



Universidad
Zaragoza

Trabajo Fin de Grado

Modelos de formas y apariencias humanas a partir
de imágenes múltiples.

Autor/es

Ibón Parra Castillo

Memoria

Director/es

Dr. Francisco José Serón Arbeloa

Escuela de Ingeniería y Arquitectura
2013

Agradecimientos

Al Dr. Francisco José Serón por su ayuda y colaboración en el proyecto .

A mi familia, por su apoyo y confianza.

Resumen del proyecto

En el contexto de la interacción hombre maquina a través de agentes conversacionales existe una problemática conocida como el Uncanny Valley (Valle desconocido) que hace referencia al rechazo que los modelos de texturización sintética producen en el espectador conforme su realismo va aumentando y mientras este no sea de una calidad prácticamente indistinguible de un ser humano.

El modelado basado en imagen es la técnica de generación de mallas 3D que permite reconstruir un modelo real en un computador a través de una serie de fotografías del mismo y es la técnica utilizada para generar mallas 3D que intenten solventar ese problema.

La memoria recoge un primer bloque en el que se hace un análisis de los conceptos básicos del modelado basado en imagen. Se estudia la aplicación de esta técnica en el desarrollo de un agente conversacional analizando sus ventajas e inconvenientes así como un análisis de la teoría del Uncanny Valley.

En un segundo parte se presenta el caso de dos modelos humanos obtenidos a través de fotografías múltiples con el software 123D Catch.

Se analizan los programas necesarios para su edición y procesado, Zbrush , así como procedimiento seguido para la adaptación de estos modelos al desarrollo de un agente conversacional humano.

Finalmente se presenta un modelo 3D con los modificadores necesarios para generar las imágenes básicas de una animación de un agente conversacional humano.

ÍNDICE

1- Introducción	6
1.1-Contexto	6
1.2-Objetivos	6
1.3- Descripción de la memoria	6
2) Fundamentos, definiciones y problemas:	7
2.1- Image based modeling (modelado basado en imágenes).	7
2.1.1 Concepto, geometría y matemáticas	7
2.1.1.1 Geometría epipolar.	8
2.1.1.2 Análisis de imágenes y correspondencia entre puntos.	11
2.1.1.3- Reconstrucción modelo 3D	12
2.1.1.4- Triangulación Delaunay	13
2.1.2- Texturizado	15
2.1.3- Análisis y ventajas del método.	15
2.2- Embodied Conversational Agent (Agente conversacional humano)	16
2.2.1- Ejemplos de agentes conversacionales	17
2.2.2.1 Aplicaciones de los ECAS:	19
2.2.2- Elementos de un agente conversacional	19
2.2.2.1- Modelo 3D	20
2.2.2.1.1- Malla 3D	20
2.2.2.1.2-Textura	20
2.2.2.1.3- Nivel de poligonación.	21
2.2.2.2- Animación	22
2.2.2.2.1-Lip Sync.	22
2.2.2.2.2-Fonemas y expresiones faciales:	23
2.2.2.2.3 Técnicas de animación	25
2.2.2.2.3.1-Morphing.	25
2.2.2.2.3.2-Animación con huesos	27
2.2.3- Retos y objetivos de las técnicas	28
2.3) Uncanny Valley	29
2.3.1-Hipótesis	30
2.3.2-Explicación	30
2.3.2.1-Errores	31
2.3.2.2-Exitos	31
2.3.3-Cómo evitarlo	32
2.3.3.1-Ejemplo Emily	32

3-Software utilizado.	34
3.1-123D Catch®.	34
3.2-Zbrush®	34
3.3-3ds Max®	34
3.4- Vray®	35
3.5- Photoshop®	35
4-Contruccion de modelo validos para ACH	36
4.1 Presentación	36
4.1.1 Modelo de un hombre.	36
4.1.2- Modelo de una mujer.	36
4.2-Procedimiento	37
4.2.1 Sesión fotográfica.	37
4.2.2-123D Catch	38
4.2.3 Edición de mallas en Zbrush	39
4.2.4 3ds max	42
4.2.5-Vray	43
5-Imágenes finales	44
5.1-Modelo 3D del hombre	44
5.2-Modelo 3D de la mujer	49
5.3-Expresiones faciales	54
6- Conclusiones	56
7- Trabajo futuro	57
8-Programación	58
 Anexo I Software utilizado	 1-12
Anexo II Procedimiento	1-30

1- Introducción

1.1-Contexto

Entre las diferentes técnicas ofrecidas por el área de los gráficos por computador que se conocen como Modelado Basado en Imagen se pueden destacar aquellas que permiten crear modelos 3D texturizados de un objeto a partir de un conjunto de imágenes de dicho objeto.

De entre sus numerosas aplicaciones, existe una que esta relacionada con obtención de Agentes Humanos Conversacionales como los que se utilizan en entornos de interfaz de usuario, realidad virtual y aumentada, videojuegos...

Las representaciones de estos agentes con apariencia humana han chocado hasta el momento con la psicología de los seres humanos que muestran un rechazo visceral a los modelos 3D texturizados sintéticos según estos van avanzando en calidad y realismo si no llegan a un nivel de calidad prácticamente indistinguible de otro ser humano.

1.2-Objetivos

Este proyecto pretende analizar la problemática asociada a este tipo de innovaciones, aprovechando la aparición de aplicaciones para usuarios capaces de ofrecer estas funcionalidades (123D Catch) y herramientas de refinamiento y modificación de mallas como Zbrush.

Los objetivos de este proyecto consisten en:

- Conocer las posibilidades de las tecnologías estudiadas.
- Utilizarlas para el desarrollo de un conjunto de visualizaciones básicas como en las que se basan aplicaciones que utilizan agentes conversacionales humanos.

1.3- Descripción de la memoria

En la primera parte de la memoria se describen los conceptos en los que se basan las técnicas de reconstrucción 3D a través de fotografías múltiples, la visión estereoscópica, la geometría epipolar y el flujo óptico.

En la segunda parte se explica que es un agente conversacional y sus elementos básicos, el modelado 3D y su animación de fonemas y expresiones.

Conocidos estos dos temas, se explicará la utilidad de la reconstrucción 3D para solventar uno de los problemas a los que se enfrentan los modelos de apariencia humana: el Uncanny Valley.

Finalmente se describe el procedimiento seguido para realizar dos modelos 3D de apariencia humana basados en modelos reales y las imágenes para su visualización.

2) Fundamentos, definiciones y problemas:

2.1) Image based modeling¹ (modelado basado en imágenes).

Dentro del modelado 3D existen varias técnicas para obtener información del mundo real y poder hacer uso de ella en un ordenador. Una de estas técnicas es el Modelado Basado en Imagen (Image Based Modeling).

Los siguientes conceptos hacen referencia a la base teórica sobre la que se sustenta y de la que hacen uso los diferentes programas de reconstrucción 3D.



Figura 1. Ejemplo de sesión fotográfica

El modelado basado en imagen es un sistema que nos permite, mediante una secuencia de fotografías desde diferentes ángulos de un mismo objeto, obtener un modelo 3D del mismo con todas sus características físicas, ya sean geometrías, de color o iluminación.

Esto se consigue por medio de operaciones y algoritmos matemáticos los cuales se encargan de relacionar dos puntos de dos fotografías diferentes y obtener el punto situado en el espacio al que hacen referencia. A partir de este planteamiento se puede interpolar la posición de la cámara y del resto de los puntos del modelo en el espacio pudiendo así obtener finalmente una nube de puntos y posteriormente un modelo 3D.

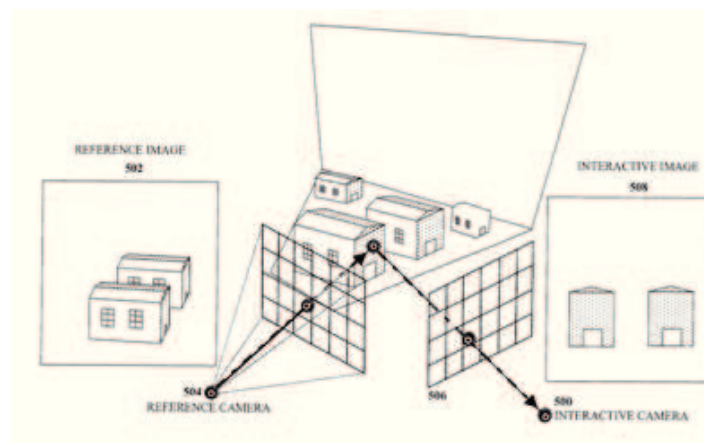


Figura 2. Esquema de una escena y las imágenes captadas.

2.1.1 Concepto, geometría y matemáticas

Esta técnica de reconstrucción 3D se basa en el concepto de la visión estereoscópica producida por la disparidad binocular

Se conoce como disparidad binocular (figura 3) o retinal a la ligera diferencia entre las imágenes generadas por los dos puntos de vista proporcionados por ambos ojos.

La disparidad binocular² es la forma de percibir profundidad y relieve más utilizada por el cerebro humano.

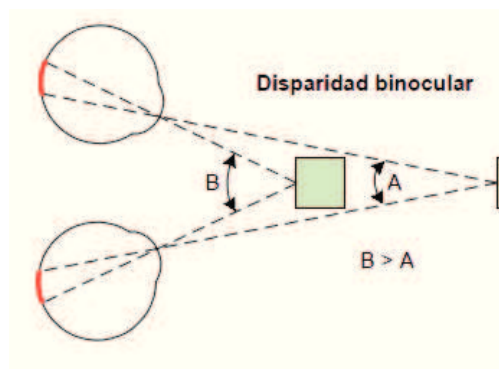


Figura 3. Esquema de disparidad binocular.

Como se puede ver en la figura 4, la imagen tridimensional, se obtiene a través del análisis, ya sea por el cerebro o por un computador, de las pequeñas diferencias que existen entre las imágenes tomadas desde dos puntos de vista diferentes del mismo objeto o escena.



Figura 4. Ejemplo de imágenes diferentes a través de visión binocular

2.1.1.1 Geometría epipolar³:

La geometría epipolar (figura 5) es la geometría de la visión estereoscópica. Cuando dos cámaras observan una escena 3D desde dos posiciones distintas, existe una relación geométrica entre los puntos 3D y sus proyecciones sobre las imágenes en 2D que conducen a las restricciones entre los puntos de la imagen.

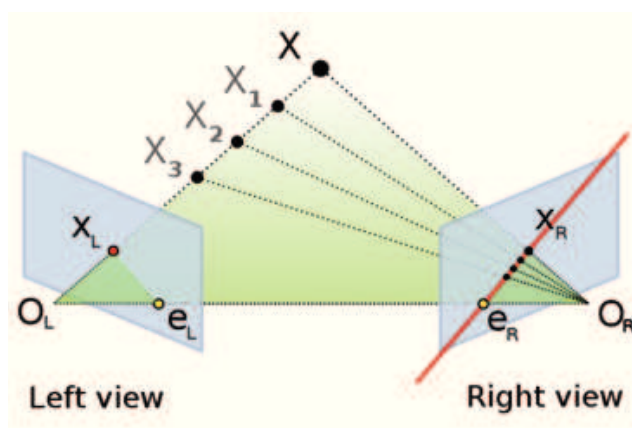


Figura 5. Geometría epipolar

En la figura 5, O_L y O_R representan los centros de proyección de las dos cámaras. X representa el punto de interés por ambas cámaras. Puntos X_L y X_R son las proyecciones del punto X en los planos de imagen.

Se define el plano epipolar para cada punto de la imagen como aquel que contiene los dos centros ópticos

(ojos, O_L y O_R) y el punto bajo estudio (objeto, X). Una vez establecido el plano epipolar, se puede determinar la recta epipolar, que se define para cada punto y cada cámara como la intersección del plano epipolar con el plano de la imagen captada.

Por tanto:

Si la posición del punto de proyección x_L es conocida, entonces la línea epipolar $e_R - x_R$ es conocida y el punto X proyectado en la imagen de la derecha, en un punto x_R que debe estar en esta línea epipolar particular. Esto significa que por cada punto observado en una imagen se debe observar el mismo punto en la otra imagen en una línea epipolar conocida. Esto proporciona una restricción epipolar a los puntos de la imagen y significa que es posible probar si dos puntos realmente corresponden al mismo punto 3D.

Si los puntos x_L y x_R son conocidos, también se conocen sus líneas de proyección. Si los dos puntos de imagen corresponden a los mismos puntos 3D las líneas de proyección deben intersectar precisamente en X . Esto significa que X puede calcularse a partir de las coordenadas de los dos puntos de la imagen, un proceso llamado triangulación.

En la figura 6, se consideran los puntos P , p_i y p_d . Dado p_i , P puede caer en cualquier punto del rayo definido por O_i y p_i . Pero dado que la imagen de este rayo en la imagen derecha es la línea epipolar a través del punto correspondiente p_d dicho punto debe estar sobre la línea epipolar. Esta correspondencia establece una aplicación entre puntos de la imagen izquierda y rectas de la imagen derecha y viceversa.

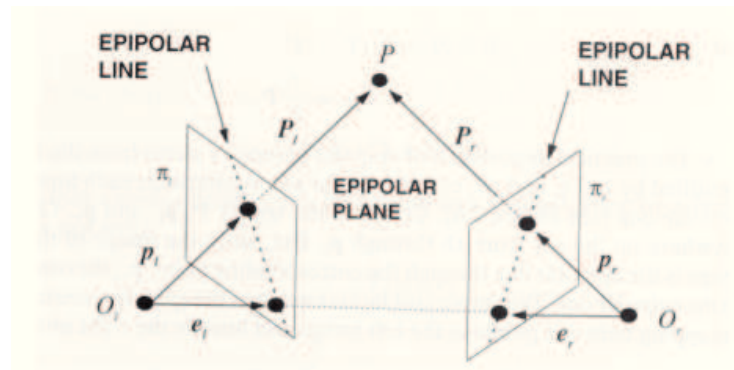


Figura 6. Esquema de la geometría epipolar con vectores.

La figura 6 muestra las dos cámaras, sus centros de proyección, O_i y O_r , y sus planos imagen o retinas en coordenadas normalizadas. Las longitudes focales de la cámara se nombran como f_i y f_d .

Cada cámara identifica un sistema de referencia 3D, el origen del cual se sitúa en el centro de proyección de la cámara y el eje Z con el eje de óptico.

Los vectores

$$P_i = [X_i, Y_i, Z_i] \quad (1)$$

$$P_d = [X_d, Y_d, Z_d] \quad (2)$$

se refieren al mismo punto en el espacio, P , como vectores en los sistemas de referencia de la cámara izquierda y derecha respectivamente.

Estos vectores definen por tanto las proyecciones del punto P en la imagen izquierda y derecha y están expresados en su correspondiente sistema de referencia. Evidentemente, para todos los puntos de las imágenes se tiene $z_i = f_i f_i$ o $z_d = f_d$ de acuerdo a la imagen que sea.

Los sistemas de referencia de las cámaras izquierda y derecha están relacionados a través de los parámetros extrínsecos. Estos definen una transformación rígida en el espacio 3D definida por un vector de traslación

y una matriz de rotación R . Dado un punto P en el espacio la relación entre P_i y P_d es por tanto

$$T = O_d - O_i \quad (3)$$

$$P_d = R(P_i - T) \quad (4)$$

Por tanto si se determina la aplicación entre puntos de la imagen izquierda y las rectas epipolares de la imagen derecha, se pueden restringir la búsqueda para el emparejamiento de p_i a lo largo de la línea epipolar correspondiente

Una cuestión importante es ¿cómo estimar la geometría epipolar? O equivalentemente ¿cómo establecer una aplicación entre puntos de una imagen y líneas epipolares de la otra?.

