

UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA

CARRERA DE INGENIERÍA ELECTRÓNICA

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN
DEL TÍTULO DE INGENIERO ELECTRÓNICO.

PROYECTO TÉCNICO CON ENFOQUE INVESTIGATIVO:

**“ANIMACIÓN FACIAL SOBRE AVATAR PARA
MEJORAR LA INTERACCIÓN DE NIÑOS CON
TRASTORNO DEL ESPECTRO AUTISTA”**

AUTOR:

JOSÉ ROBERTO TOLEDO ILLESCAS

TUTOR:

ING. PAUL ANDRÉS CHASI PESANTEZ

CUENCA – ECUADOR

2018

CESIÓN DE DERECHOS DE AUTOR

Yo, José Roberto Toledo Illescas con documento de identificación N° 010453342-5, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación: **“ANIMACIÓN FACIAL SOBRE AVATAR PARA MEJORAR LA INTERACCIÓN DE NIÑOS CON TRASTORNO DEL ESPECTRO AUTISTA”**, mismo que ha sido desarrollado para optar por el título de: *Ingeniero Electrónico*, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 15 de octubre del 2018



José Roberto Toledo Illescas

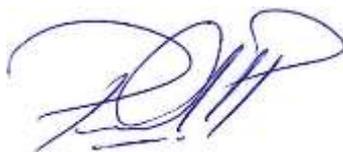
CI: 0104633425

AUTOR

CERTIFICACIÓN

Yo, declaro que bajo mi *tutoría* fue desarrollado el trabajo de titulación: **“ANIMACIÓN FACIAL SOBRE AVATAR PARA MEJORAR LA INTERACCIÓN DE NIÑOS CON TRASTORNO DEL ESPECTRO AUTISTA”**, realizado por José Roberto Toledo Illescas, obteniendo el **Proyecto Técnico con enfoque investigativo** que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, 15 de octubre del 2018



Ing. Paul Andrés Chasi Pesantez

CI: 0103652095

DIRECTOR DEL TRABAJO DE TITULACIÓN

DECLARATORIA DE RESPONSABILIDAD

Yo, José Roberto Toledo Illescas con número de cédula CI. 010463342-5, autor del trabajo de titulación: **“ANIMACIÓN FACIAL SOBRE AVATAR PARA MEJORAR LA INTERACCIÓN DE NIÑOS CON TRASTORNO DEL ESPECTRO AUTISTA”**, certifico que el total contenido del **Proyecto Técnico con enfoque investigativo**, es de mi exclusiva responsabilidad y autoría

Cuenca, 15 de octubre del 2018



José Roberto Toledo Illescas

CI: 0104633425

AUTOR

AGRADECIMIENTO

A Dios, a mis padres y hermanas por haberme apoyado en mis años de estudio, tanto económica como moralmente, por haber sido una guía en los momentos difíciles como en los momentos alegres. A Lissy por el apoyo brindado en estos años, por su amor incondicional y el esfuerzo que pusimos en nuestros años de estudio. De igual manera a mi tutor el Ing. Paul Chasi por su ayuda en la realización de este proyecto, así como al Ing. Fernando Guerrero por el apoyo y confianza brindados durante la realización de este proyecto.

José Roberto Toledo Illescas

DEDICATORIA

El valor de las personas no se puede cuantificar por la cantidad de bienes materiales que estos poseen, sino por la calidad de personas que son; es uno de los principales principios que me han inculcado mis padres, por lo cual le dedico este logro a mi abuelo Alfonso (+) que ha sido más que un abuelo un padre y un mentor, a mis padres Martín y Carmen por estar junto a mi diario caminar y a mis hermanas Belén y Carmen para que atesoren el regalo que nos han dado nuestros padres y sigan sus sueños.

José Roberto Toledo Illescas

ÍNDICE GENERAL

AGRADECIMIENTO	I
DEDICATORIA	II
ÍNDICE GENERAL.....	III
ÍNDICE DE FIGURAS	V
ÍNDICE DE TABLAS	VIII
RESUMEN	IX
INTRODUCCIÓN	X
ANTECEDENTES DEL PROBLEMA DE ESTUDIO	XI
JUSTIFICACIÓN	XII
OBJETIVOS	XIII
OBJETIVO GENERAL.....	XIII
OBJETIVOS ESPECÍFICO	XIII
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA	1
1.1 Autismo y Avatar	1
1.1.1 Avatares e interacción social.....	2
1.2.1 Avatares y entornos virtuales multiusuarios masivos	4
1.3.1 Avatares e interacción en ambientes virtuales	5
1.4.1 Avatares y aplicaciones multimedia	7
1.2 Visión por computadora.....	8
1.3 Filtro Haar	10
1.4 Local binary pattern (LBP)	14
1.5 Red neuronal artificial RNA	16
1.1.5 Perceptrón	17
1.2.5 Funcion de activación	18
1.3.5 Topologia de una red neuronal.....	20
1.6 Landmark	21
1.7 Animación facial	23
1.8 Sistema propuesto	24
CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DEL SISTEMA	27
2.1 Captación y procesamiento	27

2.1.1 Captación	27
2.2.1 Pre procesamiento	28
2.2 Procesamiento	30
2.1.2 Calculo del LBP	30
2.3 Red neuronal	31
2.1.3 Creación de patrones de entrenamiento	31
2.2.3 Diseño y entrenamiento de la RNA	33
2.4 Presentación	39
2.5 Terapias	43
2.1.5 Sistema de interacción Avatar – persona	43
2.1.5 Video chat empleando Avatares en pacientes con TEA	44
2.2.5 Reconocimiento de emociones.....	45
2.3.5 Tutor virtual para aprendizaje de expresiones faciales	46
CAPÍTULO 3: PRUEBAS Y ANÁLISIS DE RESULTADOS	48
3.1 Detección facial.....	48
3.2 Red neuronal y estimación de la ubicación de los puntos faciales	51
3.3 Animación	56
CAPÍTULO 4: CONCLUSIONES Y RECOMENDACIONES	63
4.1 Conclusiones	63
4.2 Recomendaciones.....	64
REFERENCIAS BIBLIOGRÁFICAS	65
APÉNDICES.....	71
APÉNDICE A: TARGET DE 20 PATRONES DE ENTRENAMIENTO DE LA RED NEURONAL PARA LA ESTIMACIÓN DE LA UBICACIÓN DE PUNTOS FACIALES	71
APÉNDICE B: HUESOS DE ANIMACIÓN DEL AVATAR Y SECCIONES DEL MALLADO QUE AFECTAN	73
APÉNDICE C: VALIDACIÓN DEL PROYECTO	82

ÍNDICE DE FIGURAS

Figura 1.1 Avatar en VR realizando tareas de atención a respuestas humanas sobre imágenes.....	2
Figura 1.2 Ambiente virtual para el juego de la pelota	3
Figura 1.3 Ambiente virtual para el juego de rol.....	4
Figura 1.4 Seguimiento del Avatar realizado desde el servidor	5
Figura 1.5 Avatares empleados en el sistema AMITIES	6
Figura 1.6 Clase virtual para cinco estudiantes controlada por un agente	7
Figura 1.7 Diagrama de bloques de un call center basado en Avatares	8
Figura 1.8 Estructura base del análisis de expresiones faciales	9
Figura 1.9 Características de Haar	11
Figura 1.10 Subáreas que conforman la imagen integral.....	12
Figura 1.11 Características rectangulares aplicadas sobre una ventana	13
Figura 1.12 Estructura del clasificador propuesto por Viola y Jones	14
Figura 1.13 Pasos para obtener el valor LBP de un pixel.....	16
Figura 1.14 Estructura de un perceptrón	17
Figura 1.15 Estructura de una neurona artificial	18
Figura 1.16 Función de activación escalón	18
Figura 1.17 Función de activación lineal	19
Figura 1.18 Función de activación tangente hiperbólica	19
Figura 1.19 Función de activación sigmoideal.....	20
Figura 1.20 Representación de una neurona multicapa	21
Figura 1.21 Resultados de detección de puntos faciales de Zhang.....	22
Figura 1.22 Resultados de detección de puntos faciales de Nordstrom.....	22
Figura 1.23 Resultados de detección de puntos faciales de Kazemi	23
Figura 1.24 Resultados de animación facial de Laine	24
Figura 1.25 Diagrama de bloque general del sistema propuestos	25
Figura 2.1 Imagen captada por la webcam en la etapa de adquisición	27
Figura 2.2 Imagen captada en escala de grises	28
Figura 2.3 Resultado de la detección facial basada en el filtro Haar	29
Figura 2.4 Áreas de la región de interés	30
Figura 2.5 Landmark aplicados sobre el rostro.....	32
Figura 2.6 Topología de la RNA 1	36

Figura 2.7 <i>Topología de la RNA 2</i>	37
Figura 2.8 <i>Topología de la RNA 3</i>	37
Figura 2.9 <i>Topología de la RNA 4</i>	37
Figura 2.10 <i>Topología de la RNA 5</i>	38
Figura 2.11 <i>Topología de la RNA 5 con error en su entrenamiento</i>	38
Figura 2.12 <i>Banco de Avatares para la animación facial</i>	40
Figura 2.13 <i>Aceptación de los Avatares en una población de niños de 8 a 15 años</i>	41
Figura 2.14 <i>Rostro del Avatar empleado</i>	42
Figura 2.15 <i>Esquema del sistema de interacción Avatar – persona</i>	44
Figura 2.16 <i>Esquema del sistema de video chat para pacientes con TEA</i>	45
Figura 2.17 <i>Modelo de aplicación de propuesta para reconocimiento de emociones</i>	46
Figura 2.18 <i>Modelo de aplicación de tutor de aprendizaje de emociones</i>	47
Figura 3.1 <i>Diagrama de la evaluación de la detección facial</i>	49
Figura 3.2 <i>Numero de falsos positivos obtenidos en la prueba de detección facial.</i> .	50
Figura 3.3 <i>Prueba de la RNA 1 en tiempo real</i>	54
Figura 3.4 <i>Prueba de la RNA 2 en tiempo real</i>	55
Figura 3.5 <i>Prueba de la RNA 3 en tiempo real</i>	55
Figura 3.6 <i>Prueba de la RNA 4 en tiempo real</i>	56
Figura 3.7 <i>Prueba de la RNA 5 en tiempo real</i>	56
Figura 3.8 <i>Prueba de animación: Persona normal</i>	57
Figura 3.9 <i>Prueba de animación: Avatar normal</i>	58
Figura 3.10 <i>Prueba de animación y detección facial: Persona feliz</i>	58
Figura 3.11 <i>Prueba de animación: Avatar feliz</i>	59
Figura 3.12 <i>Prueba de animación: Niño 1</i>	59
Figura 3.13 <i>Prueba de animación 1</i>	60
Figura 3.14 <i>Prueba de animación: Niño 2</i>	60
Figura 3.15 <i>Prueba de animación 2</i>	61
Figura 3.16 <i>Prueba de animación: Persona 3</i>	61
Figura 3.17 <i>Prueba de animación 3</i>	62
Figura B.1 <i>Avatar empleado en la animación facial</i>	73
Figura B.2 <i>Sección del avatar afectado por el hueso 26</i>	74
Figura B.3 <i>Sección del avatar afectado por el hueso 15</i>	74
Figura B.4 <i>Sección del avatar afectado por el hueso 16</i>	75

Figura B.5 <i>Sección del avatar afectado por el hueso 25</i>	75
Figura B.6 <i>Sección del avatar afectado por el hueso 17</i>	76
Figura B.7 <i>Sección del avatar afectado por el hueso 18</i>	76
Figura B.8 <i>Sección del avatar afectado por el hueso 19</i>	77
Figura B.9 <i>Sección del avatar afectado por el hueso 20</i>	77
Figura B.10 <i>Sección del avatar afectado por el hueso 21</i>	78
Figura B.11 <i>Sección del avatar afectado por el hueso 22</i>	78
Figura B.12 <i>Sección del avatar afectado por el hueso 23</i>	79
Figura B.13 <i>Sección del avatar afectado por el hueso 24</i>	79
Figura B.14 <i>Sección del avatar afectado por el hueso 10</i>	80
Figura B.15 <i>Sección del avatar afectado por el hueso 12</i>	80
Figura B.16 <i>Sección del avatar afectado por el hueso 7</i>	81
Figura B.17 <i>Sección del avatar afectado por el hueso 9</i>	81

ÍNDICE DE TABLAS

Tabla 2.1 Elementos de análisis de las áreas de la ROI.....	31
Tabla 2.2 Landmark que estimara cada RNA	34
Tabla 2.3 Puntos de entrenamiento equivalentes a los Landmark que estima cada RNA	34
Tabla 2.4 Ingreso y salidas de las RNAs.....	35
Tabla 2.5 Aceptación de Avatares en mujeres de 8 a 15 años	40
Tabla 2.6 Aceptación de Avatares en hombres de 8 a 15 años	40
Tabla 3.1 Datos de las pruebas de detección facial.....	51
Tabla 3.2 Distribución de los datos para entrenamiento y pruebas de las RNAs	52
Tabla 3.3 Error de las estimación de los puntos faciales en la RNA propuesta.....	53
Tabla A.1 Target de patrones de entrenamiento para la estimación de coordenadas de puntos faciales	71

RESUMEN

El siguiente proyecto nace de las condiciones que presentan las personas con autismo, siendo una limitación su interacción social; por lo que se propuso un Avatar terapéutico para la interacción indirecta con sus pares, padres y/o terapeutas.

El desarrollo del proyecto está basado en el uso de redes neuronales. Las cuales interpretan la información del rostro captadas por una webcam, para generar datos en su salida capaces de ser interpretados como indicadores de movimientos, que a su vez serán transmitidos a un Avatar para su animación. El sistema propuesto cumple con tareas de adquisición, procesamiento y presentación.

En primera instancia, se captará los datos del entorno mediante la webcam integrada en una laptop, la imagen captada será procesada cambiando su tamaño y espacio de color, obteniendo una imagen estándar para el procesamiento previo a la red neuronal y su ingreso a la misma. Esta imagen se dividirá en tres subáreas, definidas por la ubicación del tabique. Se ha seleccionado el tabique ya que este es un elemento fijo en el rostro, es decir, no presenta cambios en su forma como es el caso de los ojos y boca, que varían el tamaño de apertura entre los parpados y labios respectivamente.

En la etapa de procesamiento a las subáreas se les realizará un análisis de texturas o intensidad del espacio de color mediante el cálculo del Patrón Binario Local (LBP, por sus siglas en inglés: Local Binary Pattern), con el que se reduce el número de características que ingresan a la red neuronal sin perder información de la imagen captada. A continuación, se encuentra la red neuronal en la cual se ingresarán el vector LBP, dando como resultado en su salida las coordenadas cartesianas de los puntos faciales de la persona.

En la etapa de presentación con los datos obtenidos de la red neuronal artificial se genera el movimiento de los huesos de animación del Avatar de una posición inicial a una final, que deberá guardar parentesco con el rostro que se captó con la webcam.

INTRODUCCIÓN

En el presente proyecto “Animación facial sobre Avatar para mejorar la interacción de niños con Trastorno del Espectro Autista”, propone un sistema basado en el uso de la animación facial que busca mejorar la relación social de niños con autismo, motivados por la gran aceptación que muestran las personas con este trastorno a interactuar con herramientas tecnológicas. El eje central del sistema es una red neuronal que analiza los datos de textura de una secuencia de imágenes del niño para ser interpretadas en el proceso de animación del Avatar, transfiriendo las ubicaciones cartesianas de los puntos principales y simulando el movimiento de determinados músculos faciales.

El documento está estructurado en cuatro capítulos los que contendrán las siguientes temáticas: el Capítulo 1 presenta el sustento teórico en el que se basa el sistema, así como las tendencias que se han empleado en los últimos años; el Capítulo 2 presenta el diseño e implementación de los diversos procesos que conforman al sistema; el Capítulo 3 muestra un análisis de los resultados de diversas pruebas que se realizaron en el desarrollo e implementación del sistema; y el Capítulo 4 contiene las conclusiones y recomendaciones generadas en el desarrollo del proyecto.

ANTECEDENTES DEL PROBLEMA DE ESTUDIO

La animación facial se basa en el reconocimiento automático de las diversas expresiones faciales. En los últimos años han existido varios avances relacionados con técnicas y mecanismos usados en la detección y seguimientos de rostros, así como en el análisis de expresiones faciales y su clasificación [1].

Uno de los puntos de partida al hablar de animación facial es la visión por computadora en la cual mediante una computadora se puede analizar los movimientos y acciones de una persona. Este proceso recibe el nombre de captura del movimiento humano. Entre dichos movimientos se han analizado los relacionados con la cabeza, brazos, torso y piernas [2] [3].

El reconocimiento de las expresiones faciales parte de la visión por computadora, que ha generado una gran variedad de aplicaciones principalmente usadas en la robótica y en el desarrollo de juegos de video. Dentro de la robótica se ha centrado especialmente en el desarrollo de robots humanoides, que no será de interés para el presente trabajo. En [4] se presentan los resultados del uso de la animación facial en el desarrollo de video juegos y películas. Este artículo fue publicado por la Universidad de sur de California y NVIDIA; en él indican un método sistemático para la obtención y diseño de parámetros de entrenamiento y el uso de técnicas en el procesamiento de imágenes como el Análisis de Componentes Principales (PCA, por sus siglas en inglés: Principal Component Analysis).

Ya que los diversos gestos faciales que realiza una persona pueden ser consideradas como un sistema de comunicación, se han desarrollado herramientas que buscan replicar este principio para analizar emociones y procesos de interacción. En [5] se puede apreciar el diseño y uso de Avatares en procesos de comunicación y terapias que realizan un mayor énfasis en la interacción social.

JUSTIFICACIÓN

La animación facial posee una gran gama de aplicaciones entre las que destacan aquellas que se ligan a los procesos de comunicación y análisis del estado de una persona. En este contexto se presenta el proyecto “Animación facial sobre Avatar para mejorar la interacción de niños con Trastorno del Espectro Autista”. Con el que se busca brindar una aplicación que permita mejorar los procesos de interacción comunicación tanto receptiva como expresiva de niños con Trastorno del Espectro Autista (TEA), de quienes se conoce que presentan grandes dificultades en este ámbito específico.

Mediante el uso de un Avatar se busca que el niño interactúe de manera indirecta con otra persona, ya sea con un semejante suyo o con un terapeuta. El desarrollo del proyecto está basado en el uso de redes neuronales que permitan captar e interpretar la información proveniente de una webcam para generar datos a su salida capaces de ser interpretados como indicadores de movimientos, que a su vez serán transmitidos a un Avatar para su animación en tiempo real.

Existen tres consideraciones iniciales a tener en cuenta al trabajar con niños con este proyecto. Primero, es preciso contar con un grupo de estudio bien establecido y con los permisos necesarios otorgados por los representantes legales de los infantes. Segundo, que no existe una base de datos lo suficientemente extensa para entrenar las redes neuronales y por tal motivo es necesario generar una base datos, propia de este proyecto. Tercero, es necesario contar con la guía de un experto en el trabajo con niños, en este caso se contará con el apoyo de la Centro de Intervención Multidisciplinaria para el Autismo (CIMA).

OBJETIVOS

OBJETIVO GENERAL

- Diseñar e implementar un sistema de animación facial, capaz de imitar las diferentes gesticulaciones del rostro de niños y transmitirlos a un Avatar.

OBJETIVOS ESPECÍFICO

- Revisar el estado del arte referente a los conceptos y técnicas empleados en la animación facial, tendencias y metodologías.
- Diseñar e implementar una red neuronal capaz de procesar los datos en tiempo real de una webcam y convertirlos en datos útiles para el desarrollo de la animación del Avatar.
- Utilizar los Avatar generados en sesiones de terapia con niños de la fundación CIMA.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

1.1 AUTISMO Y AVATAR

El Trastornos del Espectro Autista constituye un grupo de alteraciones del neurodesarrollo que afectan de manera global distintas funciones cerebrales superiores del individuo, como la inteligencia, la capacidad del lenguaje y la interacción social [6].

Actualmente al autismo se lo denota como Trastorno del Espectro Autista (TEA), denominado así ya que engloba trastornos en los que se incluyen los siguientes:

- Problemas de larga duración con la comunicación e interacción social en diferentes contextos.
- Comportamientos repetitivos o no querer cambios en la rutina diaria.
- Síntomas que comienzan en la primera infancia, por lo general, en los primeros 2 años de vida.
- Síntomas que hacen que la persona necesite ayuda en la interacción social.

No todas las personas con un trastorno del espectro autista exhiben todos estos comportamientos, pero la mayoría tendrá varios de ellos [7].

El tratamiento en la actualidad abarca desde lo altamente científico y objetivo, hasta lo subjetivo e incluso místico [8]. De los diversos tratamientos y terapias, la mayoría se centran en modificar la conducta del niño o desarrollar en él hábitos que se dificultan por sus trastornos, como terapias de conducta o desarrollo cognitivo.

En la actualidad se han desarrollado varias aplicaciones basadas en la visión por computadora y la detección facial, ya sea mediante el uso de Avatares o ambientes virtuales, en donde se busca mejorar la interacción de un grupo de personas, ya sean niños con autismo o en otro tipo de terapias. Basándose en este concepto se han generado ambientes virtuales en donde, un Avatar o ente virtual realiza una interacción con un determinado individuo.

En las siguientes subsecciones se detallará trabajos realizados sobre Avatares y ambientes virtuales, con un mayor énfasis en su aplicación.

1.1.1 AVATARES E INTERACCIÓN SOCIAL

Por varios años se ha investigado la interacción de las personas con agentes virtuales como los Avatares, ya sea una aplicación que se ejecutan en un ordenador o mediante la realidad virtual.

Kim en su trabajo presenta un programa realizado mediante Realidad Virtual (VR) el que busca estudiar la interacción emocional entre humano y Avatar, para lo cual el programa consta de dos módulos control uno para la persona y otro para el Avatar, a las personas se les presento un Avatar masculino con imágenes que aparecían a su derecha o izquierda, las que podían ser casas o diseños abstractos, tras un segundo estas desaparecían y el Avatar les pedía realizar una acción con una cruz que aparecía después de la imagen, los resultados de este proyecto fue comprender la enseñanza que una persona puede realizar a un Avatar, en donde el Avatar conocía cuando la persona le mentía ya que este conocía las acciones que debía realizar la persona con cada imagen.

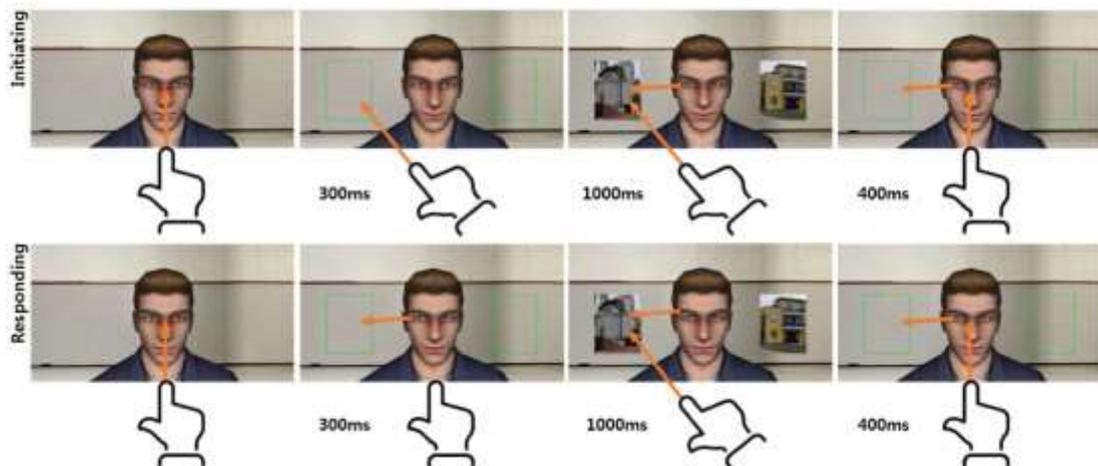


Figura 1.1 Avatar en VR realizando tareas de atención a respuestas humanas sobre imágenes
Fuente: M. Kim [9]

Ruth mediante un método experimental investigo los efectos de reducir la información social y canales de conducta mediante un entorno virtual en donde la persona era representada por un Avatar de cuerpo completos que no presentan rasgos físicos como sexo y estatura ya que los Avatares que se emplearon poseían las mismas características.

Comparando el comportamiento de las personas en la Realidad Virtual y en el mundo real se compararon dos interacciones sociales comunes, una tarea física realizada con un juego de pelota y otra tarea verbal realizada con un juego de rol. De este estudio se notó que los niveles de interacción eran más bajos en la realidad virtual resaltado mayor dificultad en la comprensión de mensajes y el afecto que se demostraba una parte en el juego de rol.

En las figuras 1.2 y 1.3 se puede apreciar el ambiente desarrollado por Roth y Lugin para realizar sus pruebas.

