

Centro Nacional de Investigación y Desarrollo Tecnológico

Subdirección Académica

Departamento de Ciencias Computacionales

TESIS DE MAESTRÍA EN CIENCIAS

**Valoración automática de la motivación y atención del
estudiante mediante expresiones faciales**

presentada por

Lic. Catalina Alejandra Vázquez Rodríguez

como requisito para la obtención del grado de
Maestra en Ciencias de la Computación

Director de tesis
Dr. Raúl Pinto Elías

Cuernavaca, Morelos, México. Enero 2017

"30 Aniversario del Centro Nacional de Investigación y Desarrollo Tecnológico"

Cuernavaca, Morelos a 19 de enero del 2017
OFICIO No. DCC/005/2017

Asunto: Aceptación de documento de tesis

C. DR. GERARDO V. GUERRERO RAMÍREZ
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial de la **Lic. Catalina Alejandra Vázquez Rodríguez**, con número de control M14CE070, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis profesional titulado "**Valoración automática de la motivación y atención del estudiante mediante expresiones faciales**" y hemos encontrado que se han realizado todas las correcciones y observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

DIRECTOR DE TESIS



Dr. Raymundo Elías
Doctor en Ciencias en la Especialidad de
Ingeniería Eléctrica
3890453

REVISOR 1



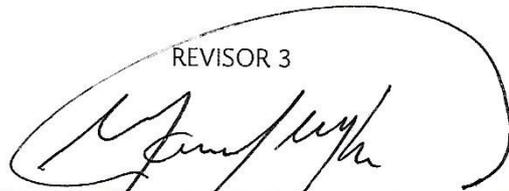
Dr. José Ruiz Ascencio
Doctor en Ciencias
5009035

REVISOR 2



Dr. Andrés Blanco Ortega
Doctor en Ciencias en Ingeniería Eléctrica
6559298

REVISOR 3



Dr. Manuel Mejía Lavalle
Doctor en Ciencias Computacionales
8342472

C.p. M.C. María Elena Gómez Torres - Jefa del Departamento de Servicios Escolares.
Estudiante
Expediente

NACS/lmz

"30 Aniversario del Centro Nacional de Investigación y Desarrollo Tecnológico"

Cuernavaca, Mor., 25 de enero de 2017
OFICIO No. SAC/093/2017

Asunto: Autorización de impresión de tesis

LIC. CATALINA ALEJANDRA VÁZQUEZ RODRÍGUEZ
CANDIDATA AL GRADO DE MAÉSTRA EN CIENCIAS
DE LA COMPUTACIÓN
PRESENTE

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Valoración automática de la motivación y atención del estudiante mediante expresiones faciales"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

"CONOCIMIENTO Y TECNOLOGÍA AL SERVICIO DE MÉXICO"



DR. GERARDO VICENTÉ GUERRERO RAMÍREZ
SUBDIRECTOR ACADÉMICO



SEP TecNM
CENTRO NACIONAL
DE INVESTIGACION
Y DESARROLLO
TECNOLÓGICO
SUBDIRECCIÓN
ACADÉMICA

C.p. M.T.I. María Elena Gómez Torres.- Jefa del Departamento de Servicios Escolares.
Expediente

GVGR/mcr

Dedicatoria

Le dedico este trabajo a:

*Mi familia y Luis García
Gracias por su apoyo incondicional.*

Agradecimientos

Agradezco al Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) por las facilidades y el apoyo proporcionado para la terminación de esta tesis.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico que me brindó durante mis estudios de maestría.

Agradezco al Dr. Raúl Pinto Elías asesor de este trabajo, por todo el apoyo brindado, por sus valiosos consejos y tiempo, los cuales contribuyeron a la culminación satisfactoria de este trabajo.

Agradezco a mi comité revisor: Dr. José Ruíz Ascencio, Dr. Andrés Blanco Ortega y al Dr. Manuel Mejía Lavalle, por sus comentarios y observaciones que me ayudaron a culminar este proyecto. Gracias por el trato amable y la confianza que siempre me brindaron.

Agradezco a mi compañera Rita Morfín por sus recomendaciones.

Agradezco a los compañeros del departamento de Electrónica por su apoyo en la creación del banco de datos de rostros y participación en las pruebas, en especial a Daniel Salazar.

Finalmente agradezco en general a todos los que de forma directa o indirecta contribuyeron a la realización de este trabajo.

Resumen

Las expresiones faciales proporcionan información no verbal acerca de los estados emocionales y mentales de las personas sin necesidad de comunicación verbal, por ello la extracción y el seguimiento automático de componentes faciales son las principales tareas que deben resolver los sistemas de visión artificial en el ámbito del comportamiento humano, detección de expresiones faciales, interfaces hombre-máquina, entre otras áreas. En este trabajo se propone utilizar Kinect y la librería *facetracking* para crear un sistema que localice automáticamente el rostro y realizar una interpretación de sus elementos para detectar la motivación en una actividad. La evaluación de las pruebas de este sistema obtuvo que, para casos donde el sistema determinó que los sujetos estaban prestando atención y realmente estaban prestando atención fue de 89.79% y para los casos donde el sistema evaluó que los sujetos de prueba no prestaron atención y en realidad no lo hicieron fue del 90.72%.

Abstact

Facial expressions provide non-verbal information about people's emotional and mental states without the need for verbal communication, so the extraction and automatic tracking of facial components are the main tasks that artificial vision systems must solve in the realm of Human behavior, detection of facial expressions, man-machine interfaces, among other areas. In this work are proposed to use Kinect and the facetracking library to create a system that automatically locates the face and perform an interpretation of its elements to detect the motivation in an activity. The evaluation of the tests of this system obtained for cases where the system determined that the subjects were paying attention was 89.79%. For the cases where the system evaluated that the test subjects did not pay attention was 90.72%.

Índice

Resumen	I
Abstact	II
Glosario	IX
CAPÍTULO 1 Introducción	1
1.1 Antecedentes	2
1.1.2 Estado del arte.....	3
1.1.2.1 Detección del rostro y elementos faciales	3
1.1.2.2 Seguimiento del rostro.....	3
1.1.2.3 Extracción de las características	4
1.1.2.4 Unidades de acción faciales.....	4
1.1.2.5 Clasificación de expresiones faciales	5
1.1.3 Descripción del problema.....	6
1.2 Antecedentes del proyecto	6
1.3 Objetivos.....	6
1.3.1 Objetivo General.....	6
1.3.2 Objetivos Específicos	6
1.4 Alcances.....	6
1.5 Limitaciones	7
1.6 Justificación	7
1.7 Marco conceptual.....	7
1.7.1 Expresiones faciales.....	7
1.7.2 Preprocesamiento.....	8
1.7.3 Segmentación.....	8
1.7.4 Extracción de características.....	8
1.7.5 Técnicas de caracterización	8
1.8 Interpretación.....	9
1.9 Organización del documento	10
CAPÍTULO 2 Análisis del problema y propuesta de solución	11
2.1 Detección del rostro y seguimiento	12
2.1.1 Función IFTFaceTracker	12
2.1.2 Función IFTImage	12
2.1.3 Función IFTResult	12