Los métodos para la estimación entre puntos de una imagen y las líneas epipolares de la otra son la matriz esencial y la matriz fundamental

Matriz esencial y Matriz fundamental⁴

La ecuación del plano epipolar a través de P puede escribirse como la condición co-planar de los vectores:

$$P_i^T T \text{ y } P_i - T \text{ ó } (P_i - T)^T (T \times P_i) = 0. \quad (5)$$

Usando la relación que liga a los vectores P_i y P_d se obtiene

$$(R^T P_d)^T (T - P_i) = 0. \quad (6)$$

Teniendo en cuenta que el producto vectorial de dos vectores se puede escribir como la multiplicación de una matriz antisimétrica por un vector, tenemos $T \times P_i = [T] \times P_i = SP_i$ donde

$$[T]_{\times} = S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \quad (7)$$

Se obtiene $P_d^T E P_i = 0$ con $E = RS$.

Por construcción E siempre tiene rango igual a 2 y se puede probar que su dos autovalores distintos de cero son iguales. La matriz E se denomina la matriz esencial y establece una unión natural entre la restricción epipolar y los parámetros extrínsecos del sistema estéreo. Ahora si consideramos las ecuaciones vectoriales de la perspectiva y sustituimos en la ecuación anterior obtenemos la ecuación que liga las proyecciones del punto P en ambos planos imagen y dividimos por $Z_i Z_d$ obtenemos

$$p_d^T E p_i = 0 \quad (8)$$

El vector que representa $u_d = E p_i$ puede ser interpretado como el vector director de la recta en que se proyecta, sobre la imagen derecha, el rayo definido por $O_i P$. Así pues, y a través de la matriz esencial se establece una aplicación entre los puntos de una imagen y las rectas epipolares de la otra. Usando la notación vectorial introducida, la ecuación anterior se escribe como:

$$p_d^T u_d = 0 \quad (9)$$

Y establece la condición de que uno de los puntos siempre se encuentra sobre la línea epipolar definida por el otro.

Este sistema trabaja en coordenadas que toman como referencia a las cámaras, sin embargo, cuando se mide los puntos de proyección en las imágenes estos están medidos en términos de píxeles, por tanto para poder sacar el máximo partido de la matriz esencial será necesario conocer la transformación desde coordenadas de la cámara a coordenadas de píxeles, es decir, conocer los parámetros intrínsecos. Esta restricción puede ser eliminada pero a costa de pagar un precio sobre la información recuperada.

La matriz fundamental

Si se estima la matriz fundamental a partir de puntos en correspondencia en coordenadas píxel se puede

reconstruir la geometría epipolar sin absolutamente ninguna información sobre los parámetros intrínsecos o extrínsecos.

Sean K_i y K_d las matrices de los parámetros intrínsecos de las cámaras izquierda y derecha respectivamente. Si notamos por p_i y p_d los puntos en coordenadas píxel correspondientes a p_i y p_d respectivamente, tenemos

$$p_i = K_i^{-1} \overline{p_i} \text{ y } p_d = K_d^{-1} \overline{p_d} \quad (10)$$

Sustituyendo en la ecuación de la matriz esencial obtenemos

$$\overline{p_i}^T F \overline{p_d} = 0 \quad (11)$$

siendo

$$F = K_d^{-T} E K_i^{-1} = K_d^{-T} R S K_i^{-1} \quad (12)$$

F se denomina la matriz fundamental.

Al igual que con $E p_i$, $F p_i$ puede ser interpretado como el vector de la recta epipolar proyectiva correspondiente al punto p_i $u_d = F p_i$

Consecuentemente, si se estima la matriz fundamental a partir de puntos en correspondencia en coordenadas píxel se puede reconstruir la geometría epipolar sin absolutamente ninguna información sobre los parámetros intrínsecos o extrínsecos del sistema estereo, en este caso la cámara de fotos.

2.1.1.2 Análisis de imágenes y correspondencia entre puntos.

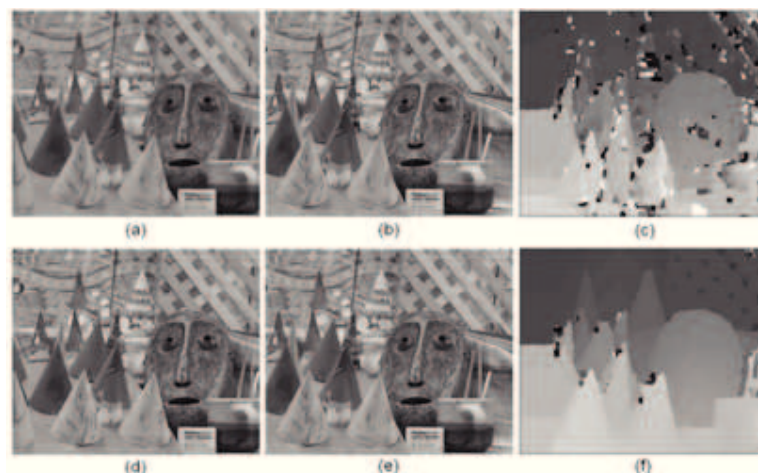


Figura 7: mapas de disparidad

Para la obtención de las coordenadas 3D de un punto del cual solo tenemos unas proyecciones 2D en una fotografía será necesario realizar una correspondencia entre esos dos puntos, es decir, localizar ese mismo punto del espacio en las dos imágenes.

Para la obtención de estos datos es necesario la utilización de unos algoritmo de correspondencia. El objetivo de cualquier algoritmo de correspondencia aplicado a visión estéreo es hallar los puntos en cada una de las imágenes que se corresponden con el mismo punto en la escena.

Un método para la búsqueda de correspondencias es el basado en el flujo óptico⁴.

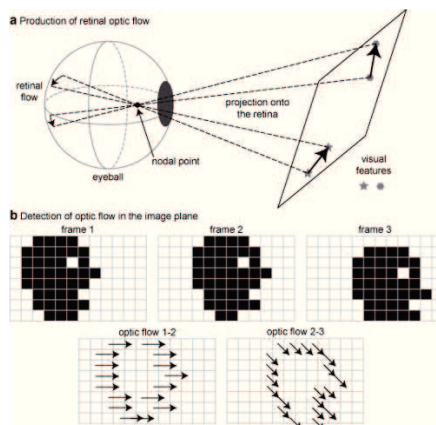


Figura 8: Flujo óptico

Como se observa en la figura 8, el flujo óptico es el patrón del movimiento aparente de los objetos, superficies y bordes en una escena causado por el movimiento relativo entre un observador (un ojo o una cámara) y la escena.

Es mediante la variación de la intensidad de los píxeles en la fotografía como se advierte el movimiento de los objetos dentro de la misma, suponiendo claro, que los puntos de la escena conservan sus propiedades inalteradas. Es decir, que no sufren ningún cambio de iluminación, brillo o contraste con el entorno.

Por tanto, un punto de la imagen conserva su intensidad al moverse por el cambio de punto de vista del observador, así que se considera que los cambios de intensidad en la imagen se deben al movimiento relativo de los objetos en la escena.

Con este procedimiento obtenemos un mapa de disparidad (figura 7), una imagen que refleja las diferencias entre los elementos de la escena y sus correspondencias.

2.1.1.3- Reconstrucción modelo 3D

Una vez conocidos los puntos de correspondencia en las fotografías se puede aproximar su posición 3D en el espacio. Este es un proceso que se conoce como reconstrucción estéreo.

El problema es encontrar las coordenadas 3D de un punto X en el espacio a partir de sus proyecciones en x_1 y x_2 en las fotografías.

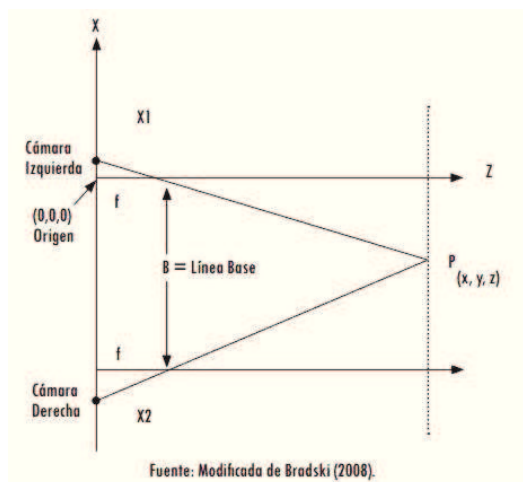


Figura 9. Esquema reconstrucción.

Si se considera, como en la figura 9, el sistema estereoscópico posicionado sobre el eje X en un sistema coordenado $O(XYZ)$, con los planos imagen sobre el plano XY , entonces los ejes ópticos son paralelos al eje

Z, como se muestra en la figura. La triangulación determina un punto sobre un objeto usando el triángulo formado entre el punto observado P y los ejes ópticos de las cámaras, separados una distancia B, también conocida como línea base o baseline. La información de profundidad del punto P está codificada en la diferencia de posición en los dos planos de imagen.

Situando el origen de coordenadas en el centro de la lente de la cámara izquierda, según la figura, se encuentra el conjunto de coordenadas (x, y, z) utilizando la idea de triangulación.

Las coordenadas reales (x, y, z) del punto de interés P se proyectan en las dos imágenes según (13), con subíndice 1 y 2 para la cámara izquierda y derecha, respectivamente. Donde f_x es el producto de la distancia focal f (mm) con el factor de conversión S_x y X_{f1} y X_{f2} están dados en píxeles

$$X_2 = \frac{X_{f2} X_z}{f_x}, X_1 = \frac{X_{f1} X_z}{f_x} \quad (13)$$

En la figura 9 la coordenada z del punto P es la misma en los sistemas de referencia de ambas cámaras; al

$$X = \frac{X_{f1} \times Z}{f_x} \quad Y = \frac{Y_{f1} \times Z}{f_x} \quad Z = \frac{f_x \times B}{X_{f2} - X_{f1}} \quad (14)$$

operar (13) se encuentran las coordenadas (x, y, z) en la escena real, según (14).

2.1.1.4- Triangulación Delaunay⁵

Una vez obtenidos todos estos puntos y sus coordenadas en el espacio, se procede a crear el modelo 3D del objeto con su representación en una malla triangular, aplicando el método de triangulación de Delaunay.

En el método de reconstrucción 3D por fotografías este sistema nos permitirá, a partir de los puntos reconstruidos de las fotografías, generar una malla optimizada y así visualizar una aproximación a la superficie real del objeto.

La triangulación de Delaunay (Figura 10) es una condición que dice que la circunferencia circunscrita de cada triángulo de la red no debe contener ningún vértice de otro triángulo.

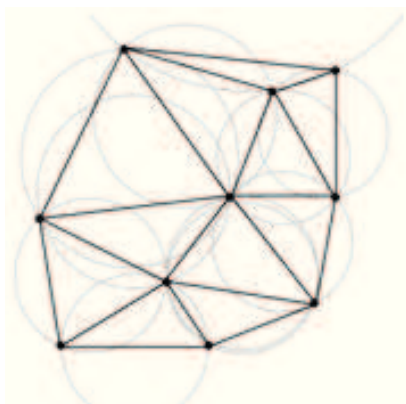


Figura 10. Triangulación 2D Delaunay.

Delaunay: Dada una nube de puntos del plano, hallar una triangulación en la que puntos próximos estén conectados entre sí por una arista. O, dicho de otro modo, en la que los triángulos sean lo más regulares posible.

Este concepto es posible ampliarlo para espacios tridimensionales (figura 11) usando la esfera circunscrita en vez de la circunferencia circunscrita. También es posible ampliarla para espacios con más dimensiones pero no se usa en la práctica.

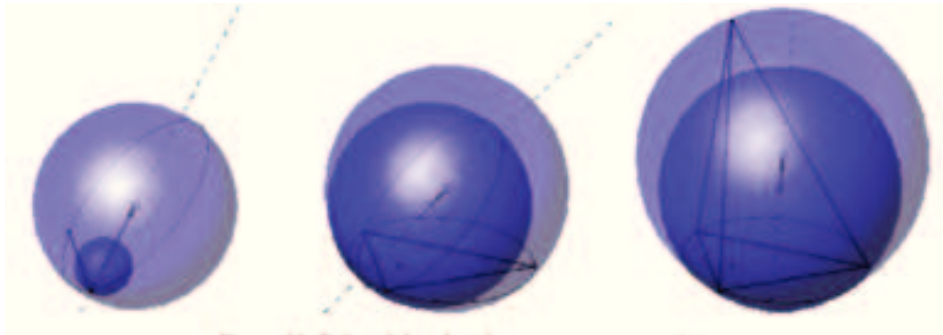


Figura 11. Triangulación 3D Delaunay.

Supóngase una esfera centrada en el punto y que llega hasta el nodo más cercano, la esfera no contiene ningún nodo interior. La esfera se hace crecer, con el nodo siempre en la superficie y el centro en la línea que une al nodo y al punto. Crece hasta encontrar un segundo nodo. En este proceso, la esfera contiene siempre al punto en el interior pues el centro está más cerca del punto que del nodo.

La esfera continúa creciendo, pero ahora con los dos nodos en la superficie y el centro alejándose por la línea de intersección de la mediatriz del segmento que une los nodos con el plano (perpendicular) definido por el punto y los nodos. Crece hasta encontrar un tercer nodo.

El punto sigue dentro de la esfera: considérese una línea por el punto y paralela al segmento que une los nodos, el centro de la esfera se alejó perpendicularmente a esa línea por lo tanto siempre estará más cerca del punto que de los nodos.

Encontrado el tercer nodo, sigue creciendo hasta encontrar un cuarto, pero con el centro en la mediatriz del triángulo de los tres nodos. Dado que el centro se aleja perpendicularmente a un plano que pasa por el punto y es paralelo al del triángulo, siempre contiene al punto.

Como resultado se halló una esfera vacía definida por el nodo más cercano al punto y otros tres y que contiene al punto.

Serán estos tres nodos los que unidos nos permitirán generar un polígono y posteriormente una malla poligonal triangular con la que visualizar el modelo 3D fotografiado. (Figura 12)

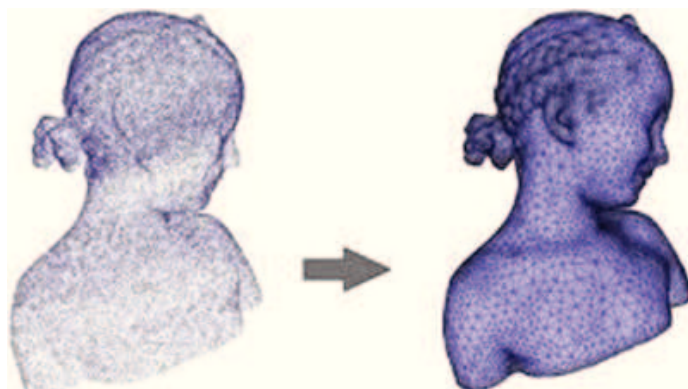


Figura 12. Generación de malla a partir de puntos.

2.1.2- Texturizado

La reconstrucción de objetos 3D por medio de fotografías implica también extraer su textura y aplicarla posteriormente a la malla del modelo. El concepto es la mezcla de las texturas tomadas en cada fotografía para su aplicación.

En su forma más simple aplicación de un material al modelo 3D es la aplicación de una textura de color, esto se puede hacer simplemente en un mapa las imágenes originales en las piezas de la 3D modelo creado por esas imágenes.

Cuando las imágenes múltiples se asignan a la misma sección de la geometría de una de las imágenes se puede elegir arbitrariamente a tener prioridad o se pueden mezclar. Si la mezcla se lleva a cabo a menudo se ponderaron en favor de las imágenes de cámaras más cercanas a la geometría y las proyecciones de la imagen más cerca de ser plana para la geometría.

