

Desarrollo de la animación facial en un agente virtual

Héctor Carrillo Randolph

Laboratorio de Cibernética, Departamento de Ingeniería Eléctrica, UAM-I
Av. San Rafael Atlixco 186, Col Vicentina, 09410, México, D.F.
hector@randolph.com.mx

Joel R. Jiménez Cruz

Laboratorio de Cibernética, Departamento de Ingeniería Eléctrica, UAM-I
Av. San Rafael Atlixco 186, Col Vicentina, 09410, México, D.F., 58044630
jcjr@xanum.uam.mx

Resumen

La animación facial por computadora es muy importante para diversas aplicaciones que involucran la comunicación verbal y no verbal y la interacción hombre-máquina. Ha tenido un crecimiento rápido y continuo desde que se inició a principios de los años 70s. Sus aplicaciones son muy variadas y se pueden usar, por ejemplo, en la atención a clientes, en multimedia educativa, chats y videotelefonía, juegos y entrenamiento, y ayuda a individuos con pérdidas auditivas o en los robots. En este artículo se presenta el desarrollo de un prototipo inicial de la animación facial de un agente virtual que es capaz de expresar emociones. Para la implementación del agente virtual se realizó una investigación de los métodos y técnicas necesarias para la representación facial, el modelado geométrico y la generación de la animación.

Palabras Claves: Agente Virtual, Animación Facial, Interfaz hombre-máquina, Modelado Geométrico.

1. Introducción

Actualmente la comunicación e interacción que tenemos con la computadora (interfaz hombre-máquina) está basada principalmente en el uso de ventanas, menús, apuntadores, teclados, ratones y monitores. Esta interacción se ha visto beneficiada por los avances realizados en las Interfaces gráficas y las interfaces perceptuales que involucran interfaces perceptivas, multimodales y multimedia [1]. También la interfaz se ha enriquecido por progresos importantes en el reconocimiento de voz, el reconocimiento visual y la capacidad de razonamiento por parte de las computadoras. Sin embargo, si se tuviera un agente virtual conversacional y emocional, la comunicación se llevaría a cabo de una manera más natural e intuitiva. Las expresiones de la cara del agente virtual, le daría una dimensión más *humana* a esta interfaz.

Se ha dado mucha atención al estudio y comprensión de la cara como un factor clave en la comunicación no verbal y verbal. En la cara se encuentran la mayoría de nuestros órganos sensoriales; los ojos, la nariz, la boca y las orejas, que no sólo participan en sus tareas principales, sino que también están involucradas en el acto de la comunicación. Las expresiones faciales posibilitan la comunicación y la integración social. Son una manifestación de la personalidad, el deseo de comunicación, el inicio de una interacción o el grado de expresividad durante la comunicación [2].

Un agente virtual cuya cara exprese emociones tiene diversas aplicaciones que incluyen al aprendizaje electrónico, el software educativo, los tutores inteligentes, la publicidad, el entretenimiento y la inclusión en los ambientes virtuales como en los chats y las películas, importantes en la industria multimedia, los asistentes en el comercio electrónico, la teleconferencia y los videoteléfonos etc. Otra aplicación importante consiste en la transmisión de imágenes sintéticas faciales con el fin de reducir el ancho de banda. Mediante la transmisión de parámetros relacionados con los movimientos en lugar de las imágenes completas, se logran reducciones importantes en la cantidad de datos que se envían a través de las líneas de comunicación [3].

En el presente trabajo se realizó una investigación de los métodos y técnicas necesarias para la generación de la animación de un rostro humano y se desarrolló un prototipo de la animación facial de un agente virtual que es capaz de representar emociones y de esta manera mejorar la interfaz entre el usuario y la computadora, esperando que se logre un uso e interacción más amigable, natural y amplia con la computadora.

La organización de este escrito es como sigue: primeramente se describen las partes constitutivas de la animación facial que incluyen la creación de la cabeza, la geometría facial y las técnicas de modelado y animación, luego se presenta la representación facial del agente virtual que se desarrolló, considerando diferentes aspectos de la piel, la anatomía esquelética y la anatomía muscular. A continuación se despliega la geometría del modelo facial del prototipo que se basó en la descripción muscular de la cara. Enseguida se describen la animación facial realizada en el agente virtual que se guió por el desempeño que se deseaba. Finalmente se realizan los comentarios y la conclusión del trabajo.