2.1.4 Función IFTModel	13
2.2 Detección del rostro	13
2.2.1 Técnicas para la detección del rostro con imagen 3D	13
2.3 Sistema de Codificación de Acciones Faciales (FACS)	14
2.5 Estudio de las unidades de acción, SDK y Toolkit de Kinect.....	16
2.6 AU's utilizadas para la interpretación de las expresiones faciales	16
2.7 Elementos para la interpretación de la atención	18
2.8 Combinación de elementos para la detección de la motivación.....	20
2.9 Videos a utilizar	22
2.10 Clasificación	22
2.10.1 Algoritmos de clasificación utilizados en el estado del arte.....	23
2.11 Propuesta de solución	24
2.12 Discusión	25
CAPÍTULO 3 Análisis, diseño e implementación de la solución	26
3.1 Análisis del sistema	27
3.2 Esquema del sistema	27
3.3 Sistema de reconocimiento de expresiones faciales.....	29
3.3.1 Etapa 1: Recopilación de los datos	30
3.3.2 Etapa 2: Detección elementos faciales e interpretación [29].....	31
3.3.2.1 FaceTracking	31
3.3.2.1.1 Función IFTFaceTracker	31
3.3.2.1.2 Función IFTImage	31
3.3.2.1.3 Función IFTResult	32
3.3.2.1.4 Función IFTModel	32
3.3.2.2 Interpretación de las unidades de acción	32
3.3.2.2.1 Histórico	36
3.3.3 Etapa 3: Clasificación Función discriminante por distancia	38
3.3.4 Etapa 4: Salida de resultado.....	40
3.4 Discusión	42
CAPÍTULO 4 Experimentación y resultados	43
4.1 Ambiente de pruebas	44
4.2 Plan de pruebas	44
4.2.1 Prueba de cambios de iluminación	45
4.2.2 Prueba a distintas distancia con respecto al sujeto de estudio	46

4.3.3 Prueba de identificación de los elementos individualmente	46
4.3.4 Pruebas de las combinaciones de los elementos faciales	48
4.3.5 Resultados de la valoración de la atención de los sujetos	54
4.3.6 Matriz de confusión	62
4.4 Discusión	64
CAPÍTULO 5 Conclusiones	67
5.1 Objetivos alcanzados	68
5.3 Logros.....	69
5.4 Conclusiones	69
5.5 Trabajos futuros	70
Referencias	71
Anexos.....	75
A. Variaciones del rostro con los que se realizó la experimentación	75
B. Video estímulo para accionar los elementos	76
C. AU's utilizadas	78
D. Configuración	79

Índice de figuras

Figura 1.1 Diversas expresiones faciales [30].	7
Figura 2.1 Músculos relacionados con AU's del rostro [29].	14
Figura 2.2 Elemento ceja correspondiente a BrowLower [31].	17
Figura 2.3 Elemento boca correspondiente a MouthCornerLift [31].	17
Figura 2.4 Elemento ojo correspondiente a UpperEyelid [31].	17
Figura 2.5 Elemento ojo correspondiente a LowerEyelid [31].	17
Figura 2.8 Rostro con movimiento el Yaw [34].	18
Figura 2.9 Elemento ojos utilizado para la detección del interés [34].	19
Figura 2.10 Elemento cejas para la detección del interés [33].	19
Figura 2.11 Elemento boca para la detección del interés [33].	19
Figura 2.12 Reglas para la evaluación de la motivación.	20
Figura 2.13. Árbol binario para evaluación de la motivación.	21
Figura 3.2 Esquema del sistema.	28
Figura 3.3 Arquitectura del sistema.	29
Figura 3.4 Funcionamiento del sistema por etapas.	30
Figura 3.5 Contenido del banco de datos.	30
Figura 3.6 Valores que toma el elemento cejas [33].	33
Figura 3.7 Valores que toma el elemento ojos [33].	34
Figura 3.8 Valores que toma el elemento boca[33].	35
Figura 3.9 Valores que toma cabeza en movimiento pitch [42].	35
Figura 3.10 Valores que toma cabeza en movimiento yaw [34].	36
Figura 3.12 Estructura del archivo .txt.	41
Figura 4.1 Captura con iluminación natural.	45
Figura 4.2 Captura con iluminación artificial.	45
Figura 4.3 Captura en diferente distancia.	46
Figura 4.4 Elemento cejas hacia abajo.	47
Figura 4.5 Elemento boca abierta.	47
Figura 4.6 Elemento ojos abiertos.	48
Figura 4.7 Elemento cabeza girada en yaw.	48
Figura 4.8 Sujeto de prueba para la rama 1 del árbol.	49
Figura 4.9 Combinaciones para rama uno.	49
Figura 4.10 Combinaciones para rama uno.	50

Figura 4.11 Combinaciones para rama dos.....	50
Figura 4.12 Combinaciones para la rama tres.	51
Figura 4.13 Combinaciones para la rama tres.	51
Figura 4.14 Combinaciones para la rama cuatro.	52
Figura 4.15 Combinaciones para la rama cuatro.	52
Figura 4.16 Combinaciones para la rama seis.	53
Figura 4.17 Combinaciones para la rama cuatro.	53
Figura 4.18 Sujeto de prueba 1.	54
Figura 4.19. Histórico de la atención del sujeto de prueba uno.	55
Figura 4.20 Evaluación de la atención sin histórico para el sujeto de prueba 1.	55
Figura 4.21 Evaluación de los elementos del un bloque.	56
Figura 4.22 Sujeto de prueba 2.	57
Figura 4.23 Histórico de la atención del sujeto de prueba uno.	57
Figura 4.24 Evaluación de la atención sin histórico para el sujeto de prueba 2.	58
Figura 4.25 Sujeto de prueba 4.....	58
Figura 4.26 Histórico de la atención del sujeto de prueba tres.	59
Figura 4.27 Evaluación de la atención sin histórico para el sujeto de prueba 3.	59
Figura 4.28 Sujeto de prueba 4.....	60
Figura 4.29 Histórico de la atención del sujeto de prueba cinco.	60
Figura 4.30 Evaluación de la atención sin histórico para el sujeto de prueba 4.	61
Figura 4.31 Sujeto de prueba 5.	61
Figura 4.32 Histórico de la atención del sujeto de prueba cinco.	62
Figura 4.31 Sujeto con gafas de marco grueso.	64
Figura 4.32 Sujeto de prueba	66
Figura 4.33 Histórico de la atención del sujeto de prueba seis.....	66
Figura A.1 Rostros para entrenamiento del sistema con variaciones de edades y sexo.	75
Figura A.2 Rostros con diversos tamaños de ojos y boca.	75
Figura B.1 Escenas del video utilizado para generar falta de motivación.....	76
Figura B.2 Sujeto de prueba desmotivado.....	76
Figura B.3 Escenas del video utilizado para generar motivación.....	77
Figura B.4 Sujeto de prueba motivado.	77