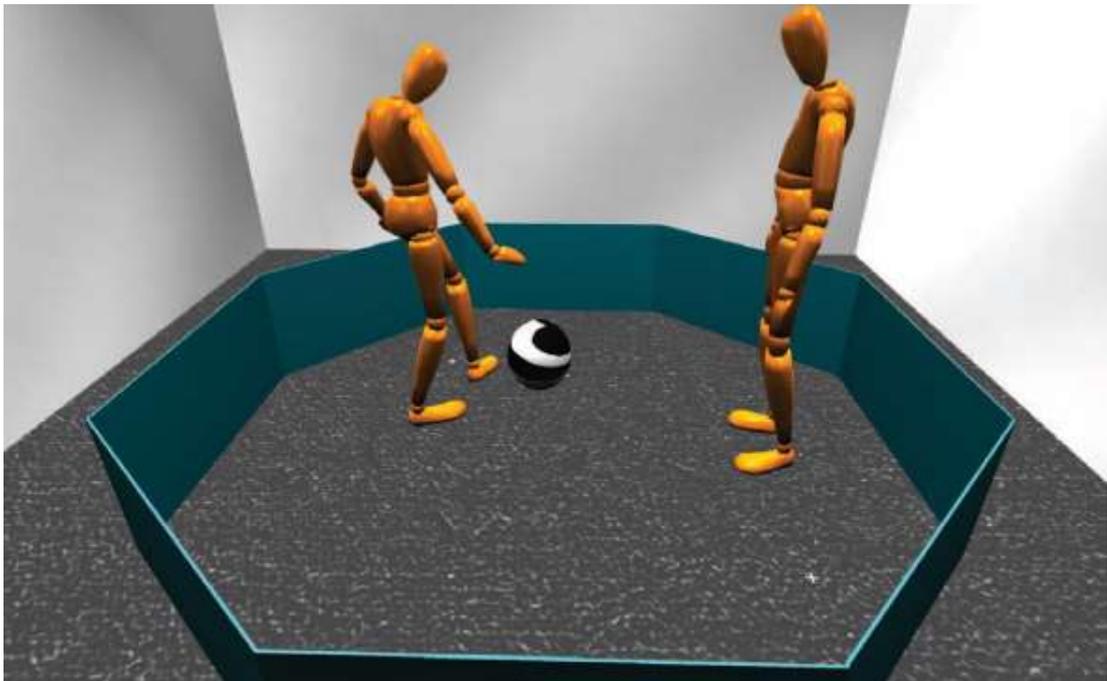


Figura 1.2 *Ambiente virtual para el juego de la pelota*
Fuente: *D. Roth & J. Lugin* [10]

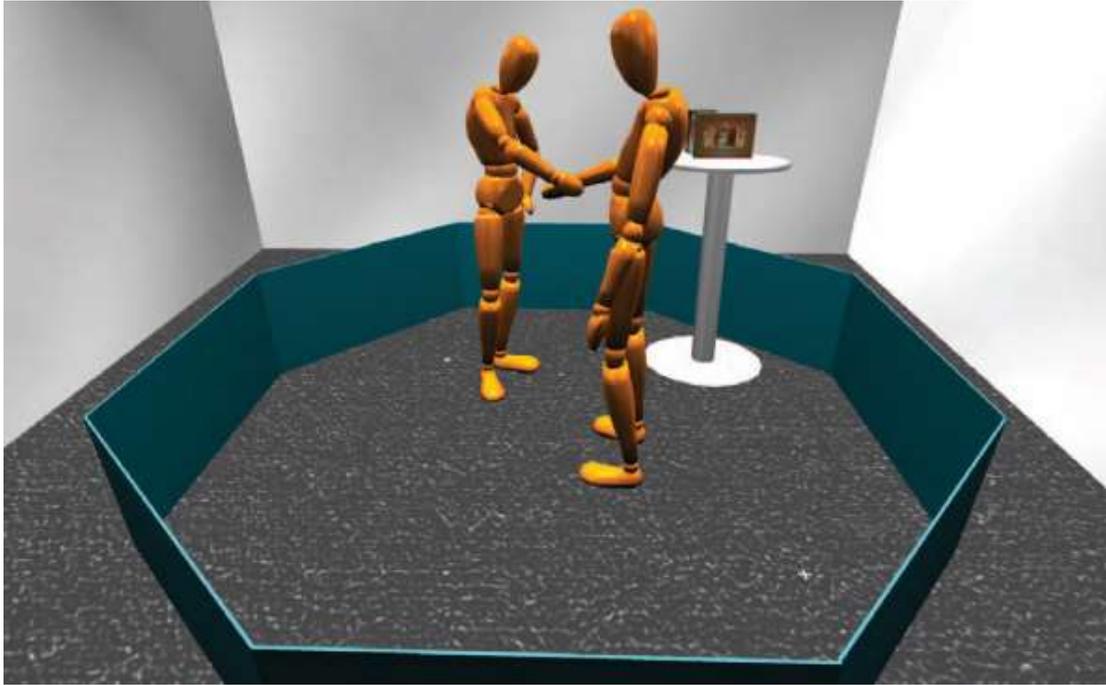


Figura 1.3 *Ambiente virtual para el juego de rol*
Fuente: D. Roth & J. Lugin [10]

1.2.1 AVATARES Y ENTORNOS VIRTUALES MULTIUSUARIOS MASIVOS

Los entornos virtuales masivos multiusuarios (MMVEs) en la actualidad han atraído la atención de las personas ya que entrega un concepto de la inclusión de una web en 3D, para lo cual como punto de partida se ha estudiado el desarrollo de estructuras P2P (Peer-To-Peer), en donde el principal problema que ha surgido en esta rama es como lograr que los Avatares interactúen bien, es decir, lograr que los Avatares tengan un buen seguimiento, provisión de sus posiciones y movimientos a un nivel más amplio que una área local sino a un nivel mundial.

Un sistema MMVE se basa en la superposición de un sistema P2P en una estructura de máquinas para el usuario sobre un servidor. Así Esch y Botev presentan dos trabajos en uno se analiza el diseño de la red en y transmisión de datos se puede apreciar en [11]; mientras que en [12] se explicara la gestión del Avatar y el servidor.

El usuario enviara datos que contienen la información referente a su posición para generar movimiento en el Avatar, dichos datos se enviaran a un servidor el cual realizara el seguimiento de los mismo para genera un movimiento fluido sin que este

se descontinúe, así este servidor permite interconectar clientes brindándoles la sensación de proximidad con otros, para evitar que se produzca esto el usuario tras enviar sus datos iniciales solo enviara actualizaciones de los mismos. En la figura 1.4 se puede apreciar el seguimiento que realizara el servidor al Avatar y usuario.

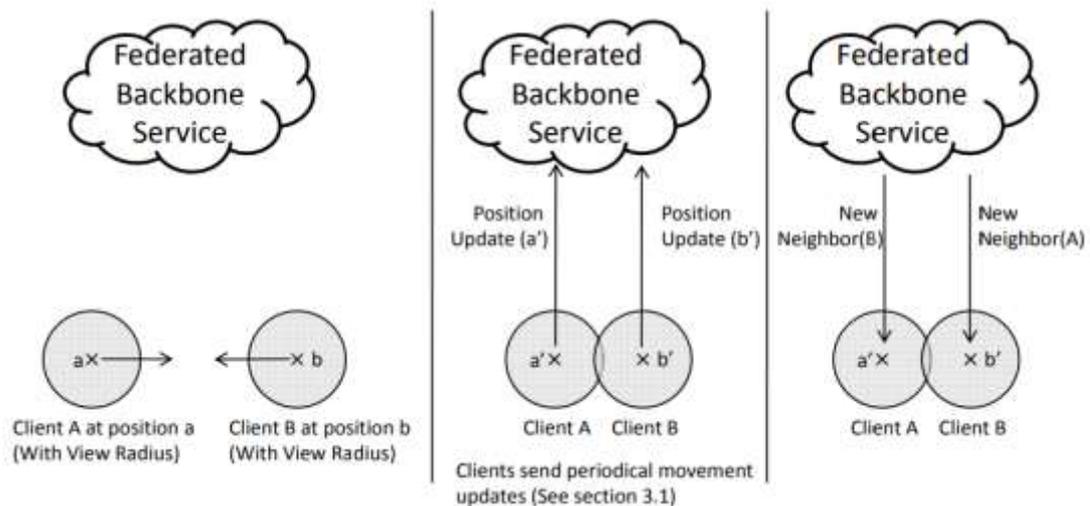


Figura 1.4 Seguimiento del Avatar realizado desde el servidor
Fuente: M. Esch & J. Botev [12]

1.3.1 AVATARES E INTERACCIÓN EN AMBIENTES VIRTUALES

Una de las principales aplicaciones de los Avatares o caracteres virtuales que se encuentran en un ambiente virtual de menor tamaño a los MMVEs, es la adaptación de estos personajes en escenarios de entrenamiento y rehabilitación. [13]

Los Avatares y ambientes virtuales presentan flexibilidad para crear un entorno o ambiente específico para una determinada tarea, pudiendo estos ser controlados por una persona o ya sea mediante un ordenador. Debido a los avances que se han realizado en el campo de la inteligencia artificial se ha logrado realizar personajes que realizan una interacción más realista con su entorno y personas.

Partiendo de este preámbulo Moghimi & Stone proponen un sistema denominado AMITIES (Avatar-Mediated Interactive Training and Individualized Experience System) que se puede definir como un sistema de entrenamiento basado en el uso de Avatares.

AMITIES se puede considerar como un sistema que vincula tres componentes que se involucran en la interacción: los Avatares, los habitantes del entorno y los

participantes. En donde las interacciones se pueden dar de diversas formas ya sea de una persona a varios Avatares o de un Avatar a varias personas.

A partir de este método de interacción se puede diferenciar la existencia de dos términos que denotan el funcionamiento del sistema Avatar y agente. Un Avatar es una entidad controlada por una persona, mientras que el agente es controlado mediante un algoritmo de computadora.

El sistema AMITIES está orientado a trabajar en ambientes que brinden un soporte o ayuda a las personas como en la educación, entrenamiento de habilidades y rehabilitación. Según el ambiente en el que se desea laborar el sistema se comportara de una determinada forma empleando diversos caracteres o personificaciones de los Avatares. En las figuras 1.5 y 1.6 se podrá apreciar los diversos Avatares desarrollados en el sistema AMITIES y un ejemplo de entorno desarrollado para la educación.

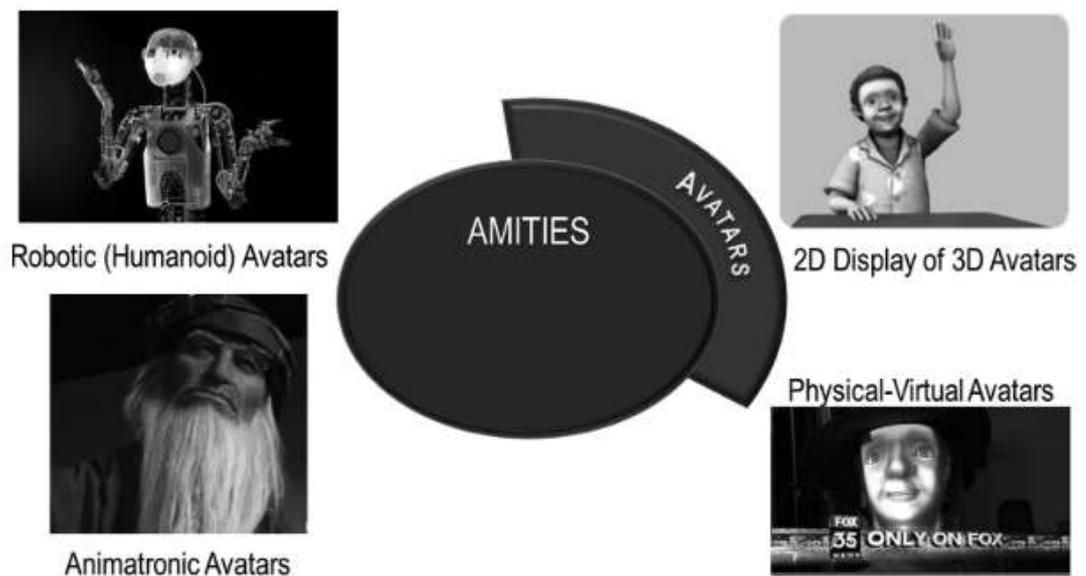


Figura 1.5 Avatares empleados en el sistema AMITIES
Fuente: M. Moghimi & R. Stone [13]



Figura 1.6 Clase virtual para cinco estudiantes controlada por un agente
Fuente: M. Moghimi & R. Stone [13]

1.4.1 AVATARES Y APLICACIONES MULTIMEDIA

Debido a los avances en el procesamiento de información, así como en el desarrollo de aplicaciones y comunicación de datos, se ha logrado la introducción de los Avatares en aplicaciones multimedia. Una característica de las aplicaciones multimedia basadas en Avatares está en el uso de interfaces que emplean la visión y voz.

Entre estas aplicaciones resaltan dos escenarios el video chat y la atención al cliente [14]. La brecha existente entre las capacidades de los sistemas y la capacidad humana para hacer un uso adecuado de las mismas se está ampliando. Para lo cual una solución a esta problemática es el desarrollo de interfaces más intuitivas para el usuario permitiendo al usuario aprovechar al máximo estas tecnologías y a su vez que el usuario elija como interactuar con el sistema.

En la figura 1.7 se puede apreciar un diagrama base de un call center basado en el uso de Avatares entre los dos usuarios del sistema.

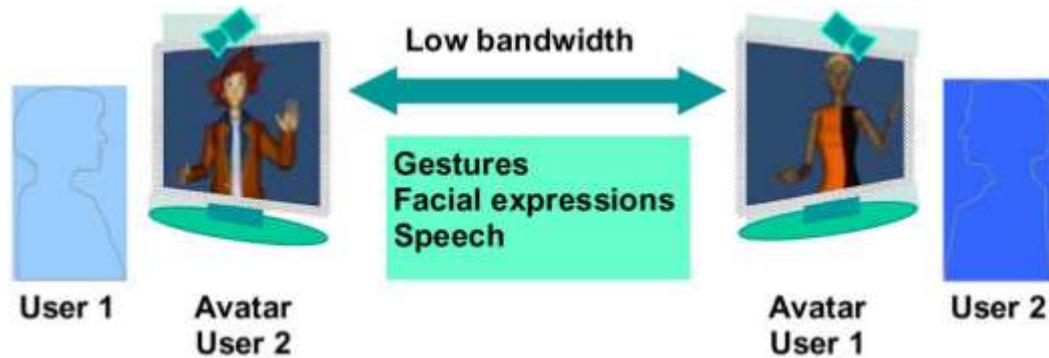


Figura 1.7 Diagrama de bloques de un call center basado en Avatares
Fuente: O. Schreer & R. Englert [14]

Como se mencionó anteriormente se busca generar interfaces que pueda emplear el usuario y a su vez que puedan correr sobre varios sistemas con diferentes prestaciones, para lo cual la tarea del diseñador recae en generar un sistema que cubra únicamente las necesidades del usuario. Así en el caso antes planteado el Avatar debe incluir una detección facial y de las manos, ya que estos dos elementos son de gran importancia al momento de brindar alguna información.

1.2 VISIÓN POR COMPUTADORA

La visión por computadora o también denominada visión artificial consiste en la deducción automática de la estructura y propiedades de un mundo tridimensional, posiblemente dinámico, a partir de una o varias imágenes bidimensionales. [15]

La visión artificial abarca diversos procesos como son la captación de imágenes, transformación y procesamiento de las mismas, buscando convertir una toma o escena de la vida, en una representación numérica.

Para iniciar la comprensión de la visión por computadora se debe definir antes un parámetro importante para esta, como es el término imagen. Una imagen es una representación de una escena del mundo tridimensional en una forma bidimensional. Las imágenes son captadas por un sensor que convierte la información de un espectro electromagnético, como es la luz, en una representación numérica.

Las imágenes se representan mediante una matriz de un determinado tamaño en donde cada elemento representa un valor de un espacio de color, comúnmente se emplea el RGB (Red Green Blue).

Al abordar el tema de visión artificial se puede abarcar una gran cantidad de aplicaciones, entre las que destacan las técnicas de procesamiento de imágenes, a continuación, se detallarán brevemente algunos trabajos relacionados a esta rama y afines al proyecto realizado.

La detección de expresiones faciales es una de estas aplicaciones, de las que se puede apreciar de una manera concisa en el trabajo realizado por B. Fasel, en [16] indica un método óptimo para esta tarea. Se puede apreciar tres etapas en este proceso. Una etapa inicial de adquisición del rostro en donde la imagen que contiene el rostro es normalizada y segmentada para su siguiente etapa que es la extracción de características faciales, la que se basa en el análisis de las imágenes empleando redes neuronales, en la cual se extraen las características rostro ya sea mediante reconocimiento de patrones o extracción de puntos característicos. Para ser procesados en una etapa final de clasificación en donde cada representación indica una acción facial que será interpretada como una emoción en su etapa final. En la figura 1.8 se presenta el modelo presentado por Fasel.

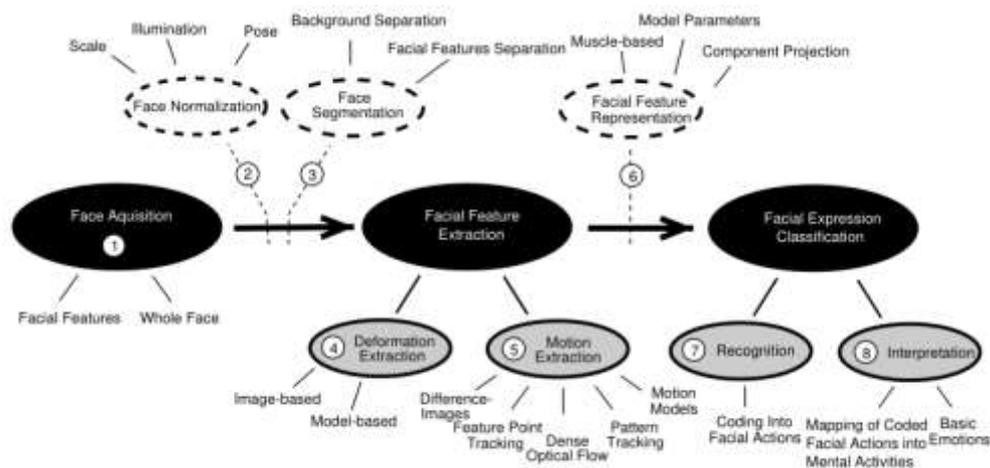


Figura 1.8 Estructura base del análisis de expresiones faciales
Fuente: B. Fasel [16]

La estimación de la pose de una persona tiene un principio similar al análisis de expresiones faciales. En este proceso se busca determinar o identificar como el cuerpo de una persona está en una determinada escena o imagen. Para este proceso es común el uso de un modelo geométrico del cuerpo, el uso de este modelo tiene el

propósito analizar y comprender al cuerpo humano y su movimiento. El modelo mencionado puede emplearse ya sea en dos o tres dimensiones, según la aplicación que se desee realizar. Este proceso se puede aplicar a un reconocimiento estático o dinámico del cuerpo, variando la complejidad de los datos que emplean estos sistemas; como resultado de este proceso es una imagen en donde cada pixel contiene el número de movimientos realizados a partir de la imagen raíz. [3]

En las siguientes subsecciones se detallarla ciertas técnicas empleadas en la visión artificial que se consideraron para el desarrollo del presente proyecto, así como las características que definieron su uso en el mismo.

1.3 FILTRO HAAR

Los filtros Haar están conformados por una red neuronales, empleados para la detección de objetos, con un procesamiento rápido de imágenes [17] el cual se basa en la introducción del término de imagen integral, la que contiene todas las características primordiales que definirán al detector o filtro, permitiendo procesar a una imagen con pocas operaciones por pixel.

El detector Haar se encuentra basado en el AdaBoost o adaptaciones que mejoran el modelo original propuesto por Viola como se indica en [18]; el que se basa en la selección de un pequeño número de características visuales de un gran número de clasificadores, en donde se van descartando ventanas que no contienen información útil para el clasificador.

Un aporte importante que realizan Viola y Jones es un método de combinación de clasificadores en una estructura en cascada, los que a diferencia de sus partes individuales pueden realizar una tarea más compleja e incrementar la velocidad con la que el sistema procesa la información hasta llegar a brindar un resultado útil.

El clasificador propuesto por Viola y Jones se basa en un entrenamiento supervisado por el operador o diseñador. El termino supervisado hace referencia a que el diseñador seleccionara los elementos que se emplearan en el enteramiento

En el ámbito de la detección de rostros al emplear dos características de tipo Haar se puede obtener un 1% de falsos negativos y un 40% de falsos positivos, para lo cual

trabajo realizado por Viola busca reducir este resultado empleando el detector en cascada antes mencionado.

La detección del objeto se basa en la extracción de características mas no en el análisis de cada pixel directamente. Un pixel contiene únicamente los valores de intensidad del espacio de color empleado, si suponemos que se emplea el espacio de color RGB, el pixel estará representado por un valor de intensidad para cada color que compone este espacio; mientras una característica puede contener mayor información que la intensidad, pudiendo ser la relación existente entre varios pixeles o la información de una ventana de pixeles sobre una máscara binaria como el caso del cálculo de los patrones binarios locales (LBP).

El valor de una característica se puede definir según áreas rectangulares, la característica de dos rectángulos es el valor de la diferencia entre la suma de dos pixeles dentro de dos áreas rectangulares, mientras que la de tres rectángulos está dada por la suma de dos rectángulos exteriores que se resta la suma del rectángulo central. La característica más compleja es la de cuatro rectángulos la que se calcula mediante la diferencia entre las diagonales del rectángulo, definiendo así a un detector con un tamaño de 24x24.

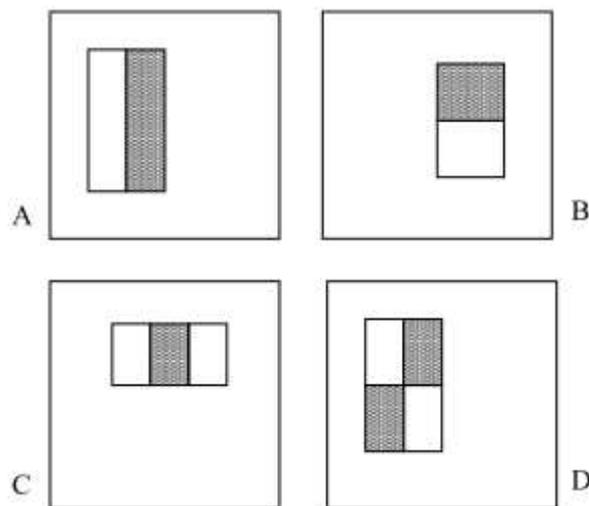


Figura 1.9 Características de Haar
Fuente: P. Viola & M. Jones [17]

Mediante las características antes mencionadas se puede obtener la imagen integral, la cual contendrá la información de los pixeles que componen esta característica.

La imagen integral se basa en la sumatoria de los pixeles adyacentes al punto que se desea analizar, Así el valor en un punto está dado por los pixeles más cercanos a este en la imagen original, que formaran posteriormente una imagen que contendrá estos valores.

Así mediante la imagen integral un área rectangular puede ser representada en cuatro subáreas. En donde cada área intervendrá en el cálculo de la imagen integral del área más próxima a esta. De la figura 1.10 se puede decir que el valor de las imágenes integrales de cada sub área será:

- El valor de 1 está dado por la suma de los pixeles que conforman al área A
- El valor de 2 está dado por la suma de los pixeles que conforman al área A y B
- El valor de 3 está dado por la suma de los pixeles que conforman al área A y C
- El valor de 4 está dado por la suma de los pixeles que conforman al área A, B, C y D

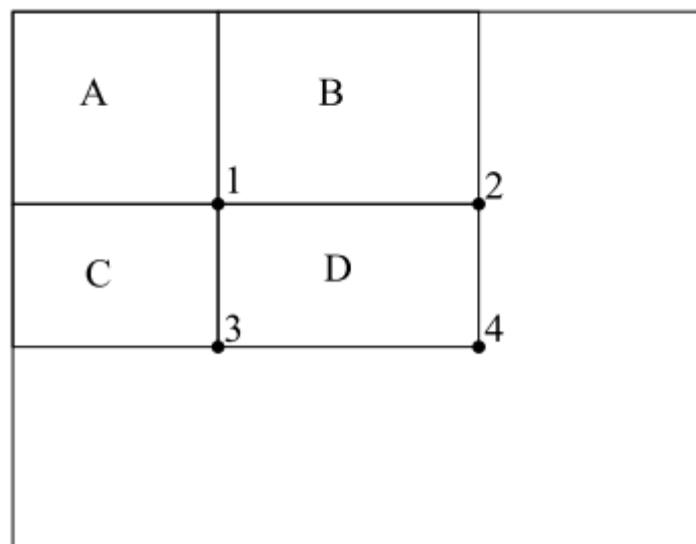


Figura 1.10 *Subáreas que conforman la imagen integral*
Fuente: P. Viola & M. Jones [17]

A diferencia de los filtros, las características rectangulares, aunque sensibles a los bordes, proporcionan una representación de la imagen más rica para procesos de aprendizaje en redes neuronales.