Más allá de esta técnica simple existen muchos otros, tales como:

- Mapeo: se mezcla las imágenes en función del ángulo de visión, de tal manera que las imágenes tomadas desde un ángulo similar al espectador tienen prioridad. Esto puede ser muy eficaz como detalles demasiado pequeños para ser capturados por el modelo 3D puede obtener una sensación de profundidad en la cuenta de que se queden de la dirección correcta.
- Súper Resolución, implica la combinación de las imágenes para crear una mayor resolución de imagen.
- El material modelado por parámetros va más allá del color simple y medición los detalles que definen cómo un material cambia de color dependiendo del ángulo de visión y las condiciones de observación. Esto requiere muchas muestras del material y no es práctico en una escena del mundo real. Las muchas muestras requeridas se puede conseguir por segmentar el modelo por el material, por lo que obtener muchas muestras de muchas direcciones para cada material.

2.1.3- Análisis y ventajas del método.

El modelo 3D que se obtiene de este proceso de reconstrucción está totalmente texturizado y ya que el programa hace uso de las propias fotografías del objeto para la textura del modelo. Es la unión de la geometría y la texturas, ambas tomadas del objeto real la que nos permiten obtener un modelo 3D con alto realismo en cuanto a forma, color e iluminación.

Tras el análisis de la técnica podemos apreciar que tenemos una serie de ventajas al utilizar este método para realizar un modelo 3D.

- 1-Todos los datos están tomados directamente del mundo real, haciendo que aunque con un margen de error, muchas veces sea más exacto y rápido obtener un primer modelo de esta forma que modelar desde cero.
- 2-Obtenemos el modelo texturizado, su detalle, color, sombras iluminación dentro del contexto de la escena en la que este. El color de la piel o la luz reflejada en un objeto cercano serán elementos que obtendremos directamente de la escena y no deberemos generar después.
- 3-Nos proporciona un modelo primitivo y proporcionado con el que empezar a trabajar a la hora de desarrollar un modelo más complejo como pudiera ser un agente conversacional.

2.2) Embodied Conversational Agent (Agente conversacional humano)

Un agente conversacional humano es una interfaz de interacción hombre-máquina diseñada para mantener una comunicación con el interlocutor a través de su apariencia y movimientos similares a los humanos, y tiene como objetivo convertir esta relación en lo más natural posible.

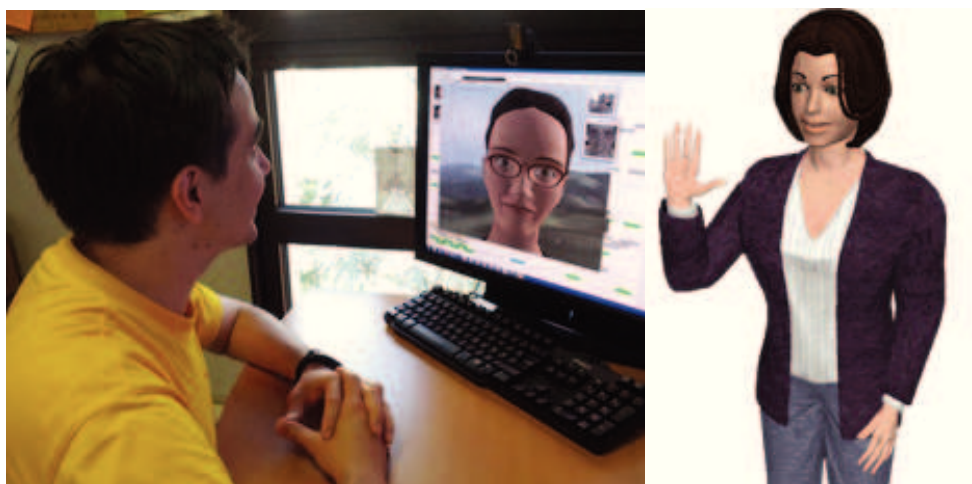


Figura 13. Agentes conversacionales humanos.

Entendemos pues, que un agente conversacional es ese modelo 3D de apariencia humana que mediante una pantalla, proyector u holograma permite interactuar con una máquina mediante una conversación similar a la que se tendría con un humano. Esta conversación se puede llevar a cabo mediante texto, de forma oral o incluso una conversación no verbal.

En la relación de interacción entre hombre y un ordenador hay algo que el usuario no puede obviar, y es que finalmente está interactuando con una máquina. Es en este punto donde los agentes conversacionales buscan minimizar esa sensación de frialdad o lejanía con la máquina haciendo que esta relación, finalmente, sea lo más humana posible.

Los agentes conversacionales o ECA llevan a cabo esta comunicación humanizada con el usuario mediante una morfología, gestos, expresiones y lenguaje lo más realista posibles.

Agentes conversacionales animados o ECA (Embodied conversational Agents) son interfaces gráficas capaces de comunicarse de manera verbal y no verbal con el usuario. Mediante el uso de movimientos y expresiones faciales básicos, un sintetizador de voz y la sincronización de esta con el movimiento de los labios (Lip Sync) son capaces de responder a los requerimientos hechos por el usuario manteniendo así una conversación a nivel gestual, escrita o verbal. Todo esto unido a una representación 3D del modelo humano a imitar generan un conjunto que es capaz de comunicarse en varios niveles con el usuario de la interfaz.

Realizar un ECA es una tarea que reúne a muchas disciplinas, incluye desde problemas tradicionales de la Inteligencia artificial, de cómo procesa la información recibida y como responde, a los propios de las Ciencias sociales, qué responde y de qué forma.

Es por tanto que el lenguaje verbal, gestual y el contenido mismo de la conversación afectan directamente a la relación del usuario con el ECA.

2.2.1- Ejemplos de agentes conversacionales

Podemos encontrar algunos ejemplos de ECAS y se pueden ordenar según su complejidad, tanto de modelo 3d, como de animaciones o nivel de su conversación.

En un primer nivel de complejidad podemos encontrar ejemplos de ECA con pequeñas animaciones:

- Asistente de microsoft office (figura 14).

Fue utilizado para proporcionar la ayuda al usuario del paquete Office de Microsoft. Carecía de forma humana realista siendo generalmente una representación amigable del propio menú de ayuda a través de un personaje cartoon.

La comunicación era escrita haciendo preguntas básicas al usuario que podía responder haciendo click en algunas en las opciones que le ofrecía. Contaba con algunas animaciones básicas que buscaban captar la atención del usuario. Se ha hecho famoso por en ocasiones ser algo molesto por su insistencia.



Figura 14. Agente conversacional Microsft Office.

Si avanzamos en complejidad, ya no tanto de las animaciones si no de la relación con el usuario podemos encontrar:

-Anna de Ikea (figura 15).



Figura 15. Agente conversacional Anna de Ikea.

Se trata de un avatar virtual que actúa a modo de asistente en la búsqueda de productos dentro de la página web de IKEA . El usuario de la página web puede expresar el motivo de la búsqueda en un campo de texto como si de una conversación se tratara, con una pregunta de estructura completa y esperar que el asistente le conteste de manera acorde.

Reconoce la estructura e intención de la consulta ofreciendo una respuesta en el mismo nivel de comunicación. Se trata de un avatar con un número limitado de expresiones dado su campo de aplicación, el mundo web, donde

prima la velocidad de descarga y respuesta sobre la calidad del mismo.

Aun así, las intenciones y reacciones del avatar animado van siempre en relación con el contenido de la conversación, expresando por ejemplo, alegría o repulsión en función de la conversación que se esté manteniendo.

-Mobile secretary (figura 16)



Figura 16. Mobile secretary

En el siguiente nivel complejidad aparecen ejemplos como Mobile Secretary, un avatar que gestiona y proporciona avisos de los eventos del teléfono móvil, y que acompaña sus mensajes con gestos, imitando el lenguaje corporal que seguiría una persona en esas situaciones concretas.

Un lenguaje corporal adecuado, además de complementar visualmente el discurso, proporciona numerosos elementos de naturalidad -aunque el parpadeo del avatar no proporcione información, le hace parecer más realista- que provocan en el interlocutor una mayor sensación de confianza, lo que incrementa automáticamente el nivel de credibilidad.

-Proyecto RED⁶. (Figura 17)

Se trata de una gente conversacional desarrollado para su uso en hospitales que busca reducir la tasa de re-hospitalización utilizando avatares en representación de enfermeros para realizar un mejor diagnóstico de los pacientes y sus síntomas

Nos encontramos ante un ECA de alta complejidad técnica en cuanto al nivel de la conversación que ha de llevar a cabo pero que tiene su punto más flaco en el aspecto realista de la interacción, tanto el avatar como sus gestos y expresiones son todavía muy primitivas.



Figura 17. Proyecto RED

2.2.2.1 Aplicaciones de los ECAS:

En esencia un ECA es una interfaz, y como tal tiene múltiples aplicaciones. Si pensamos en la cantidad de veces que interactuamos con una máquina en una jornada normal, muchas de ellas pueden estar apoyadas o sustituidas por un ECA. Desde la máquina de billetes de autobús, un cajero o una ventanilla de cualquier administración.

Por hacer un ejemplo de clasificación, podríamos hablar de:

-Asistentes Virtuales Sociales: Su propósito es conversar libremente acerca de algo con cualquier usuario, así como lo haría con un amigo. Usados principalmente para el entretenimiento.

-Asistentes Virtuales Educativos: Su propósito es ayudar a aprender acerca de algo, como un nuevo lenguaje, historia, geografía, etc. Tienen uso en centros educativos.

-Asistentes Virtuales Orientados a Servicios: Estos generalmente son utilizados por empresas que ofrecen servicios. Su propósito es ayudar a los clientes a encontrar el camino en el sitio web o en el uso concreto de algún dispositivo así como contestar preguntas acerca de sus productos y servicios.

Por lo tanto, si vamos a introducir un ECA en alguna de las facetas de la vida diaria, a la hora de llevar una conversación con una persona deben ser capaces de realizar acciones comunes como:

-Representar rasgos en su personalidad. De ser un elemento amigable, cariñoso, gracioso o por el contrario autoritario y estricto debe ser capaz de representarlo a través de su actitud tanto en lenguajes verbal como gestual.

-Emociones, objetivos e intenciones. A la hora de llevar una conversación un ECA debería ser capaz de expresar emociones que hagan referencia al contenido de la conversación, así como posibles objetivos o intenciones de la misma.

-Representación de la conversación. La representación gestual de la conversación debe ser acorde con el contenido de la misma. Es por esto que los gestos faciales deberán ir sincronizados con los sonidos o el texto emitido por el agente conversacional.

Como podemos observar, en todos los elementos que el ECA debe ser capaz de llevar a cabo en una conversación entran en juego muchas expresiones tanto faciales como corporales apoyadas en un lenguaje verbal.

2.2.2- Elementos de un agente conversacional

Una vez analizados algunos tipos de ECA podemos apreciar que tienen 3 grandes apartados y que son los que forman finalmente la totalidad del agente conversacional.

-La conversación:

Hace referencia a cómo interpretan lo que el usuario les dice y posteriormente a cómo llevan a cabo su respuesta, en cualquiera de los niveles en los que se esté desarrollando la conversación, verbal, escrita o gestual. Es un campo que afecta más a la inteligencia artificial, reconocimiento de voz, rostro, etc.

-Su apariencia humana. (Modelo 3D)

La apariencia real que toma el avatar. Sus formas y colores que le llevan a ser similar a un humano. Afectan al modelado y texturizado 3D.

-Sus movimientos. (Animación 3D)

Enmarca las animaciones de las que hace uso el avatar para llevar a cabo la conversación. Van desde los corporales, como pueden ser los movimientos con las manos o las piernas hasta las posiciones que adopta la cara al generar un fonema o una expresión concreta.

2.2.2.1- Modelo 3D

Al desarrollar un humano en 3D tenemos que tener en cuenta que cualquier modelo 3D estático, sin animación, tiene las siguientes características o componentes.

2.2.2.1.1- Malla 3D

Comprende todos los polígonos que generan el modelo del personaje. La acumulación de los mismos y la posición relativa entre ellos acaba dando forma a un modelo. Esta malla está compuesta por tres elementos básicos. El punto, la recta o arista y el polígono. De tratarse de una malla poligonal triangular, cada polígono estará delimitado por tres segmentos o aristas y cada una de estas por dos vértices o puntos.



Figura 18. Modelo humano 3D.

2.2.2.1.2- Textura

Hace referencia al color que toman los polígonos del modelo. La aplicación de esta textura sobre el modelo le permite a éste adoptar un color u otro en la visualización final del modelo.

Esta textura es una representación 2D de una superficie 3D y por esto responde a una coordenadas determinadas y sufre una deformación como podemos ver en la figura 19.

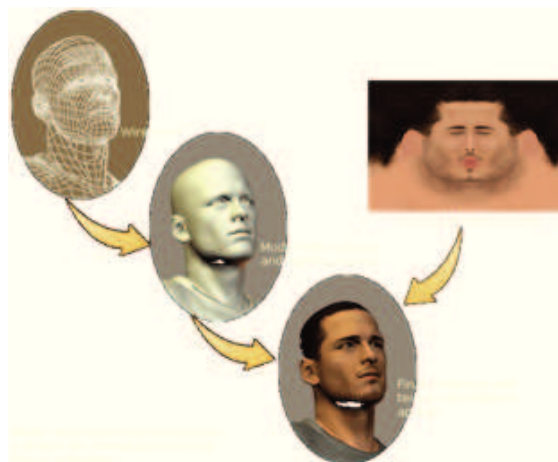


Figura 19: Esquema de un modelo 3D.

Todo esto se hace mediante la aplicación de un mapa UV. El proceso de creación de un UV consiste en la asignación de los píxeles de la imagen 2D a la superficie de los polígonos 3D. Esto se hace mediante un desplegado de los polígonos del modelo 3D en una superficie 2D. En este proceso se produce una rotura del mismo a través de unas costuras para conseguir situar los polígonos sobre una superficie bidimensional.

Una vez que el modelo está desenvuelto, se puede aplicar una textura en cada triángulo individualmente o a todos en conjunto para que cuando el modelo sea reconstruido, esta textura se aplique a modo de calco sobre el modelo 3D.

En muchos casos este mapa UV puede estar generado automáticamente, pero lo más normal es que a la hora del modelado se elijan las zonas de rotura del modelo para obtener así una mejor aplicación de las texturas (ejemplo rotura de cabeza)

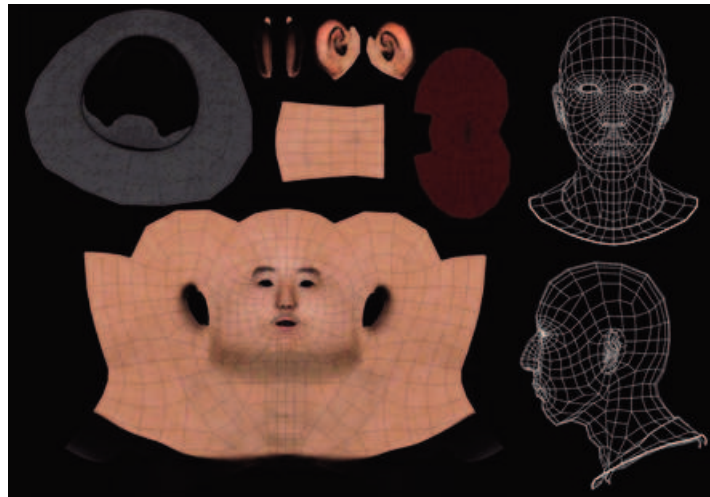


Figura 20. Textura y mapa UV.

2.2.2.1.3- Nivel de poligonación.

En un agente conversacional, al tratarse de un elemento animado, se debe tener especial cuidado con el número de polígonos que conforman el modelo final.

La calidad de visualización de un avatar 3D depende directamente de la cantidad y tamaño de los polígonos utilizados, siendo más detallado cuando más cantidad y más pequeños son.

