno de la reconstrucción obtenida.
- `Sujeto`: Número de sujeto que se deja fuera de entrenamiento.
- `LV`: Variables latentes.
- `HayPerfil`: Variable que contiene los contorno extraídos del vídeo, correspondientes a los ángulos `Angulos_in`.

Parámetros de salida:

- `Contorno_R`: Contorno de ángulo pitch `Angulos_out` extraído de la superficie reconstruida.
- `R`: Coordenadas (x,y,z) de la superficie reconstruida.

CreaPredictores

Función que prepara la matriz de Predictores. Es decir, función que extrae los contornos de ángulos pitch `Angulos` de los rostros de los sujetos sintéticos.

Parámetros de entrada:

- `Angulos`: Ángulos pitch para crear la matriz de Predictores (información de entrenamiento).
- `BD`: Superficie 3D de la que se quiere extraer los contornos predictores.

Parámetros de salida:

- `P2New`: Matriz que contiene los contornos de ángulos pitch `Angulos` extraído de los sujetos sintéticos.

Distance

Función que calcula la distancia entre los contornos 2D reales y simulados.

Parámetros de entrada:

- `profile_simulation`: Contorno 2D extraído de la superficie 3D generada.
- `profile_real`: Contorno 2D extraído del vídeo.

Parámetros de salida:

- `distance`: Distancia euclidiana entre los contornos 2D de entrada.

Euclidean_Distance

Función que calcula la distancia entre los contornos 2D reales y simulados para el vídeo indicado. Esta distancia se calcula con la ayuda de la función `Distance`.

Parámetros de entrada:

- `video`: Número del vídeo

Parámetros de salida:

- `distance`: Distancia euclidiana entre los contornos 2D (reales y simulados) para los distintos ángulos de pitch estudiados.