La detección de objetos basada en los principios antes mencionados se caracteriza por el uso de ventanas que van comparando las diversas regiones de una imagen hasta encontrar una que guarde un determinado parecido con las provenientes de los

patrones de entrenamiento, para la obtención de la ROI (Region of Interest) se analizan las imágenes en diferentes cuadrantes por las cuales se emplearan ventanas de diferentes tamaño, las cuales descartaran los cuadrantes que sean considerados como negativos, es decir no contengan información relevante, hasta obtener la ROI.

En la figura 1.11 se puede apreciar el ejemplo diseñado por Viola en el diseño de la cascada Haar.

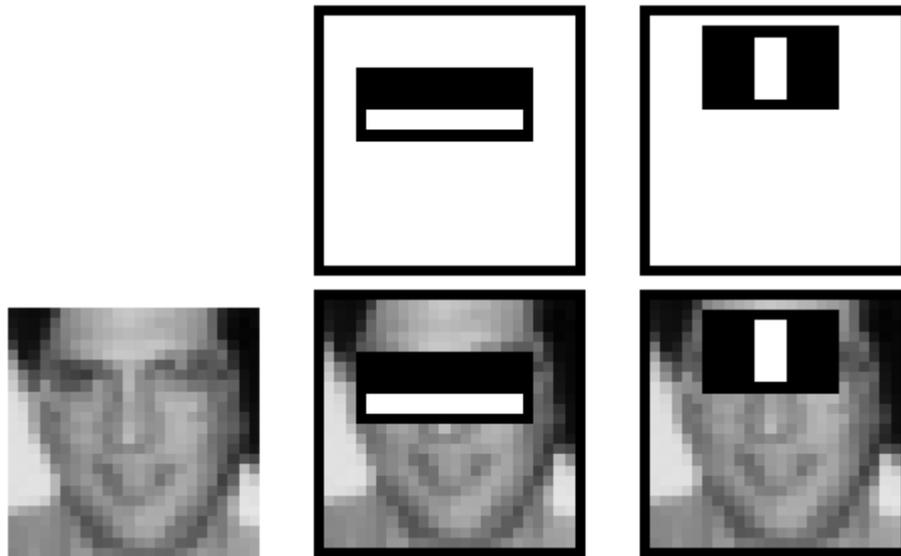


Figura 1.11 Características rectangulares aplicadas sobre una ventana
Fuente: P. Viola & M. Jones [17]

Como se mencionó en un principio este detector se basa en un sistema en cascada de detectores más sencillos, los cuales tomaran las diversas ventanas con tamaños diferentes las cuales serán procesadas por cada uno de estos, pudiendo presentar dos casos uno en que la ventana contenga información falsa y sea descartada, y otro verdadero en donde la ventana pasara al siguiente detector y así sucesivamente hasta alcanzar un resultado de mayor exactitud que al emplear un solo detector. En la figura 1.12 se presenta la estructura que sigue el detector en cascada propuesto por Viola.

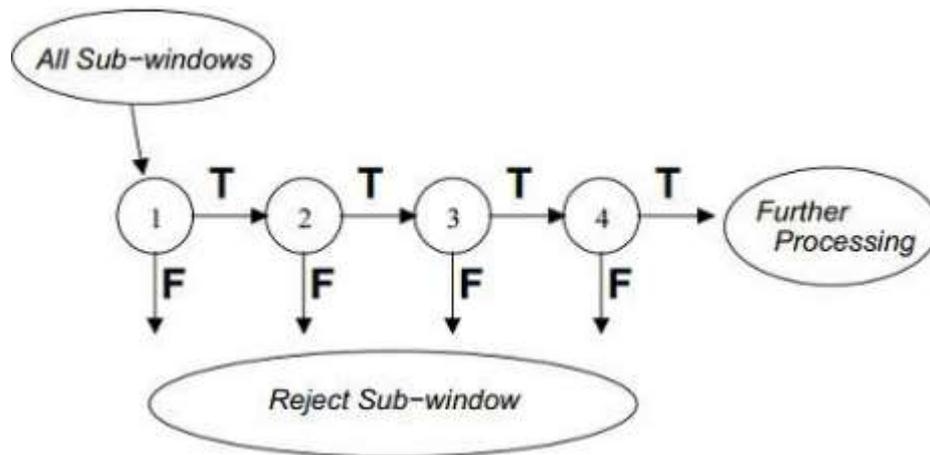


Figura 1.12 Estructura del clasificador propuesto por Viola y Jones
Fuente: S. Soo [19]

A partir de la publicación de Viola se han generado varias aplicaciones basadas en este detector como se puede apreciar en: [19] [20] [21] [22]

La captura de las características faciales ya sean las diversas poses que puede generar un modelo se debe replicar mediante una representación digital de este, la cual debe imitar o realizar movimientos a esto de lo denomina animación facial.

El proceso de captación de esta información ya ha sido solventada por los procesos anteriores, actualmente se emplea las redes neuronales en este fase ya que estas permiten al diseñador poseer flexibilidad en el diseño de sus valores de salida y a su vez permiten poseer un mayor grado de robustez, siendo la influencia de la luz la principal interferencia o ruido presente en este proceso ya que esta puede alterar las propiedades de los datos de ingreso, pudiendo aumentar o disminuir la intensidad de los diversos tonos de colores presentes en la imagen.

1.4 LOCAL BINARY PATTERN (LBP)

LBP es un operador de textura simple y eficiente que etiqueta cada píxel de la imagen analizando su vecindario, estudiando si el nivel de gris de cada píxel supera un determinado umbral y codificando dicha comparación mediante un número binario. Debido a su bajo coste computacional y al gran poder discriminativo que ha demostrado tener, LBP se ha convertido en los últimos años en una de las soluciones más utilizadas en numerosas aplicaciones relacionadas con textura. La característica más importante de LBP en aplicaciones reales es la robustez que ofrece frente a

variaciones en la intensidad del nivel de gris, causado, entre otras muchas cosas, por diferencias en la iluminación. [23]

Para calcular el valor de LBP se debe considerar los siguientes valores:

- Numero de pixeles vecinos P
- Tamaño del vecindario R
- Valor del pixel central g_c
- Valor del pixel a analizar g_p

Conociendo estos términos mediante la siguiente ecuación se puede obtener el valor LBP que representara el área a analizar.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \quad s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

De la ecuación anterior se puede denotar que el valor de la función s puede tomar un valor binario en donde si el pixel que se va a analizar es mayor al del central tomará un valor de uno, caso contrario será cero. El cual tomara el valor de una máscara binaria definida por el diseñador, el cual puede determinar la ubicación del pixel inicial de análisis en el vecindario, pero es recomendable que este criterio se conserve en el análisis de imágenes de mayor escala ya que se obtendrá un valor normalizado.

En la figura 1.13 se puede apreciar el cálculo del valor LBP en un vecindario de 8 pixeles, los que están denotados de color rojo.

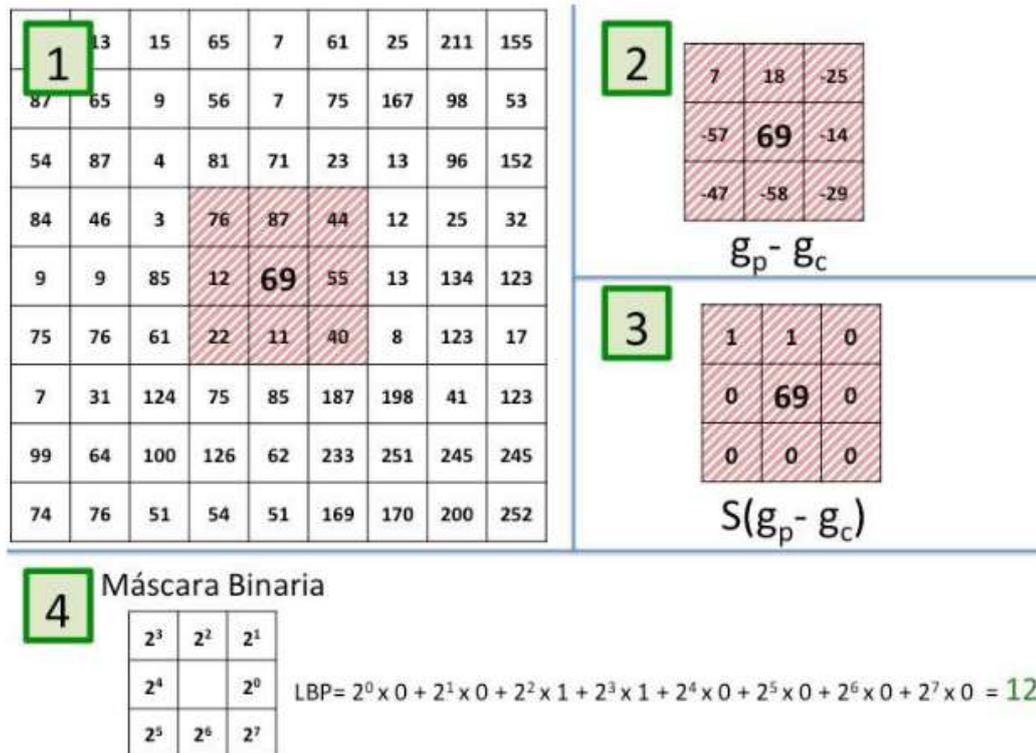


Figura 1.13 Pasos para obtener el valor LBP de un pixel
Fuente: O. García-Olalla [23]

De la figura 1.13 se puede concluir cuatro pasos para la obtención del valor LBP de un pixel.

1. Determinar el pixel a analizar o pixel central y su vecindario
2. Obtener la diferencia del valor entre el pixel a analizar del vecindario y el pixel central
3. Si este valor es mayor a cero tomará un valor de 1 caso contrario será 0
4. Multiplicar el valor obtenido por la máscara binaria a emplear.

1.5 RED NEURONAL ARTIFICIAL RNA

Las redes neuronales artificiales están inspiradas en el sistema nervioso biológico, partiendo de la unidad de este, la neurona, de la conexión de diversas neuronas se determina el comportamiento del sistema.

Las redes neuronales se basan en generalizar información extraída de datos. [24] Las redes neuronales se entrenan de modo que una determinada entrada conduce a una respuesta basada en los datos que se emplearon en su entrenamiento.

El entrenamiento de una red neuronal puede ser supervisado o no, en donde en el primer caso el diseñador le brinda a la RNA los parámetros de entrenamiento, es

decir le indica que respuesta debe dar está a un determinado estímulo en su ingreso; mientras en el caso de un aprendizaje no supervisado la red aprende por sí misma.

El entrenamiento de estas redes comúnmente está basado en el algoritmo de retro propagación del error que emplea el método de gradiente descendiente, en donde los pesos de la RNA se actualizan mediante el uso de los datos de entrenamiento y la comparación de estos con el resultado de la RNA.

1.1.5 PERCEPTRÓN

Para comprender una red neuronal se debe iniciar conociendo la unidad de esta, que es el perceptrón. El perceptrón consiste en la suma de las señales de entrada en producto con los pesos sinápticos. En la figura 1.14 se puede apreciar la estructura de un perceptrón.

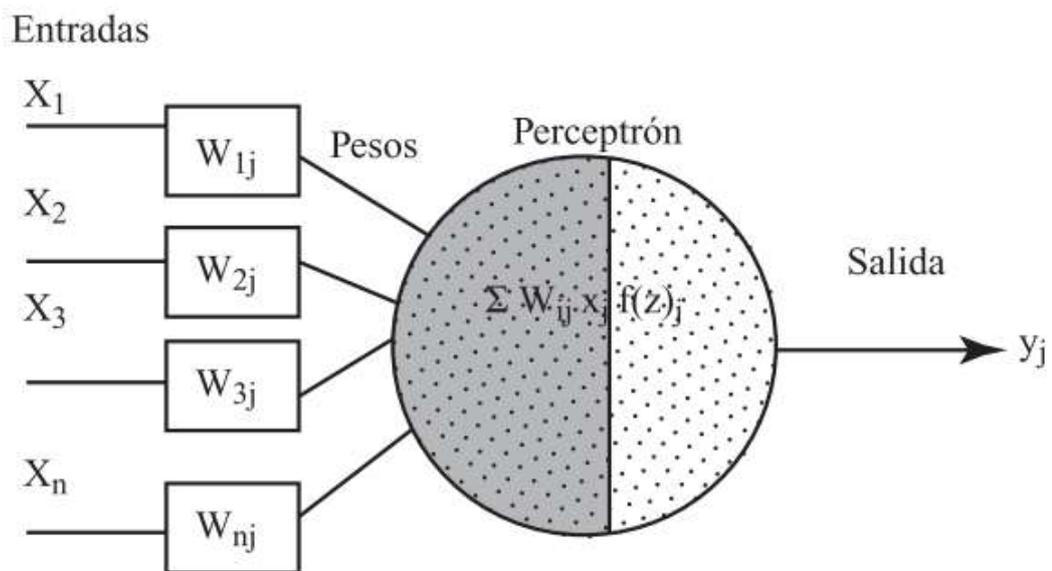


Figura 1.14 Estructura de un perceptrón
Fuente: P. Ponce [24]

En donde la salida del perceptrón estará dada por el valor de una función de activación del producto del ingreso o entrada con el peso sináptico, adicionalmente las redes neuronales pueden contar con el bias que es un valor que se agrega dentro de la función para generar que la salida tenga un valor con un desplazamiento, En la figura 1.15 se puede apreciar una neurona artificial con y sin bias y su representación matemática.

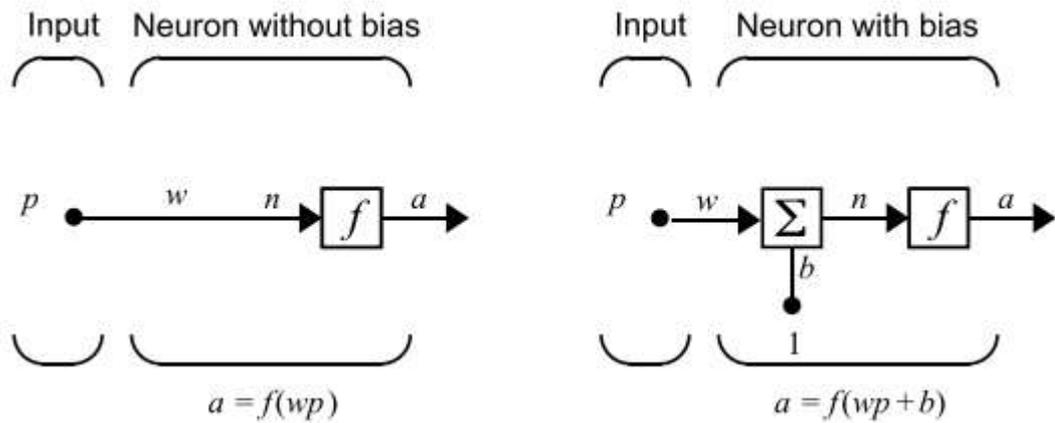


Figura 1.15 Estructura de una neurona artificial
Fuente: H. Demuth [25]

1.2.5 FUNCION DE ACTIVACIÓN

La función de activación determina el nivel de activación que tendrá una neurona artificial en su salida, basándose en el valor en la entrada de esta. Estas funciones pueden ser sigmoideal, tangente hiperbólica, tansig, hardlim, purelin entre otras.

La función escalón se emplea a neuronas binarias, es decir esta brindara dos estados únicos en su salida ya sea 0 o 1, en el caso de un escalón simétrico 1 y -1

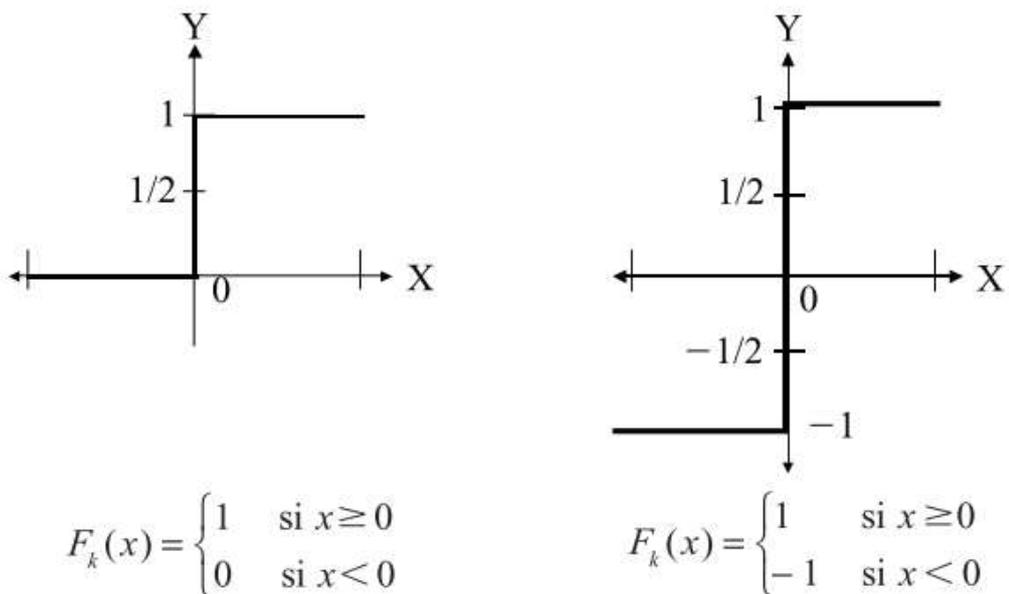


Figura 1.16 Función de activación escalón
Fuente: P. Ponce [24]

La función de activación lineal o identidad es aquella que en su salida refleja el valor que está en su ingreso, a estas también se les puede aplicar límites con los cuales se estable un rango que puede tomar la salida.

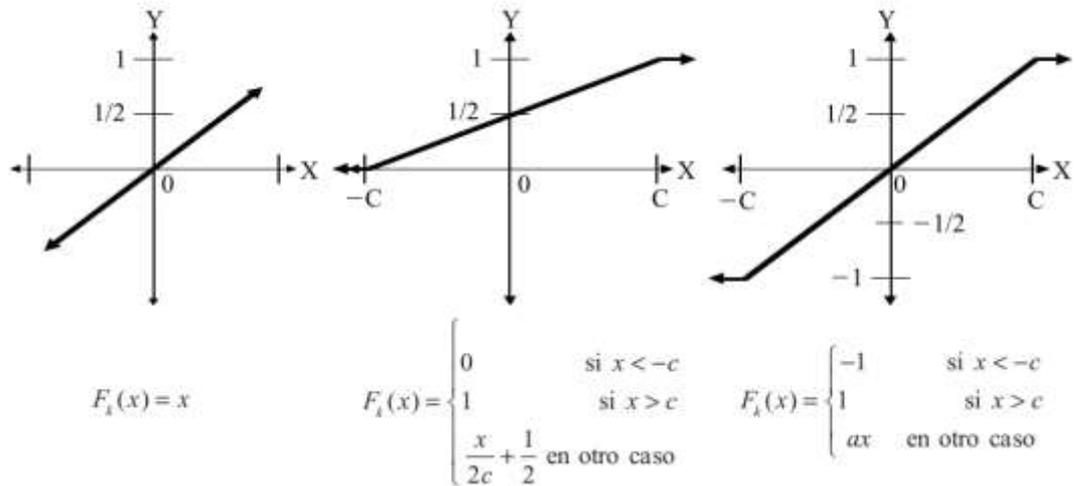


Figura 1.17 Función de activación lineal

Fuente: P. Ponce [24]

La función de activación tangente hiperbólica se aplica a ingresos que presentan variaciones sutiles ya sean positivas o negativas.

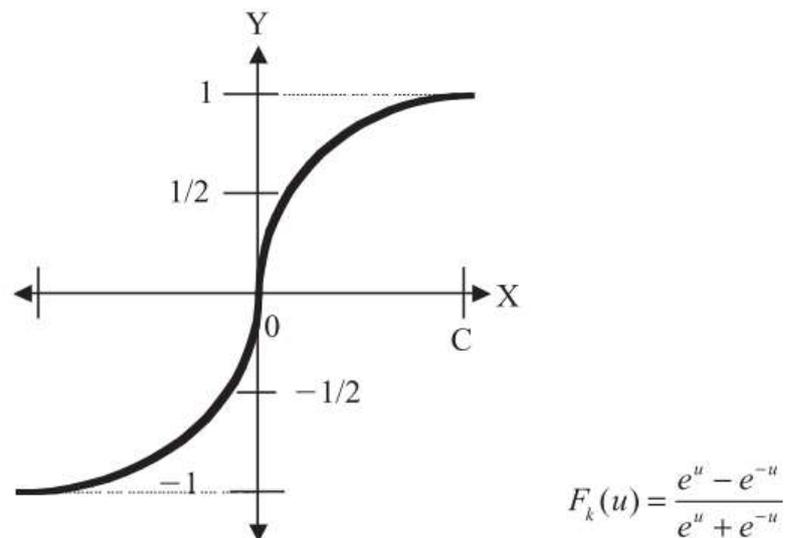


Figura 1.18 Función de activación tangente hiperbólica

Fuente: P. Ponce [24]

La función sigmoïdal da un valor cercano a un valor asintótico, permitiendo que el valor en su salida este comprendido en la zona alta o baja de la misma.

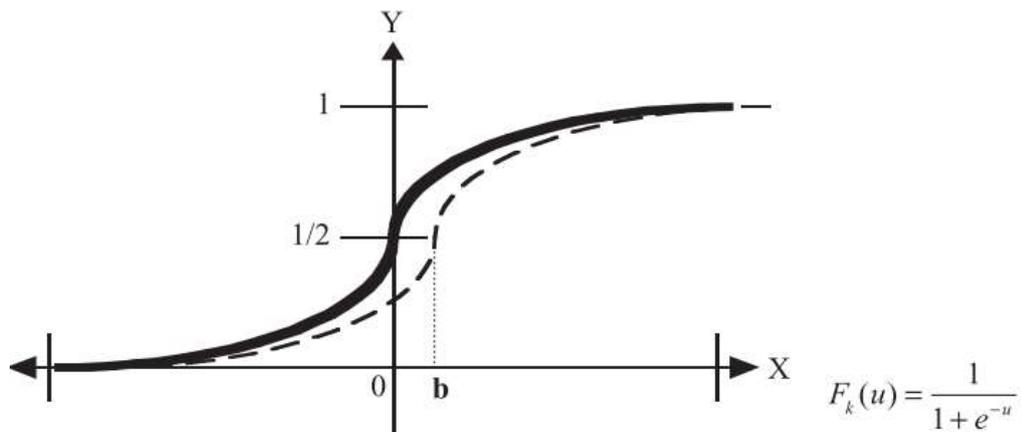


Figura 1.19 *Función de activación sigmoïdal*
Fuente: P. Ponce [24]

1.3.5 TOPOLOGIA DE UNA RED NEURONAL

Una red neuronal se puede definir como el conjunto de neuronas artificiales conectadas entre sí que se transmiten información. Una red neuronal está conformada por varias capas una de ingreso en la cual están los estímulos que generaran una respuesta en su capa de salida y capas ocultas que realizaran la generalización de los ingresos para estimar la salida.

En la figura 1.20 se puede apreciar una representación de una red multicapa, con la cual se puede comprender la relación entre las diversas capas. Partiendo de una capa de ingreso que tomará los valores de P y una primera capa los procesará y dará un resultado a1 en su salida, la cual se considera como el ingreso de la siguiente capa y así sucesivamente hasta llegar a su capa de salida.

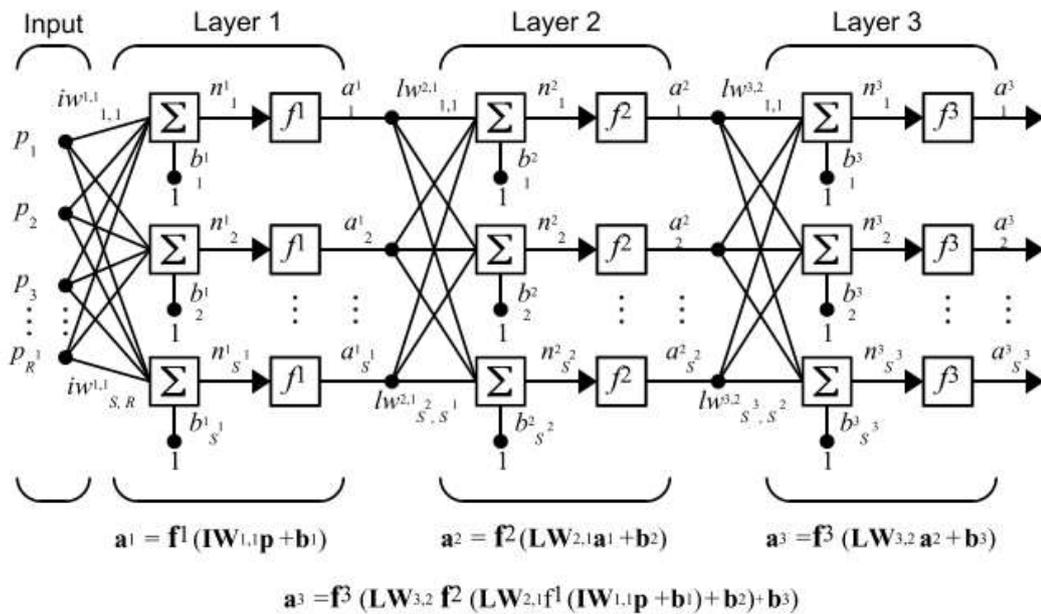


Figura 1.20 Representación de una neurona multicapa
Fuente: H. Demuth [25]

Una RNA sin importar el número de capas e ingresos o salidas que presente contara con los siguientes elementos:

- Neuronas o unidades de procesamiento
- Pesos sinápticos o conexiones entre unidades
- Funciones de activación en cada neurona

1.6 LANDMARK

El termino en facial landmark hace referencia a puntos faciales que determinan sector del rostro de una persona. Al hablar de este tema no se habla únicamente de determinar la ubicación de un determinado rasgo facial, sino de la estimación de la pose de la cabeza reconocimiento de diversos atributos faciales.