Esto está enfrentado directamente con la capacidad de computación y la facilidad a la hora de animar un personaje.

Cuanto mayor sea la cantidad de polígonos, más compleja y costosa computacionalmente será llevar a cabo la animación.



Figura 21: Nivel de poligonación

2.2.2.2- Animación

En conjunto con el modelo 3D con apariencia humana va la animación del mismo.

Su movimientos la sincronía de los mismos no harán sino incrementar más el realismo de la animación.

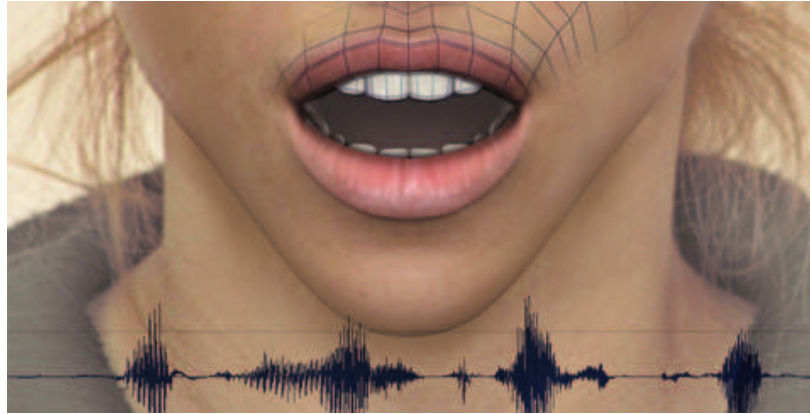


Figura 22. LipSync

En la animación de personajes y muy en concreto en aquellos con los que se quiere realizar una simulación de habla o conversación se ha de tener en cuenta un concepto como es el Lip Sync o sincronía de labios.

2.2.2.2.1-Lip Sync?

El lip sync es el concepto que hace referencia a la sincronía de los labios y si se quiere, de la cara en general, con el audio que acompaña la animación.

Se trata de una técnica cuyos conceptos básicos nos van a permitir realizar un acercamiento a las técnicas de animación de personajes y más en concreto a las referentes a la animación facial.

No existen las letras en los diálogos.

A la hora de realizar una animación que acompaña a un diálogo, lo primero que tenemos que tener en cuenta y debemos observar es que un sonido no es una letra, y por lo tanto una letra no implica directamente una posición concreta en la boca.

Las letras son símbolos representados en una página (con un sonido correspondiente arbitrariamente asignado) que cuando se colocan juntas forman palabras, que comunican un pensamiento. Pero las letras no están hechas para el habla. No hay letras en el dialogo, solo sonidos, y la figura (forma) que nuestra cara toma para hacer dichos sonidos.

Las formas de la boca para los sonidos deben ser animadas en contexto.

La forma de sonido que precede afecta a la forma de sonido actual. De igual modo, la siguiente forma de sonido es anticipada por la forma de sonido actual.

Así las formas deben presentarse todas en contexto con la forma/sonido que las precede y la que la sigue. Es animar el “flujo” de formas que son necesarias para representar los sonidos dentro de lo que se esta comunicando.

La mejor estrategia es interpretar el diálogo, capturar los elementos esenciales de la comunicación como fueron grabados en la pista de sonido.

Interpretar la animación como un todo.

Se deben interpretar las formas principales dejando un poco de lado los pequeños detalles.

Una animación donde las aperturas de boca llevan una relación correcta con el audio es una animación, que aun sin grandes detalles parece sincronizada, cumpliendo totalmente el objetivo.

Captar las poses abiertas y cerradas de forma correcta y construye a partir de ellas.

Gran parte de la sincronización labial reside en colocar correctamente las formas abiertas y cerradas dentro de la animación. El resto, lo hace la conjunción del audio con las transiciones que se realizan entre los momentos claves de la animación.

Elementos a tener en cuenta

- 1) Realizar una transición suave entre formas abiertas y cerradas.
- 2) No mantener las formas de sonido estáticas. Una cara esta en constante movimiento aunque no esté realizando sonidos.
- 3) Anticipar las formas de los fonemas a que se escuche finalmente el sonido. Los humanos vemos las cosas más rápidas que lo que escuchamos, por lo que tomamos las referencias de la forma antes que el sonido.
- 4) Rompe los ángulos y simetría de la boca. Podemos hablar y sonreír, hablar y bostezar, hablar y estar serios, etc.
- 5) Los dientes superiores no se mueven. Están unidos al cráneo.
- 6) No olvidar el movimiento de las mejillas y nariz cuando se adoptan los movimiento de los fonemas. Las mejillas y la nariz son estupendos conectores en la animación facial, cruzando el puente desde la animación de la boca a los ojos y a la animación de las cejas. Manteniendo la nariz y las mejillas en acción tú unes la cara del personaje, creando un personaje mucho más creíble el cual puede actuar.

2.2.2.2-Fonemas y expresiones faciales:

Fonemas

Existen una serie de fonemas que debemos tener en cuenta a la hora de hacer una sincronización labial.

Cada fonema tendrá una posición diferente de labios, dientes y lengua.

Hay que tener en cuenta que son formas de referencia y en cada animación o personaje puede haber pequeñas modificaciones.

Fonema A-E:

Cara, reto, acto, plato, escena



Fonema I:

Vivo, frio, pirado, higo.



Fonema O:

Olor, odio, loco, exotico



Fonema U.

Funda, universo, tu



Fonema C, CH, G, J, K, LL

Casa, chapa, jirafa



Fonema F:

Fonema, feo, fantástico



Fonema: C, Z:

Cero, caza, cepo, zoo



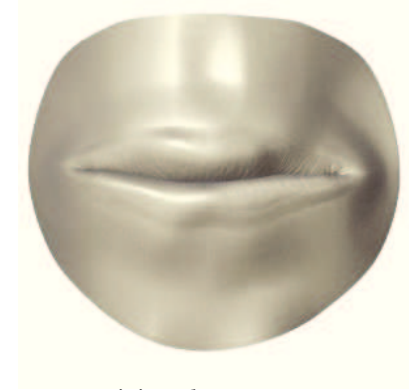
Fonema L:

Elección, alegrar, Lucia



Fonema, M, B, P:

Mar, bien, perro



Fonema: G, W:

Guapa, whisky



Resto de fonemas y posición de transición:



Figura 23. Fonemas

Expresiones faciales

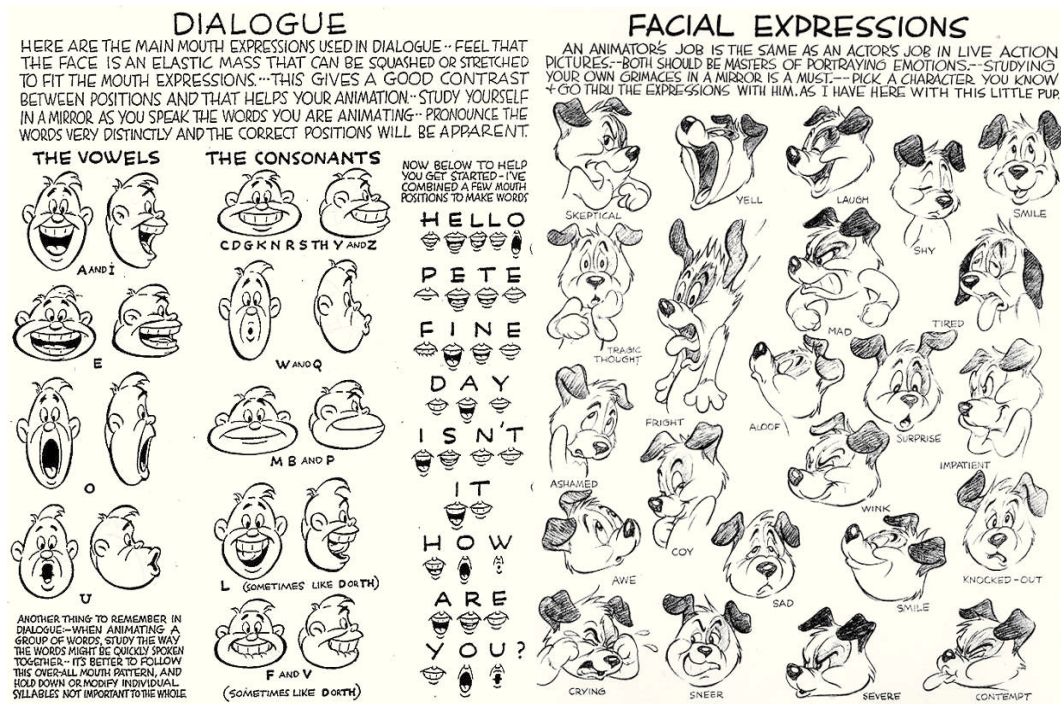


Figura 24. Ejemplos de expresiones en animación 2D.

Durante la animación de un personaje con el cual se busca simular una conversación se ha de tener en cuenta que el agente conversacional a de ser capaz de adoptar una serie de expresiones faciales que permiten comunicar en su totalidad el mensaje pretendido.

Es por esto que se han de tener en cuenta ejemplos de expresión facial tales como: Sorpresa, sospecha, duda, cansancio, indiferencia...

En caso de necesitar una expresión en concreto, la mejor forma de obtenerla es observar cuales son los movimientos que se producen en la cara de un ser humano real l cuando la realiza.



Figura 25. Ejemplos de expresiones

2.2.2.2.3 Técnicas de animación

2.2.2.2.3.1-Morphing.

Comúnmente se conoce con el nombre Morphing a una técnica de manipulación de imágenes digitales que permite obtener una animación -transición controlada- entre dos imágenes diferentes.



Figura 26. Ejemplo de morphing 2D

Si se traslada el concepto de morphing al modelado 3D es la transformación de un objeto tridimensional en otro. A nivel de objetos se suele denominar “deformación”, pero no deja de ser un mismo concepto de “cambio a un nuevo estado de forma”.

En animación el concepto de morphing hace referencia a las posibles deformaciones que sufre un objeto o modelo 3d hasta alcanzar la deformación total de un modelo con el mismo número de polígonos, que se ha señalado como objetivo.



Figura 27: Morphing 3D

Es decir, teniendo dos modelos iguales tanto en tamaño como en número de polígonos, podemos, modificando uno de ellos que el otro adopte todas las deformaciones intermedias hasta llegar a la que es su objetivo.

Esta técnica nos permite, teniendo una malla base, aplicar toda una serie de deformaciones objetivo como expresiones faciales o fonemas obteniendo así también todas las posiciones intermedias y negativas.

En el morphing, las deformaciones son combinables ya que se aplican de manera absoluta sobre la posición y tamaño de los polígonos permitiendo que se solapen de manera eficaz.

Es el desarrollo de estas deformaciones en un espacio de tiempo lo que nos permite desarrollar una animación.

Técnica

El “objetivo morph” es una versión transformada de una forma. Cuando se aplica a una cara humana, por ejemplo, la cabeza se modela primero con una expresión neutra y una se crea una deformación de destino a continuación, para cada otra expresión.



Figura 28. Ejemplos de morph

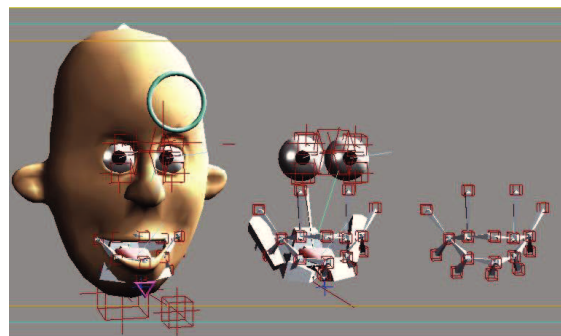


Figura 29. Animación facial con huesos

Así pues, cuando la cara es animada pueden aplicársele varios morph para que haga transiciones entre las expresiones. Ejemplos típicos de objetivos morph utilizados en la animación facial es una boca sonriente, un párpado cerrado o una ceja levantada, pero la técnica también se puede utilizar para una animación facial completa con LipSync y expresiones complejas.

Ventajas e inconvenientes

Hay ventajas en el uso de la animación morph objetivo sobre la animación de esqueleto o huesos. A la hora de desarrollar una expresión se tiene más control sobre la posición final de los elementos de la malla cuando estos son modificados manualmente frente a la animación con huesos que funciona con zonas y gradientes de aplicación de deformación.

Esta técnica es muy útil para la animación de tela, piel, y las expresiones faciales, ya que esas deformaciones específicas son difíciles de alcanzar con la animación con huesos.

Sin embargo, también hay desventajas. Al tener que deformar la malla polígono a polígono es generalmente mucho más costosa que la animación con esqueletos de huesos. Por esta razón, en muchas ocasiones se hace una utilización combinada de ambos.

Un caso de uso combinado en animación facial sería el caso en el que para los labios, cara, y ojos se utilizan objetivos morph mientras que para la lengua, que necesita un movimiento más libre se utilizan huesos.

2.2.2.3.1-Animación con huesos

Es una técnica de animación que permite, mediante la construcción de una cadena de ‘huesos’ dentro del modelo 3D modificar la posición de la malla.

Cada hueso tiene una posición tridimensional (que incluye su posición, escala y orientación), que puede ser modificada mediante una serie de manejadores.

A medida que el objeto 3D es animado, los huesos modifican la posición y deformación de malla. Cada hueso en el esqueleto está asociado con alguna porción de la representación visual del modelo.

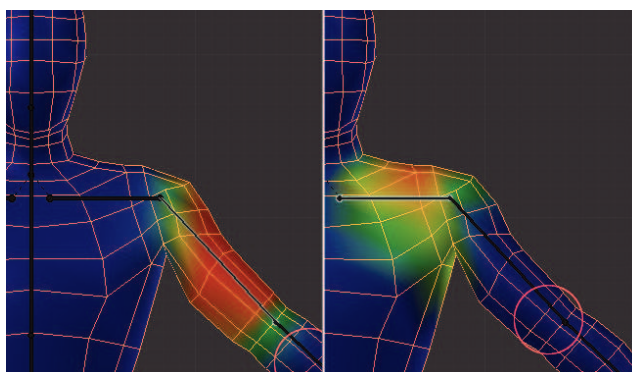


Figura 30. Animación con huesos.

En el caso más común de una malla poligonal, el hueso está asociado con un grupo de vértices, por ejemplo, en un modelo de un ser humano, el hueso ‘muslo’ estaría asociado con los vértices que constituyen los polígonos en el modelo de muslo. Las porciones de la piel del personaje normalmente se pueden asociar con múltiples huesos, cada una de ellas con un factor de escala llamado peso de vértice.

El movimiento de la piel cerca de las articulaciones de dos huesos, por lo tanto, puede estar influenciada por los dos huesos.

Por tanto para el cálculo de la posición final del vértice, cada transformación de hueso se aplica a la posición de vértice, escalado por su correspondiente peso.

Ventajas:

- Cada hueso representa a un grupo de vértices lo que permite una mejor comprensión del modelo.
- El animador controla un menor número de características del modelo ya que solo debe encargarse del movimiento de los huesos.

Una animación puede ser definida por los movimientos simples de los huesos, en lugar de vértice por vértice como en el caso de los morphers.

Inconvenientes:

- Puede no proporcionar un movimiento realista de la piel o músculo de un modelo orgánico.
- Exige de un gran trabajo previo de pesado de vértices. Se trata de una animación menos directa.

2.2.3- Retos y objetivos de las técnicas

El conjunto de todas estas técnicas reunidas alrededor de un agente conversacional tienen una serie de retos y objetivos a cumplir :

-Ser capaz de reproducir una sensación de ambiente o relación social . Hacer creíble la relación del usuario con el ECA. Estar lo más cercano posible a la verdadera dinámica social.