2. Las partes constitutivas de la animación facial

La implementación de la animación facial sigue una serie de pasos que se ilustran en la Fig. 1. El proceso de la animación comienza con la creación del modelo de la cabeza. La geometría facial se crea a partir de los datos de entrada que se adquieren de imágenes o de movimientos faciales. Esta entrada se acondiciona y se filtra para crear una malla de la cara que sea reconfigurable. Luego se aplica alguna técnica de modelado basada en la geometría para detallar las partes de la cara como el cráneo y la piel, entre otras. Los diferentes atributos de la cara se modelan utilizando técnicas de parametrización, de interpolación o de descripción muscular. Después del modelado se puede añadir la textura correspondiente.

Para desarrollar una animación facial realista y mejorar el desempeño del modelo, se agregan otros atributos de movimiento que pueden estar guiados por imágenes, o el

habla o el desempeño que se desee. Finalmente se realiza la simulación y el renderizado de la cara para comprobar su funcionamiento.

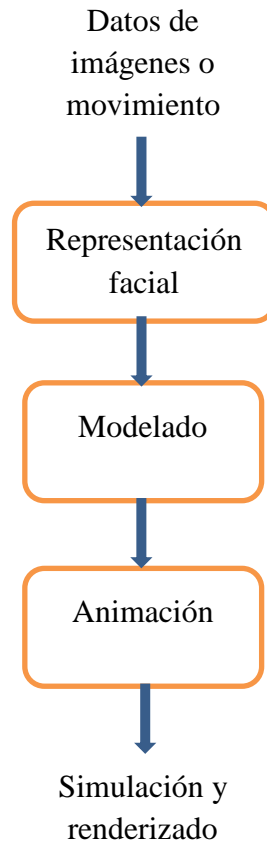


Fig. 1. Los componentes de la animación facial por computadora.

Un ejemplo de las partes constitutivas y las diferentes técnicas que se pueden utilizar para crear un personaje artificial se describen en la tesis de Medina Bello [4]. En ella, se presentan las bases teóricas y el desarrollo de una aplicación web que permite reproducir expresiones en un avatar 3D. En su aplicación utilizó un sistema de codificación de acción facial (FACS) y las unidades de acción (AU) que utiliza tal codificación. Incluye también información sobre el modelado facial en 3D que utiliza el estándar de codificación de

objetos audiovisuales MPEG-4. Describe también información general de las técnicas básicas de modelado y animación facial del avatar.

En el trabajo que aquí se presenta, se explican, de manera similar, los pasos realizados en esta tesis.

3. La representación facial

La expresión facial es el medio más rico e importante para expresar emociones y diferentes estados de ánimo que dependen del contexto y del estado emocional de la interacción. El ser humano tiene cuarenta y dos músculos faciales diferentes. Dependiendo de cómo los mueva, expresa determinadas emociones [5]. Las diferentes expresiones faciales son internacionales, dentro de diferentes culturas ya que existe un lenguaje similar de comunicación. Las investigaciones han determinado la existencia de seis expresiones faciales principales: alegría, tristeza, enojo, sorpresa, temor e interés.

Para la creación de un modelo de la cara con expresiones se necesitan estudiar los aspectos de la anatomía y fisiología de la cara. Para la representación computarizada de la cabeza se toman en cuenta la superficie de la piel, la anatomía esquelética, la anatomía muscular y el sistema vascular.

La piel del rostro como superficie externa está compuesta por la epidermis, la dermis y la hipodermis. Su textura puede cambiar, entre otros factores, por la edad, el género, las enfermedades, la raza, o el medio ambiente. El modelo de la cara debe reflejar estas características de textura lo más apegado a la realidad considerando la velocidad del renderizado en una interacción en tiempo real.

La anatomía esquelética de la cabeza proporciona la estructura para colocar la piel y los músculos. La deformación de la cabeza determinará las diferentes formas de la cara. La cabeza está compuesta básicamente por el cráneo donde se encuentra el cerebro y los huesos móviles de las mandíbulas. La forma y atributos de la cabeza dependen de los mismos factores internos y externos que afectan a la piel.