Así mediante la ubicación de estos puntos se puede determinar si una persona abre la boca, determinar la distancia entre sus pupilas entre otros atributos del rostro de una persona.

Los puntos faciales de mayor interés en diversos casos de estudio han sido los ojos y los extremos de la boca, puesto que estos dos elementos son los que transmiten mayor información sobre un mensaje que desea transmitir una persona.

Conociendo los puntos de mayor análisis y utilidad se han generado sistemas que determinan la ubicación de estos puntos, las cuales en su mayor parte basan su funcionamiento en redes neuronales variando su arquitectura para obtener resultados en un periodo de tiempo menor.

El número de puntos faciales dependerá primordialmente de la tarea a realizar pudiendo partir de dos puntos para los ojos hasta 64 puntos que cubren el contorno del rostro. En las figuras 1.21, 1.22 y 1.23 se puede apreciar los resultados obtenidos por varios autores en esta rama.



Figura 1.21 *Resultados de detección de puntos faciales de Zhang*
Fuente: Z. Zhang & P. Luo [26]

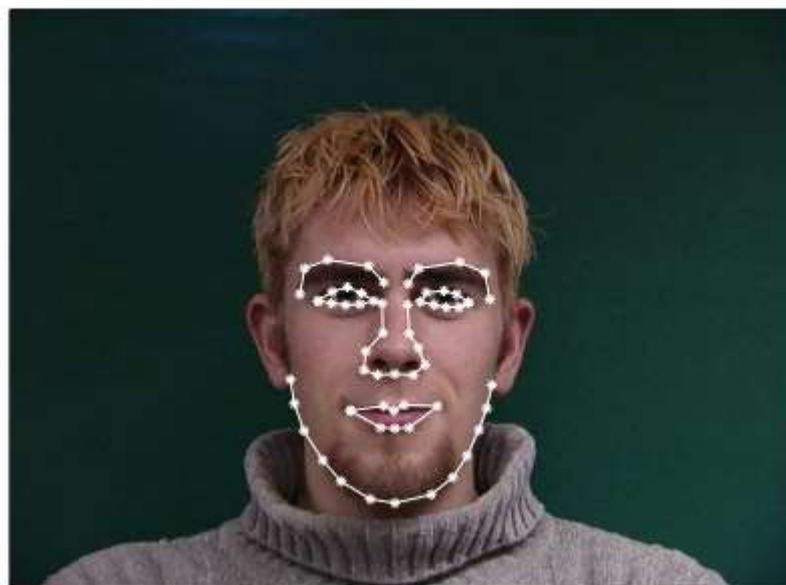


Figura 1.22 *Resultados de detección de puntos faciales de Nordstrom*
Fuente: M. Nordstrom [27]

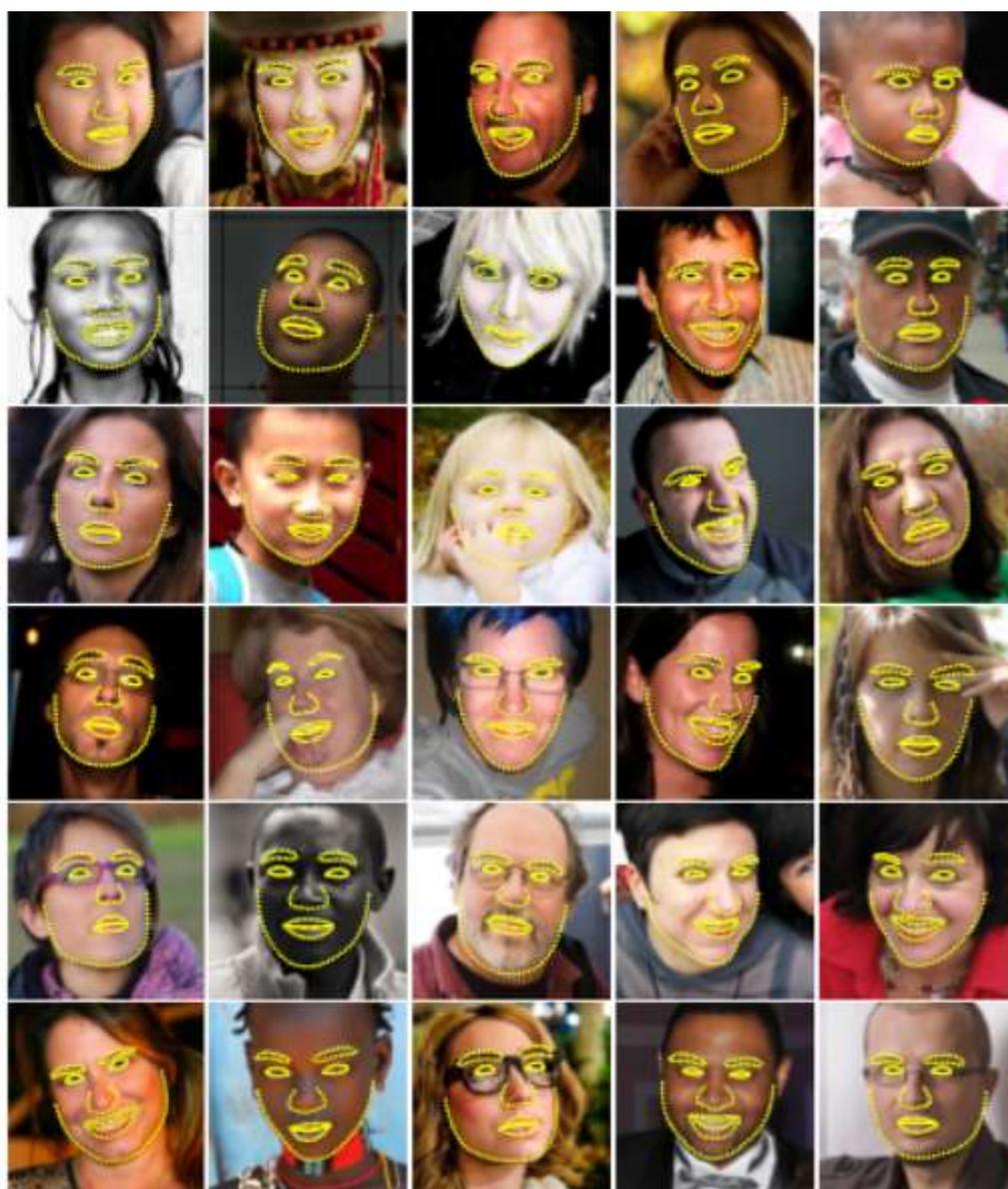


Figura 1.23 Resultados de detección de puntos faciales de Kazemi
Fuente: V. Kazemi & J. Sullivan [28]

1.7 ANIMACIÓN FACIAL

La animación facial ha tenido gran acogida en los últimos años a partir de los diversos avances en la rama de la computación, que han producido equipos capaces de generar y procesar grandes cantidades de información y trabajar de manera paralela con varios procesos, permitiendo tener un mejor detalle en la animación generando mejores texturas y por consiguiente un mejor acabado en la salida de este

proceso. En la figura 1.24 se puede apreciar los resultados obtenidos por los investigadores de NVIDIA y la Universidad de California.

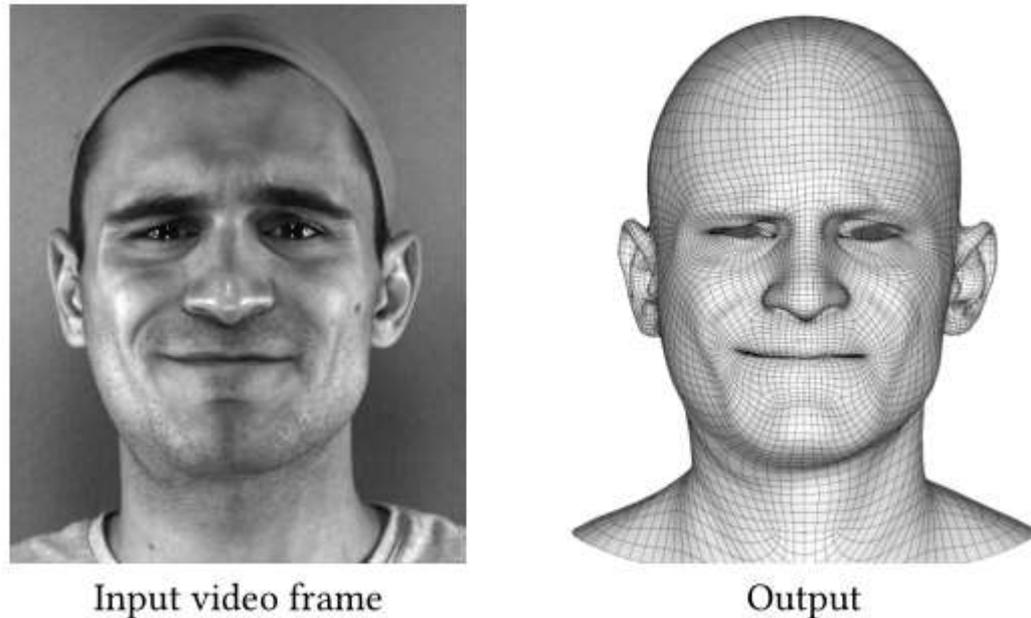


Figura 1.24 *Resultados de animación facial de Laine*
Fuente: S. Laine [4]

De manera similar a los filtros Haar se han generado varias aplicaciones y algoritmos, los cuales poseen un principio similar pero diversas etapas en la fase de procesamiento de la información ya sea mediante técnicas de procesamiento de imágenes o mediante redes neuronales, en los siguientes artículos se podrá apreciar trabajos relacionados: [29][30][31][32][33].

1.8 SISTEMA PROPUESTO

Diversas incapacidades y trastornos impiden que la interacción entre personas se pueda realizar de una manera cómoda y amigable, tal es el caso del autismo, un trastorno caracterizado por la dificultad en los procesos de socialización, comunicación e imaginación; a partir de esta problemática nace el proyecto AVATAR, el cual busca facilitar la inclusión de los niños con autismo en la sociedad.

Para solventar esta problemática, como ya se ha mencionado, el presente proyecto busca generar un sistema capaz de imitar las diversas gesticulaciones que realice un individuo mediante un Avatar. El sistema propuesto se basa en el uso de redes neuronales, que controlaran las etapas de procesamiento de información; el sistema

cuenta con diversas etapas las cuales parten desde la captación de las imágenes hasta la generación del movimiento del Avatar, a continuación, se muestra un diagrama de bloque que describe de manera general su funcionamiento y los conceptos involucrados en este.

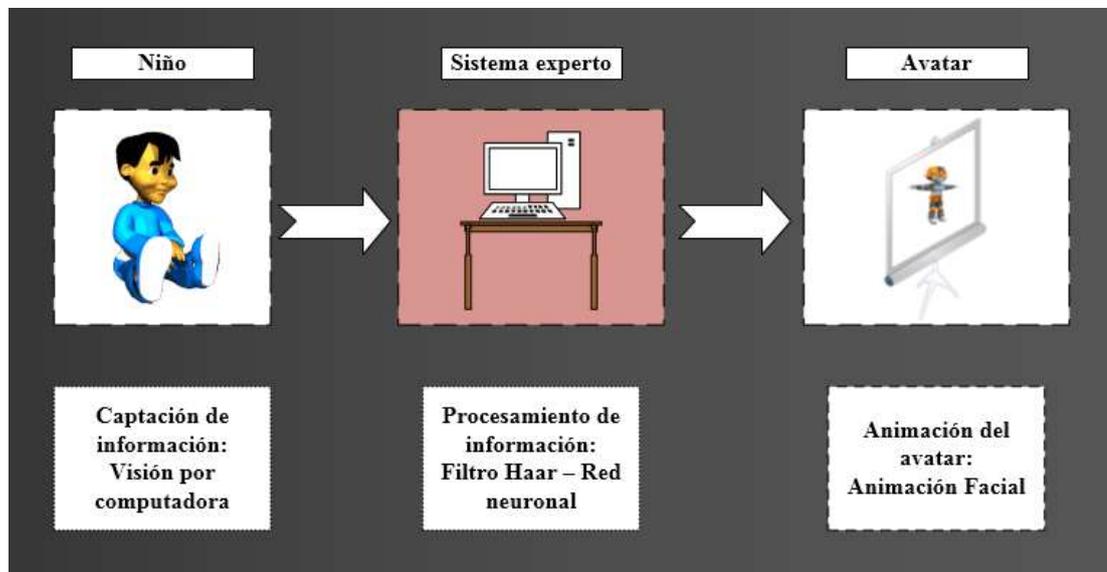


Figura 1.25 Diagrama de bloque general del sistema propuestos

Como se pudo apreciar en la figura 1.25 el proyecto está basado en la implementación de los siguientes conceptos:

Visión por computadora: fase encargada de la captación de la información necesaria, para esta etapa se empleará la webcam propia del ordenador.

Filtro Haar: filtro de imágenes basado en regresares, es una aplicación de la visión por computadora, mediante la cual se obtendrá la ROI (Region of Interest), es decir el área de la imagen correspondiente al rostro.

Red neuronal artificial: para el procesamiento de la ROI se empleará una red neuronal, cuyo entrenamiento se basa en patrones diseñados y obtenidos para este caso, cuyo ingreso será las ROI obtenida del filtro Haar y su salida será un conjunto de valores que representes los diversos landmarks del rostro (puntos específicos que diferencian un área de otra).

Animación facial: mediante los datos resultantes de la red neuronal se diseñará un algoritmo capaz de generar los diversos movimientos en los Avatares, para lo cual se empleará el software BLENDER.

La visión por computadora se puede considerar como la etapa inicial o de captación de información para diversas aplicaciones, entre estas se puede resaltar el análisis de gestos, movimiento de extremidades, a su vez a partir de este se puede extraer varias características ya sea la detección del color de la piel, cálculo de la velocidad de movimiento extremidades o cuerpos. [34][35]

En el presente proyecto no se puede emplear los modelos de aplicación desarrollados en publicaciones como [34][35][36] ya que estas publicaciones se basan en la detección de la piel ya que este método detecta la ubicación de la piel en la ROI mas no permite analizar de una manera más directa los diversos gestos que se generan.

En el siguiente capítulo se podrá apreciar con mayor detalle cada una de las etapas y subsistemas que conforman el presente proyecto.

CAPÍTULO 2: DISEÑO E IMPLEMENTACIÓN DEL SISTEMA

Para la implementación del presente sistema se contemplaron varias sub etapas las cuales iniciaran con la imagen captada por la webcam y terminaran en el movimiento del Avatar. A continuación, se detallarán los procesos realizados en cada etapa.

2.1 CAPTACIÓN Y PROCESAMIENTO

En esta etapa se tomará la información del individuo y a su vez se procesará dicha información para su ingreso en el sistema experto, que está formado por la red neuronal interactuando con el software Blender.

2.1.1 CAPTACIÓN

En una fase inicial se tomará la información del entorno del sujeto a partir de una webcam, para lo cual se empleará la librería de OpenCV que permite controlar a la webcam del computador. En la figura 2.1 se puede apreciar un ejemplo de este procedimiento.



Figura 2.1 *Imagen captada por la webcam en la etapa de adquisición*

2.2.1 PRE PROCESAMIENTO

Una vez captada la imagen se debe cambiar al espacio de color a emplear es decir se convertirá la imagen RGB a su equivalente en escala de grises.



Figura 2.2 Imagen captada en escala de grises

Una vez se ha convertido la imagen a su equivalente en escala de grises se realizará la detección facial, para lo cual se empleará un clasificador Haar, con el cual se obtendrá una nueva imagen que contendrá únicamente el área en la que está el rostro.



Figura 2.3 Resultado de la detección facial basada en el filtro Haar

Tras la detección facial se debe escalar la imagen a un valor único, lo cual permitirá que esta pueda ser empleada en las siguientes etapas. Para el análisis de los puntos faciales en las siguientes etapas, se debe segmentar la imagen en tres subáreas según la posición del tabique, este valor se determinó estableciendo la posición del mismo en mil imágenes de las que se obtuvo su valor promedio. Así la primera área contendrá la imagen del ojo derecho, la segunda la del ojo izquierdo y la tercera contendrá la boca y mandíbula.

En la figura 2.4 se puede apreciar las áreas antes mencionadas.

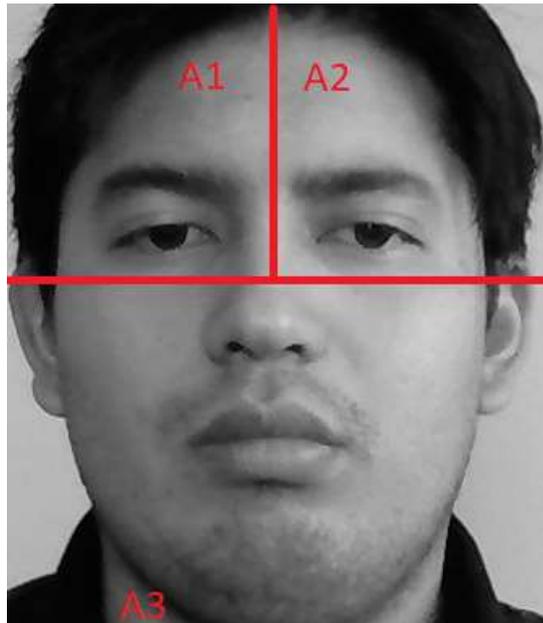


Figura 2.4 Áreas de la región de interés

2.2 PROCESAMIENTO

El procesamiento de la información captada mediante la webcam consta de dos etapas el cálculo del LBP y la red neuronal.

2.1.2 CALCULO DEL LBP

Mediante el cálculo del LBP se analizará la textura de cada vecindad, la que está conformada por un pixel central y ocho vecinos, a partir del resultado de este cálculo se formará una matriz que contendrá las características de textura de la imagen.

Al analizar el valor de texturas se elimina los efectos de los cambios de luz en la fotografía, debido a que se analiza la relación de intensidad existente entre el pixel central y sus vecinos, en el caso de variar la intensidad de la fuente de luz esta afectara a toda la imagen de manera uniforme.

Al emplear el análisis de las vecindades se logra que el sistema emplee menos características en el procesamiento de la red neuronal, pasando de un valor próximo a 2548 pixeles en la subárea uno a un valor de 272 características, reduciendo el valor de ingresos que poseerá la red neuronal a una décima parte.

En la tabla 2.1 se realiza una comparación de los elementos de análisis en las áreas de la región de interés.

Tabla 2.1 Elementos de análisis de las áreas de la ROI

Área	Numero de pixeles	Características LBP
1 Ojo derecho	2548	272
2 Ojo izquierdo	2303	256
3 Boca	4950	512

2.3 RED NEURONAL

El procesamiento que realiza el sistema experto se encuentra en la aproximación que realiza la RNA, tomando los datos de textura obtenidos mediante el cálculo LBP para interpretarlos como coordenadas de los puntos faciales.

2.1.3 CREACIÓN DE PATRONES DE ENTRENAMIENTO

Para el proceso de entrenamiento de esta red se empleó un total de 13 casos con una población variada de niños de 8 a 15 años, de los cuales 7 son mujeres y 6 varones.

Cada miembro de la población cuenta con un total de 1500 datos de entrenamiento, imágenes captadas por las etapas anteriores. En la sección de apéndices se puede apreciar una porción de los datos de entrenamiento.

Como se mencionó en las primeras secciones este proyecto es la base del proyecto AVATAR, para lo cual se diseñaron dos tipos de patrones uno que brinda las coordenadas de los puntos faciales y otra en la que se aproximara la ubicación de los huesos de animación en el Avatar en Blender.

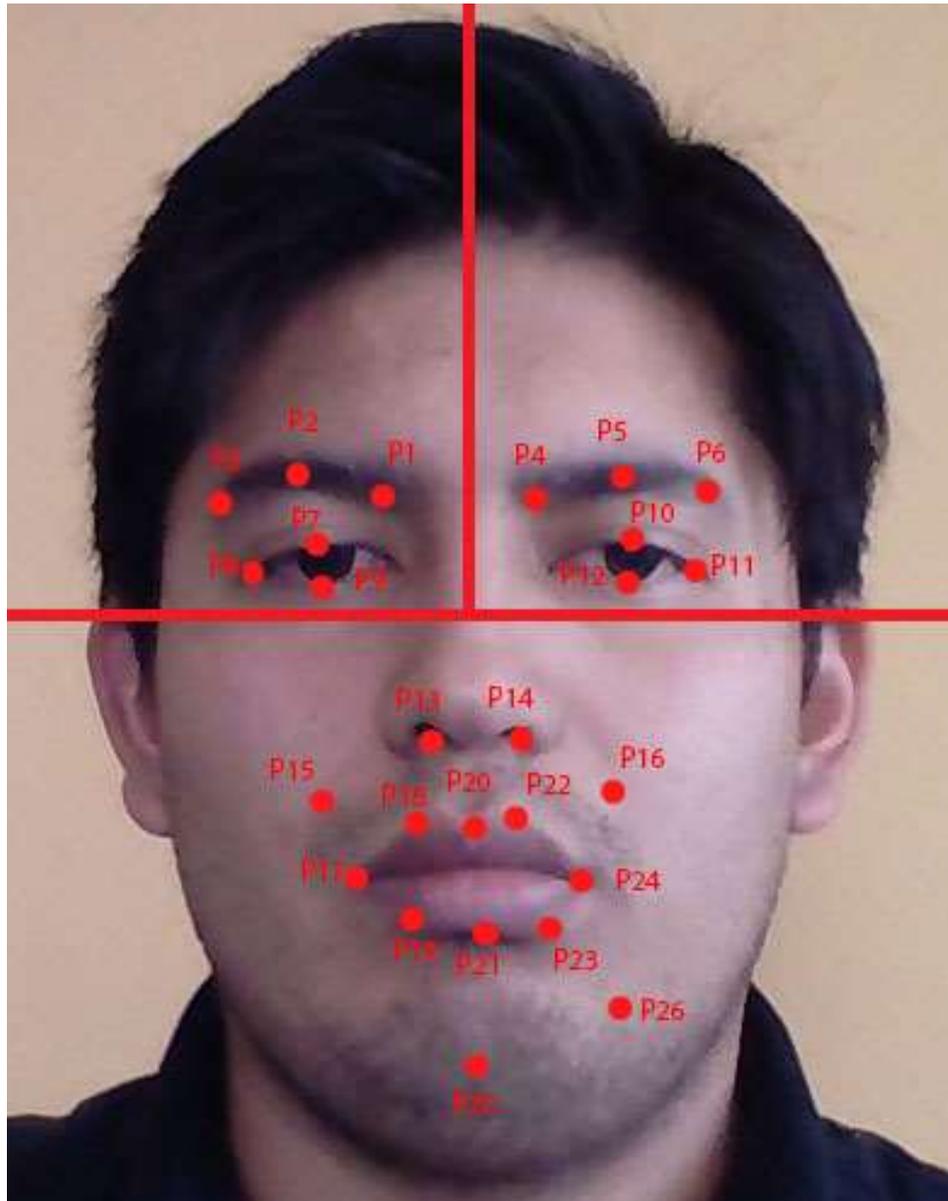


Figura 2.5 *Landmark aplicados sobre el rostro*

Los puntos faciales y la ubicación de los huesos del Avatar se establecieron de manera manual, es decir, se determinó la ubicación de cada uno de estos colocando puntos de referencia, para determinar de manera visual la ubicación de estos se empleó un software de edición fotográfica. En la figura 2.5 se indican los puntos faciales que estimara la RNA. Estos datos se almacenaron en un archivo que se empleara como el target o valor deseado para el entrenamiento de la red neuronal.

De un total aproximado de 34 mil imágenes para el entrenamiento, se clasificaron aquellas que no poseían información repetida, aquellas que se repetía una posición del rostro se descartaron, reduciendo así el valor de estas hasta alcanzar las 100

imágenes por cada caso. El sobrante de imágenes se empleó en las pruebas realizadas a la red neuronal.

2.2.3 DISEÑO Y ENTRENAMIENTO DE LA RNA

El diseño de una RNA recae primordialmente en la experiencia del diseñador, y a su vez a las limitaciones del hardware que se empleara en dicho proceso.