-Contar con credibilidad en sus movimientos y expresiones.

-Ajuste de simultaneidad entre expresiones, faciales gestuales corporales con el tiempo real así como con la comunicación verbal entre usuario-máquina.

Por lo tanto, en la búsqueda de esa credibilidad, los gestos y el lenguaje no consiguen su efecto si realmente el agente conversacional no tiene los rasgos y formas que lo puedan caracterizar al agente conversacional como humano.

Es aquí donde la generación del modelo 3D del ECA tome una gran importancia.

Para que la relación humano-interfaz humana 3D sea completa, debe tenerse en cuenta que el modelo 3D debe ser lo más fiel posible al modelo real o por el contrario, distar mucho del mismo.

2.3- Uncanny Valley⁸

La hipótesis del valle inexplicable en el campo de la robótica y la animación de gráficos por ordenador hace referencia a la relación que existe entre el realismo de un modelo gráfico o robótico que representa a un humano real y la reacción que esta produce en el observador.

El término fue acuñado por el profesor experto en robótica, Masahiro Mori, en 1970. La hipótesis afirma que cuanto más se parece un robot a un humano tanto en su apariencia como en sus movimientos más empatía genera en los humanos.

Aun así, existe un momento en que se entra en un “Valle Inexplicable” en el que el robot o avatar genera repulsión en el observador (figura 31).

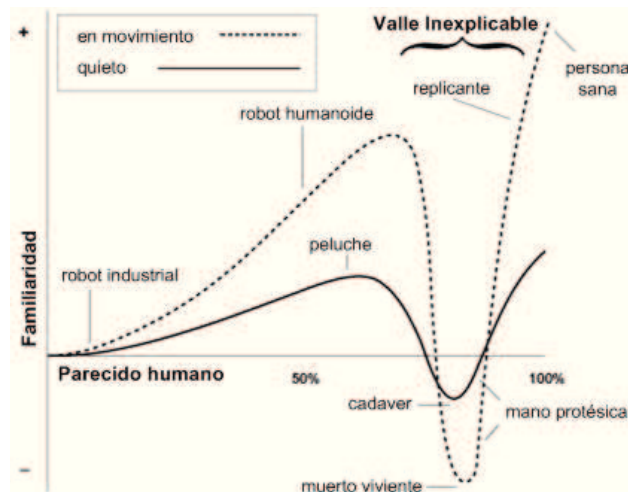


Figura 31: Gráfica valle desconocido.

Como ejemplo podemos nombrar un peluche con forma de oso, tiene unas características muy alejadas de un humano real y no corre riesgo de entrar en el valle y por lo tanto resulta muy agradable y simpático. Por el contrario un peluche que sea muy realista en su forma humana puede llegar a producir un sentimiento de rechazo al observador y sentimientos enfrentados.

No existe una teoría científica que respalde con experimentos la hipótesis, pero algunos estudios sugieren que la repulsión ante un elemento humanoide muy realista pero con ciertos defectos parte de la base de que somos reacios a mantener contacto con entes de apariencia enferma o que pueden no ser idóneos genéticamente.

Algunos robots humanoides, en la búsqueda de un realismo mayor han caído en el valle de una forma que los hace parecer enfermos, muertos o simplemente, no humanos.



Figura 32. Robots humanoides

2.3.1-Hipótesis

La hipótesis formulada por el experto japonés expone que la apariencia humana de un robot está ligada íntimamente con la respuesta que esta genera en el observador. La respuesta según avanza la similitud entre robot y humano es positiva hasta llegar a un momento en el cual los sentimientos que levanta este robot son de repugnancia y rechazo.

Sin embargo, si la apariencia sigue avanzando en la similitud con humanos conforme más difícil es apreciar las diferencias estos niveles de aceptación se van recuperando hasta retomar los anteriores niveles de empatía.

Es a este bache al que se conoce como valle inexplicable se produce tanto con la apariencia como con los comportamientos casi humanos.

Enmarcando todo esto dentro del mundo de los gráficos por ordenador y en concreto de los ECAS podemos hablar de que el modelo 3D de un humano que busca tener un gran nivel de similitud debe tener cuidado de que sus formas, proporciones y movimientos no caigan en el Uncanny Valley para evitar producir una sensación de rechazo en el espectador.

El observador centra su atención en los pequeños detalles que permiten diferenciar al robot y son estos, los que frente al resto del humanoide indistinguible, lo que nos produce un sentimiento de rechazo. Es el enfrentamiento entre lo humano y lo no humano.

2.3.2-Explicación

Se puede explicar el fenómeno diciendo que una entidad muy diferente al ser humano resalta las pocas cualidades que comparte con él.

Por otro lado, si la entidad es casi humana, lo que aparecerá resaltado serán los elementos que lo diferencian del ser humano acrecentando el sentimiento de que se trata de un elemento extraño y por lo tanto generando un sentimiento de repulsión.

Las anomalías que apreciamos en elementos humanoides con una gran índice de realismo están ligados a las diferencias que notamos entre humanos vivos y sanos y los que están enfermos o muertos y que por lo tanto causan la misma sensación de rechazo.

Anomalías en los movimientos, formas y expresiones de la entidad humanoide son los que nos sitúan en la predisposición de rechazarlos.

Las anomalías de comportamiento y sincronización motriz son también indicativos de enfermedad, y desórdenes mentales o neurológicos y de nuevo evocan sentimientos negativos acertados.

Algunos detalles a tener en cuenta a la hora de animar un modelo humano que hacen referencia al valle desconocido son:

- El modelo 3d debe estar en constante movimiento: Las personas, por muy quietas que intentemos estar, siempre tenemos una serie de movimientos involuntarios. La cara es un elemento crítico en el que por ejemplo se debe tener en cuenta las pequeñas variaciones en las mejillas que produce la respiración.

- Color: la piel humana es un elemento translúcido y como tal tiene una serie de rebotes internos de luz y variaciones de color según la zona del cuerpo en la que se sitúe.

- Formas: las proporciones humanas son un elemento importantísimo, aunque en cada individuo puedan ser diferentes, siempre guardan una relación entre ellas.

2.3.2.1-Errores

En el mercado existen claros ejemplos de películas, videojuegos o animaciones que en la búsqueda de un gran realismo en sus personajes han caído irremediabilmente en el valle desconocido.

Ejemplos como Tom Hanks y sus representación en la película The Polar express son claros referentes de personajes que han incurrido en el mundo del Uncanny Valley.



Figura 33. Tom Hanks en Polar Express

2.3.2.2-Exitos

En el otro extremo existen modelos 3D y animaciones que han conseguido solventar el problema del valle desconocido y situarse por tanto en un nivel de realismo tal que hace muy difícil la diferenciación entre personaje real y virtual.

Es el caso por ejemplo del modelo 3D de la actriz coreana Song hye kyo⁹.

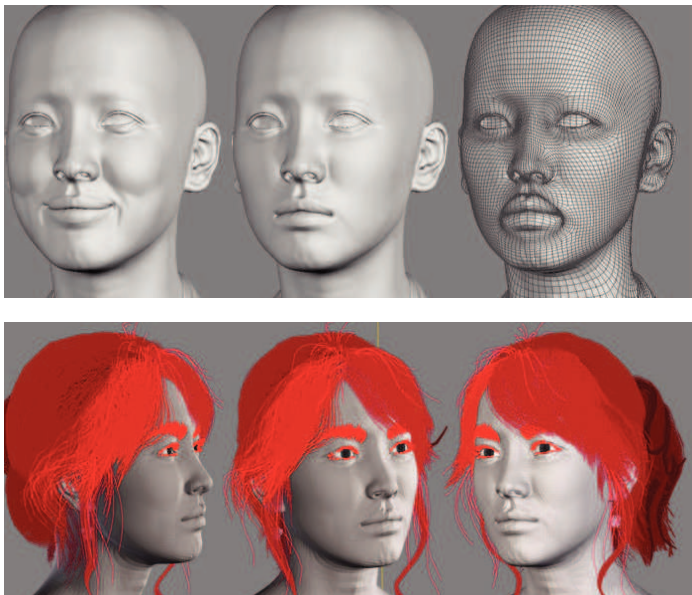


Figura 34. 3D de la actriz Coreana Song Hye Kyo

2.3.3-Cómo evitarlo

Uno de los métodos utilizados para evitar el valle inexplicable es la reconstrucción 3D a través de fotografías.

2.3.3.1-Ejemplo Emily¹⁰

Uno de los más claros ejemplos de animación facial que ha evitado a todos los niveles el valle inexplicable es el desarrollo de un actor virtual fotorealista como es Emily.

“Es absolutamente increíble. Yo soy uno de los críticos más duros de la captura de la cara, e incluso tengo que admitir que estos chicos han dado en el clavo. Esta es la primera secuencia animada humano virtual que evita por completo todas mis advertencias subconscientes. Me da la sensación de Emily como persona. Toda la sutileza está ahí. Esto no es un trabajo publicidad, es la cosa real ... Yo declaro oficialmente que Image Metrics finalmente ha construido un puente sobre el valle inquietante, y trasladado al otro lado.”

Peter Plantec, “ El Ojo Digital: Image Metrics intentos de Salto del valle inquietante “, VFXWorld, 07 de agosto 2008

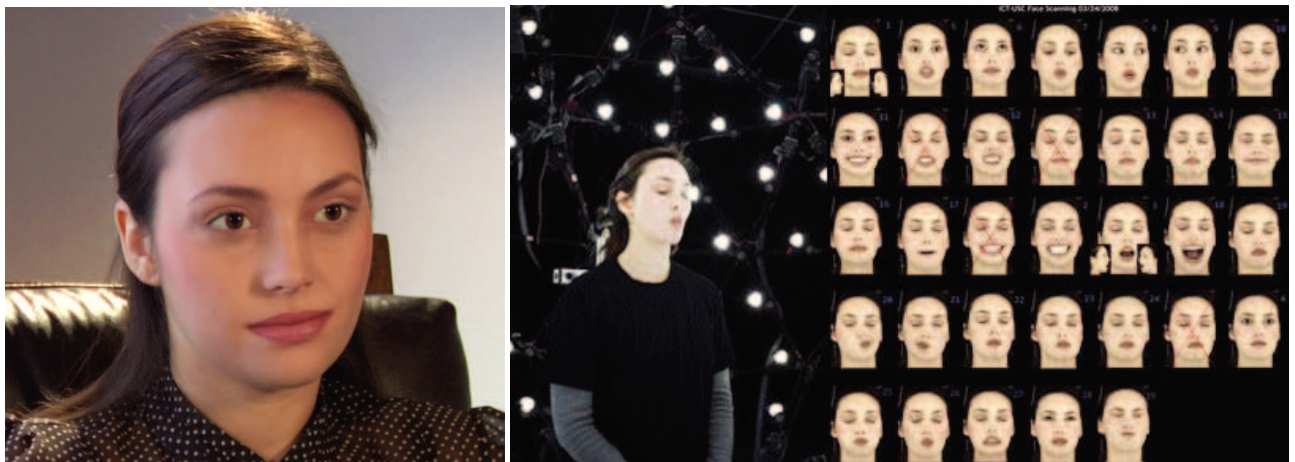


Figura 35.Emily y las fotografías captadas

Este proyecto buscaba generar un modelo 3D fotorealista mediante la captura de las expresiones de una actriz real en un entorno controlado y mediante el uso de múltiples fotografías.

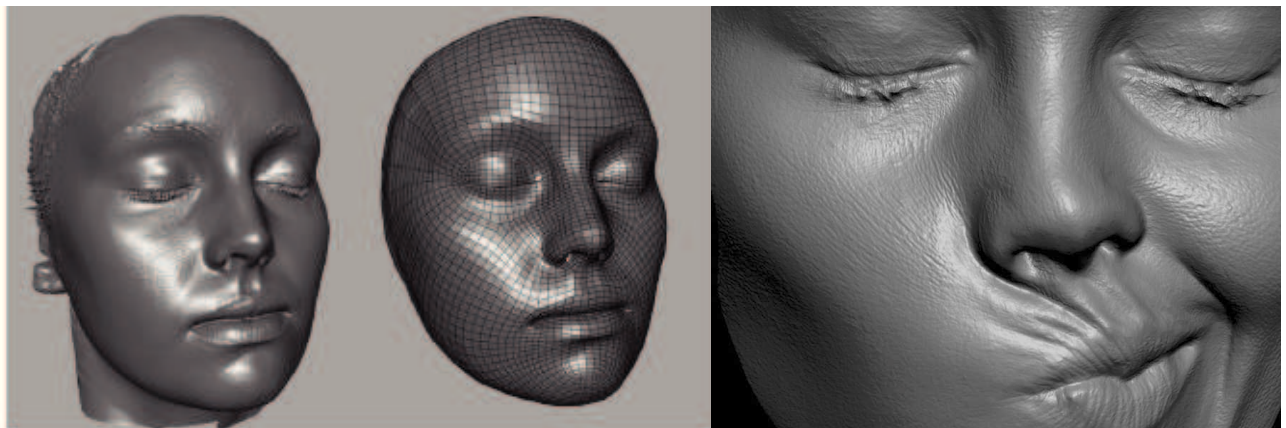


Figura 36. Malla base de Emily



Figura 37. Expresiones Emily

Con estas fotografías buscaban captar la totalidad de los tonos de piel de la actriz en diferentes estados lumínicos así como algunas de las expresiones más características.

Posteriormente, generando una geometría de alta resolución y otra de baja resolución realizan una animación facial mediante el uso de morphers.

La animación de la actriz generada por ordenador, se incrusta en un vídeo real tomando como referencia el marco de la cara de la actriz real previamente grabada.

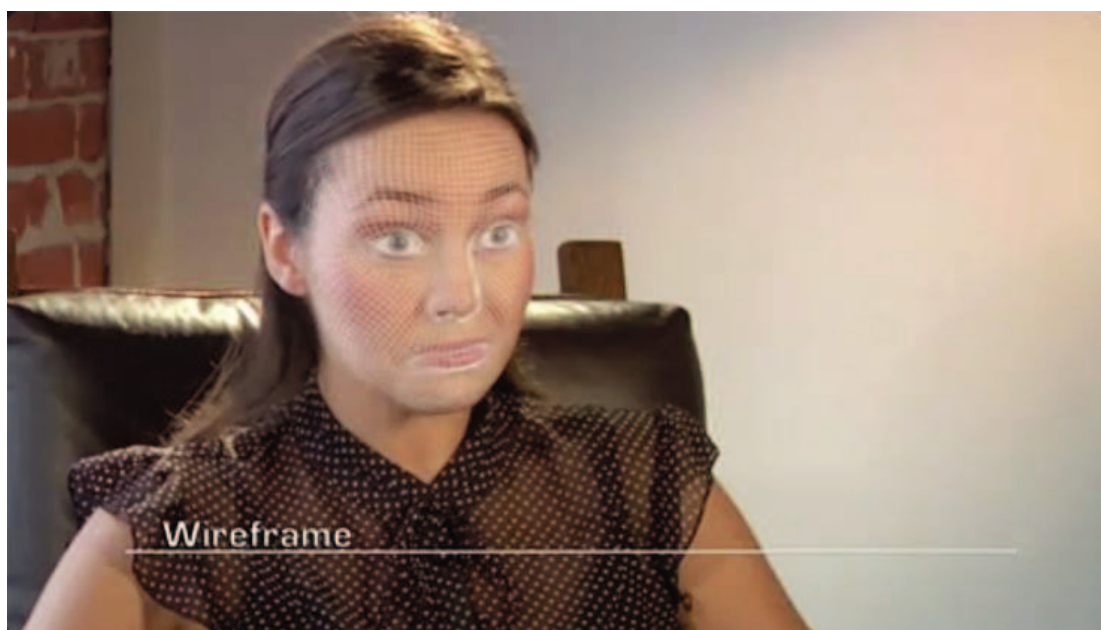


Figura 38. Fotograma animación Emily

Todo este conjunto, la información captada del modelo físico, su forma tridimensional, su color de piel, las deformaciones de la cara en cada una de las expresiones es lo que finalmente hacen que esta animación sea un ejemplo de buenas prácticas y un referente en cuanto a cómo hacer una animación facial que busca solventar el valle inexplicable.