La anatomía muscular proporciona las bases para las funciones fisiológicas que dan como resultado las diversas expresiones faciales. Los músculos faciales son de diferentes tipos; lineal, hoja y esfínter. Existe una tendencia en el modelado facial de utilizar una representación muscular, como la que se utilizó en el prototipo que se presenta.

En un futuro se puede considerar, en la animación facial, la simulación del sistema vascular para proporcionar un mayor realismo ya que la cara puede sonrojarse cuando el suministro de sangre aumenta o disminuye en función de la exposición ante diferentes situaciones [6].

Para hacer la representación facial más apegada a la realidad, normalmente se incluyen otros atributos como los ojos, labios, lengua, dientes. En la Fig. 2 se muestra la representación del agente virtual desarrollado. En esta representación inicial que se ha hecho, solo se ha considerado el modelo de los ojos.



Fig. 2. La representación facial del modelo que se construyó.

4. El modelado facial

En los campos de la visión o las gráficas por computadora, la cara se puede representar utilizando técnicas geométricas de dos o tres dimensiones. Estas técnicas generan la superficie facial y la manipulan y deforman de acuerdo a los requerimientos de modelado. Esta parte del modelado es importante ya que la estructura facial determina los potenciales de la animación. En la Fig. 3 se muestra el modelo geométrico desarrollado en el agente virtual.

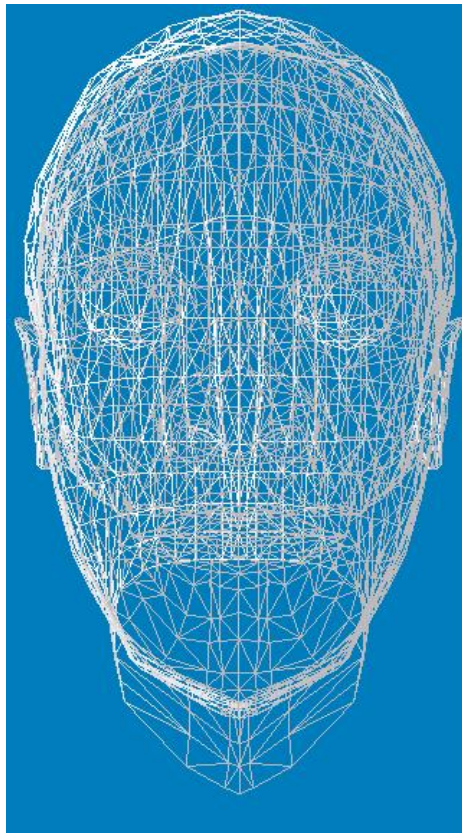


Fig. 3. Modelo geométrico del prototipo actual.

Las técnicas geométricas que se pueden usar para el modelado facial incluyen la interpolación, la parametrización o la descripción muscular de la cara.

En la interpolación se crea una malla poligonal de la cara para aproximar las expresiones utilizando varias sub-mallas. La interpolación puede ser bilineal o tipo spline. Entre las aplicaciones de la interpolación se encuentran la rotación o el escalamiento de la cara [7].

En la parametrización, permanece la idea de la interpolación, pero la geometría de la cara se divide en partes, donde cada parte es controlada por un conjunto de parámetros o puntos de control. El estándar MPEG-4 [8] utiliza los parámetros de animación facial (FPAs) o las unidades de acción (AU) para crear la animación de diferentes expresiones faciales. Este sistema de codificación se conoce sistema de codificación de la acción facial (FACS). Por ejemplo, en este sistema, la unidad de acción número 1 representa el levantamiento de la ceja interna, el número 5, el levantamiento del labio superior. Los parámetros dependen de la topología de la maya de la cara y por ende para otra cara, los parámetros se calibran, generalmente, en forma manual, aunque también existen métodos automáticos para la generación de los parámetros.

El modelado que depende de la estructura muscular comenzó en los 80s con los trabajos de Platt y Badler [9] quienes propusieron un modelo de masa resorte que conecta la piel, los huesos y los músculos. La red de resortes interconectan los músculos con las unidades de acción correspondientes. A finales de los 80s Waters propuso un modelo más elaborado basado en vectores. Clasificó los músculos en tres tipos, en base a su funcionamiento: lineales, de hoja y tipo esfínter. También utilizó el sistema FACS para relacionar las expresiones faciales con la actividad muscular [10]. En los 90s Terzopoulos y Waters introdujeron un modelo de malla de resortes en tres capas que incorpora a la piel, la anatomía muscular y a los tejidos grasos con lo cual se logra un mayor realismo anatómico de los comportamientos faciales [11].