La principal limitante en el diseño de la red neuronal está en el hardware, ya que esta red neuronal realiza una aproximación compleja de la posición de los puntos faciales, ya que el entrenamiento se realizó en MATLAB, se tenía una diversidad al momento de crear la red, pero se debía considerar que la misma correría sobre Blender conjuntamente con la animación, por lo cual esta no debería emplear excesivo tiempo de procesamiento en el cálculo computacional que realiza.

Por lo cual se empleó un perceptrón multicapa, que sigue los principios indicados en el capítulo anterior, al tratar de aproximar un valor se empleó una función de activación lineal, lo que permite a la red brindar un gran rango de valores y no limitarse a un rango entre 0 y 1 como en el caso de los clasificadores.

El número de capas ocultas a emplear fue determinado por la capacidad de la maquina en el entrenamiento, ya que MATLAB internamente genera una matriz que almacena los datos del entrenamiento como los pesos y valores de bias que se realizan en cada iteración para reducir de mejor manera el error, basado en este criterio se determinó que el número máximo de capas ocultas que permitía para las tres áreas era de dos capas ocultas.

Para determinar el número de neuronas que tendría cada capa se emplearía la selección del error mínimo que presente la red variando el valor del número de neuronas y capas, concluyendo en un total de tres capas con cincuenta neuronas como base para cada área del rostro.

Se entrenaron un total de cinco redes neuronales que conformarían al sistema experto, mediante las cuales se logró minimizar aún más el error, ya que estas tratarían de estimar valores próximos entre sí. En la tabla 2.2 se puede apreciar la sección del rostro que cubriría cada red neuronal, así como los respectivos landmark que estas estimarían.

Tabla 2.2 Landmark que estimara cada RNA

RNA	Estima la posición de
1	Ceja y ojo derecho
2	Ceja y ojo izquierdo
3	Mentón y mejillas
4	Labio superior
5	Labio inferior

Tabla 2.3 Puntos de entrenamiento equivalentes a los Landmark que estima cada RNA

RNA	Punto target para RNA	Punto facial
1	1	2 Mitad de la ceja
	2	3 Extremo derecho de la ceja
	3	7 Parpado superior
	4	9 Parpado inferior
2	1	4 Mitad de la ceja
	2	5 Extremo izquierdo de la ceja
	3	10 Parpado superior
	4	12 Parpado inferior
3	1	15 Mejilla derecha
	2	16 Mejilla izquierda
	3	25 Mentón eje X
	4	25 Mentón eje Y
	5	26 Mejilla izquierda parte inferior

4	1	17 Labio extremo derecho
	2	18 Labio superior centro derecho
	3	20 Mitad del labio superior
	4	22 Labio superior centro izquierdo
	5	26 Labio extremo izquierdo
5	1	19 Labio inferior centro derecho
	2	21 Mitad del labio inferior
	3	23 Labio inferior centro izquierdo

Tabla 2.4 Ingreso y salidas de las RNAs

RNA	Entrada	Número de salidas que estima la RNA
1	Sub área 1	4
2	Sub área 2	4
3	Sub área 3	5
4	Sub área 3	5
5	Sub área 3	3

Así cada red estima puntos cercanos con un valor similar entre sí. Como se mencionó para el entrenamiento se empleó MATLAB el comando feedforwardnet para el diseño de la red y su posterior entrenamiento, el que pertenece al toolbox para el entrenamiento de redes neuronales.

En el entrenamiento se emplearon un total antes mencionado del set de entrenamiento, mil datos, de los cuales se debían segmentar en tres partes, a las que se les asignaría un valor del total de los datos de entrenamiento, los cuales fueron:

- Entrenamiento 80%
- Testeo 10%
- Validación 10%

En las figuras 2.6 a 2.10 se puede apreciar la topología de la RNA, las cuales constan con una capa de ingreso, una de salida y una capa oculta. Por los motivos antes expuestos se empleó una red con una sola capa oculta, ya que los resultados obtenidos con dos capas son similares al que se obtiene con una capa, estos datos se pueden apreciar en el capítulo de pruebas. Cada una de estas redes varia por el valor que se presenta en su ingreso, es decir según la sub área que sea el ingreso de esta; mientras que el número de neuronas en la capa de salida está dado por el número de puntos que estimara cada RNA lo que se indica en la tabla 2.2. Se empleó un valor de 60 neuronas en las capas ocultas, ya que este es límite máximo que soporta la memoria empleada en el entrenamiento, ya que si se realiza un entrenamiento con un valor mayor a este la memoria se desborda lo que produce que el entrenamiento se cancele.

Como las redes diseñadas no son clasificadores no se empleó una función escalón, las redes al tratar de estimar o aproximarse al valor de la ubicación se empleó una función lineal o PURELIN, con lo cual se obtuvo el comportamiento deseado de la red.

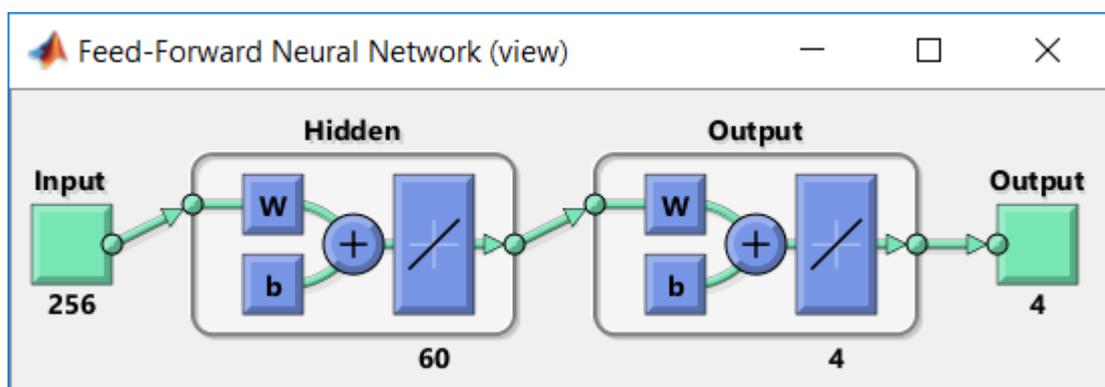


Figura 2.6 Topología de la RNA 1

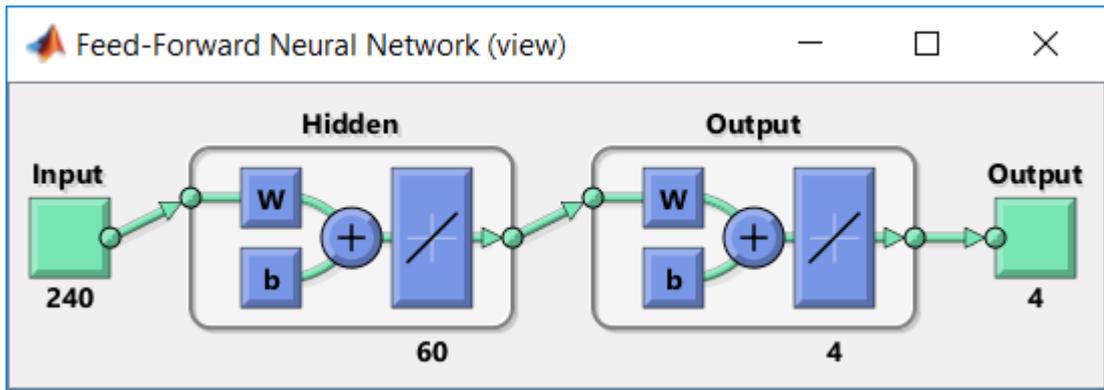


Figura 2.7 Topología de la RNA 2

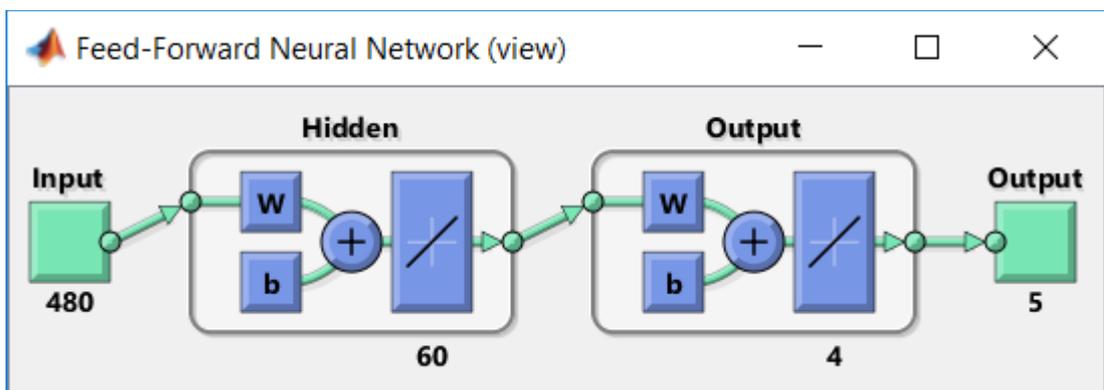


Figura 2.8 Topología de la RNA 3

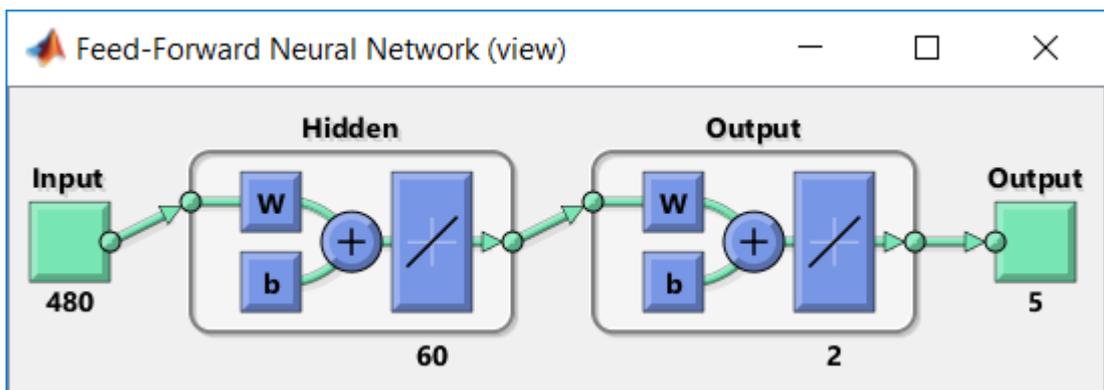


Figura 2.9 Topología de la RNA 4

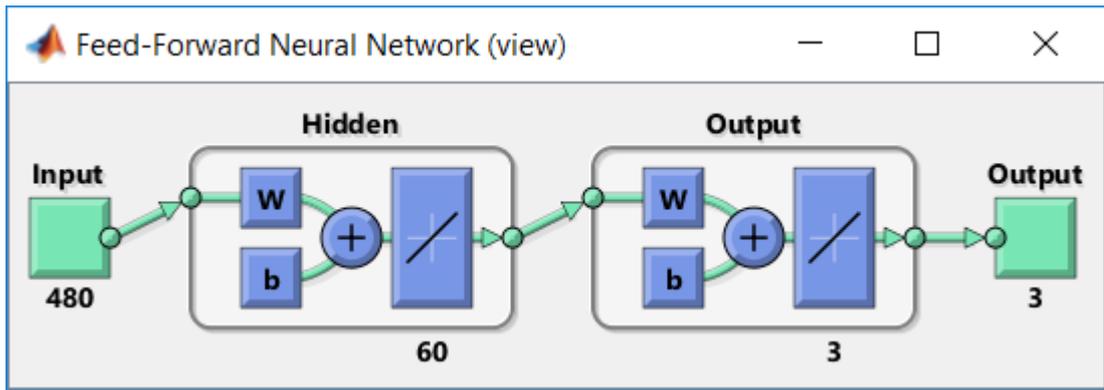


Figura 2.10 Topología de la RNA 5

En el caso de que se entrene una RNA con un numero de neuronas superiores que empleara una mayor cantidad de memoria en su entrenamiento, MATLAB trata de realizar este entrenamiento, para lo que elimina conexiones de las capas, lo que produce que la red no estime bien el valor ya que al no tener esta conexión se impondrá valores aleatorios para el entrenamiento, en la figura 2.11 se puede apreciar la topología de entrenamiento de la RNA 5 cuando se elimina una conexión entre capas en el entrenamiento, produciendo que la capa de salida no posea neuronas.

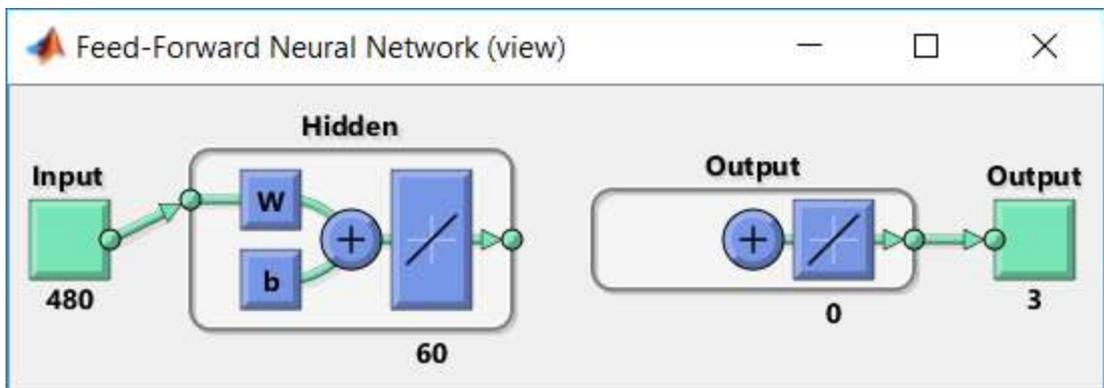


Figura 2.11 Topología de la RNA 5 con error en su entrenamiento

Estos valores se escogieron debido al tiempo de entrenamiento y error, de lo cual se concluyó que mayor número de datos de entrenamiento este sería más lento, pero brindaría un error menor, pero si se emplea un valor más alto para el testeo y validación que los indicados el entrenamiento duraría menos, pero se obtendría una red deficiente que aproximaría ciertos valores, pero ante algunos casos dará una respuesta errónea.

2.4 PRESENTACIÓN

Con los datos brindados de la etapa de procesamiento se traslada la ubicación de los huesos de una posición inicial a una nueva posición, esta posición está dada por los datos que se captaron mediante la webcam, en esta etapa convergen todas las anteriores y se realiza la integración de las etapas previas a la RNA que se realizaron en PYTHON y la red neuronal realizada en MATLAB. En el software BLENDER corre sobre una base en PYTHON se logró integrar las partes antes mencionadas.

Para la selección del Avatar a emplear se realizaron ciento uno encuestas que determinaron la aceptación de estos, para una población de niños que poseen una edad entre un rango de ocho a quince años. Estos Avatares se seleccionaron en un principio por poseer características similares al rostro de una persona.



Figura 2.12 *Banco de Avatares para la animación facial*

En las encuestas realizadas se solicitó seleccionar los Avatares que más llamaran su atención, con un máximo de tres elementos, de estos datos se puede apreciar que los Avatares que representan a personas poseen mayor aceptación, mientras algunos robots poseen una aceptación menor a esta, pudiendo considerar estos dos tipos para la implementación de diversos Avatares en las terapias, el resultado de estas encuestas se presenta en las tablas 2.5 y 2.6 según la edad de cada segmento de la población.

Tabla 2.5 Aceptación de Avatares en mujeres de 8 a 15 años

Avatar	8 años	9 años	10 años	Mayores de 10 años	Total
1	3	2	1	1	7
2	11	9	1	0	21
3	5	0	0	3	8
4	8	4	8	7	27
5	13	12	10	6	41
6	2	3	1	3	9
7	14	12	9	9	44
8	3	2	3	0	8

Tabla 2.6 Aceptación de Avatares en hombres de 8 a 15 años

Avatar	8 años	9 años	10 años	Mayores de 10 años	Total
1	8	12	5	6	31
2	6	6	6	3	21
3	4	3	1	1	9
4	4	5	3	3	15
5	3	3	3	2	11
6	2	4	2	2	10
7	3	2	4	3	12
8	8	7	8	6	29

De los datos expuestos en las tablas 2.5 y 2.6 se puede apreciar que en la población estudiada el Avatar que tuvo mayor aceptación es el número 7, a partir de este se desarrolló la etapa de presentación, modificando a este para que cumpla con los requerimientos del sistema, es decir agregando los diversos huesos de animación y delimitando la región del mallado del Avatar que este afectara.

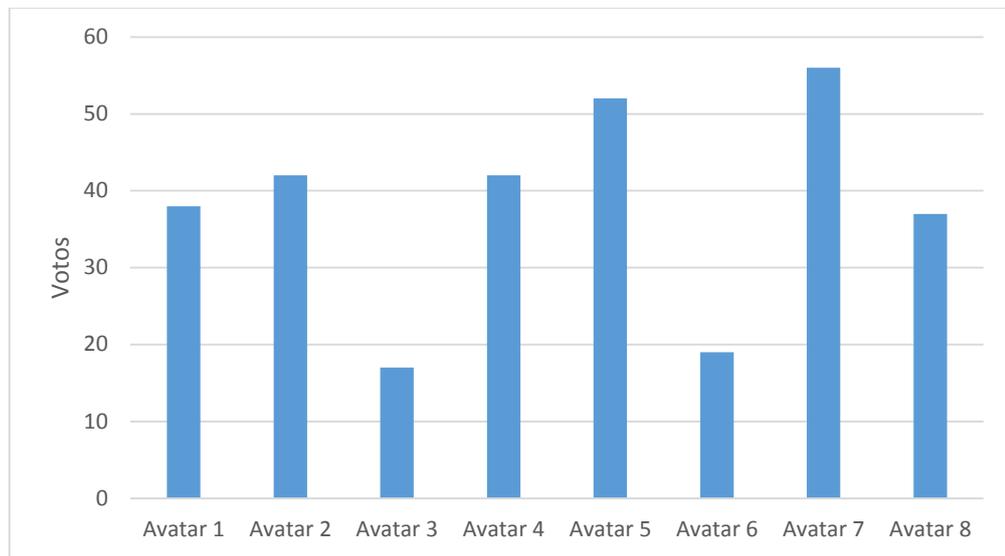


Figura 2.13 *Aceptación de los Avatares en una población de niños de 8 a 15 años*

Para la animación se empleó el modo juego de Blender el cual a diferencia del modo de render permite captar eventos que se pueden realizar ya sea con el teclado o ratón, permitiendo así el diseño de botones que controlen al sistema.

El Avatar está compuesto por huesos o huesos de animación, mediante los cuales, al variar su rotación, ubicación se puede generar el movimiento de una determinada sección del acatar, dicho hueso de animación afectara en cierto grado la estructura del Avatar, pudiendo mover una determinada sección del mallado del Avatar, para lo cual se debe considerar la fisionomía del rostro de una persona y los movimientos que esta pueda realizar con un determinado musculo. En la sección de apéndices se puede encontrar la información referente a los 26 huesos empleados como su designación en el Avatar y la sección del mismo que este afectara.

El proyecto al estar destinado a trabajar con niños se empleó un Avatar que no les cause temor y a su vez posea características físicas similares a las de una persona para que el niño se sienta familiar con el mismo, como ya se mencionó anteriormente

este se seleccionó de un banco de Avatares mediante una serie de encuestas. En la figura 2.14 se puede apreciar el rostro del Avatar empleado.

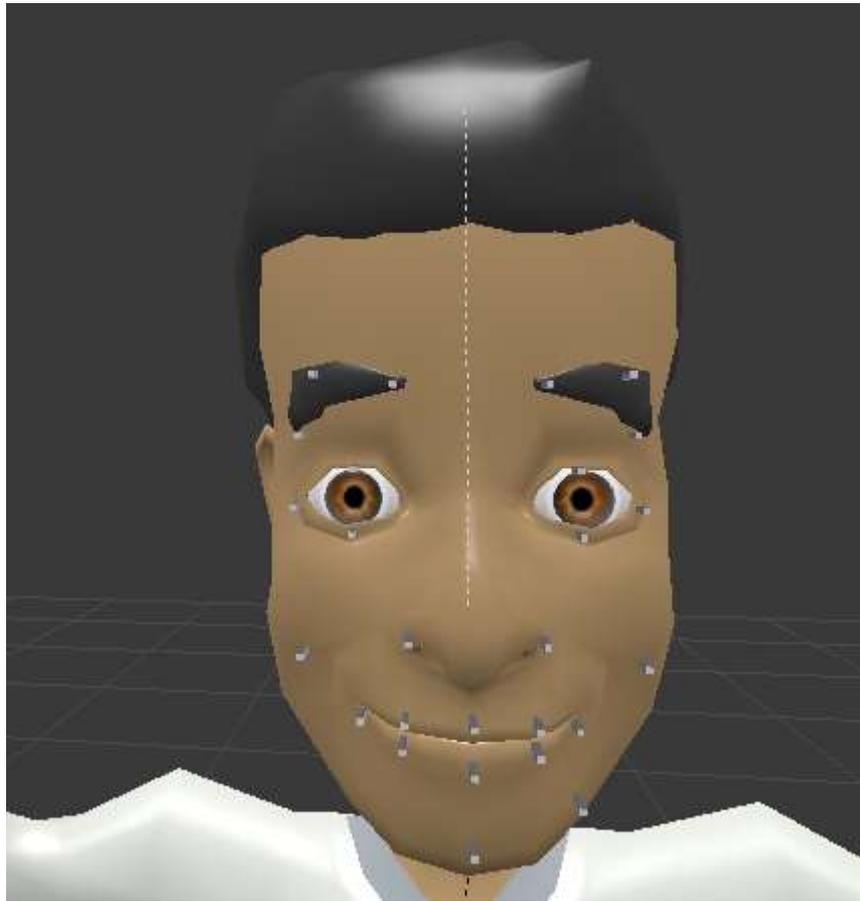


Figura 2.14 Rostro *del Avatar empleado*

Mediante cambios en la ubicación de un extremo del bone se generó el movimiento del Avatar y con la aplicación de límites en el movimiento de los mismo se generó que este no tenga movimientos anormales, exagerados y se mantenga dentro de la capacidad real que tiene una persona para mover cierta parte del cuerpo.

Un detalle importante de este proyecto es el uso de una postura frontal en la etapa de captación ya que si se inclina la cabeza más de 10° hacia cualquier lado, genera una pérdida de información del lado contrario. En el caso que en un trabajo futuro se deseara realizar la animación de todo el rostro sin importar su posición se deberá considerar un sistema de cámaras que compense la falencia antes mencionada al usar una sola cámara.

2.5 TERAPIAS

Como se ha mencionado en secciones anteriores, el proyecto AVATAR que se realiza en la Universidad Politécnica Salesiana, busca brindar soporte terapéutico en el tratamiento de niños con Trastorno del Espectro Autista mediante el uso de agentes virtuales.

El presente proyecto es la base o núcleo del proyecto antes mencionado, ya que consta con dos etapas que serán el armazón del proyecto AVATAR, dichas etapas son:

- Diseño de una RNA para la detección facial
- Diseño de un sistema de animación facial

En esta sección se detallará la aplicación de estas etapas en el proceso de terapia que se realizará en el CIMA, así como la aplicación que surgirá a partir de estas.

Los Niños con TEA muestran gran aceptación a interactuar con Avatares. El desarrollo de Avatares como herramienta de soporte terapéutico tiene ventajas sobre otras tecnologías. Por ejemplo, al desarrollar una herramienta basada en la robótica o algún sistema embebido, estamos limitados por las características físicas del sistema; en cambio, los Avatares presentan características propias tales como la maleabilidad, adaptabilidad, controlabilidad, replicabilidad y la posibilidad de implementar intervenciones individualizadas. [37]

2.1.5 SISTEMA DE INTERACCIÓN AVATAR – PERSONA

Como se analizó en el capítulo 1, en la actualidad se han generado diversos sistemas que estudian la interacción de una persona con un agente virtual, en este caso se busca emplear un agente virtual basado en inteligencia artificial como método terapéutico para mejorar la interacción de una persona con este agente.

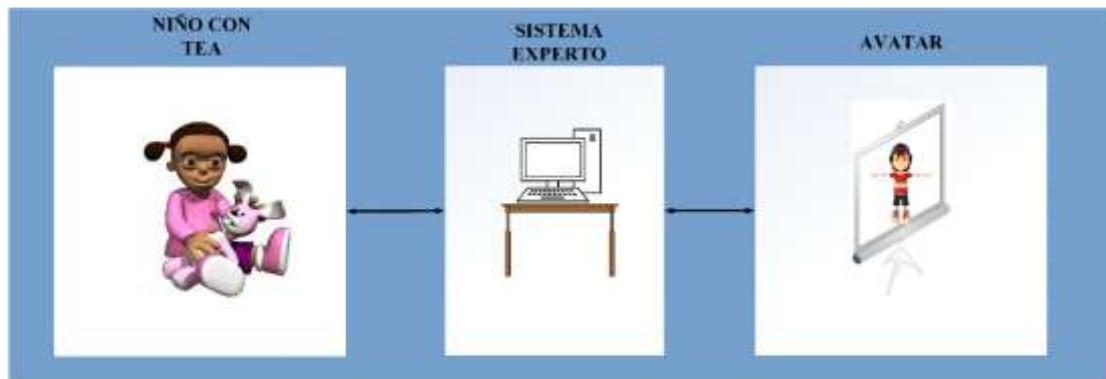


Figura 2.15 Esquema *del sistema de interacción Avatar – persona*

Para realizar esta interacción el agente virtual o Avatar empleara una inteligencia artificial similar a la de un chat bot, siendo un sistema capaz de brindar una determinada respuesta ante un determinado estímulo.