3-Software utilizado.

Durante el proyecto se ha hecho uso de varios tipos de software, los cuales se describen brevemente a continuación. Los detalles sobre sus herramientas y características se encuentran en el anexo I de este proyecto.

3.1-123D Catch®¹¹.

Existen diferentes aplicaciones con la capacidad de generar modelos 3D a través de múltiples imágenes. Durante el desarrollo de este proyecto se ha utilizado 123D Catch.

Se trata de un programa en fase Beta, que enmarcado dentro de la suite 123d App de Autodesk, permite a cualquier usuario obtener un modelo 3D mediante una serie de fotografías de una escena u objeto.

Se trata de un programa preparado para un usuario con poco nivel técnico así que con unos conocimientos básicos sobre como tomar las fotografías se puede conseguir un modelo 3D con un buen nivel de detalle.

Se ha escogido este programa por ser un programa de descarga gratuita, su sencillez de uso y su gran capacidad de procesamiento.

Durante el proyecto permitirá generar reconstrucciones 3D haciendo uso de las fotografías de los modelos escogidos exportandolo todo después al programa de edición.

3.2-Zbrush®¹²

ZBrush es un software de modelado 3d, escultura y pintura digital que constituye una nueva forma de afrontar los trabajos de modelado debido a su característica forma de trabajo.

ZBrush es un programa con el que es posible esculpir detallados modelos de un modo semejante a pintar o trabajar la arcilla y es lo que ha popularizado entre los artistas 3D de las industrias del cine, videojuego e ilustración.

El software se basa en el concepto de esculpir pintando sobre un objeto 3D y la posibilidad de aumentar cómodamente la cantidad de polígonos según vamos añadiendo detalle. Tiene la capacidad de trabajar con modelos de varios millones de polígonos, cosa impensable en otros software similares.

En el proyecto es utilizado como el principal programa de edición y procesamiento de los modelos generados por la reconstrucción 3D.

Se ha escogido este programa por ser capaz de manejar mallas de gran cantidad de polígonos con mucha fluidez y por su característico flujo de trabajo.

3.3-3ds Max®

3ds max es un software de modelado y animación 3D comercializado por la empresa Autodesk. Se trata de un programa generalista que permite desarrollar todas las tareas necesarias para crear modelos, texturizarlos, animarlos y renderizarlos.

Durante el desarrollo del proyecto ha permitido unir en un único modelo los diferentes componentes del agente conversacional generados así como su texturización y edición.

Se ha escogido este software por ser uno de los programas con más herramientas y posibilidades del mercado a la hora de resolver las necesidades del proyecto.

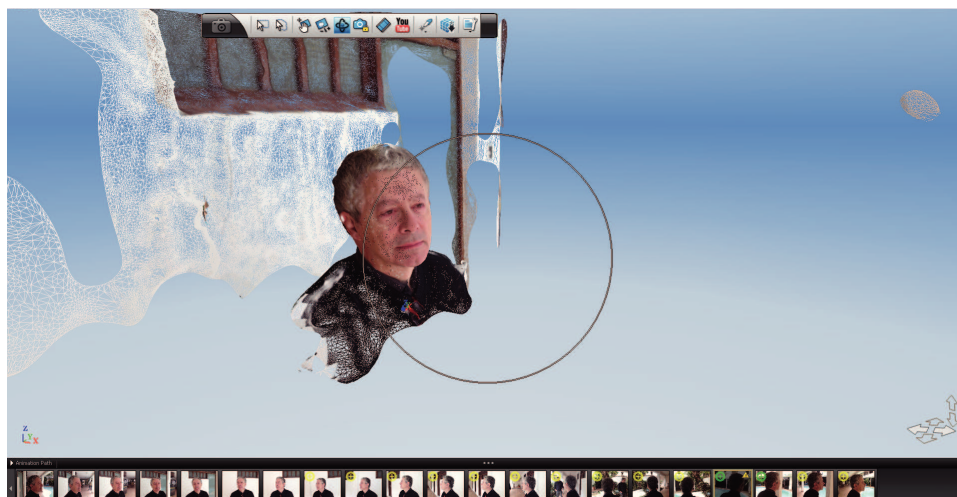


Figura 39. Ejemplo de visualización de un modelo en 123D Catch

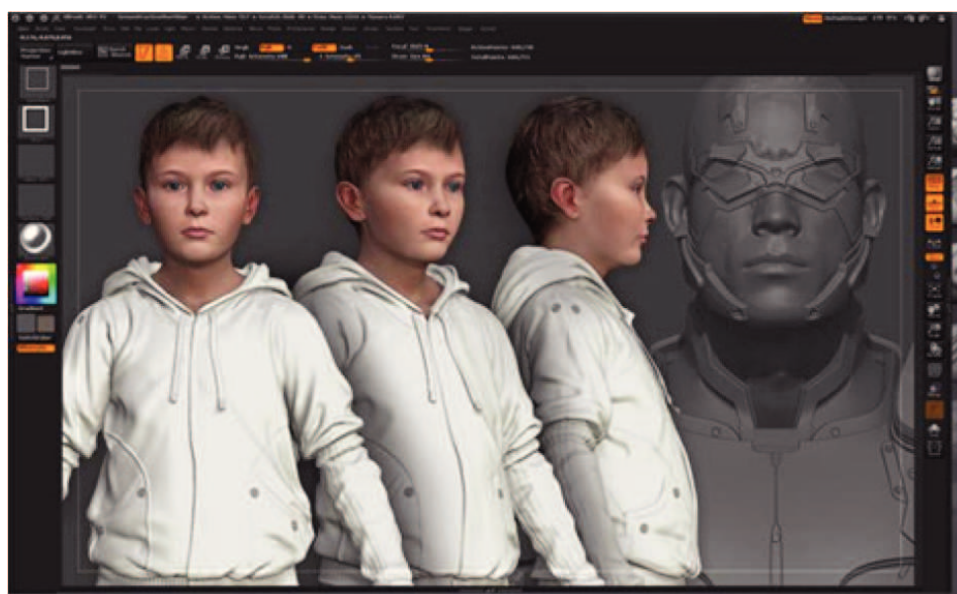


Figura 40. Ejemplo modelo complejo en Zbrush

3.4- Vray®

V-Ray es el motor de render que funciona como plugin de programas como 3ds Max y que permite generar imágenes de los modelos 3D generados en dicho software.

Se ha utilizado en conjunto con 3ds Max para generar la visualización del modelo final del agente conversacional.

3.5- Photoshop®

Photoshop es un software de creación y edición de imágenes comercializado por la empresa Adobe. Se ha escogido este programa por estar ya familiarizado con su uso y características generales.

Durante el proyecto se ha utilizado para la edición y corrección de las fotografías utilizadas en la reconstrucción y para la generación y edición de mapas de textura para los modelos.

4-Contruccion de modelo validos para ACH

4.1 Presentación

En el desarrollo de este proyecto se ha llevado a cabo el desarrollo de 2 modelos 3D enfocados a generar las imágenes básicas necesarias para animar un agente conversacional humano.

4.1.1 Modelo de un hombre.

Se trata de un modelo humano de un hombre cuyas fotografías se realizaron con un iPad en un entorno de iluminación exterior no controlada.

Las imágenes, no fueron tomadas con el propósito de ser utilizadas en este proyecto y tenían como fin ser usadas en una prueba del software 123D Catch. Son por tanto un caso extremo de fotografías tomadas por un usuario no experimentado y como tal no han sido retocadas. Son un total de 21 fotografías, tomadas en torno al modelo humano a una altura constante.



Figura 41. Imágenes del modelo del hombre.

4.1.2- Modelo de una mujer.

En este caso se trata de una mujer a la que se le hicieron 56 fotografías a diferentes alturas y desde todos los ángulos.

Se realizaron en un entorno con luz natural controlada y con un apoyo de luz artificial. Fueron tomadas con una cámara reflex Olympus E-1 en formato RAW aunque finalmente se utilizarán convertidas en JPEG por ser el único formato admitido por 123D Catch.



Figura 42. Modelo Mujer

4.2-Procedimiento

A continuación se detalla el proceso seguido para generar las mallas 3D de ambos modelos y las tareas realizadas. Para mas detalles sobre el proceso seguido durante este desarrollo consultar anexo II.

4.2.1 Sesión fotográfica.

Procedimiento: (Apartado 1.1, anexo II)

-Toma de fotografías: Unicamente se pudo realizar sesión fotográfica en un entorno controlado para el modelo de la mujer. El procedimiento utilizado responde a los directrices proporcionadas por la propia distribuidora del software y a los diferentes experimentos realizados.

-Conversión de las fotografías RAW-JPEG: Las fotografías fueron tomadas en formato RAW dada su mayor calidad, para ser procesadas por 123D catch deberán ser convertidas a formato JPEG.

-Parámetros de entrada

Una sesión fotográfica en entorno controlado y con equipo fotográfico profesional:

-Parámetros de salida:

Un conjunto de 56 imagenes en formato JPEG para el modelo de la mujer.

Uno de 21 fotografías en formato JPEG para el modelo del hombre



Figura43. Fotografías sesión fotográfica



Figura 44. Fotografías sesión fotográfica

4.2.2-123D Catch

Procedimiento: (Apartado 1.3, anexo II)

-Generación de modelo 3D:

Con las fotografías de los dos modelos se han generado las reconstrucciones 3D de ambas personas.

-Eliminación de elementos sobrantes de la escena y corrección de errores:

En las dos reconstrucciones aparecen elementos externos al modelo que se deben eliminar.

-Exportación modelo 3d:

El software permite exportar el modelo en formato Obj generando un archivo JPEG para la textura

-Parámetros de entrada

Conjunto de fotografías de ambos modelos, tanto del hombre como el de la mujer.

-Parámetros de salida:

Mallas 3D texturizadas en formato Obj y una textura en JPEG



Figura 45. Modelos 123d Catch y sus texturas

.2.3 Edición de mallas en Zbrush

Procedimiento: (Apartado 2.1, anexo II)

-Importación de las mallas de 123D Catch a Zbrush

Los modelos 3D obtenidos se deben importar a ZBrush para ser editados.

-Selección de zona de trabajo:

De ambos modelos seleccionaremos una parte del modelo para trabajar. En ambos casos se trata de la zona de la cara siendo lo menos importante el pelo y la zona de la nuca por tener menos detalle.

-Adecuación de la malla al programa.

Zbrush trabaja con malla basada en cuadrados mientras que la importada está basada en triángulos.

-Corrección de errores en la geometría y el color.

Ambos modelos tienen una serie de errores comunes como son, la geometría de los ojos, las orejas, los labios y la nariz.

En el caso del modelo del hombre es especialmente complejo ya que las fotografías fueron tomadas con una postura determinada y ha tenido que ser corregida.

-Traspaso de información a un modelo animable. El modelo que finalmente animaremos debe tener unas características geométricas básicas como son las líneas de expresión y la simetría. Se deberá pasar la información del modelo escaneado al modelo animable.

-Generación de detalle. Cuando la información ha sido traspasada al modelo animable se puede generar detalles en la textura tales como imperfecciones en la piel o arrugas.

-Creación de mapas UV y mapas de textura.

Los modelos que se están editando en ZBrush no tienen mapa UV y por lo tanto deben ser generados para poder exportar las texturas del modelo a otro programa.

-Generación de morphers objetivo.

Para el desarrollo de los diferentes fonemas y expresiones del agente conversacional habrá que desarrollar las formas básicas de dichas expresiones. Estas formas son las que conoceremos como morphers objetivo.

Los morphers objetivo generados están clasificados de la siguiente manera:

- Movimiento de mandíbula: Tanto para fonemas abiertos como fonemas cerrados.
- Movimiento de parpados y cejas.
- Movimientos de la cabeza.

-Exportación de los morphers objetivo:

Estas formas deben ser exportadas en formato Obj para poder generar en 3Ds Max un modelo animable mediante los correspondientes modificadores de forma.

-Modelado de dientes y encías

Como el modelo va a realizar movimientos de vocalización, debe tener generada la geometría del interior de la boca. Estos elementos serán modelados en ZBrush partiendo de sus formas primitivas y haciendo uso de sus herramientas de modelado y pintura hasta conseguir un modelo y su textura.

-Parámetros de entrada

2 mallas de 123D Catch y sus texturas

-Parámetros de salida:

Modelos animables y mapas de texturas

Morphers objetivo de cada uno de los modelos

Modelos del interior de la boca

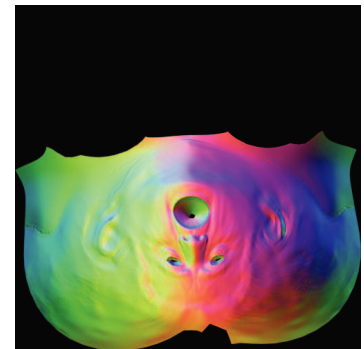
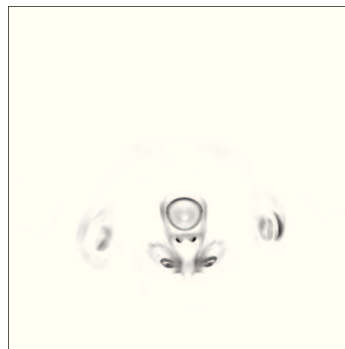
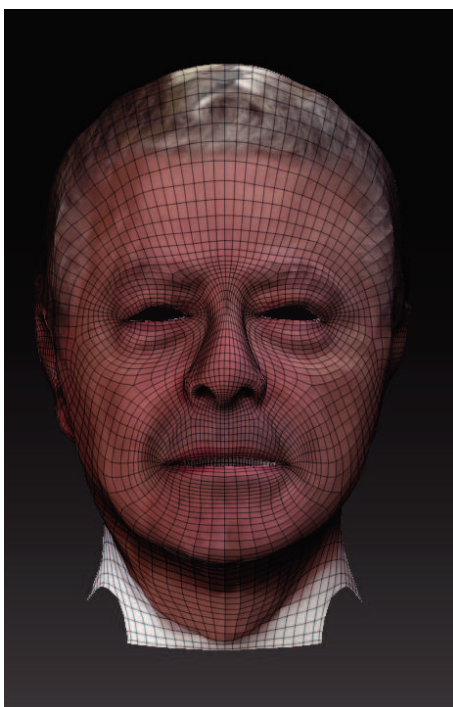