En el proyecto del agente virtual que se está desarrollando se ha empleado el lenguaje C++, utilizando la técnica de la programación orientada a objetos y utilizando librerías de graficación de OpenGL. Actualmente se está utilizando una librería con el motor gráfico

de Irrlicht que ha permitido mejorar la apariencia del agente virtual y su entorno. En la Fig. 4, se muestran los músculos que pueden afectarse para producir diferentes expresiones faciales.

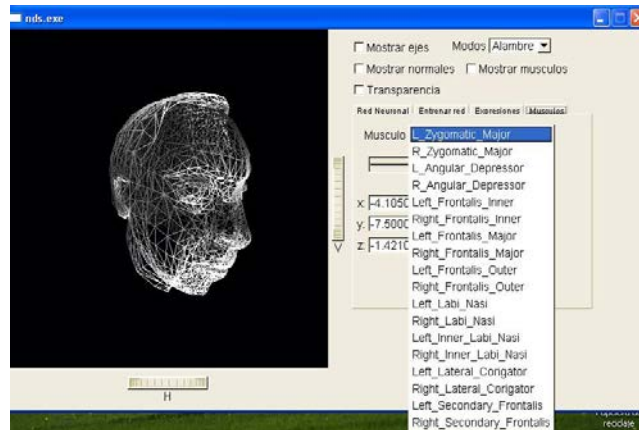


Fig. 4. Malla del modelo facial del prototipo actual que se basa en la representación muscular.

5. La animación facial

La expresión facial depende casi en su totalidad de los músculos de la cara. Los músculos están normalmente unidos al hueso en un extremo y a la fascia en el otro. Una combinación de los tres tipos primarios de músculos faciales (lineal, hoja y esfínter) coordinan la generación de las expresiones faciales. Entre los músculos faciales, los músculos lineales son los más comunes. Un ejemplo de un músculo lineal es el cigomático mayor, que se une en la esquina de la boca. Los músculos lineales tienen dos puntos de fijación, al momento de contraerse, tiran de la fascia hacia el punto de fijación en el hueso, produciendo un movimiento facial.

El músculo tipo hoja, funciona de una manera algo diferente. Funciona en una sola dirección como el músculo lineal, pero es más plano y ancho y se adhiere en forma más amplia. Un ejemplo de este tipo es el epicranius. Este músculo tiene dos partes: el músculo frontal anterior y el occipital posterior. Un tendón plano conecta las dos hojas musculares. La contracción de uno u otro de los músculos epicraneales tira del cuero

cabelludo hacia atrás o hacia adelante. Finalmente, los músculos tipo esfínter rodean normalmente los orificios del cuerpo. Son en forma de anillo y amplían o reducen la apertura mediante la relajación o la contratación [12].

Las técnicas de animación facial basados en datos proporcionan una animación más exacta, detallada y rica en detalles. Estas técnicas están guiadas por imágenes, el habla o el desempeño [13].

En la técnica basada en imágenes, la superficie facial y la posición de los datos se reconstruyen a partir de diferentes vistas de las imágenes faciales. En esta técnica se puede considerar el morphing (cambio suave y gradual de una imagen a otra) y combinación entre fotografías, la textura y el sistema vascular.

La técnica basada en el habla trata de sintetizar la animación visual del habla con el habla. Se mapean los fonemas con la animación para producir la sincronización de los labios de la boca. En esta técnica se necesita una base de datos pregrabada para realizar la síntesis. También se pueden incorporar emociones en concordancia con la sincronización de los labios y los sonidos que producen.

La animación conducida por desempeño involucra la adquisición de los datos del movimiento por medio de dispositivos de entrada como cámaras que detectan los movimientos de caras reales y las transfieren en forma digitalizada. Se utilizan marcadores en la cara y las cámaras escanean la cara y siguen a los marcadores. Existen desafíos para producir la animación en tiempo real considerando los movimientos estáticos y dinámicos de la cara de una persona y en como extraer los parámetros de seguimiento para conducir la animación.