Índice de tablas

Tabla 2.1 Comparativa sobre detección de rostro en estado del arte.....	14
Tabla 2.2 Combinaciones de elementos la parte superior e inferior del rostro [17].	15
Tabla 2.3 AU's de los elementos seleccionados.	16
Tabla 2.4 Descripción de videos utilizados.	22
Tabla 2.5 Clasificadores del estado del arte.....	23
Tabla 3.1 Rango de valores para las AU's y elementos.	33
Tabla 3.2 Atención y porcentaje.	42
Tabla 4.1. Matriz de confusión para la atención.	62
Tabla 4.2. Desglose de los datos utilizados en la matriz de confusión.	63
Tabla 4.3 Atención y elemento detonan para la clasificación.....	65
Tabla C.1. 5 Unidades de acción para el reconocimiento de expresiones faciales básicas ...	78

Glosario

2D	2 dimensiones
3D	3 dimensiones
AU	Action Unit (Unidad de Acción Facial)
AU's	Action Units (Unidades de Acción Facial)
Candide	Malla parametrizada desarrollada específicamente para la codificación basada en el modelo de rostros humanos
CENIDET	Centro Nacional de Investigación y Desarrollo Tecnológico
Facetracking	Librería del SDK de Kinect para detección y seguimiento del rostro y sus componentes
FACS	<i>Facial Action Coding System</i> (Sistema de Codificación de Acciones Faciales)
FPS	<i>Frames Per Second</i> (Fotogramas por segundo)
Frame	Fotograma
IR	Infra Rojo
OpenCV	Librería libre de visión artificial
Kinect	Dispositivo de juego desarrollado por Microsoft
RGBD	<i>Red, Green, Blue, Depth</i> (Rojo, Verde, Azul, Profundidad)
SDK	Software Development Kit (Kit de desarrollo de software)
Toolkit	Kit de herramientas
<i>Engagement</i>	Compromiso con actividad realizada
<i>Pitch</i>	Movimiento de la cabeza de manera ascendente-descendente
<i>Yaw</i>	Movimiento de la cabeza arriba abajo

CAPÍTULO 1 Introducción

En esta sección se describen los fundamentos necesarios así como el marco teórico para la comprensión de este trabajo. Se describe el problema general y los problemas específicos de esta tesis, los alcances, limitaciones y la justificación. Se incluye el estado del arte así como los antecedentes del mismo. Además se introduce a las etapas que forman este trabajo para resolver la problemática del reconocimiento automático de la motivación.

1.1 Antecedentes

La comunicación entre personas puede darse de diversas maneras, sin embargo una comunicación precisa dependerá no sólo de aspectos verbales, por tanto se va adaptando acorde con las expresiones corporales. El análisis del comportamiento no verbal es muy útil en diversas áreas: en cursos y talleres presenciales, en los procesos de entrevista laboral, en la investigación del comportamiento del consumidor al seleccionar un producto (por mencionar algunos). La comunicación no verbal transmite información importante con respecto a la actividad que se está realizando.

La Motivación es la técnica que procura que los estudiantes muestren interés por los nuevos conocimientos que imparte el docente, para contribuir al desarrollo académico de los nuevos entes de conocimientos [1]; la falta de interés es señalada como una de las principales causas de fallo en el proceso de aprendizaje, numerosas investigaciones han mostrado que sin interés no hay aprendizaje [2]; tanto teóricos del aprendizaje como profesores están de acuerdo en que los estudiantes que se presentan motivados, aprenden con mayor rapidez y eficacia que los que no lo están [3].

El interés es un sentimiento o emoción que hace que la atención se centre en un objeto, un acontecimiento o un proceso, es por ello que es importante saber si un estudiante se encuentra interesado en la clase ya que a mayor interés su desempeño será mejor; sin embargo la motivación no siempre será expresada verbalmente, también puede ser expresada mediante expresiones faciales, postura y diversas características que pueden dar información acerca de las personas [4].

Las expresiones faciales son la manera más natural e inmediata en que los seres humanos comunican sus emociones. Para reconocer las expresiones faciales mediante técnicas computacionales es fundamental realizar una buena extracción de características [5]. Aunque el reconocimiento de estas expresiones resulta natural para los seres humanos, dotar a una máquina de esta habilidad resulta bastante difícil. Dos tipos de problemas tienen que ser resueltos para llevar a cabo esta tarea: extracción de características y clasificación de la expresión facial [6].

La extracción e interpretación de expresiones faciales es un tema que se ha estado trabajando en CENIDET se desarrollo el tema “Localización, seguimiento y caracterización de expresiones faciales” [7] por Juan José Rosas Varela de maestría. El cual consiste en detectar el cansancio en el rostro aplicado a conductores de automóviles. Otro trabajo realizado es el tema “caracterización de expresiones faciales con un sensor RGBD” llevado a cabo por Rita Alejandra Morfin Villatoro [8] el que consiste en la detección de las 6 emociones básicas mediante un sensor RGBD.

1.1.2 Estado del arte

En este capítulo además se describe el estado del arte de las técnicas utilizadas en las etapas de esta tesis para la localización, seguimiento del rostro, FACS y clasificación.

1.1.2.1 Detección del rostro y elementos faciales

La detección del rostro es la primera etapa del reconocimiento de expresiones faciales, y por lo tanto, su localización debe ser precisa. Los trabajos revisados se clasificaron en dos categorías: los que procesaron imágenes 2D y los que procesan imágenes 3D incluyendo las imágenes RGBD con las cuales se trabajó en esta tesis.

Para imágenes 2D se encontró documentando el uso de clasificador de características Haar para detectar el rostro y sus elementos faciales [9].

Para imágenes 3D y específicamente del tipo RGBD en la tesis “Seguimiento y Caracterización de Componentes del Rostro para la Detección de Expresiones Faciales” se detectó la región más grande de piel y de esta manera ubicar el rostro en la escena [10].

Una nube de puntos es un conjunto de vértices en un sistema de coordenadas tridimensional. Estos vértices se identifican habitualmente como coordenadas X , Y , y Z y son representaciones de la superficie externa de un objeto. Las nubes de puntos tienen múltiples aplicaciones, entre las que se incluye la detección de la forma y textura de rostro [11][12][13].

La detección del rostro es posible llevarse mediante un proceso inverso, es decir, eliminar todo lo que no sea rostro, al eliminar el fondo de la escena el rostro queda detectado para trabajar con sus elementos faciales [35][36].

SDK es un kit de desarrollo de software Microsoft el cual permite crear aplicaciones que puedan rastrear rostros humanos en tiempo real. El motor de rastreo de rostro de Face Tracking SDK analiza la información de entrada de una cámara Kinect, deduce la postura de la cabeza y expresiones faciales, y pone esa información a disposición de una aplicación en tiempo real para la detección de expresiones faciales [21] [22] [23][24].

Estos trabajos se enfocaron en reconocer los elementos y la pose expresiones faciales.

1.1.2.2 Seguimiento del rostro

El seguimiento del rostro es una parte importante para el análisis de los componentes faciales, una vez que se ha detectado el rostro para algunas aplicaciones es necesario dar seguimiento del mismo, el seguimiento del rostro es aplicado a diversos ámbitos de investigación desde cuestiones de seguridad hasta aplicaciones educativas. Algunas de las tecnologías utilizadas en el seguimiento de

rostro son: IFTFaceTracker, IFTImage, IFTResult, IFTModel estas tecnologías son parte del SDK de Kinect [37].

1.1.2.3 Extracción de las características

La extracción de las características de los elementos faciales es muy importante para analizar los parámetros e identificar su estado. Los trabajos revisados se clasificaron en dos: los que procesaron imágenes 2D y los que procesaron imágenes 3D.