Figura 46. Modelo del hombre y mapas de textura

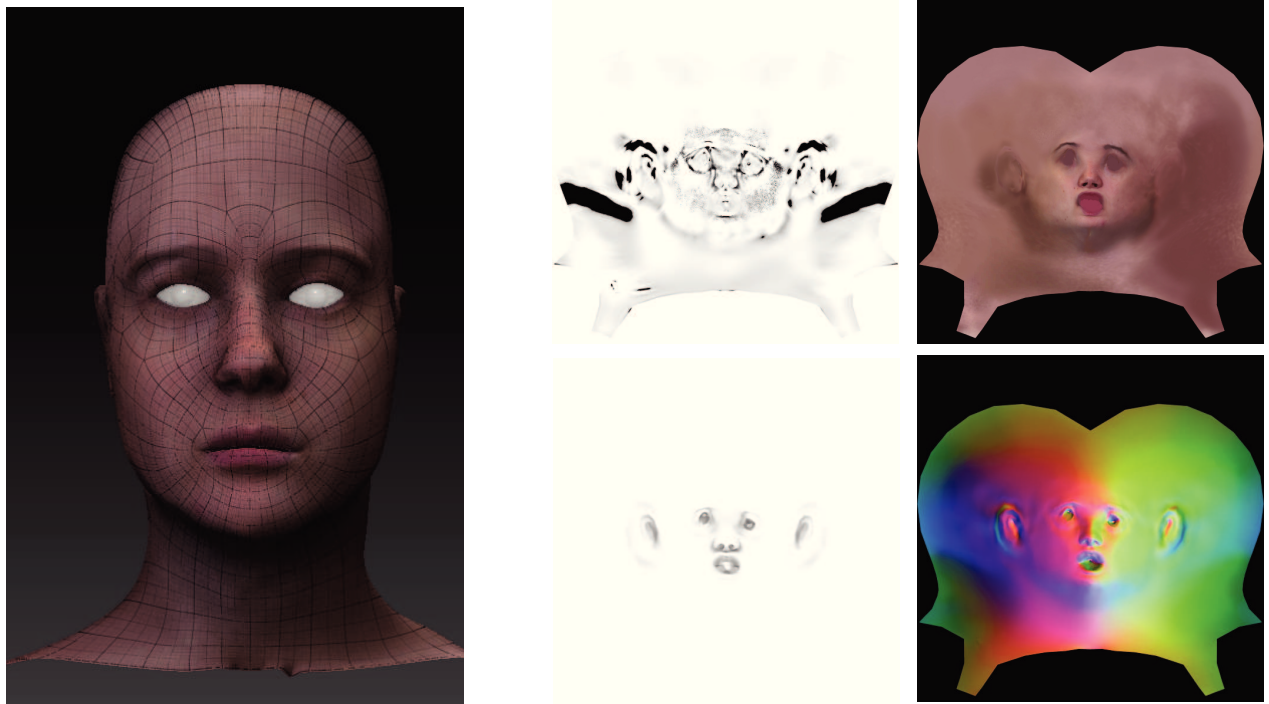


Figura 47. Modelo de la mujer y mapas de textura

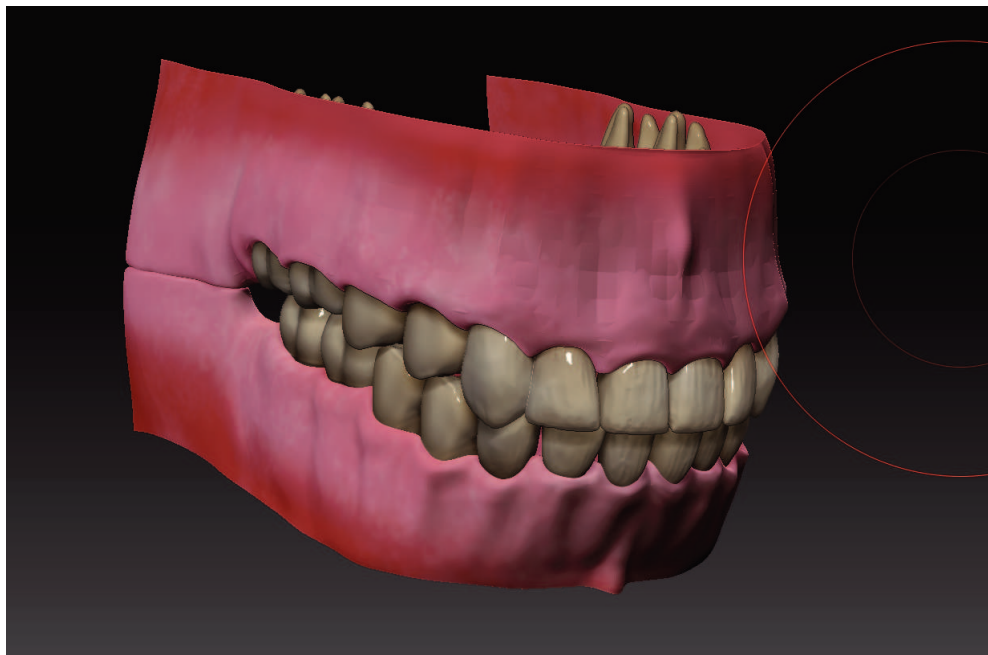


Figura 48. Modelo del interior de la boca

4.2.5-Vray

Para la visualización de las imágenes finales del modelo 3D se hace uso de un motor de render como V-Ray. Este motor permite generar una escena 3D en la que controlar parámetros de iluminación y color así como los de la cámara virtual que finalmente toma la imagen.

Procedimiento:

-Generación de la escena:

Se genera un fondo continuo que hará no tenga sombras tras el modelo.

-Iluminación:

Se ilumina la escena con 3 luces de V-Ray para la correcta visualización de la textura y la geometría.

- Render:

Finalmente se lanza el motor de render generando las imágenes correspondientes.

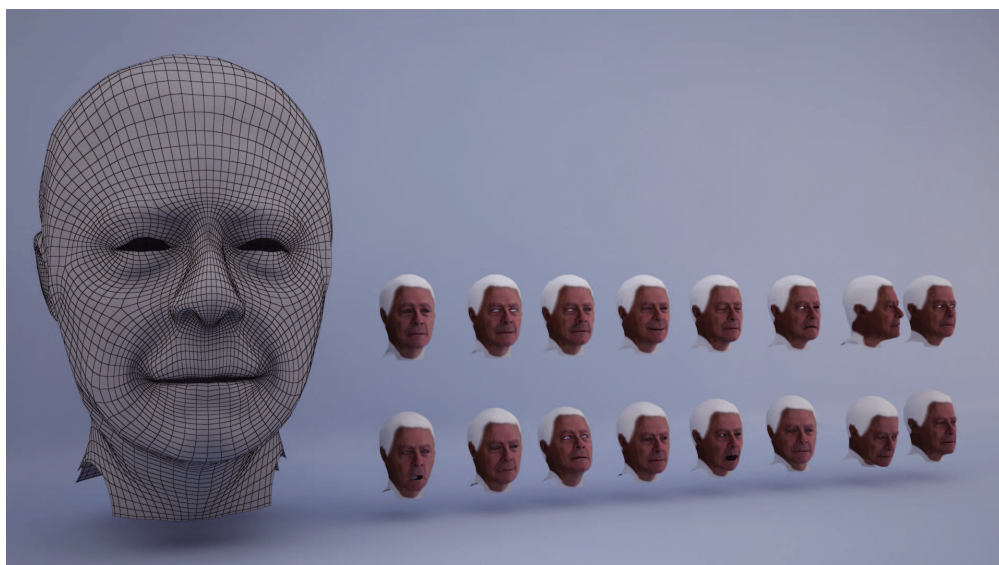
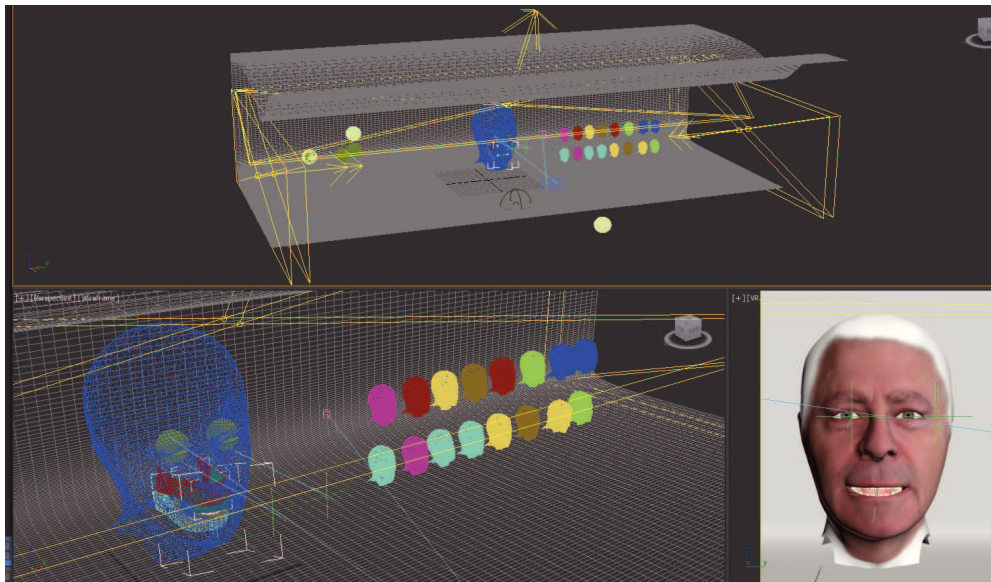


Figura 50 Modelo base con todos los morphers objetivo

5-Imágenes finales

A continuación se presentan las imágenes finales obtenidas mediante cada uno de los modelos 3D generados así como todos sus objetivos de forma.

5.1-Modelo 3D del hombre

-Malla 3D de base

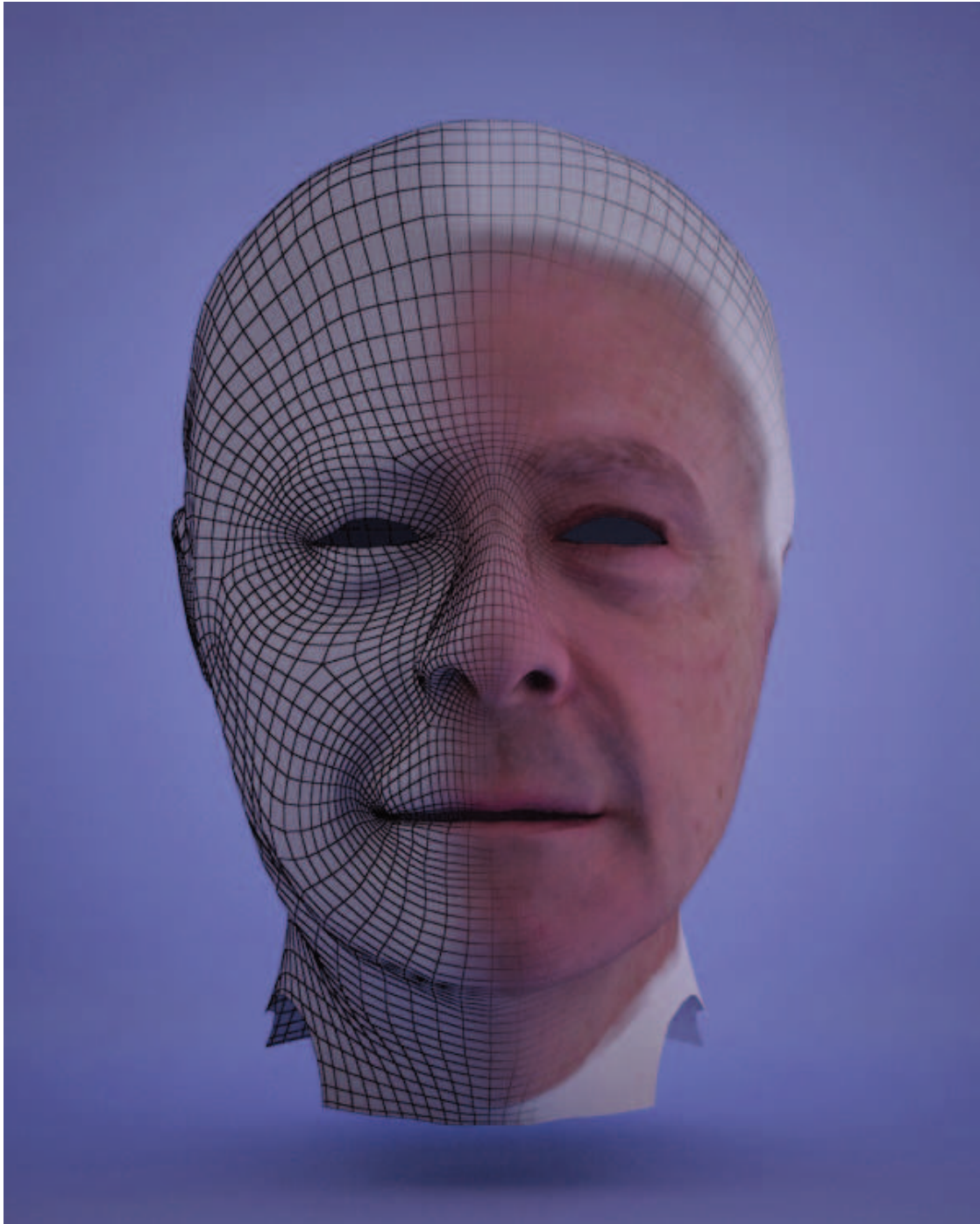


Figura 51. Modelo base hombre

-Movimientos de la cabeza



Figura 52. Mirando a la izquierda.

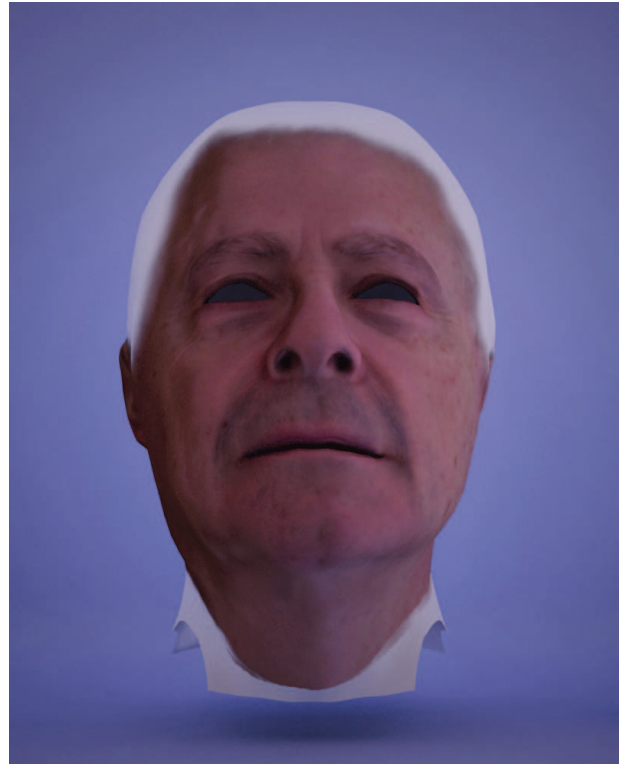


Figura 53. Mirando a arriba



Figura 54. Mirando abajo



Figura 55 Mirando a la derecha

-Movimientos de parpados y cejas



Figura 56. Parpadeo inferior



Figura 57. Cejas arqueadas.



Figura 58. Parpadeo normal



Figura 59. Cejas arqueadas

-Movimientos de mandíbula y labios



Figura 60. Fonema abierto

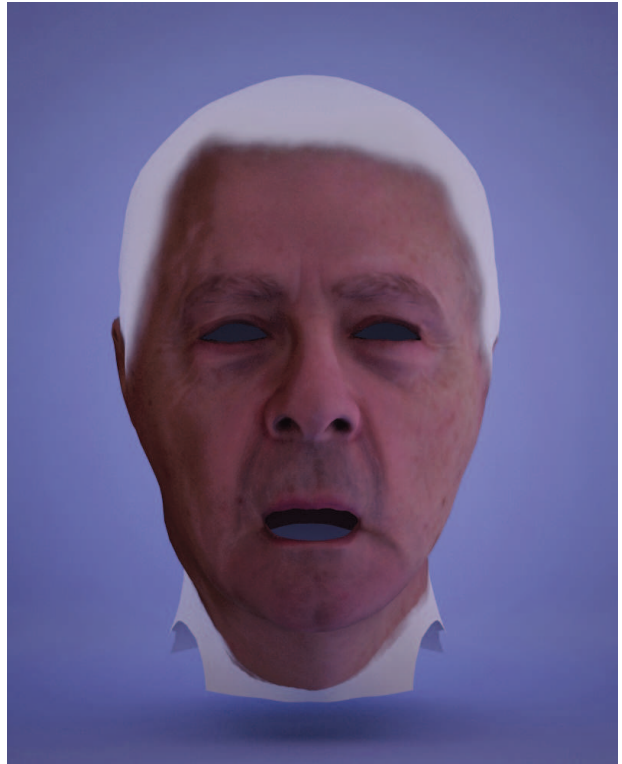


Figura 61. Fonema cerrado



Figura 62. Fonema F, labio inferior hacia atrás.