En la Fig. 5 se muestra una expresión de desagrado que se ha generado en el prototipo con la técnica basada en desempeño. Hasta el momento se han manipulado en forma manual grupos de diferentes músculos con el fin de probar las emociones básicas.

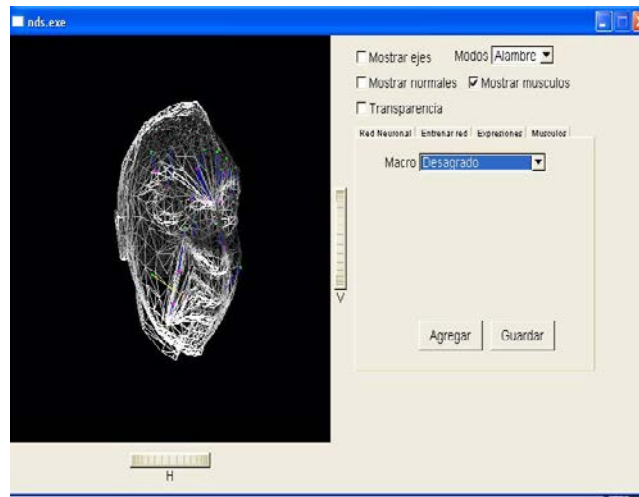


Fig. 5. En esta imagen se muestra la animación del prototipo actual con la expresión de **desagrado**.

6. Comentarios

En este escrito se han comentado varias técnicas utilizadas para lograr la animación facial: la representación facial, el modelado geométrico facial y la animación basada en desempeño.

La implementación de la animación facial es compleja e involucra varias etapas. El agente virtual desarrollado que se ha descrito se encuentra en sus fases de prueba iniciales. Hasta el momento se pueden generar diferentes expresiones en forma manual. También se han realizado pruebas iniciales con el entrenamiento de una red neuronal basada en retropropagación que puede, a partir de activar una entrada, generar la expresión correspondiente.

Por otro lado, falta Mejorar la eficiencia de los algoritmos gráficos y de la animación facial y extender la capacidad para importar modelos de diversas aplicaciones como 3D studio, blender, maya, etc. También falta automatizar la interacción con el personaje.

La siguiente fase del proyecto consiste en agregar al software, el manejo de músculos circulares para el movimiento de labios y ojos. También existe la posibilidad de agregar

movimiento en los párpados, cuello y mandíbula. También es posible agregar un módulo de síntesis de voz que se sincronice con el movimiento de la boca [14]. Otro módulo necesario es la captura y reconocimiento de la imagen del usuario con el fin de manipular los músculos y la expresión facial del agente virtual para entablar una mejor interacción y comunicación.

7. Conclusión

Las aplicaciones de personajes virtuales son muy extensas. Se pueden usar, por ejemplo, en la atención a clientes, en multimedia educativa, chats y videotelefonía, juegos y entrenamiento, y ayuda a individuos con pérdidas auditivas.

También hay una necesidad inherente de agentes sociales o avatares en la web para mejora la experiencia del usuario. La interacción puede ser muy simple, como una reacción a una acción simple como hacer clic en un enlace, a algo mucho más complejo como actuar como un asistente personal para realizar una compra. Con las mejoras en las tecnologías de inteligencia artificial, estos "bots" se verán con mayor frecuencia.

Con el fin de representar y hacer un modelo creíble y realista de la cara debe haber una integración de animación facial, la voz sintética, las emociones y el reconocimiento del interlocutor para la interfaz hombre-máquina sea más eficiente en la comunicación no verbal y verbal.

El futuro de la interacción depende en gran medida en que tan naturalmente nos podemos comunicar con los diferentes dispositivos electro-mecánicos ya sean estáticos o móviles. Estos dispositivos tendrán muy pronto caras que escuchen, hablen e interactúen con los humanos.

Deseamos crear una cara virtual creíble que actúe naturalmente. Esta cara se puede colocar en un robot que mejore la comunicación, la interacción y el entendimiento con los humanos.

8. Agradecimientos

Extendemos los más altos agradecimientos a Luis Antonio González González, Einstein Román González Montañez e Isaac De la Vega Gutiérrez quienes colaboraron en diferentes etapas iniciales del proyecto.