Los estímulos antes mencionados serán determinadas frases a las que el Avatar deberá responder con una frase que de la ilusión de que la persona está tratando con otra.

La combinación de la inteligencia artificial tras el chat bot con un sistema de animación facial producirá que el Avatar pueda realizar una conversación con la persona de una manera más fluida y natural, al emplear la animación facial basada en el aprendizaje de los movimientos que realiza una persona al hablar, combinado con un correcto entrenamiento del chat bot se podrá lograr una terapia de interacción social similar a una conversación con una persona.

2.1.5 VIDEO CHAT EMPLEANDO AVATARES EN PACIENTES CON TEA

Esa terapia a diferencia de la anterior empleara la animación facial sobre dos personas, de las cuales la una es el niño con TEA y la otra será un terapeuta o ya sea otra persona, con lo que se busca que el niño con trastorno del espectro autista pueda interactuar con otra persona sin conocerle directamente.

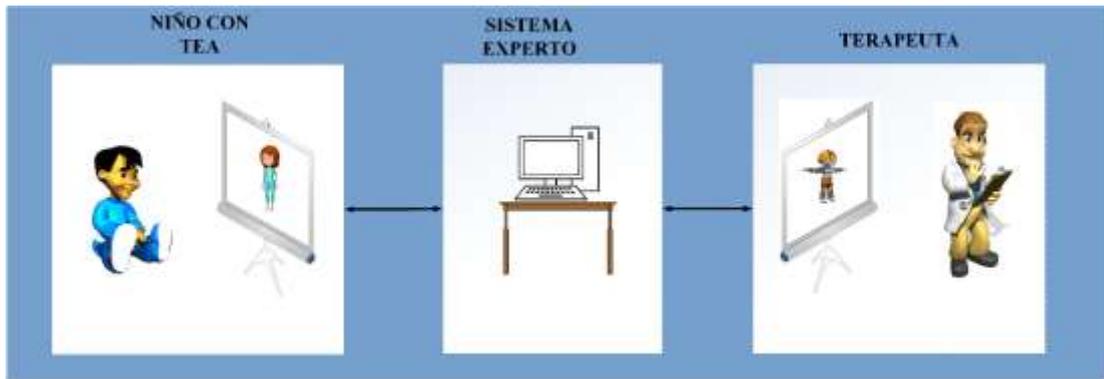


Figura 2.16 *Esquema del sistema de video chat para pacientes con TEA*

El sistema en este caso estará conformado por elementos que comúnmente constituyen un sistema de video llamada, es decir dos ordenadores que cuenten con una webcam, con el adicional de que estos contarán no con un convencional programa para realizar esta video llamada, sino que con el agregado de la inclusión de la animación facial sobre un Avatar.

2.2.5 RECONOCIMIENTO DE EMOCIONES

LA aceptación de un agente virtual como media terapéutica ha tenido gran aceptación, por lo en esta terapia se propone un sistema que adiciona al anterior una etapa de reconocimiento de emociones, para lo cual se reemplaza la interacción Avatar - persona por dos personas interactuando a través del Avatar.

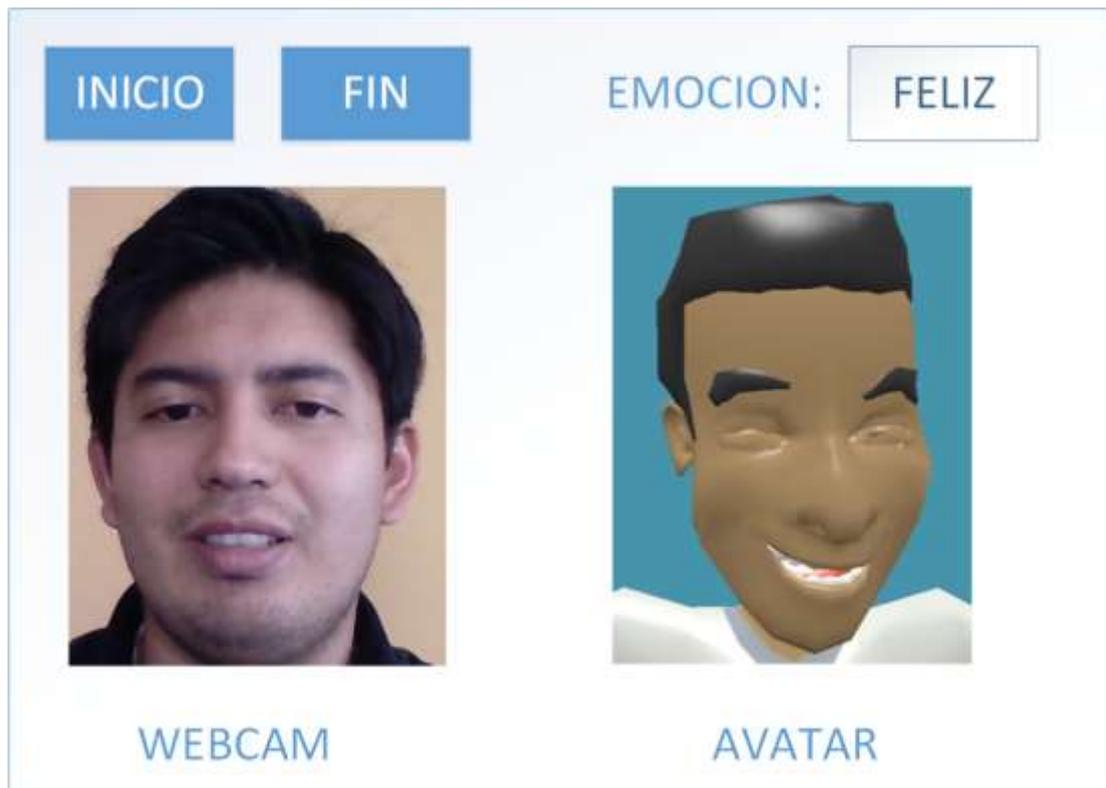


Figura 2.17 *Modelo de aplicación de propuesta para reconocimiento de emociones*

En esta terapia se busca emplear una red neuro difusa tras la detección facial para identificar las emociones que transmite la persona en las sesiones de terapia, manteniendo un registro de dichas emociones.

Con los datos de la RNA empleada para el reconocimiento facial en su salida se deberá emplear lógica difusa para realizar un clasificador de estos datos para obtener una aproximación a un valor numérico de esta emoción.

2.3.5 TUTOR VIRTUAL PARA APRENDIZAJE DE EXPRESIONES FACIALES

Los ambientes virtuales controlados buscan apoyar en el aprendizaje de algún aspecto de una determinada población, como se ha mencionado entre los desórdenes que sufren las personas con autismo se puede encontrar su dificultad en la interacción social.

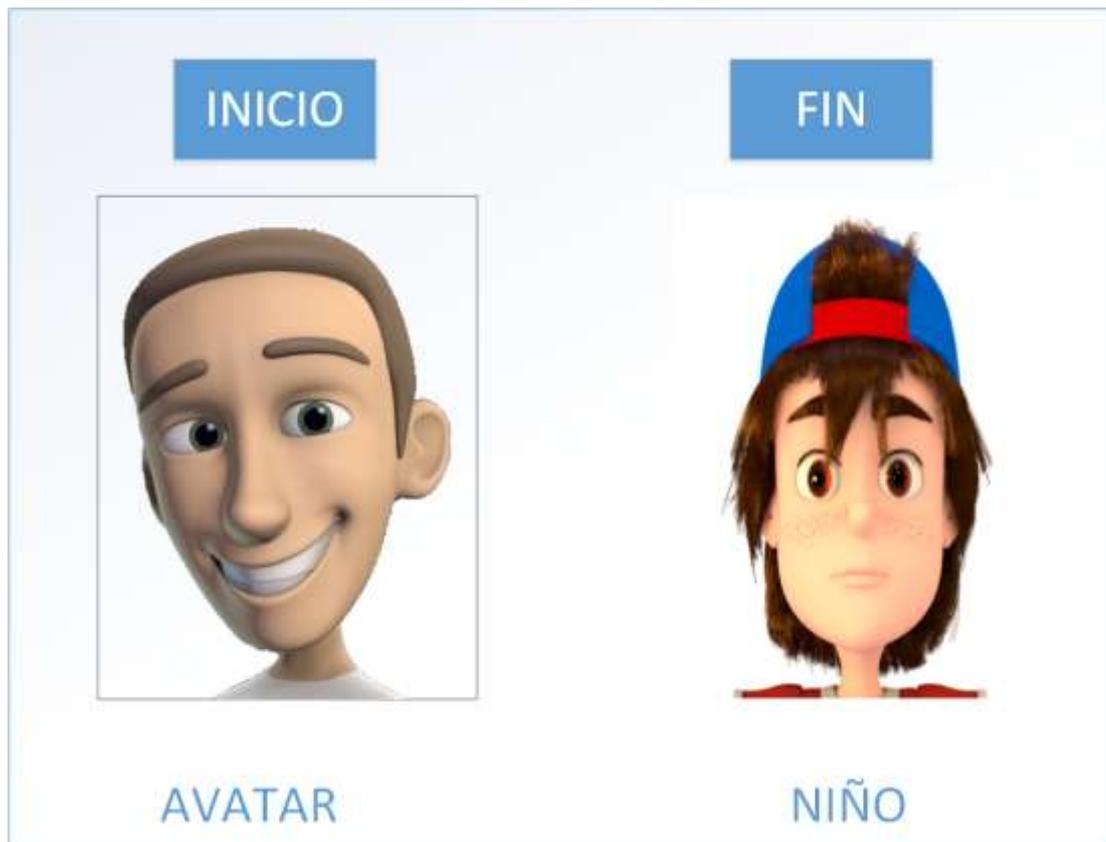


Figura 2.18 Modelo de aplicación de tutor de aprendizaje de emociones

En este caso se busca emplear un tutor, pudiendo este ser un agente virtual o un Avatar guiado por el terapeuta, que brinde al niño indicaciones sobre la terapia, es decir, este le explicara y realizaran juntos gesticulaciones que representaran emociones, indicando a este que reacción debe tener ante una determinada situación.

CAPÍTULO 3: PRUEBAS Y ANÁLISIS DE RESULTADOS

En este capítulo se analizarán las pruebas realizadas a las etapas del sistema de manera principal al procesamiento y presentación. En estas pruebas se incorporaron dos casos adicionales, los cuales no entraron en el proceso de entrenamiento de la red neuronal.

Las pruebas realizadas no solo se representan mediante un valor numérico, también requieren de la percepción visual del diseñador, como en el caso de la red neuronal que busca aproximar la ubicación de los puntos faciales, por lo cual para obtener una mejor apreciación de estos datos es recomendable colocar dichos puntos sobre la región de interés.

3.1 DETECCIÓN FACIAL

En la detección facial se analizará la funcionalidad del detector Haar, para lo cual se diseñó un algoritmo que tomara varias fotografías por un lapso de tiempo, tras tomar una fotografía este algoritmo detectara la ubicación del rostro, permitiendo un máximo de tres rostros en la misma imagen, considerando al de mayor tamaño como el rostro de mayor relevancia. En la figura 3.1 se puede apreciar un diagrama de flujo que sigue el algoritmo de testeo de detección facial.

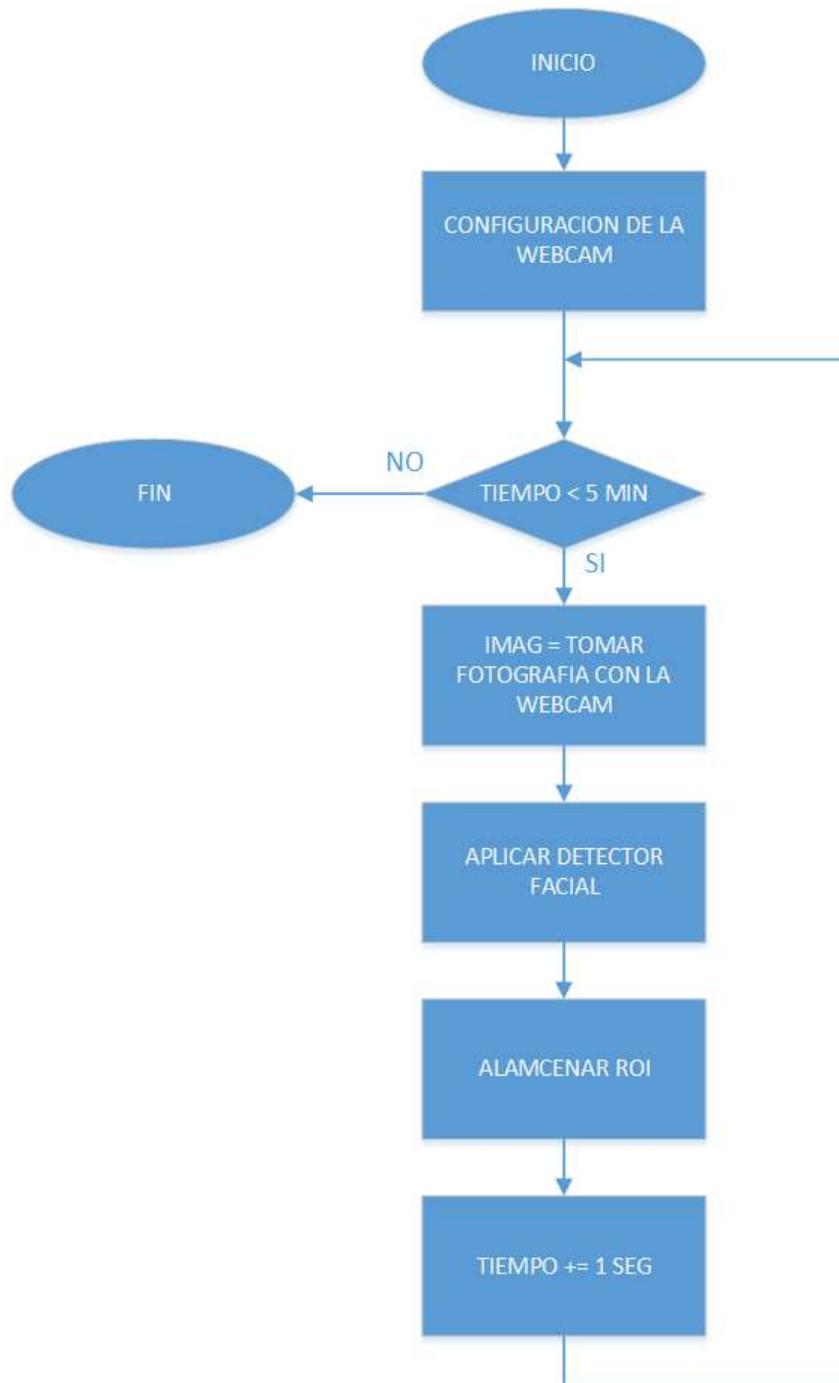


Figura 3.1 Diagrama de la evaluación de la detección facial

Entre los datos que capta este algoritmo se debe considerar dos posibles resultados que son: valores positivos que contienen la información del rostro y los falsos positivos. Se consideran a los valores que no contienen información del rostro como falsos, porque estos al no ser útiles son descartados por la cascada de detectores. Los valores falsos positivos hacen referencia a los resultados que el filtro considera como positivos, pero no contiene información de un rostro. En la figura 3.2 se puede

apreciar el resultado de estas pruebas en donde el eje horizontal representa el caso que se testeó y el eje vertical el porcentaje de falsos positivos obtenidos de las ROIs de cada caso, el valor total de estas se puede observar en la tabla 3.1.

Para considerar un resultado como falso positivo se consideró aquellos que no contienen información del rostro, contienen una sección del mismo y la existencia de elementos que no se consideran como rostro como secciones de ropa o elementos con una forma similar al rostro.

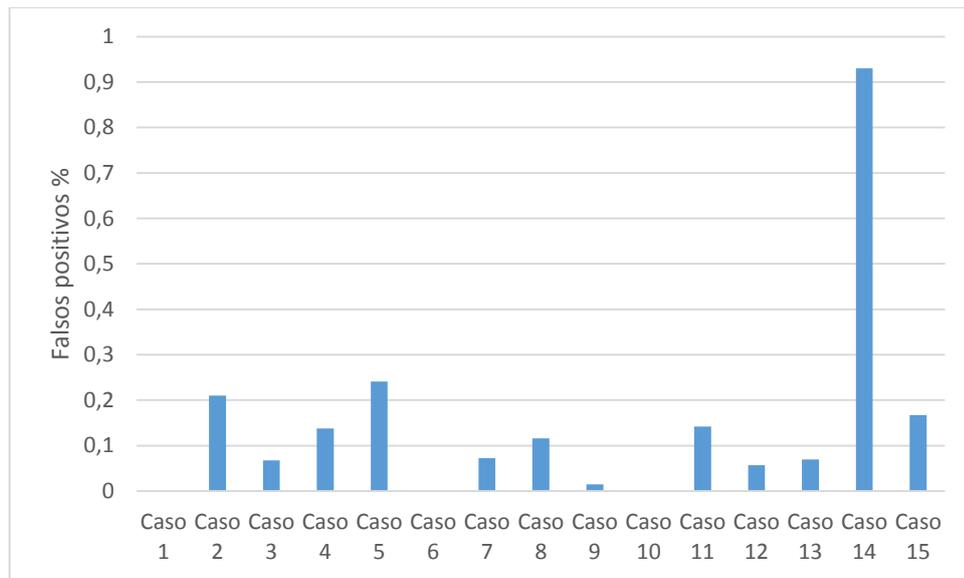


Figura 3.2 Numero de falsos positivos obtenidos en la prueba de detección facial.

Un dato relevante en este proceso que influirá en las siguientes etapas es el número de cuadros que puede captar el sistema en una unidad de tiempo, en esta prueba se empleó un tiempo promedio entre dos minutos a cinco minutos, según la edad del caso de prueba y el nivel de comodidad con el mismo. A partir de estas consideraciones se concluyó que el sistema puede procesar un total de 10 fps (Fotogramas por segundo). En la tabla 3.1 se puede apreciar el número de imágenes captadas en cada caso y su valor equivalente en fps. El valor de los fps que se procesa en cada caso dependerá de la similitud que presente este con los datos empleados en el entrenamiento, es decir, si se busca detectar el rostro de una persona con bigote este proceso tomara más tiempo que el determinar la existencia del rostro de un niño, ya que el entrenamiento del clasificador se centró en el uso de imágenes de niños con una edad comprendida entre los 8 a 15 años.

Tabla 3.1 Datos de las pruebas de detección facial

Caso	Sexo	Falsos positivos	ROI captadas	Tiempo [minutos]	FPS
1	Hombre	0	186	0.5	6.2
2	Mujer	3	1428	3.5	6.8
3	Hombre	1	1478	3.5	7.03809524
4	Mujer	2	1449	3.5	6.9
5	Mujer	8	3317	5	11.0566667
6	Mujer	0	4488	5	14.96
7	Mujer	4	5514	5	18.38
8	Mujer	5	4309	6	11.9694444
9	Mujer	1	6588	6	18.3
10	Hombre	0	1574	4	6.55833333
11	Hombre	2	1408	3.5	6.7047619
12	Hombre	1	1755	4	7.3125
13	Hombre	1	1429	3.5	6.8047619
14	Mujer	15	860	2	7.16666667
15	Hombre	2	1198	2.5	7.98666667

A partir de los datos antes expuestos se puede considerar que los resultados brindados en esta etapa poseen un alto índice de confiabilidad, se debe contemplar que el valor de fotogramas por segundo que captará el sistema completo se verá reducido debido a que los procesos de procesamiento y presentación se realizan de manera secuencial a este, aumentando el tiempo entre las capturas de las ROIs.

3.2 RED NEURONAL Y ESTIMACIÓN DE LA UBICACIÓN DE LOS PUNTOS FACIALES

Para evaluar la RNA se tomó un segmento del set de entrenamiento de cada caso, adicionalmente se empleó todos los datos de los dos casos para pruebas. En la tabla 3.2 se puede apreciar el número de regiones de interés que se emplearon en cada caso para el entrenamiento de la red neuronal y las pruebas de la misma.

Tabla 3.2 Distribución de los datos para entrenamiento y pruebas de las RNAs

Caso	ROI captadas	ROI usadas en el entrenamiento	ROIS usadas para pruebas de la RNA
1	186	98	88
2	1428	130	1298
3	1478	116	1362
4	1449	128	1321
5	3317	114	3203
6	4488	104	4384
7	5514	52	5462
8	4309	120	4189
9	6588	84	6504
10	1574	50	1524
11	1408	74	1334
12	1755	32	1723
13	1429	62	1367
14	860	0	850
15	1198	0	1100

En estas pruebas se testaron cada uno de los puntos faciales, para lo cual se realiza un conjunto de valores esperados, los cuales serán comparados y se obtendrá un valor de error representado por la diferencia entre la posición del punto obtenido y su valor ideal. Para obtener los valores ideales para validar la red neuronal, se diseñaron nuevos patrones con los datos que no se emplearon en el entrenamiento de la RNA.

El valor de error de estimación en las redes neuronales está dado por la diferencia entre el valor esperado y el estimado. Se debe considerar que a la salida de la red neuronal se tiene un número con decimales, mientras que los píxeles se representan mediante un número entero, para lo cual se debe redondear este valor,

En la tabla 3.3 se puede observar el error promedio de las pruebas realizadas a las redes neuronales. A partir de los datos expuestos en la tabla 3.3 se concluye que la estimación de los puntos faciales se realiza de manera idónea, puesto que el valor de error promedio obtenido es inferior a los 3 píxeles.

Tabla 3.3 Error de las estimación de los puntos faciales en la RNA propuesta

RNA	Punto facial	Error promedio[Pixel]
1 Ceja y ojo derecho	Mitad de la ceja	0,011178944
	Extremo derecho de la ceja	0,016365335
	Parpado superior	0,040769866
	Parpado inferior	0,014917803
2 Ceja y ojo izquierdo	Mitad de la ceja	0,062142934
	Extremo izquierdo de la ceja	0,017148435
	Parpado superior	0,042729381
	Parpado inferior	0,020128064
3 Mentón y mejillas	Mejilla derecha	0,402555068
	Mejilla izquierda	1,097213115
	Mentón eje X	0,896039883
	Mentón eje Y	0,880521044
	Mejilla izquierda parte inferior	1,114816689
4 Labio superior	Labio extremo derecho	0,11155721
	Labio superior centro derecho	0,004924307
	Mitad del labio superior	0,153811341
	Labio superior centro izquierdo	0,214222137
	Labio extremo izquierdo	0,145774697
5 Labio inferior	Labio inferior centro derecho	0,070415586
	Mitad del labio inferior	1,144056709
	Labio inferior centro izquierdo	2,485352053

Para el diseño de una RNA para la detección facial en tiempo real, se desarrolló un módulo de Python que toma una fotografía mediante la webcam y sigue el proceso hasta estimar los puntos faciales, este módulo en su salida presenta una ventana que contiene la imagen del rostro ya procesado, es decir en escala de grises y escalado a un tamaño de 100x100 píxeles, sobre el cual se colocaran líneas perpendiculares a la coordenada de interés de cada punto, para algunos puntos faciales la información del

eje Y es de mayor relevancia por lo tanto el modulo diseñado dibujara una línea paralela al eje X cruzando por la coordenada de interés, algunos puntos de la detección facial no se han considerado a partir de esta etapa ya que no generan un movimiento, es decir son partes fijas del rostro, como es el caso de las fosas nasales.

La velocidad de procesamiento a partir de este caso se ve reducida debido a que en este punto la red neuronal realizara la estimación de los puntos respectivos, lo que conlleva una carga computacional mayor, viéndose reducida a un valor de 5 a 8 fps lo cual dependerá si la persona posee características similares a las de las personas en el set de entrenamiento, en caso de que esta no posea estas características la estimación tomara más de tiempo.