En imágenes 2D se utilizó la detección del contorno de los componentes faciales mediante operadores morfológicos y umbralizado para detectar cejas, ojos, nariz y boca, mentón. En estos trabajos se realizó un ajuste del contraste de la imagen para obtener un mejor resultado [38] [39].

En imágenes 3D para extraer la información de ojos, cejas, espacio entre las cejas, surcos alrededor de la boca y boca utilizan ICP - *Iterative Closest Points* (utilizan puntos destacados) y colocan manualmente los puntos, posteriormente los ubican dentro de cada región del rostro [17] [18].

Usando sólo la imagen de profundidad se logra extraer puntos característicos del rostro, específicamente de cejas, ojos y boca mediante el algoritmo de reconocimiento de forma facial básica (BFSC) [12] [19].

Utilizando la imagen de profundidad se aplica ecualización del histograma adaptativo para una mejor definición de la imagen y mediante los momentos de Zernike lograr identificar cejas, ojos y boca [22].

Para el análisis de cejas, boca y mandíbula inferior se utilizan el algoritmos *EigenFaces* de *FaceTracking* del SDK de Kinect [23] [11] [40].

Estos trabajos tienen como finalidad encontrar en tiempo real expresiones faciales.

1.1.2.4 Unidades de acción faciales

Los trabajos revisados relacionados con las unidades de acción que se utilizaron para clasificar las expresiones faciales se clasificaron en dos categorías: los que procesaron imágenes 2D y los que utilizaron 3D.

Las siguientes AU's: 1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25 y 26 se utilizan para la detección de todas las expresiones básicas y el estado neutral analizando los ojos, cejas, nariz, mentón y boca mediante el uso de imágenes 2D [37] [11] [40].

En imágenes 3D con cámara RGBD se detectaron todas las expresiones básicas analizando frente, cejas, ojos, nariz, boca y mandíbula inferior, espacio entre las cejas, surcos alrededor de la boca mediante el uso de AU's de la librería *facetracking*, las cuales trabajan los 6 estados de animo básicos (felicidad, tristeza, miedo, ira, sorpresa y repugnancia).

1.1.2.5 Clasificación de expresiones faciales

Los trabajos revisados relacionados con las unidades de acción que se utilizaron para clasificar las expresiones faciales se clasificaron en dos categorías: los que procesaron imágenes 2D y los 3D.

Para imágenes 2D se utilizó un sistema experto basado en reglas y con eliminación de información ambigua con un porcentaje de reconocimiento del 90% [37].

Creando reglas del tipo IF - THEN se pueden evaluar los parámetros o variables del vector característico de información del rostro. El algoritmo C5.0 se basa en árboles de decisión para clasificar el vector en alguna clase, mediante el algoritmo C5.0 se obtuvo un porcentaje de reconocimiento del 90% [11].

Las máquinas de soporte vectorial (SVM), son un conjunto de algoritmos de aprendizaje supervisado estas máquinas son utilizadas en problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento se etiquetan las clases y se puede construir un modelo que prediga la clase a la que pertenece un nuevo dato, Kotsia [11] utilizó una SVM con kernel de base radial y polinomial con un porcentaje de reconocimiento del 93% para kernel de base radial y 97% para kernel polinomial. Ardabilian [41] obtuvo un reconocimiento de 75% con SVM y kernel lineal. Azazi [40] agregó a la SVM estimaciones de probabilidad con la integración las AU's para maximizar los límites de clase y la variación de las clases fuera mayor, obteniendo una clasificación del 84% correctamente. En Boqing [38] utilizó SVM con kernel lineal con un porcentaje de reconocimiento de 73%, y 89% con kernel sigmoide. Para finalizar con los trabajos que utilizan SVM. Li [39] utilizó kernel de base radial y se consiguió una clasificación del 87% de instancias correctamente clasificadas.

Si se crea una serie de condiciones para definir umbrales es posible diferenciar las expresiones faciales. Luca Stocchi documenta un 89% de clasificación correcta mediante las reglas que creó [17].

El paradigma del algoritmo Eigen Faces supone que cualquier rostro deseado puede ser reconstruido a partir de la imagen media, mediante la adición de una serie de detalles creando un conjunto de vectores llamados vectores propios y los cuales se obtienen utilizando la matriz de covarianza y se aplicó un análisis de componentes principales con el fin de producir la imagen-media, entonces cualquier rostro nuevo se compara utilizando el cálculo de distancias. De esta manera Seddik se consiguió un reconocimiento de 64% [12].

Ruiz trabajó la clasificación de las expresiones faciales mediante árboles de decisión con un porcentaje de reconocimiento de 97% para clasificar la alegría y 96% para clasificar el disgusto [23].

1.1.3 Descripción del problema

El problema que se propuso resolver es la valoración automática de la motivación y atención, mediante la clasificación en: motivado desmotivado o neutral, mediante expresiones faciales. La resolución de este problema conlleva un trabajo en varias etapas; primeramente es necesario detectar el rostro, posteriormente detectar cada uno de los elementos faciales con los que se trabajará para extraer su información; de esta manera se puede realizar una interpretación de los elementos faciales para una clasificación futura.

1.2 Antecedentes del proyecto

La extracción e interpretación de expresiones faciales es un tema que se ha estado desarrollando en CENIDET con diversos fines. Algunos de los trabajos relacionados: “Localización, seguimiento y caracterización expresiones faciales” [7], el cual se enfoca en detectar el cansancio en el rostro, aplicado a conductores de automóviles. Otro trabajo relacionado es el tema “caracterización de expresiones faciales con un sensor RGBD” el cual consiste en la detección de las 6 emociones básicas mediante un sensor RGBD [8].

1.3 Objetivos

1.3.1 Objetivo General

Estudiar e implementar técnicas para el reconocimiento de expresiones faciales (o corporales) y asociarlas a estados de ánimo: motivado, desmotivado o neutral.

1.3.2 Objetivos Específicos

1. Evaluar el sensor RBGD del Kinect y su correspondiente SDK para el reconocimiento de expresiones faciales.
2. Seleccionar elementos del rostro que denoten motivación y desmotivación (tomando en cuenta la motivación desde el enfoque referente al *Engagement*).
3. Obtener bancos de datos de rostros.

1.4 Alcances

1. Evaluar el sensor RBGD del Kinect y su correspondiente SDK para el reconocimiento de expresiones faciales.
2. Elegir entre 2 a 5 elementos faciales o corporales para su estudio.
3. Interpretar los elementos faciales seleccionados.

1.5 Limitaciones

1. No es requisito trabajar con las FACS.
2. No se realizará identificación de personas.
3. No se garantiza la correcta identificación, interpretación y clasificación del rostro si las personas están usando accesorios (entiéndase como accesorios: lentes, gorra, maquillaje excesivo, tapaboca, bigote, barba, etc.).
4. Se pretende que el funcionamiento del sistema de visión artificial sea en tiempo real, sin embargo eso dependerá de las características del hardware y de los algoritmos a utilizar.