9. Referencias

- [1] D. Diaz. User Experience in Human-Technology Interaction. Communication, context and evaluation methodology. Tesis (Doctoral), E.T.S.I. Telecomunicación (UPM). 2014. <http://oa.upm.es/28992/>.
- [2] A. López Get, "Individuación y lenguaje no verbal en la CMC". *Revista de Ciencias Sociales (Cr)*, vol. I-II, núm. 131-132, 2011, pp. 169-18. <http://www.redalyc.org/pdf/153/15323166011.pdf>.
- [3] H. Yu Ping, L. Nurliyana Abdullah, P. Suhaiza Sulaiman, y A. Abdul Halin. "Computer Facial Animation: A Review". *International Journal of Computer Theory and Engineering*, Vol. 5, No. 4, August 2013. <http://ijcte.org/papers/770-A011.pdf>.
- [4] P. N. Medina Bello. Aplicación web para el modelado de expresiones faciales en un avatar. Tesis de Licenciatura. Ingeniería en Sistemas y Tecnologías de Información. Departamento de Computación, Electrónica y Mecatrónica, Escuela de Ingeniería, Universidad de las Américas Puebla, 2013. http://catarina.udlap.mx/u_dl_a/tales/documentos/lst/medina_b_pn/index.html.
- [5] S. Vargas. "Músculos de la expresión facial". SlideShare, Aug 20, 2011. <http://www.slideshare.net/sindyvargas/msculos-de-la-expresin-facial>.
- [6] J. Jimenez, T. Scully, N. Barbosa, C. Donner, X. Alvarez, T. Vieira, P. Matts, V. Orvalho, D. Gutierrez, T. Weyrich. "A Practical Appearance Model for Dynamic

- Facial Color”. ACM Transactions on Graphics, Vol. 29(5), SIGGRAPH Asia 2010. <http://www.iryoku.com/skincolor/>.
- [7] V. Peña Araya. Mejoramiento y evaluación de técnicas de interpolación para la animación de mallas faciales. Memoria para optar por el título de ingeniero civil en computacion. Universidad de Chile, abril 2012. <http://tesis.uchile.cl/handle/2250/111345>.
- [8] MPEG-4. *Wikipedia*, 3 jun 2014. <http://es.wikipedia.org/wiki/MPEG-4>.
- [9] S. Platt, N. Badler. “Animating facial expressions”. ACM SIGGRAPH Computer Graphics 08/1981; 15(3): 245-252. http://www.researchgate.net/publication/234810531_Animating_facial_expressions.
- [10] K. Waters. “A Muscle Model for Animating Three-Dimensional Facial Expression”. Computer Graphics, Volume 21, Number 4, July 1987. <http://excelsior.biosci.ohio-state.edu/~carlson/history/PDFs/ani-papers/waters-muscle.pdf>.
- [11] D. Terzopoulos, K. Waters. “Physically-Based Facial Modeling, Analysis, and Animation”. Journal of Visualization and Computer Animation, 1990, 1(2): 73–80. <http://www.csie.ntu.edu.tw/~r96944046/paper/face/vca90.pdf>.
- [12] Y. Zhang, E. Sung, E.C. Prakash, “3D modeling of dynamic facial expressions for face image analysis and synthesis”, International Conference on Vision Interface, Canada, 2001. <http://www.cipprs.org/papers/VI/VI2001/pp001-008-Zhang-et-al-2001.pdf>.
- [13] S. Villagrasa, J. Duran, J. Clares. “Animación facial: cómo (re)crear el perfecto rostro humano 3D”. Octava Conferencia Iberoamericana en Sistemas, Cibernética e Informática: CISCI 2009. <http://www.iiis.org/CDs2009/CD2009CSC/CISCI2009/PapersPdf/C715HZ.pdf>.
- [14] J.F. Ferreira, F. De la Rosa. “Una Propuesta de Integración de Animación Facial y Voz Sintética”. XXX Conferencia Latino-americana de Informática CLEI'2004,

Arequipa, Perú, Septiembre 2004, pp. 180-191.
http://www.loop.la/noticias/computacionvisual/pdfs/animacion_facial.pdf [4].

10. Autores

Héctor Carrillo Randolph Se dedica al desarrollo de software. Hizo sus estudios de Licenciatura en Computación en la UAM-I.

Joel R. Jiménez Cruz es profesor del Departamento de Ingeniería Eléctrica en la UAM-I, y cultiva el área de conocimiento de la Cibernética.