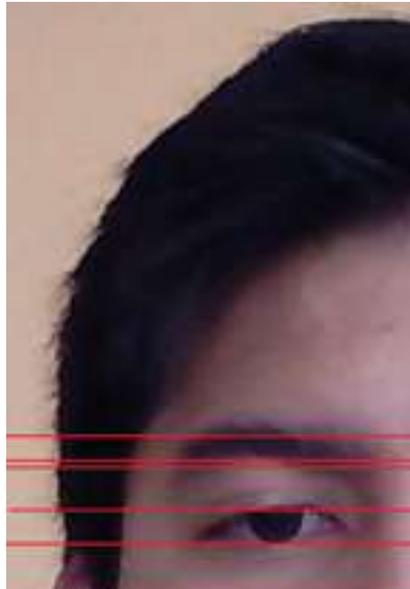


Figura 3.3 *Prueba de la RNA 1 en tiempo real*

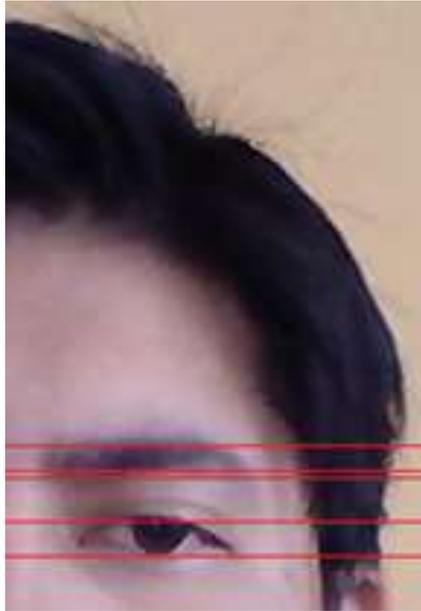


Figura 3.4 *Prueba de la RNA 2 en tiempo real*

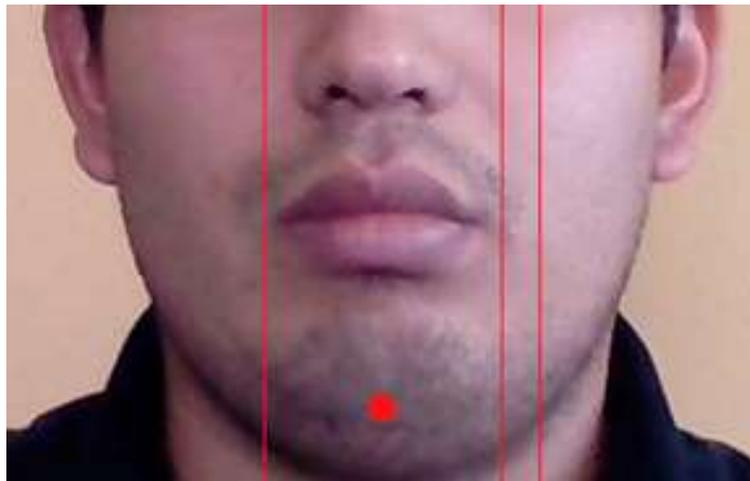


Figura 3.5 *Prueba de la RNA 3 en tiempo real*

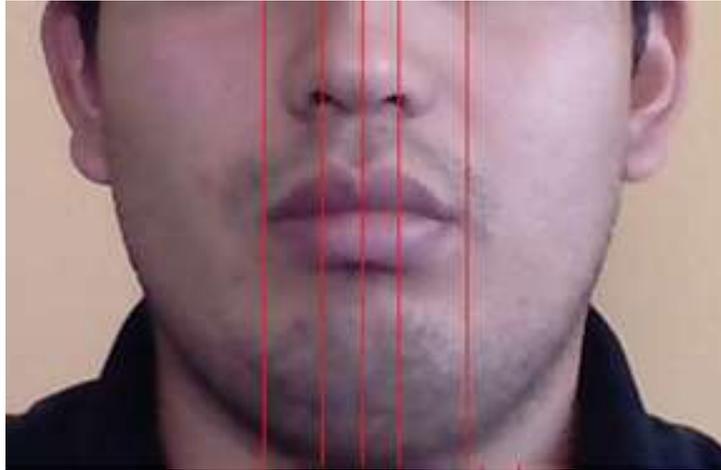


Figura 3.6 *Prueba de la RNA 4 en tiempo real*

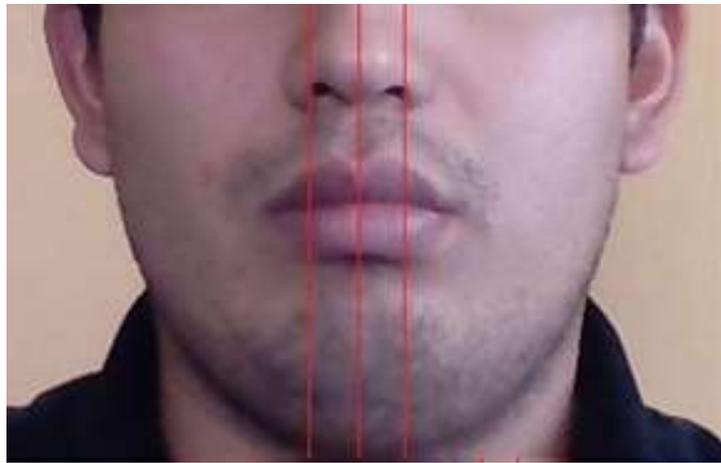


Figura 3.7 *Prueba de la RNA 5 en tiempo real*

3.3 ANIMACIÓN

Para la evaluación de la animación se considera la capacidad del sistema en presentar un resultado que se asemeje a la posición del rostro que es captado por la webcam; para lo que se han analizado diversas emociones y se comparara visualmente la similitud de estas.

En estas pruebas se centrará en analizar las emociones básicas de una persona: un estado neutro, alegre, triste e ira. Adicionalmente a estos estados se analizará la capacidad del sistema de variar únicamente un componente del rostro, es decir su capacidad de respuesta ante el parpadeo o mantener un ojo abierto, con lo cual se busca determinar la existencia de un falso positivo en la boca, considerando a esta

como un ojo. Para evitar esto se empleó las subáreas en la etapa de captación, eliminando la posibilidad de confundir estos dos elementos.

Para las pruebas se empleó el principio de la terapia mediante video chat, en donde se analizaría la aceptación del Avatar ante los niños y a su vez la interacción que estos tienen con los mismos, en donde ellos ven reflejados su rostro en el Avatar.



Figura 3.8 *Prueba de animación: Persona normal*



Figura 3.9 *Prueba de animación: Avatar normal*



Figura 3.10 *Prueba de animación y detección facial: Persona feliz*



Figura 3.11 *Prueba de animación: Avatar feliz*

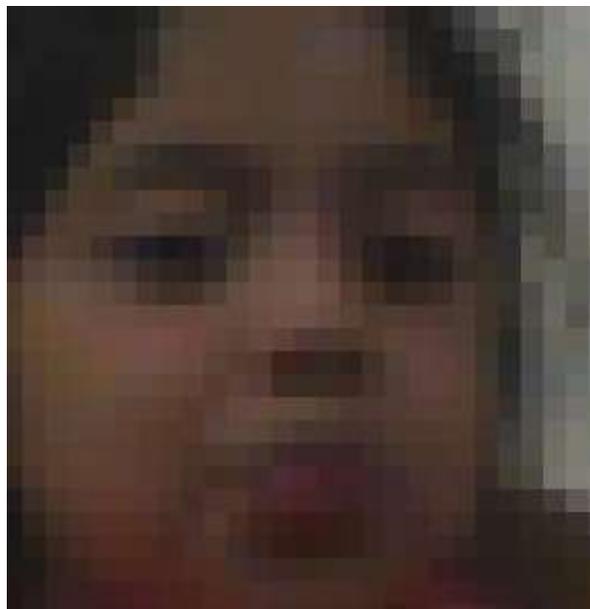


Figura 3.12 *Prueba de animación: Niño 1*



Figura 3.13 *Prueba de animación 1*

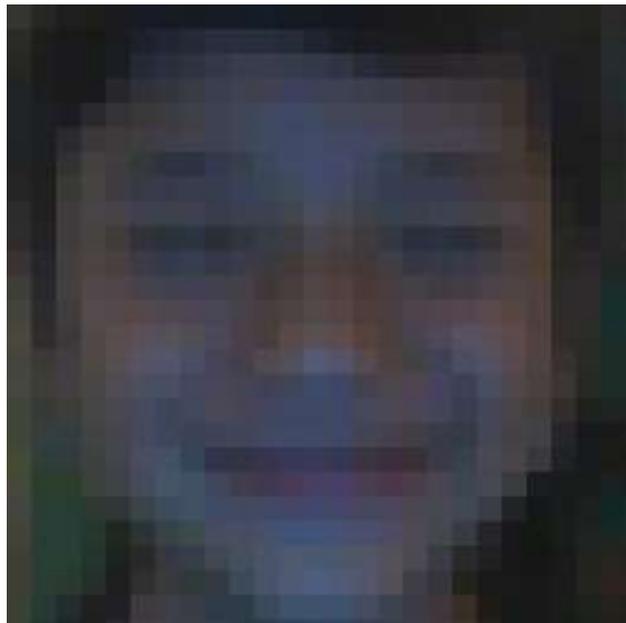


Figura 3.14 *Prueba de animación: Niño 2*



Figura 3.15 *Prueba de animación 2*

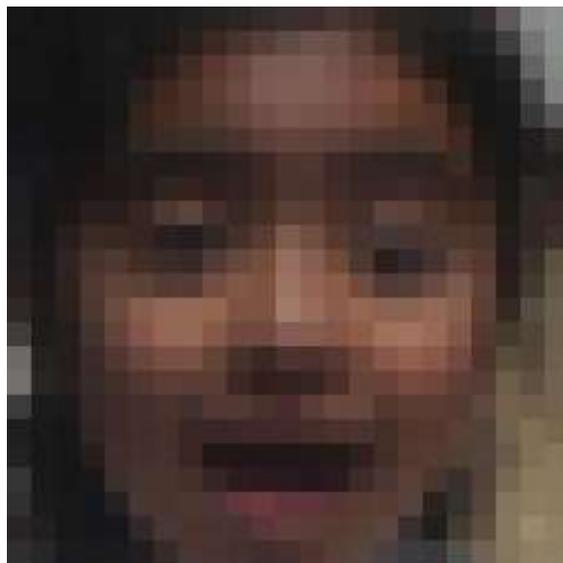


Figura 3.16 *Prueba de animación: Persona 3*



Figura 3.17 *Prueba de animación 3*

CAPÍTULO 4: CONCLUSIONES Y RECOMENDACIONES

4.1 CONCLUSIONES

El uso de un sistema de detección facial basado en una cascada de detectores con características más simples facilita el procesamiento de la imagen, para la obtención de una región de interés ROI que contiene el área del rostro de una persona, permite que el sistema emplee menos recursos en el cálculo, ya que la cascada de detectores va descartando la información que considera innecesaria.

Para mejorar tiempos de entrenamiento, se procedió a segmentar el rostro en 3 partes. Esto ayudo a reducir el error en la aproximación de la RNA y evitar la eliminación de conexiones sinápticas entre capas.

Se utilizó la función PURELINE, esto es debido a que el sistema requiere un regresor ya que se utiliza un valor que representa coordenadas mas no una respuesta booleana.

Para mejor el entrenamiento de la RNA se realizó un proceso sistemático en el que se fueron variando el número de neuronas en las capas ocultas, partiendo de 3 capas ocultas con 40 neuronas hasta llegar a un número idóneo de una capa con 40 neuronas, obteniendo así una RNA con un error mínimo.

Al ser un proyecto enfocado a niños, no existe una base de datos de rostros y gestos disponible. Por lo fue necesario generar una base de datos para el entrenamiento de la red. No se puede utilizar una base de datos de adultos o adolescentes ya que su estructura del rostro varia notablemente con la de un menor.

El movimiento que realiza el avatar busca imitar los datos que obtiene la webcam del niño. Partiendo de la información que se capta el sistema genera datos a la salida de la RNA, que son indicadores de movimiento del avatar. Para que la variación de estos datos sea interpretada como movimiento se requiere que los huesos que conforman la armadura del avatar cambien su localización según como varían los datos a la salida de la RNA.

Para obtener mayor flexibilidad en el manejo del avatar se empleó el modo juego de BLENDER. El software de animación BLENDER presenta tres modos de operación la cíclica, el render y el juego, de estos modos sobresalen el modo juego para aplicaciones de este tipo, ya que permite al diseñador emplear al teclado y ratón como sensores que controlan un determinado proceso de la animación.

El movimiento del avatar en BLENDER está dado por la variación de parámetros de los huesos de animación. Los huesos de animación poseen propiedades que al variarlas permiten realizar un movimiento en la estructura del avatar, pudiendo variar su rotación, posición o localización se podrá generar un movimiento según la estructura de los huesos y como estos afecten al mallado del avatar.

4.2 RECOMENDACIONES

El uso del software BLENDER es de gran ayuda, ya que permite la inclusión de algoritmos en PYTHON para realizar tareas complejas.

Se recomienda emplear una misma versión de PYTHON tanto en BLENDER como en el desarrollo de librerías y módulos, evitando así errores de compatibilidad en estos.

El uso de avatares con rasgos humanoides facilita el proceso de animación por lo cual se debe considerar emplear avatares con rasgos marcados, para que se pueda percibir de mejor manera las expresiones que este realiza.

La selección de los parámetros que necesita el sistema se debe considerar el ingreso y la salida de la RNA, como el tamaño que tendrán estos para evitar que el ordenador realice tareas complejas, evitando así que el software desarrollado se congele.

REFERENCIAS BIBLIOGRÁFICAS

- [1] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, 2000.
- [2] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognit.*, vol. 25, no. 1, pp. 65–77, 1992.
- [3] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Comput. Vis. Image Underst.*, vol. 81, no. 3, pp. 231–268, 2001.
- [4] S. Laine *et al.*, "Production-Level Facial Performance Capture Using Deep Convolutional Neural Networks," 2016.
- [5] I. Saldaña and C. Del Río Blanco, "Modelado y animación facial de un Avatar virtual realista mediante formas clave," 2013.
- [6] F. Mulas, G. Ros-Cervera, M. G. Millá, M. C. Etchepareborda, L. Abad, and M. Téllez De Meneses, "Modelos de intervención en niños con autismo," *Rev. Neurol.*, vol. 50, no. SUPPL. 3, pp. 77–84, 2010.
- [7] Instituto Nacional de la Salud Mental, "Trastornos del espectro autista," *Natl. Inst. Ment. Heal.*, vol. 1, pp. 221–237, 2016.
- [8] M. Vera, "Método terapéutico aplicado en los trastornos de la comunicación a niños autistas en el Departamento Nacional de Logopedia y Foniatría.," 2007.
- [9] M. Kim, W. Jang, and K. Kim, "A Virtual Joy-Stick Study of Emotional Responses and Social Motivation in Children with Autism Spectrum Disorder," *J. Autism Dev. Disord.*, vol. 45, no. 12, pp. 3891–3899, 2015.
- [10] D. Roth *et al.*, "Avatar realism and social interaction quality in virtual reality," *Proc. - IEEE Virtual Real.*, vol. 2016–July, pp. 277–278, 2016.
- [11] M. Esch and J. Botev, "Distance-Aware Avatar Interaction in Online Virtual Environments," *2010 Second Int. Conf. Adv. Futur. Internet*, pp. 56–62, 2010.
- [12] M. Esch, J. Botev, H. Schloss, and I. Scholtes, "P2P-based Avatar interaction in massive multiuser virtual environments," *Proc. Int. Conf. Complex, Intell. Softw. Intensive Syst. CISIS 2009*, pp. 977–982, 2009.
- [13] M. Moghimi, R. Stone, P. Rotshtein, and N. Cooke, "Influencing Human

- Affective Responses to Dynamic Virtual Environments,” *Presence Teleoperators Virtual Environ.*, vol. 25, no. 2, pp. 81–107, 2016.
- [14] O. Schreer, R. Englert, P. Eisert, and R. Tanger, “Real-time vision and speech driven Avatars for multimedia applications,” *IEEE Trans. Multimed.*, vol. 10, no. 3, pp. 352–360, 2008.
- [15] E. Alegre and G. Pajares, *Conceptos y Métodos en Visión por Computador*, vol. 1. 2016.
- [16] B. Fasel and J. Luettin, “Automatic facial expression analysis: A survey,” *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, 2003.
- [17] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR 2001*, vol. 1, p. I-511-I-518, 2001.
- [18] R. Lienhart and J. Maydt, “An extended set of Haar-like features for rapid object detection,” *Proceedings. Int. Conf. Image Process.*, vol. 1, p. I-900-I-903.
- [19] S. Soo, “Object detection using Haar-cascade Classifier,” *Inst. Comput. Sci. Univ. Tartu*, vol. 2, no. 3, pp. 1–12, 2014.
- [20] D. Gay, “Implementación de detección del parpadeo para raton de cabeza,” Universitat Rovira Virgili, 2016.
- [21] Y. Yang and S. Newsam, “Comparing sift descriptors and gabor texture features for classification of remote sensed imagery,” *Proc. - Int. Conf. Image Process. ICIP*, pp. 1852–1855, 2008.
- [22] T. Soukupová and J. Cech, “Real-time eye blink detection using facial landmarks,” *21st Comput. Vis. Winter Work.*, 2016.
- [23] O. García-Olalla and E. Alegre, “Descripción de textura en imágenes utilizando Local Binary Pattern (LPB),” in *Conceptos y Métodos en Visión por Computador*, no. 1979, 2016, pp. 115–130.
- [24] P. Ponce, *Inteligencia artificial con aplicaciones a la ingeniería*. 2010.
- [25] H. Demuth, “Neural Networks,” *Mathworks inc*, vol. 19, no. 1, pp. 1–7, 2006.
- [26] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, “Learning Deep Representation for Face Alignment with Auxiliary Attributes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 918–930, 2016.
- [27] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann, “The IMM face database,” *Environment*, vol. 22, no. 10, pp. 1319–1331, 2003.

- [28] V. Kazemi and J. Sullivan, “One Millisecond Face Alignment with an Assemble of Regression Trees,” *27th IEEE Conf. Comput. Vis. Pattern Recognit.*, no. August, pp. 1867–1874, 2014.
- [29] V. Blanz and T. Vetter, “A morphable model for the synthesis of 3D faces,” *Proc. 26th Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '99*, pp. 187–194, 1999.
- [30] K. Waters, “A muscle model for animation three-dimensional facial expression,” *ACM SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 17–24, 1987.
- [31] T. Costigan, M. Prasad, and R. McDonnell, “Facial retargeting using neural networks,” *Proc. Seventh ...*, pp. 31–38, 2014.
- [32] Y. Lee, D. Terzopoulos, and K. Walters, “Realistic modeling for facial animation,” *Proc. 22nd Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '95*, pp. 55–62, 1995.
- [33] P. Hong, Z. Wen, and T. S. Huang, “Real-time speech-driven face animation with expressions using neural networks,” *IEEE Trans. Neural Networks*, vol. 13, no. 4, pp. 916–927, 2002.
- [34] K. K. Kim, K. C. Kwak, and S. Y. Chi, “Gesture analysis for human-robot interaction,” *2006 8th Int. Conf. Adv. Commun. Technol.*, vol. 3, pp. 1824–1827, 2006.
- [35] V. A. Oliveira and A. Conci, “Skin Detection using HSV color space,” *H. Pedrini, J. Marques Carvalho, Work. Sibgrapi*, pp. 1–2, 2009.
- [36] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, “A survey of skin-color modeling and detection methods,” *Pattern Recognit.*, vol. 40, no. 3, pp. 1106–1122, 2007.
- [37] L. F. Guerrero-Vasquez, J. F. Bravo-Torres, and M. Lopez-Nores, “AVATAR ‘autism: Virtual agents to augment relationships in children,’” *Proc. 2017 IEEE 24th Int. Congr. Electron. Electr. Eng. Comput. INTERCON 2017*, 2017.
- [38] M. Rezaei and R. Klette, “3D cascade of classifiers for open and closed eye detection in driver distraction monitoring,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6855 LNCS, no. PART 2, pp. 171–179, 2011.
- [39] H. A. I. Tao, “A Piecewise Bézier Volume Deformation Model and Its

- Applications in Facial Motion Capture,” *Image Process.*, vol. D, no. c.
- [40] J. Ahlberg, “An Experiment on 3D Face Model Adaptation using the Active Appearance Algorithm,” pp. 1–9.
- [41] J. Ondras, O. Celiktutan, E. Sariyanidi, and H. Gunes, “Automatic replication of teleoperator head movements and facial expressions on a humanoid robot,” *RO-MAN 2017 - 26th IEEE Int. Symp. Robot Hum. Interact. Commun.*, vol. 2017–Janua, pp. 745–750, 2017.
- [42] D. Moore, Yufang Cheng, P. Mcgrath, and N. J. Powell, “Collaborative Virtual Environment Technology for People With Autism,” *Focus Autism Other Dev. Disabl.*, vol. 20, no. 4, pp. 231–243, 2005.
- [43] A. D. Kelley, M. C. Malin, and G. M. Nielson, “Computer Graphics Volume 15, Number 3 August 1981,” *Computer (Long. Beach. Calif.)*, vol. 15, no. 3, pp. 263–268, 1981.
- [44] C. Zalapa, “Cómputo tangible en apoyo en las terapias sensoriales para niños con autismo,” CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN SUPERIOR DE ENSENADA, BAJA CALIFORNIA, 2014.
- [45] E. Carrillo Zambrano and C. M. Pachón Meneses, “Creación, Diseño e Implantación de Plataforma E-Learning utilizando Mundos 3D para los niños con trastorno del espectro autista,” *Rev. Educ. y Desarro. Soc.*, vol. 5 No. 1, pp. 70–80, 2011.
- [46] M. Rezaei, “Creating a Cascade of Haar-Like Classifiers : Step by Step,” pp. 1–8, 2014.
- [47] M. Milne, M. H. Luerssen, T. W. Lewis, R. E. Leibbrandt, and D. M. W. Powers, “Development of a virtual agent based social tutor for children with autism spectrum disorders,” *2010 Int. Jt. Conf. Neural Networks*, pp. 1–9, 2010.
- [48] S. V. Ioannou, A. T. Raouzaïou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis, and S. D. Kollias, “Emotion recognition through facial expression analysis based on a neurofuzzy network,” *Neural Networks*, vol. 18, no. 4, pp. 423–435, 2005.
- [49] M. B. Stegmann, “Extraction of MPEG-4 Facial Animation Parameters with Active Appearance Models Aim & Method,” *Technology*, 2001.
- [50] M. Rezaei, “Face and Eye Detection Using OpenCV : Step by Step,” pp. 1–14.
- [51] Z. Zhang, P. Luo, C.-C. Loy, and X. Tang, “Facial Landmark Detection by

- Deep Multi-task Learning,” *Eccv*, 2014.
- [52] R. Green, “Human Motion Capture and Representation,” no. Ivcnz, pp. 209–214, 2009.
- [53] P. Potocnik, “Neural Networks : MATLAB examples,” p. 91, 2012.
- [54] D. Terzopoulos and K. Watters, “Physically-based Facial modelling, analysis and animation,” *Vis. Comput. Animat.*, vol. 1, pp. 73–80, 1990.
- [55] Y. Fu, S. Member, R. Li, T. S. Huang, and L. Fellow, “Real-Time Multimodal Human – Avatar Interaction,” *IEEE Trans. Circ. Syst. Video Technol.*, vol. 18, no. 4, pp. 467–477, 2008.
- [56] T. Nose and Y. Igarashi, “Real-Time Talking Avatar on the Internet Using Kinect and Voice Conversion,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 12, 2015.
- [57] P. A. G. Forbes, X. Pan, and A. F. Antonia, “Reduced Mimicry to Virtual Reality Avatars in Autism Spectrum Disorder,” *J. Autism Dev. Disord.*, vol. 46, no. 12, pp. 3788–3797, 2016.
- [58] S. Anila and N. Devarajan, “Simple and Fast Face Detection System Based on Edges,” *Intl J. Univers. Comput. ...*, vol. 1, no. May 2014, pp. 54–58, 2010.
- [59] S. Parsons and S. Cobb, “State-of-the-art of virtual reality technologies for children on the autism spectrum,” *Eur. J. Spec. Needs Educ.*, vol. 26, no. 3, pp. 355–366, 2011.
- [60] C. C. P *et al.*, “Técnicas de reconocimiento facial mediante redes neuronales,” *2000 IEEE Int. Symp. Circuits Syst. Emerg. Technol. 21st Century Proc. IEEE Cat No00CH36353*, vol. 5, no. 2, p. 135, 2003.
- [61] S. J. Weeks and R. P. Hobson, “The Salience of Facial Expression for Autistic Children,” *J. Child Psychol. Psychiatry*, vol. 28, no. 1, pp. 137–152, 1987.
- [62] S. A. A. Jin, “The virtual malleable self and the virtual identity discrepancy model: Investigative frameworks for virtual possible selves and others in Avatar-based identity construction and social interaction,” *Comput. Human Behav.*, vol. 28, no. 6, pp. 2160–2168, 2012.
- [63] D. Esteban, G., Fernández, C., Casimiro, A. and A. Perdro, J., Pedro y Vega, “TIC y Envejecimiento de la Sociedad. Interacción háptica,” *Cuad. la Cátedra Telefónica*, vol. 1, pp. 27–46, 2012.
- [64] L. F. Guerrero-vasquez, D. Landy-rivera, G. Ávila, J. F. Bravo-torres, and M. López-nores, *World Congress on Medical Physics and Biomedical*

Engineering 2018, vol. 68/2. Springer Singapore, 2019.