1.6 Justificación

La motivación es una emoción que no siempre será expresada verbalmente, pudiendo ser observada mediante el comportamiento de los elementos faciales, tales como: postura y diversas características que pueden representar si una persona está motivada o desmotivada durante el proceso de aprendizaje. Diversos estudios psicológicos han mostrado que en ocasiones los alumnos por miedo a ser criticados no realizan preguntas y se quedan con dudas y tras una acumulación de dudas se va perdiendo la motivación en la actividad realizada y por tanto el interés.

1.7 Marco conceptual

En esta sección se presentan los conceptos y temas relacionados, necesarios para la comprensión de esta tesis.

1.7.1 Expresiones faciales

Las expresiones de la cara son probablemente la variable que más se observa para obtener información de las emociones de los interlocutores. Las expresiones faciales de estas emociones son comunes en todas las culturas, e innatas, es decir, se nace con la capacidad de emitirlas, no se tiene que aprender a hacerlo [14]. En la Figura 1.1 se muestran diversos estados de ánimo, definidos únicamente por las características de los elementos faciales.



Figura 1.1 Diversas expresiones faciales [30].

1.7.2 Preprocesamiento

El objetivo del preprocesamiento es conseguir una mejora en la calidad y/o la apariencia de la imagen original. Resalta ciertas características de una imagen: bordes, contraste y, se ocultan o eliminan otras por ejemplo ruido. El preprocesamiento es una etapa necesaria para otras fases posteriores de análisis tales como: segmentación, extracción de características, reconocimiento, e interpretación o clasificación. Los principales tipos de preproceso son el realzado y filtrado.

1.7.3 Segmentación

La segmentación es un proceso por el cual se extrae de la imagen cierta información subyacente para su posterior uso. La segmentación está basada en dos principios fundamentales: discontinuidad y similitud.

La segmentación orientada a bordes (discontinuidad) y orientada a regiones (similitud). Un borde es una línea que separa dos regiones por tanto de diferentes propiedades por ello se le conoce como discontinuidad. Una región en general es un área de la imagen en la que sus píxeles poseen propiedades similares ya sea de intensidad, de color u otras. Tanto la detección de bordes como la de regiones implican una manipulación de la imagen original, que supone en definitiva una transformación de la imagen original de forma que los valores de los píxeles originales son modificados mediante ciertas funciones.

1.7.4 Extracción de características

Una vez segmentada la imagen, es necesario obtener las características de las distintas regiones de interés. La extracción de características basadas en el contenido de la imagen, tanto en la parte de creación como de recuperación del banco de datos trabaja a partir de imágenes en color; para realizarlo es necesario extraer patrones que permitan representarlo y describirlo ya sea por su color, tamaño, posición, entre otros. Para hacer una buena descripción de las característica debe tratarse de una descripción única y completa por lo que no deberá presentar ambigüedades.

1.7.5 Técnicas de caracterización

Para reconocer las expresiones faciales es fundamental realizar una buena extracción de características. En este trabajo se presentan diferentes técnicas de extracción de características, especialmente aquellas que se aplican en forma global u holística; se hace una comparación entre métodos, en términos de sus porcentajes de reconocimiento, se describen las técnicas más usadas para esta tarea tales como: seguimiento de movimientos, análisis espacial estadístico, y análisis espacio/frecuencia y se presentan los requerimientos que debe tener un sistema automático de reconocimiento de expresiones faciales.

Entre las características faciales a obtener se pueden distinguir dos tipos: las transitorias que abarcan diferentes clases de arrugas y abultamientos y se producen debido a las expresiones faciales y las no transitorias, que siempre están presentes en el rostro, entre ellas se encuentran los ojos, las cejas y la boca, y a diferencia de las primeras siempre están presentes en las expresiones faciales.

Dentro de los métodos que aproximan movimientos quizás el más estudiado ha sido el flujo óptico, este refleja los cambios de la imagen durante un intervalo de tiempo. La estimación del flujo óptico es un tema que ha sido abordado básicamente de dos formas: calculando gradientes o realizando segmentación, en este documento se comentará la primera forma por ser la primera y hasta ahora la más usada. El método de flujo óptico basado en gradientes consiste en determinar los cambios de niveles de gris (intensidad) sobre el plano de la imagen y sobre el tiempo, y a partir de ellos obtener la aproximación del movimiento. Este método se basa en la información que entrega los cambios de intensidad de la imagen asociados al movimiento. Los algoritmos que determinan el flujo a partir del gradiente se basan en los trabajos de Horn and Shunck [12] quienes fueron los primeros en proponer una forma de aproximación de movimientos con imágenes, partiendo que la intensidad de un pixel particular es constante y no cambia con el tiempo.

En la extracción de características por medio de procesos holísticos espaciales se utilizan por lo general imágenes en niveles de grises entre estos estudios se encuentran: Analisis de componentes principales (por sus siglas en inglés PCA) e análisis de componentes independientes (por sus siglas en inglés ICA) los cuales usan representaciones del rostro, encontrados por métodos típicos estadísticos no supervisados que encuentran un conjunto de imágenes base y representan los rostros como una combinación lineal de estas imágenes base.

La idea del PCA es encontrar la base de vectores que mejor exprese la distribución de las imágenes de las caras dentro del espacio completo. Estos vectores describen la base del sub-espacio de las imágenes de caras. Cada vector de dimensión $d=N \times L$ describe una imagen de dimensión $N \times L$, y es una combinación lineal de los vectores base del sub-espacio. Este procedimiento se asemeja a encontrar las componentes ortogonales de un vector en un espacio, es decir, es como encontrar un conjunto de vectores que, combinados en forma lineal, representen a los elementos del conjunto. Estos elementos son los vectores propios de la matriz de covarianza correspondiente al espacio original de las imágenes, y en la literatura se les denomina caras propias o *eigenfaces* [15].

El ICA es una generalización de PCA y a diferencia de este el ICA provee una mejor representación de la imagen ya que no se basa únicamente en estadísticas o relaciones entre pixeles.

El ICA sobre imágenes puede realizarse basándose en dos diferentes arquitecturas, la primera trata la imagen como una señal aleatoria y los pixeles como salidas y una segunda la cual trata a los pixeles como variables aleatorias y las imágenes como salidas. La primera arquitectura produce imágenes base estadísticamente independientes, sin embargo los coeficientes que codifican cada rostro no son necesariamente independientes, por lo que se requiere de una segunda arquitectura que use ICA para encontrar una representación en la cual los coeficientes usados para codificar las imágenes sean estadísticamente independientes [16].

1.8 Interpretación

La interpretación de los elementos faciales es una etapa en la cual cada elemento, boca, ojos, rostro, ceja. Es evaluado con base a su valor numérico, cada uno de estos atributos tiene diversos valores a evaluar y el conjunto de todos ellos da una decisión de si se está interesado o no.

1.8.1 Interpretación de la boca

Evalúa si está abierta o cerrada, esto para analizar si la boca se encuentra en una posición neutral o se está bostezando ya que una persona que esta bostezando, y que lo hace constantemente podría no estar interesada.

1.8.2 Interpretación de la cabeza

El rostro es analizado con base a dos movimientos cabeceo (mejor conocido por su termino en inglés *pitch*) y guiñada (mejor conocido por su termino en inglés *yaw*). El movimiento de *pitch* evalúa si la cabeza está volteando hacia arriba o hacia abajo, mientras que el movimiento de *yaw* evalúa si la cabeza está girada hacia los costados. Son evaluados estos atributos del rostro para concluir si una persona tiene la cabeza girada en otra posición que no corresponde a lo que debería estar atendiendo.