Figura 63. Sonrisa



Figura 64. Movimiento lateral de mandíbula



Figura 65. Movimiento lateral de mandíbula

5.2-Modelo 3D de la mujer

-Malla 3D de base



Figura 65- Malla base de la mujer

-Movimiento de la cabeza.



Figura 66. Giros de la cabeza



Figura 67. Ladeos de la cabeza



Figura 68. Mirada arriba y abajo

-Movimientos de labios y mandíbula



Figura 69. Fonemas abiertos y cerrados



Figura 70. Labios apretados y sonrisa

-Movimientos de cejas y parpados



Figura 71. Parpadeo normal



Figura 72. Ceño fruncido



Figura 73. Guiño

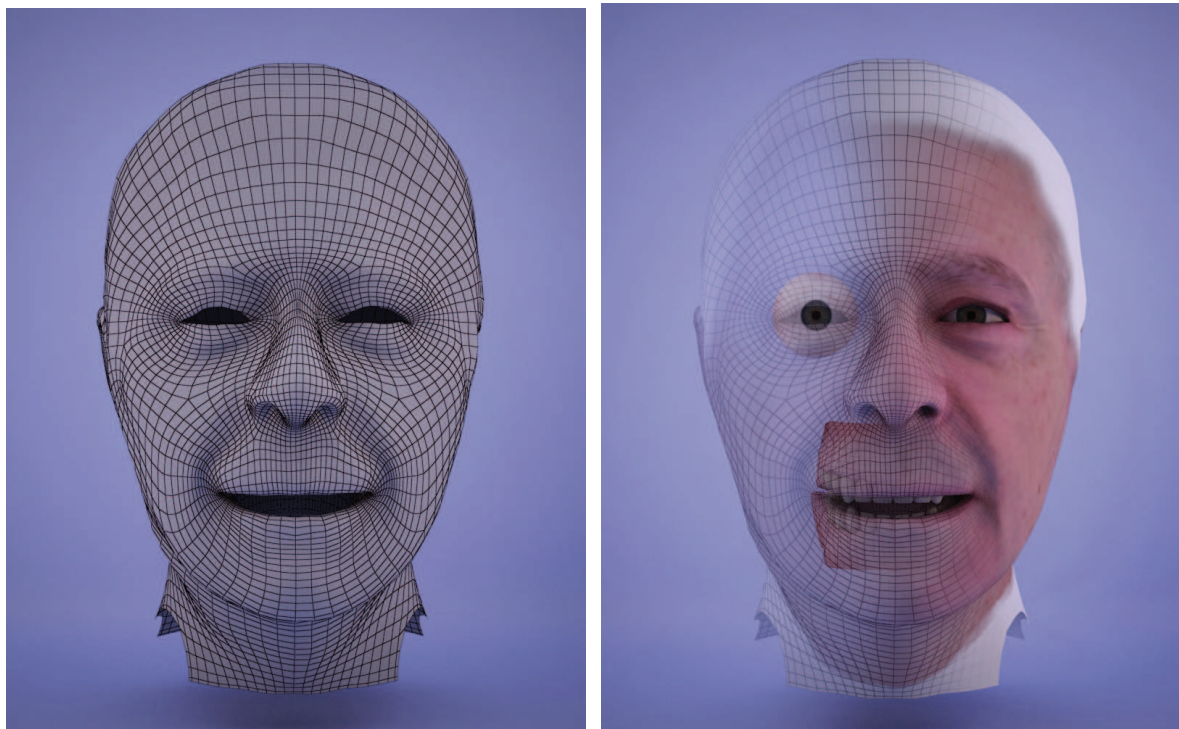


Figura 74. Wireframe del modelo y superposición de ojos, boca y piel.

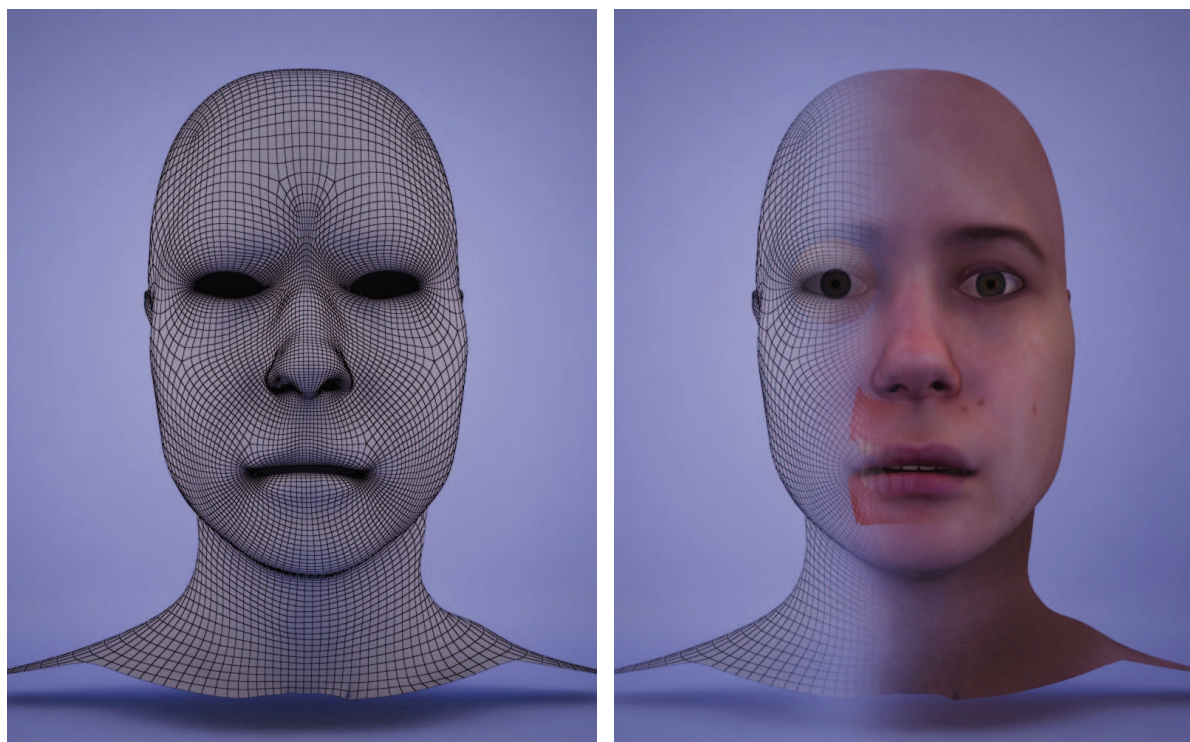


Figura 75. Wireframe del modelo y superposición de ojos, boca y piel.

5.3-Expresiones faciales



En esta posición de asustada se han utilizado:

- El movimiento de los ojos hacia una esquina de la escena.
- El movimiento de arqueado de cejas
- Apertura de la boca
- Movimiento hacia atrás y a un lado de la cabeza

Figura 76. Mirada de terror.



En esta posición de abatida se han utilizado:

- El movimiento de los ojos hacia la parte baja de la escena
- El movimiento negativa del morpher de la sonrisa
- El 50% del parpadeo normal.

Figura 77. Mirada de cansancio.



Figura 78. Carcajada

En esta posición de carcajada se han incluido los siguientes morphers:

- El parpadeo total.
- La apertura de la boca.
- La sonrisa y la media sonrisa.
- Arqueo de cejas.



Figura 79. Mirada sospechosa.

En esta posición de mirada sospechosa se han utilizado:

- El arqueado de cejas.
- 25% de parpadeo normal y 40% de parpadeo inferior
- La sonrisa y la media sonrisa.
- El movimiento de los ojos.

6- Conclusiones

Este proyecto busca analizar la problemáticas asociadas al desarrollo de un modelo 3D para un agente conversacional a través la técnica del modelado basado en imagen. Aprovechando la aparición de aplicaciones para usuarios capaces de ofrecer estas funcionalidades (123D Catch) y herramientas de refinamiento y modificación de mallas como Zbrush.

Los objetivos de este proyecto consisten en:

- Conocer las posibilidades de las tecnologías estudiadas.
- Utilizarlas para el desarrollo de un conjunto de visualizaciones básicas como en las que se basan aplicaciones que utilizan agentes conversacionales humanos.

Para obtener los objetivos del proyecto se han realizado las siguientes tareas

- Estudio de las técnicas y conceptos del modelado basado en imagen y las posibilidades para el desarrollo de un modelo para un agente conversacional.
- Definición y análisis de los elementos y posibles aplicaciones de los agentes conversacionales como elementos de interacción hombre-maquina.
- Estudio de las problemáticas asociadas al desarrollo de modelos de apariencia humana, como el Uncanny Valley, y el rechazo que estos pueden producir según aumenta su realismo.

Para el desarrollo de los modelos 3D de apariencia humana se ha seguido un procedimiento consistente en:

- Generar una reconstrucción 3D del modelo humano a través de fotografías mediante el uso de 123D Catch
- Someterlo a un proceso de edición y adaptación en Zbrush que permita hacer uso de la información obtenida en la reconstrucción.
- Desarrollo de fonemas y expresiones necesarias para un agente conversacional.
- Generación de mapas de textura y UV para los modelos.
- Generación de elementos 3D que no se habían obtenido en la reconstrucción 3D tales como los ojos y el interior de la boca.
- Unión de todos los elementos en 3Ds Max.

Como resultado final del proyecto se han obtenido dos mallas 3D con los modificadores necesarios para poder desarrollar un animación mediante morphing de fonemas y expresiones de un agente conversacional humano.

Este proyecto ha hecho uso de conocimientos adquiridos durante el Grado en Ingeniería de Diseño Industrial y Desarrollo de Producto, en concreto las asignaturas de interacción usuario-producto, fotografía o entornos 3D han aportado los elementos básicos para llevarlo a cabo.

Por otro lado ha sido necesario adquirir otros conocimientos como de anatomía humana, animación 2D y 3D así como los conocimientos para hacer uso de herramientas informáticas como 123D Catch, Zbrush, 3DsMax, Vray o Photoshop.

Las principales dificultades en el desarrollo del proyecto se han encontrado en el aprendizaje de las técnicas necesarias para conservar la información del modelo obtenido con las fotografías y hacer que esta fuera realmente útil. La edición en Zbrush para conseguirlo ha sido la fase mas laboriosa y que mas dificultades ha generado.

7- Trabajo futuro

De poder seguir desarrollando los modelos que se han obtenido durante este proyecto los esfuerzos deberían centrarse en maximizar el detalle obtenido en la primera reconstrucción en 123D Catch. Las fotografías deben estar aun mas controladas tanto en iluminación como en posición para poder conseguir minizar aun mas los errores.

Otro aspecto al que también se debería prestar atención sería el desarrollo de los detalles del modelo tales como el pelo y las pestañas que en este proyecto se han quedado sin profundizar. Son elementos que unidos al máximo detalle posible en el modelo de la cabeza y su textura marcaran una gran diferencia en cuanto a calidad se refiere.



Figura 80. Ambos modelos en reposo

Programación

	OCT.	NOV.	DIC.	ENE.	FEB	MAR.	ABR.	MAY.	JUN.	JUL.	AGO.	SEP.	OCT.	NOV.	DIC.
Estudio alternativas TFG															
Estudio de conceptos básicos															
Modelado basado en imágenes															
Uncanny valley															
Agentes conversacionales															
Aprendizaje 123D Catch															
Experimentos fotografías															
Aprendizaje Zbrush															
Edición y simplificación de modelos															
Creacion de morphers															
Aprendizaje 3dsMax															
Generación del modelo final															
Aprendizaje Vray															
Imágenes finales															
Generacion de documentacion															

	DURACIÓN EN SEMANAS	DURACIÓN EN HORAS	
Estudio alternativas TFG	1	30	
Estudio de conceptos básicos	3	50	
Modelado basado en imágenes	3	65	
Uncanny valley	2	45	
Agentes conversacionales	3	30	
Aprendizaje 123D Catch	4	55	
Experimentos fotografías	3	100	
Aprendizaje Zbrush	4	150	
Edición y simplificación de modelos	8	100	
Creacion de morphers	8	120	
Aprendizaje 3dsMax	3	120	
Generación del modelo final	5	75	
Aprendizaje Vray	3	40	
Imágenes finales	2	75	
Generacion de documentacion	8	100	
TOTAL	52	1115	

- [1] **3D reconstruction from multiple images** [En línea] <ftp://ftp.esat.kuleuven.ac.be/psi/visics/konijn/ICVSS08/vangool.pdf>
- [2] **Disparidad binocular y 3D** [En línea] <http://www.wisphysics.es/2008/02/disparidad-binocular-y-3d>
- [3] **Epipolar geometry** [En línea] http://en.wikipedia.org/wiki/Epipolar_geometry
Geometría epipolar [En línea] <http://ie.fing.edu.uy/investigacion/grupos/gti/cursos/egvc/material/tema-4.pdf>
- [4] **Reconstrucción 3D de escenas mediante un sistema de visión estéreo** [En línea] http://www.scielo.org.co/scielo.php?pid=S0123-21262012000200011&script=sci_arttext
- [5] **Triangulación de Delaunay** [En línea] <http://www.dma.fi.upm.es/mabellanas/voronoi/delone/delone.html>
Triangulación Delaunay 3D [En línea] <http://www.cimec.org.ar/ojs/index.php/cimec-repo/article/viewFile/3736/3658>
Agentes conversacionales. [En línea] http://www.chatbots.org/embodied_conversational_interface_agent/
- [6] **Proyecto RED.** [En línea] <https://www.bu.edu/fammed/projectred/index.html>
- [7] **Lip Sync** [En línea] <http://www.escardo.com/charanim/?p=15>
 Animación de personajes [En línea] <http://isilanimacion.blogspot.com.es/2010/01/ejercicio-de-lip-sync-sincronizacion-de.html>
Lip Sync [En línea] <http://www.escardo.com/charanim/?p=33>
Principios de sincronismo labial [En línea] <http://keithlango.com/tutorials/old/Principios%20del%20Sincronismo%20Labial.pdf>
Animating Facial Features & Expressions
 Bill Fleming y Darris Dobbs
 ISBN 1-886801-81-9
- [8] **Uncanny Valley** [En línea] http://es.wikipedia.org/wiki/Valle_inquietante
Uncanny Valley [En línea] http://www.arteyefectos.com/articulo_cont.asp?artic=15
- [9] **Making of the Korean Actress “Song Hye Kyo”** [En línea] <http://www.cgarena.com/freestuff/tutorials/max/songhyekyo/>
- [10] **The Digital Emily Project** [En línea] <http://gl.ict.usc.edu/Research/DigitalEmily/>
- [11] **Descarga de 123D Catch** [En línea] <http://www.123dapp.com/Catch>
- [12] **ZBrush** [En línea] <http://pixologic.com/>
ZBrush Central [En línea] <http://www.zbrushcentral.com/forum.php>
ZBrush Character Sculpting V1
 Aaron Sims
 ISBN 978-0-9551530-8-2
- [13] **Stop Staring. Facial Modeling and Animation Done Right**
 Jason Osipa
 ISBN 978-0-470-60990-3
- [14] **The Animator’s survival kit**
 Richard Williams
 ISBN 978-0571202287