- [65] T. Roosendaal and S. Selleri, “Blender 2.3 La Suit Abierta de creacion 3D,” p. 469, 2015.

APÉNDICES

APÉNDICE A: TARGET DE 20 PATRONES DE ENTRENAMIENTO DE LA RED NEURONAL PARA LA ESTIMACIÓN DE LA UBICACIÓN DE PUNTOS FACIALES

Tabla A.1 Target de patrones de entrenamiento para la estimación de coordenadas de puntos faciales

Imag	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
PX0	57	52	56	52	56	53	50	46	46	53	51	49	53	51	47	46	48	47	49	50
PY0	50	53	50	49	48	47	46	48	49	50	41	50	49	50	49	50	46	45	53	49
PX1	25	21	16	19	23	22	19	17	18	21	20	20	17	17	19	18	20	21	20	17
PY1	31	31	31	34	26	25	35	39	40	35	39	38	27	35	34	35	38	39	41	36
PX2	34	33	29	26	33	32	27	25	26	28	28	28	29	29	28	28	28	29	28	29
PY2	27	27	28	30	24	23	32	33	34	31	33	34	25	31	30	32	33	34	35	31
PX3	45	43	42	36	41	40	34	35	36	38	41	37	39	39	38	36	36	37	38	37
PY3	29	30	28	30	25	24	30	34	35	30	31	34	26	31	33	32	33	34	34	31
PX4	35	32	30	29	35	34	28	27	28	33	28	28	26	32	31	29	29	30	26	31
PY4	40	39	42	44	40	39	45	46	47	43	44	46	38	44	45	44	46	47	52	44
PX5	26	24	22	23	24	23	22	20	21	24	22	23	16	23	20	24	22	23	21	23
PY5	43	41	44	45	40	39	46	49	50	46	47	49	43	47	46	45	49	50	57	47
PX6	34	33	30	29	34	33	27	25	26	30	28	28	27	30	28	30	28	29	25	29
PY6	37	36	40	41	37	36	43	45	46	41	43	43	33	43	42	42	43	44	47	42
PX7	35	33	31	30	33	32	27	27	28	32	30	28	27	31	28	30	28	29	25	30
PY7	46	43	46	46	43	42	46	47	48	47	45	48	45	48	47	47	48	49	58	47
PX8	63	61	63	62	63	62	60	58	59	65	54	62	66	61	60	58	59	58	62	60
PY8	33	32	27	27	31	30	25	27	28	27	28	25	28	28	27	28	27	27	29	29
PX9	73	73	76	70	73	72	73	66	67	73	66	71	76	68	68	67	69	67	74	69
PY9	32	31	22	24	31	30	24	23	24	25	24	24	29	26	27	26	23	24	24	26
PX10	81	84	85	79	78	77	82	76	77	81	76	79	84	80	78	74	81	78	91	79
PY10	34	36	22	25	34	33	25	24	25	27	24	26	35	27	27	26	25	23	24	28
PX11	69	69	77	82	69	68	72	68	69	71	72	73	72	70	70	69	72	79	72	70
PY11	42	41	35	38	46	45	37	37	38	38	35	38	41	38	39	39	37	37	42	38
PX12	78	78	87	74	76	75	80	80	81	81	79	81	80	80	79	75	79	80	90	78
PY12	43	42	35	38	90	89	36	37	38	39	36	38	47	39	39	40	38	39	43	39
PX13	72	70	79	74	70	69	73	70	71	72	74	73	73	72	71	69	72	73	77	71
PY13	39	37	33	36	44	43	34	35	36	35	34	35	38	36	37	38	34	35	35	37
PX14	71	69	79	74	68	67	73	69	70	72	74	73	72	71	72	70	73	74	77	71
PY14	45	46	37	39	48	47	38	40	41	40	36	39	48	42	41	42	38	39	46	41
PX15	51	44	51	47	48	47	46	42	43	48	46	44	44	47	45	43	45	44	39	47
PY15	61	63	60	60	53	52	59	60	61	59	59	60	59	99	59	60	58	59	72	60

PX16	61	56	64	59	59	58	57	55	56	59	57	57	58	58	56	53	56	55	50	57
PY16	64	64	58	59	56	55	57	57	58	58	56	58	61	58	57	59	56	55	69	58
PX17	43	38	39	37	40	39	34	34	35	38	37	35	34	39	38	35	36	35	34	37
PY17	63	63	64	65	53	52	60	61	62	61	59	64	62	61	61	61	62	61	76	62
PX18	65	62	74	69	64	63	65	66	67	66	69	67	63	66	64	61	66	65	64	65
PY18	65	65	59	61	59	58	57	55	56	57	54	57	65	58	58	58	56	55	72	59
PX19	38	34	37	37	30	29	32	33	34	37	32	35	27	39	37	35	35	34	33	38
PY19	78	78	82	81	71	70	85	87	88	78	82	87	82	79	77	77	86	85	99	78
PX20	64	59	76	71	62	61	79	75	76	73	76	76	64	71	67	65	76	75	74	68
PY20	78	81	78	79	77	76	79	80	81	75	79	82	89	75	75	75	79	78	94	76
PX21	43	39	50	47	36	35	48	47	48	47	49	50	38	50	44	44	47	46	40	47
PY21	84	85	97	94	85	84	98	99	99	89	98	97	95	88	87	87	96	95	100	86
PX22	61	57	71	67	57	56	64	62	63	68	63	66	56	62	60	56	66	65	68	63
PY22	83	85	94	91	88	87	97	96	97	87	97	95	94	86	84	85	94	93	100	84
PX23	41	39	42	40	33	32	36	36	37	41	38	35	33	42	39	36	36	35	35	40
PY23	73	76	77	77	68	67	81	83	84	74	80	82	78	75	73	73	81	80	96	73
PX24	52	47	54	51	48	47	48	46	47	52	51	48	45	50	49	46	46	45	43	51
PY24	71	73	70	71	63	62	70	72	73	69	68	71	75	69	68	70	68	67	86	69
PX25	60	53	66	58	57	56	59	56	57	59	59	56	54	57	56	54	55	54	53	58
PY25	72	74	69	70	65	64	68	68	69	67	66	68	74	67	68	68	64	63	84	68
PX26	64	59	77	69	64	63	74	73	74	70	76	73	66	69	68	66	71	70	70	70
PY26	74	78	73	74	74	73	73	74	75	71	72	77	82	72	71	71	73	72	92	71
PX27	57	50	63	59	54	53	60	49	50	59	63	59	53	58	56	54	61	60	52	56
PY27	77	79	77	76	72	71	84	87	88	73	84	85	80	73	74	74	83	82	95	73
PX28	49	45	56	52	45	44	50	60	61	51	53	49	44	49	49	46	50	49	42	49
PY28	77	79	77	76	71	70	86	86	87	73	85	86	80	74	74	73	85	84	97	74

APÉNDICE B: HUESOS DE ANIMACIÓN DEL AVATAR Y SECCIONES DEL MALLADO QUE AFECTAN

El Avatar cuenta con un total de 26 huesos empleados para la animación facial, en la figura inferior se puede apreciar la numeración de estos.

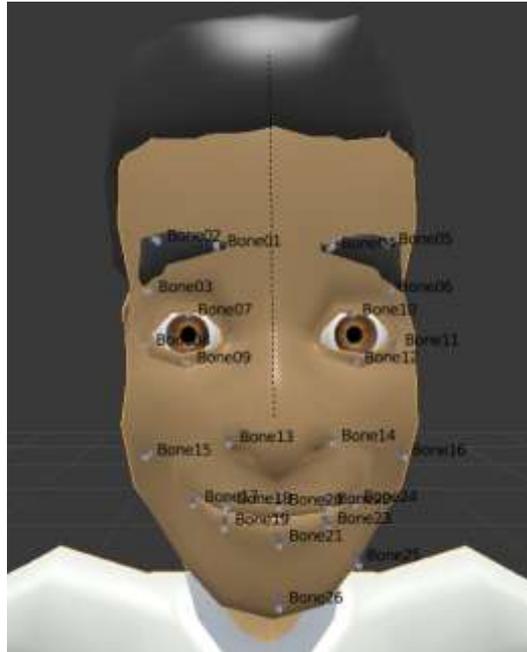


Figura B.1 Avatar empleado en la animación facial

Cada hueso de animación afecta una sección del malla del Avatar, por lo cual para realizar el movimiento de un musculo, como puede ser el ojo, se ha de considerar el movimiento de los huesos que componen al parpado.

La mandíbula y mentón son controladas por el movimiento del bone 26 en los ejes XY, en la siguiente figura se aprecia la sección de malla que este afecta.



Figura B.2 *Sección del avatar afectado por el hueso 26*
En el caso de los cachetes este movimiento estará dado por los huesos 15, 16 y 25 lo que poseen un movimiento único en el eje X

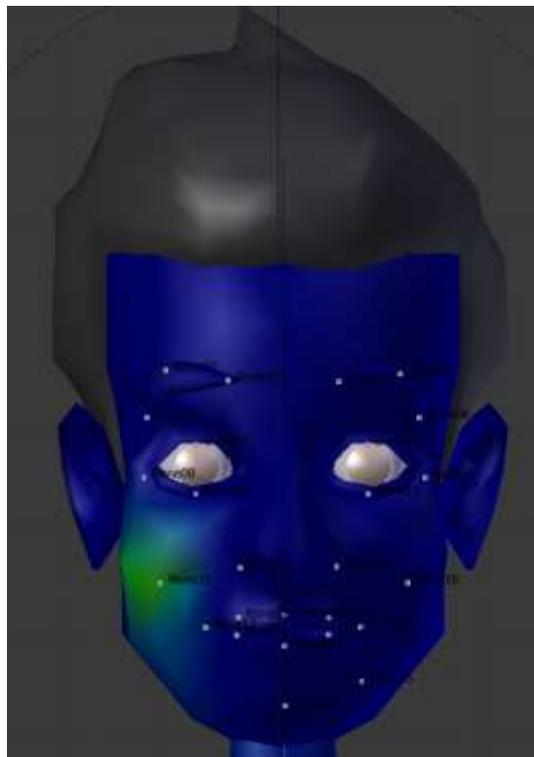


Figura B.3 *Sección del avatar afectado por el hueso 15*



Figura B.4 Sección del avatar afectado por el hueso 16



Figura B.5 Sección del avatar afectado por el hueso 25

El movimiento que abarca una mayor cantidad de huesos está dado por la boca el cual emplea un total de ocho huesos, los que presentan un movimiento en el eje Y, estos huesos abarcan desde el 17 al 24.

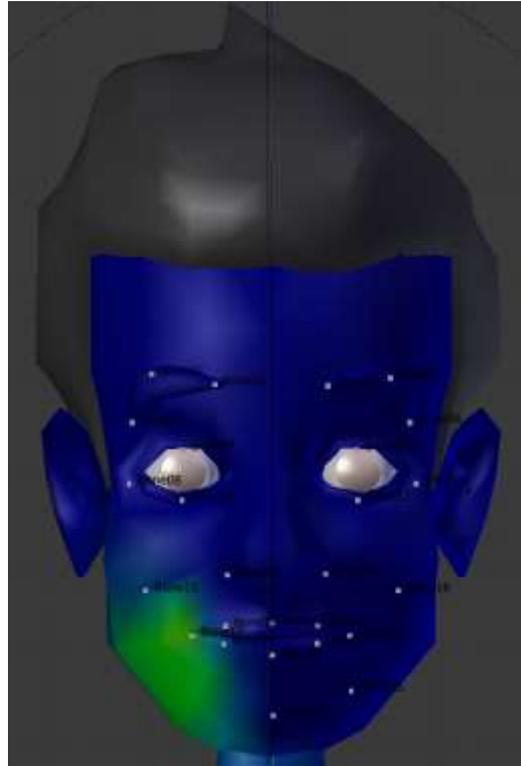


Figura B.6 Sección del avatar afectado por el hueso 17

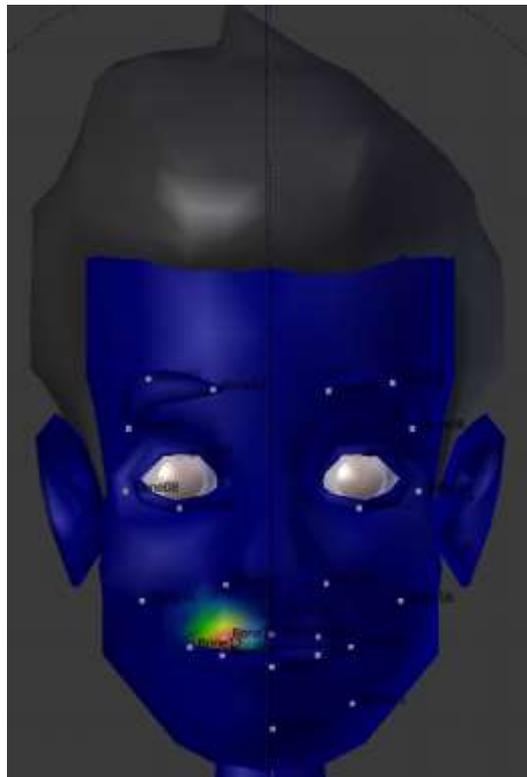


Figura B.7 Sección del avatar afectado por el hueso 18



Figura B.8 Sección del avatar afectado por el hueso 19

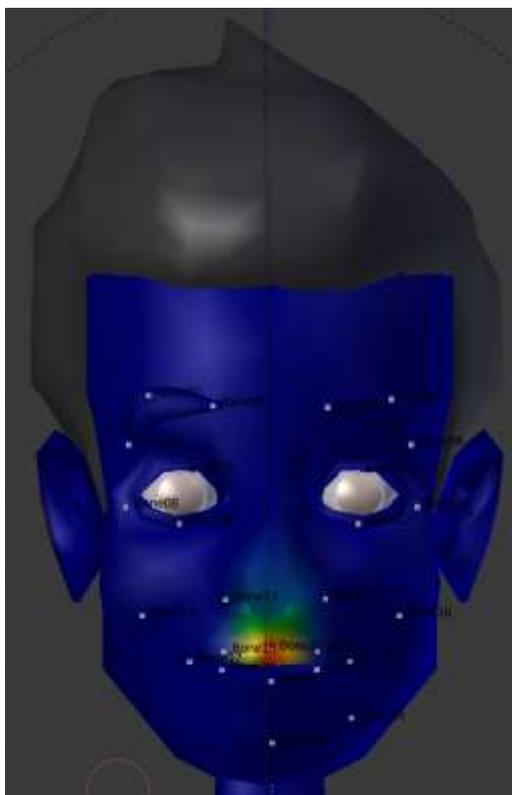


Figura B.9 Sección del avatar afectado por el hueso 20



Figura B.10 Sección del avatar afectado por el hueso 21



Figura B.11 Sección del avatar afectado por el hueso 22



Figura B.12 Sección del avatar afectado por el hueso 23

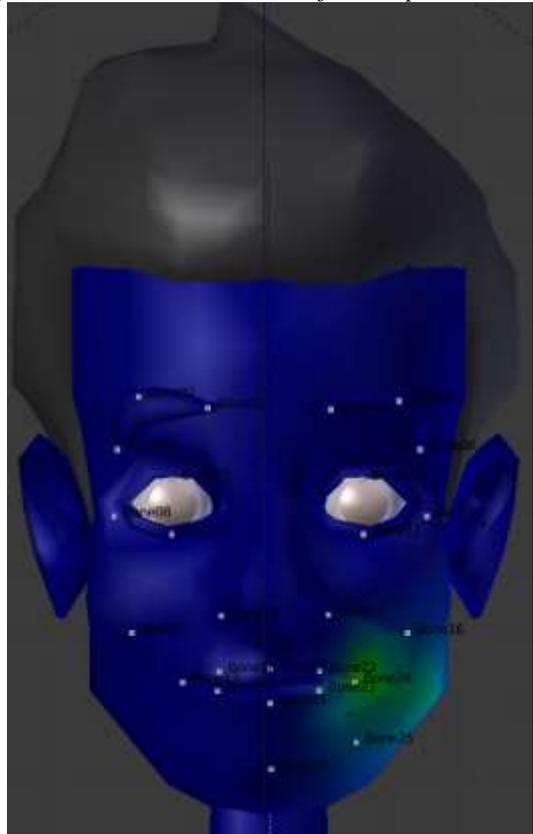


Figura B.13 Sección del avatar afectado por el hueso 24

En el caso del ojo izquierdo su movimiento estará dado por los parpados de este, los cuales se mueven empleando los huesos 10 y 12, lo que se mueven únicamente en el eje Z.

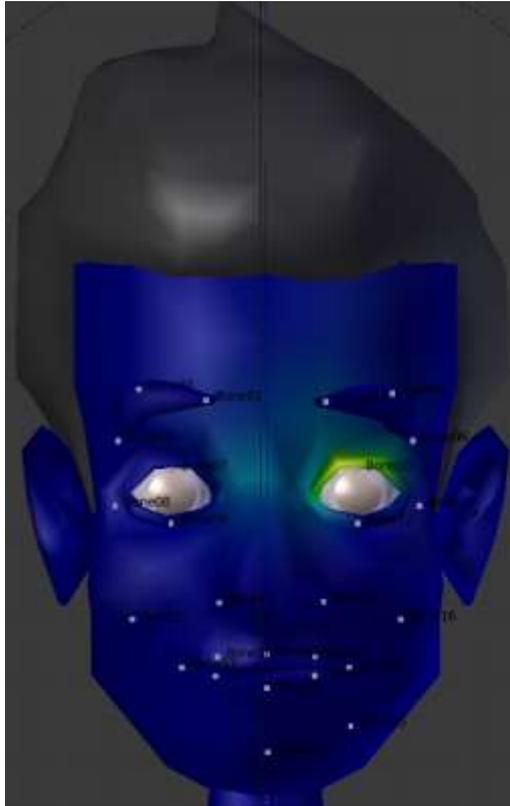


Figura B.14 Sección del avatar afectado por el hueso 10

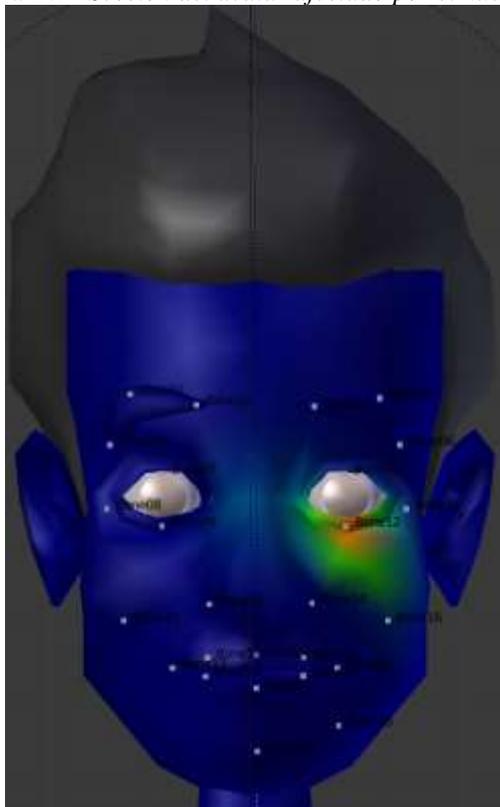


Figura B.15 Sección del avatar afectado por el hueso 12

De manera similar para el ojo su movimiento se dará mediante los huesos 7 y 9.

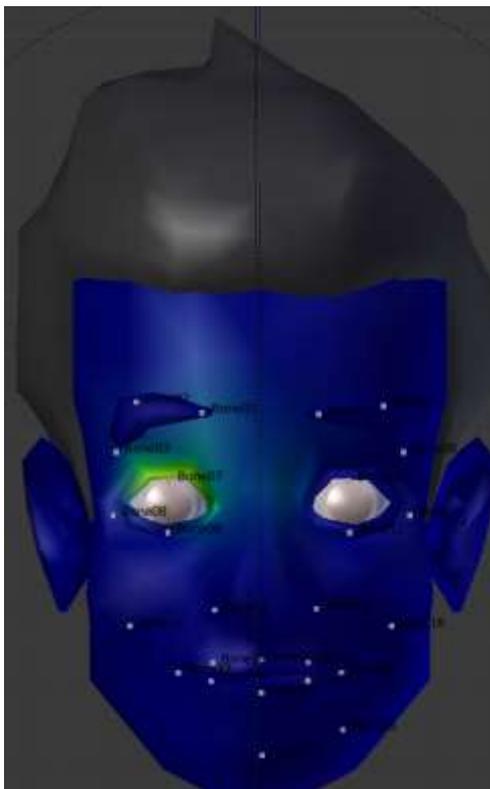


Figura B.16 Sección del avatar afectado por el hueso 7

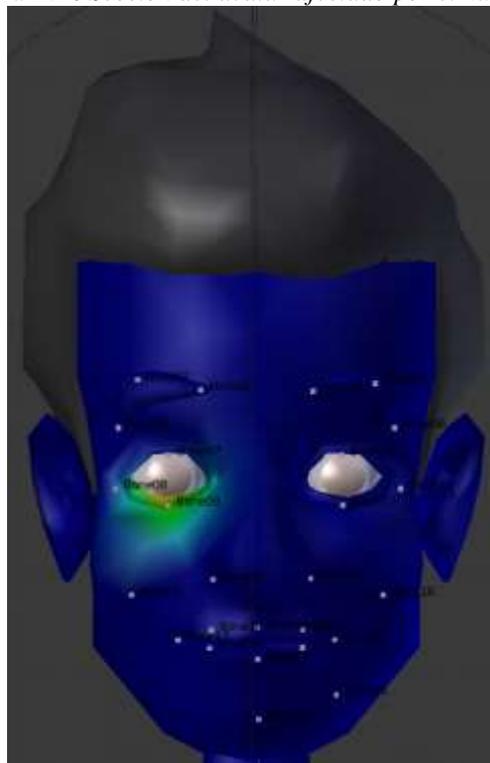


Figura B.17 Sección del avatar afectado por el hueso 9

APÉNDICE C: VALIDACIÓN DEL PROYECTO

Yo Lcda. Viviana Chumbi Mgt, Psicóloga Educativa, Magister en Educación y Desarrollo del Pensamiento con CI 0102918216, actualmente Coordinadora del Departamento de Consejería estudiantil distrito educativo 01d06 por medio de la presente y después de la revisión del trabajo de titulación previo a la obtención del título de ingeniero electrónico del estudiante José Roberto Toledo Illescas **"ANIMACIÓN FACIAL SOBRE AVATAR PARA MEJORAR LA INTERACCIÓN DE NIÑOS CON TRASTORNO DEL ESPECTRO AUTISTA"**, se considera valido el mismo para ser aplicado como terapia alternativa para niños y niñas diagnosticados con trastorno de espectro autista considerando que el trastorno del espectro autista (TEA) es una condición neurológica y de desarrollo que comienza en la niñez y dura toda la vida. Afecta cómo una persona se comporta, interactúa con otros, se comunica y aprende, considerando que si a temprana edad se detecta este trastorno con este tipo de terapia se podría maximizar la capacidad del niño para crecer y aprender nuevas habilidades, ya que cuanto antes se comience, mayores son las probabilidades de tener más efectos positivos en los síntomas y las aptitudes, siempre considerando que cada niño y niña presentara síntomas y la profundidad del mismo dependerá de ello.

Es todo cuanto puedo certificar en honor a la verdad, pudiendo la parte interesada hacer uso del mismo como creyere conveniente.

Atentamente,


Lcda. Viviana Chumbi Mgt.
C.I. 0102918216

Yo Psc. Jaqueline Muñoz Torres Mgt, Psicóloga Clínica, Magister en Educación y Orientación familiar con CI 0301183687, actualmente Coordinadora del Departamento de Consejería Estudiantil Distrito Educativo 01d06 por medio de la presente y después de la revisión del trabajo de titulación previo a la obtención del título de ingeniero electrónico del estudiante José Roberto Toledo Illescas **"ANIMACIÓN FACIAL SOBRE AVATAR PARA MEJORAR LA INTERACCIÓN DE NIÑOS CON TRASTORNO DEL ESPECTRO AUTISTA"**, se considera válido el mismo para ser aplicado como terapia alternativa para niños y niñas diagnosticados con trastorno de espectro autista considerando que el trastorno del espectro autista (TEA) es una condición neurológica y de desarrollo que comienza en la niñez y dura toda la vida. Afecta cómo una persona se comporta, interactúa con otros, se comunica y aprende, considerando que si a temprana edad se detecta este trastorno con este tipo de terapia se podría maximizar la capacidad del niño para crecer y aprender nuevas habilidades, ya que cuanto antes se comienza, mayores son las probabilidades de tener más efectos positivos en los síntomas y las aptitudes; siempre considerando que cada niño y niña presentará síntomas y la profundidad del mismo dependerá de ello.

Es todo cuanto puedo certificar en honor a la verdad, pudiendo la parte interesada hacer uso del mismo como creyere conveniente.

Atentamente,


Psc. Jaqueline Muñoz Torres Mgt
C.I. 0301183687