1.8.3 Interpretación de ojos

Los ojos son analizados para saber si se encuentran abiertos o cerrados, ya que una persona con los ojos cerrados definitivamente no está prestando atención incluso podría estar durmiendo.

1.8.4 Interpretación de cejas

Las cejas dicen bastante sobre las emociones, se ha encontrado en la literatura que las cejas en un estado fruncido de la parte interior, suelen denotar concentración por lo cual se decidió analizar si las cejas están fruncidas.

1.9 Organización del documento

Capítulo 1. Contiene los temas que introducen este trabajo.

Capítulo 2. Contiene el análisis del problema de este trabajo el cual comprende el marco teórico, el estado del arte y la discusión correspondiente.

Capítulo 3. Este capítulo comprende el diseño e implementación de la solución para el problema de tesis, se muestra el desarrollo de la herramienta computacional o software asociado al tema de tesis.

Capítulo 4. Describe el objetivo, el diseño y los resultados de las pruebas que se realizaron a lo largo del desarrollo de la tesis.

Capítulo 5. Contiene las conclusiones generales, la discusión sobre el tema de tesis, trabajos futuros y aportaciones.

Anexos. Contienen algunos elementos importantes para el desarrollo de esta tesis, los cuales son: AU's, imágenes, gráficas, tablas etc.

CAPÍTULO 2 Análisis del problema y propuesta de solución

En este capítulo se realiza un breve repaso sobre la problemática a abordar en este trabajo, y la propuesta de cómo resolverlo así como los temas necesarios, los conceptos están divididos en grupos: localización y seguimiento del rostro, FACS y algoritmos de clasificación. En este capítulo además se describe el estado del arte de las técnicas utilizadas en cada una de las etapas siguiendo la arquitectura general de un sistema para el reconocimiento de expresiones faciales.

2.1 Detección del rostro y seguimiento

En esta sección se explica lo que es la detección y seguimiento del rostro aunque no es un objetivo o alcance definido en este trabajo es parte de la problemática a resolver, por tanto se buscó alguna función y librería que realizará dicho proceso para concentrarse en la problemática real a resolver. La librería *facetracking* es una librería que se encarga de la localización y seguimiento del rostro, a continuación se describe y las funciones utilizadas en el estado del arte para localizar y seguir el rostro mediante esta librería.

2.1.1 Función *IFTFaceTracker*

La función *IFTFaceTracker*, es una función de tipo booleano la cual se encarga de la interfaz principal para inicializar la detección del rostro en forma síncrona mediante el paso de *frames* de color y profundidad.

Verifica que tanto los *frames* de color como los de profundidad llegan desde el sensor Kinect.

2.1.2 Función *IFTImage*

Es una función de tipo booleano, esta función se encarga del almacenamiento en memoria del proceso de seguimiento del rostro, la función carga y libera memoria para el buen funcionamiento del proceso de reconocimiento de rostro. Dentro de esta función se encuentran métodos encargados del acceso a la información acerca de la imagen: el formato, la altura, la anchura, el tamaño de la imagen, y los bytes por pixel. Además, *IFTImage* proporciona acceso a la memoria intermedia. Métodos de ayuda para la copia de imágenes y dibujo de líneas de depuración son también parte de esta interfaz.

2.1.3 Función *IFTResult*

Es una función de tipo decimal. Esta función está encargada de verificar que se haya detectado un rostro de forma exitosa, está encargada del seguimiento del rostro.

IFTResult proporciona acceso a la siguiente información a *GetFaceRect* y *Get2DShapePoints*.

GetFaceRect en el fotograma de video ofrece información de las coordenadas del cuadro delimitador alrededor del rostro en seguimiento. *Get2DShapePoints* 2D (x , y) proporciona coordenadas de los puntos clave en el rostro del fotograma de video. *IFTResult* realiza un seguimiento de los puntos en 2D.

2.1.4 Función IFTModel

Es una función de tipo decimal, esta función crea la malla *Candide* sobre el rostro. La interfaz *IFTModel* proporciona una manera de convertir los resultados de rastreo a una malla de vértices en 3D en el espacio de la cámara. La interfaz proporciona diversos métodos para obtener varias propiedades del modelo los cuales se describen a continuación:

1. *GetSUCount*, *GetAUCount* - devuelven respectivamente el número de unidades de forma (SU) o unidades de acción (AU), utilizados en el modelo lineal 3D (malla *Candide*).
2. *GetTriangles* - devuelve triángulos del modelo de la malla 3D (índices de vértices). Cada triángulo tiene 3 índices de vértices que figuran en el sentido horario.
3. *GetVertexCount* - devuelve el número de vértices del modelo de la malla 3D.

Se puede observar que para la detección del rostro son necesarios los *frames* de color y de profundidad. Para extraer las características sólo son necesarios los *frames* de color. Las funciones de *FaceTracking* utilizadas en esta tesis, se encuentran implementados en lenguaje C++.

2.2 Detección del rostro

Esta es la primera etapa del reconocimiento de expresiones y por lo tanto, su localización debe ser precisa. Los trabajos revisados se clasificaron en dos categorías: los que procesaron imágenes 3D para la localización del rostro y los que utilizaron una cámara RGBD.

2.2.1 Técnicas para la detección del rostro con imagen 3D

En la literatura se encontraron algunas técnicas para la detección de rostro mediante sensor RGBD, mismo, que se utilizó en este trabajo.

Todos los trabajos presentados en la Tabla 2.1 tienen como finalidad encontrar expresiones faciales. Las aplicaciones que utilizaron la librería *FaceTracking* realizan el proceso en tiempo real. En la columna técnica se mencionan los algoritmos utilizados. La columna elementos detectados, hace mención de los elementos con los que se trabajó. Finalmente en la columna de descripción se tiene una breve descripción de la técnica utilizada en el artículo.

Tabla 2.1 Comparativa sobre detección de rostro en estado del arte.

Técnica	Elementos detectados	Descripción
Algoritmo de reconocimiento de forma facial básica (BFSC) Gong, 2009 [17]	Cejas, ojos y boca	Extraen puntos característicos del rostro. Sólo usan la imagen de profundidad
Momentos de Zernike [19]	Cejas, ojos y boca	Utiliza la imagen de profundidad aplican un pre-proceso a la imagen de profundidad (ecualización del histograma adaptativo)
Algoritmos EigenFaces [20]	Frente, cejas, ojos, nariz, boca y mentón	El algoritmo se aplica sobre la imagen de profundidad
Herramientas de SDK de Kinect: <i>FaceTracking</i> [21]	C e j a s , b o c a y mandíbula inferior	Proceso en tiempo real
Herramientas de SDK de Kinect: <i>FaceTracking</i> [22]	Cejas, boca y mentón	Proceso en tiempo real
Momentos de Zernike[23]	Cejas, ojos y boca	Sólo utiliza la imagen de profundidad
Herramientas de SDK de Kinect: <i>FaceTracking</i> [24]	Cejas, boca y mentón	Proceso en tiempo real
Herramientas de SDK de Kinect: <i>FaceTracking</i> [25]	Cejas, ojos, nariz, boca y mentón	Proceso en tiempo real

2.3 Sistema de Codificación de Acciones Faciales (FACS)

FACS por sus siglas en inglés *Facial Action Coding System*, es un sistema de codificación desarrollado por Ekman y Friesen, publicado en 1978 [26]. Con este sistema se identifican los movimientos musculares y marcas faciales o líneas de expresión que producen cambios momentáneos en la apariencia del rostro. Estos movimientos son llamados Unidades de Acción (AU). En la Figura 2.1 se muestra la relación entre la anatomía muscular y el movimiento muscular de la parte superior del rostro [11].

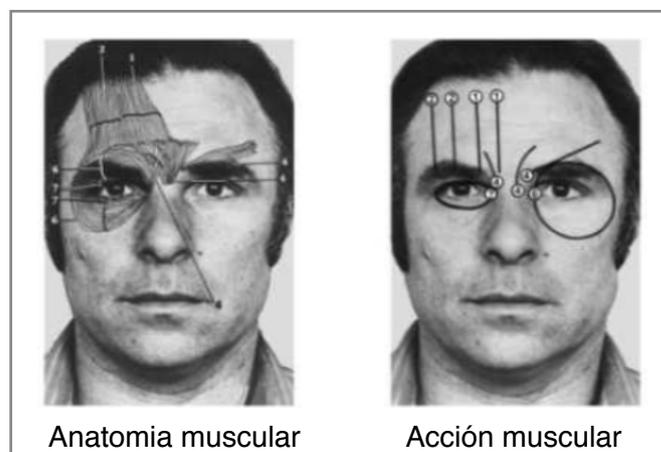
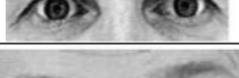


Figura 2.1 Músculos relacionados con AU's del rostro [29].

En la Tabla 2.2 se muestran algunas AU's y combinaciones de AU's relacionadas con la parte superior e inferior del rostro.

Tabla 2.2 Combinaciones de elementos la parte superior e inferior del rostro [17].

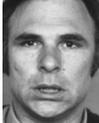
AU	Descripción	Imagen
AU1	Elevación de la parte interna de las cejas	
AU2	Elevación de la parte externa de las cejas	
AU5	Elevación del párpado superior	
AU6	Elevación de mejillas	
AU7	Elevación del párpado inferior	
AU1+4	Elevación de la parte interna de la ceja + generación de esfuerzo	
AU4+5	Disminución de altura y acercamiento de las cejas + elevación de párpado superior	
AU1+2	Elevación de la parte interna y externa de las cejas	
AU1+2+4	Las cejas son empujadas hacia arriba con esfuerzo	
AU1+2+5+6+7	Elevación de cejas, párpado y mejillas	
Neutral	Relajación de ojos cejas y mejillas	
AU9	Elevación del centro del labio superior. Presencia de arrugas en la parte superior de la nariz	
AU10	Elevación del labio superior sin presencia de arrugas	
AU20	Movimiento hacia abajo de manera lateral de los labios	
AU15	Movimiento de las esquinas de los labios hacia abajo	
AU17	Elevación de la region de la barbilla	

La librería *Facetracking* sólo maneja seis AUs relacionadas con las cejas, la boca y el mentón de las cuales se utilizan cuatro en este trabajo para el análisis de la boca, cejas y ojos.

2.5 Estudio de las unidades de acción, SDK y *Toolkit* de Kinect

En este apartado se muestra una tabla de las unidades de acción, las cuales pueden representar a los elementos faciales que son utilizados para la interpretación y clasificación de las expresiones faciales. Los elementos tomados en cuenta son: cejas, boca, ojos y cabeza con diversas posiciones. En la Tabla 2.3 se muestran algunas AU, su correspondiente AU en Kinect, una breve descripción e imagen.

Tabla 2.3 AU's de los elementos seleccionados.

AU	AU Kinect	Descripción	Imagen
AU 5	AU5	Mover ceja inferior hacia abajo (denota concentración)	
AU+1	AU5	Contracción en los músculos de la ceja (denota frustración)	
AU14	AU4	Esquina del labio abajo (denota frustración)	
AU15	AU4	Boca ligeramente abierta (denota desconcierto)	
AU16+25	AU2	Correspondencia de plantillas	
AU27	AU1	Mandíbula hacia abajo (denota bostezo)	

2.6 AU's utilizadas para la interpretación de las expresiones faciales

En el presente trabajo se utilizaron 4 AU's del Toolkit de Kinect: *BrowLower*, *MouthCornerLift*, *LowerEyelid* y *UpperEyelid*. Las cuales corresponden a cuatro de los elementos utilizados para la interpretación de la motivación de los sujetos. *BrowLower*, es la AU que corresponde a las cejas y la cual se utiliza para medir la elevación de las cejas. En la Figura 2.2 se muestra el elemento correspondiente a la AU de *BrowLower*.



Figura 2.2 Elemento ceja correspondiente a *BrowLower* [31].

MouthCornerLift, corresponde a la boca y se utiliza para medir la apertura de la boca mediante las comisuras de la misma. En la Figura 2.3 se muestra el elemento correspondiente a *MouthCornerLift*.

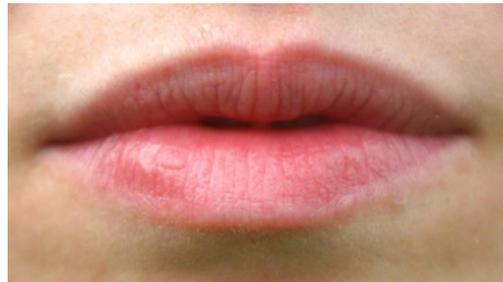


Figura 2.3 Elemento boca correspondiente a *MouthCornerLift* [31].

LowerEyelid, corresponde a los párpados al igual que *UpperEyelid*, con dichas AU's se evalúa la apertura del ojo. En la Figura 2.4 se muestra el elemento correspondiente a *Upper Eyelid* y en la Figura 2.5 *Lower Eyelid*.

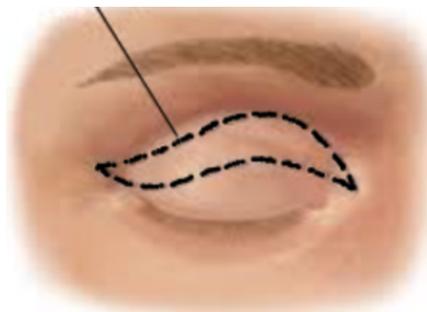


Figura 2.4 Elemento ojo correspondiente a *UpperEyelid* [31].

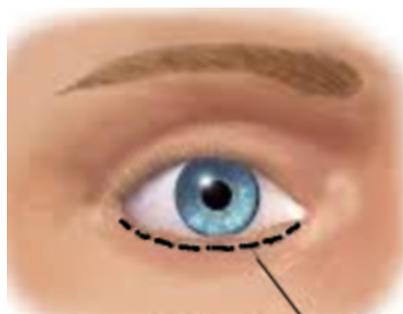


Figura 2.5 Elemento ojo correspondiente a *LowerEyelid* [31].

2.7 Elementos para la interpretación de la atención

Los elementos utilizados como objeto de estudio en esta tesis forman parte de las FACS, que representan expresiones faciales, mas no una interpretación general.

Tras analizar la información, la Psicóloga Verónica Sotelo Boyás enfatizó que algunos de ellos son muy importantes para saber si una persona está interesada en una actividad y por esto podrían ser considerados para reconocer un estado de motivación, desmotivación por tanto se utilizaron: ojos, boca, ceja y posición de la cabeza como elementos para la detección de la motivación e interés en los sujetos.

En la Figura 2.7 se muestra el elemento rostro con el movimiento *Pitch* el cual consiste en la cabeza girada hacia arriba y abajo, este elemento y este movimiento fue tomando en cuenta en el presente trabajo debido a que cuando una persona no está de frente atendiendo la actividad podría ser un indicador de falta de interés.

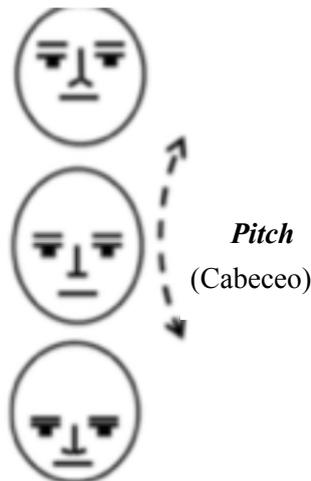


Figura 2.7 Rostro con movimiento en *Pitch* [32].

En la Figura 2.8 se muestra el elemento rostro con el movimiento *Yaw* el cual consiste en la cabeza girada hacia los costados, este elemento con este movimiento fue tomando en cuenta en este trabajo debido a que si una persona no está de frente atendiendo la actividad podría ser un indicador de falta de interés.

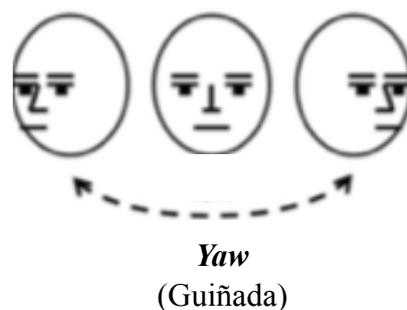


Figura 2.8 Rostro con movimiento el *Yaw* [34].

En la Figura 2.9 se muestra el elemento ojos, este elemento fue tomado en cuenta en este trabajo debido a que si una persona tiene los ojos cerrados podría ser un indicador de falta de interés, incluso si la persona está con los ojos cerrados por un tiempo prolongado podría haberse quedado dormida.



Figura 2.9 Elemento ojos utilizado para la detección del interés [34].

En la Figura 2.10 se muestra el elemento cejas, las cuales fueron tomadas en cuenta ya que dan bastante información sobre las emociones, se ha encontrado en la literatura que, las cejas en un estado bajo o fruncido, suelen denotar concentración por lo cual se decidió analizar las cejas.



Figura 2.10 Elemento cejas para la detección del interés [33].

En la Figura 2.11 se muestra el elemento boca la cual se evalúa si está abierta o cerrada, ya que una persona podría estar bostezando y si esta acción es repetitiva podría ser una característica de falta de interés.



Figura 2.11 Elemento boca para la detección del interés [33].

2.8 Combinación de elementos para la detección de la motivación

Los elementos presentados en el apartado anterior sobre la atención de los sujetos, son utilizados combinados entre ellos, ya que, en general, no es posible con elementos aislados determinar si realmente se está prestando atención o no, por lo cual se realizó un análisis de dichas expresiones en conjunto y sus combinaciones para llegar a un estado de interés o desinterés lo cual se vería reflejado en la motivación de las personas en la actividad que estén realizando.

En la Figura 2.12 se muestra el árbol de reglas para la evaluación del interés.

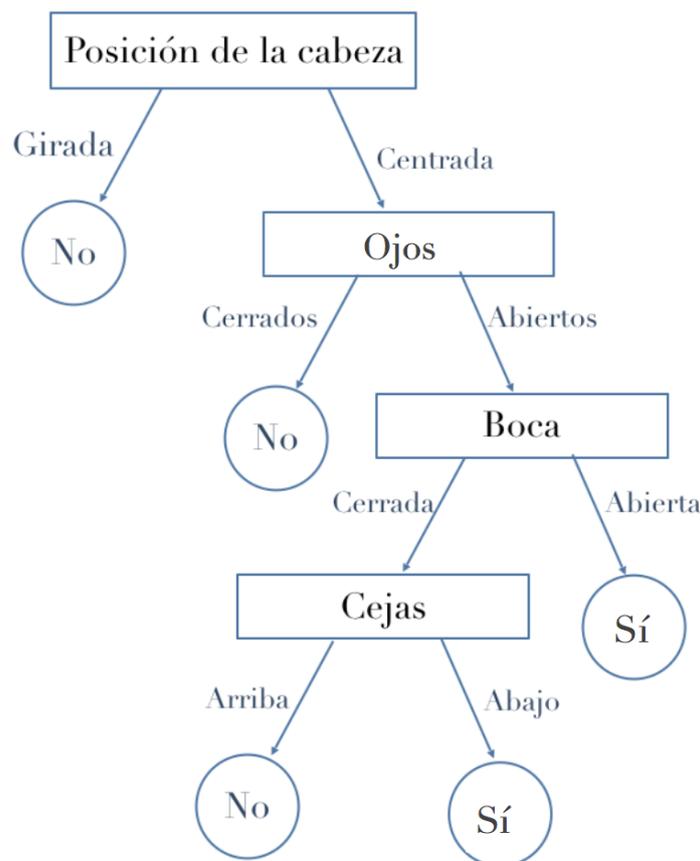


Figura 2.12 Reglas para la evaluación de la motivación.

El primer elemento a evaluar es la cabeza; si la cabeza se encuentra girada no es necesario evaluar los demás elementos debido a que el elemento cabeza es la raíz del árbol y con éste se descarta la atención del sujeto de prueba si está girada. Sin embargo si la cabeza esta centrada se procede a analizar los ojos.

En el segundo nivel del árbol se tiene elemento ojos; donde se evalúa si están abiertos o cerrados, en caso de estar cerrados no es necesario analizar otros elementos, sin embargo si los ojos están abiertos entonces se analizará el elemento boca.

El elemento boca se encuentra en el nivel tres y es evaluada de la siguiente manera: abierta o cerrada; en caso de estar abierta se puede concluir que está atento el sujeto debido a las combinaciones con los otros elementos.

Finalmente las cejas son evaluadas, están en el cuarto nivel del árbol, y este es evaluado si y sólo si la cabeza está centrada, los ojos abiertos y la boca cerrada. Si la ceja está arriba entonces se concluye que no está atento. En el caso de que la ceja esté abajo se concluye que se está poniendo atención. Mediante estas combinaciones y su valor binario es posible determinar si un individuo está prestando atención realmente ya que con un solo elemento no es posible atender una actividad requiere de varios elementos de atención.

Las reglas para la evaluación de la motivación de la Figura 2.12 se basan en un árbol binario, pero fue simplificado para un mejor entendimiento, en la Figura 2.13 se muestra el árbol binario del cual procede el árbol de la Figura 2.12.

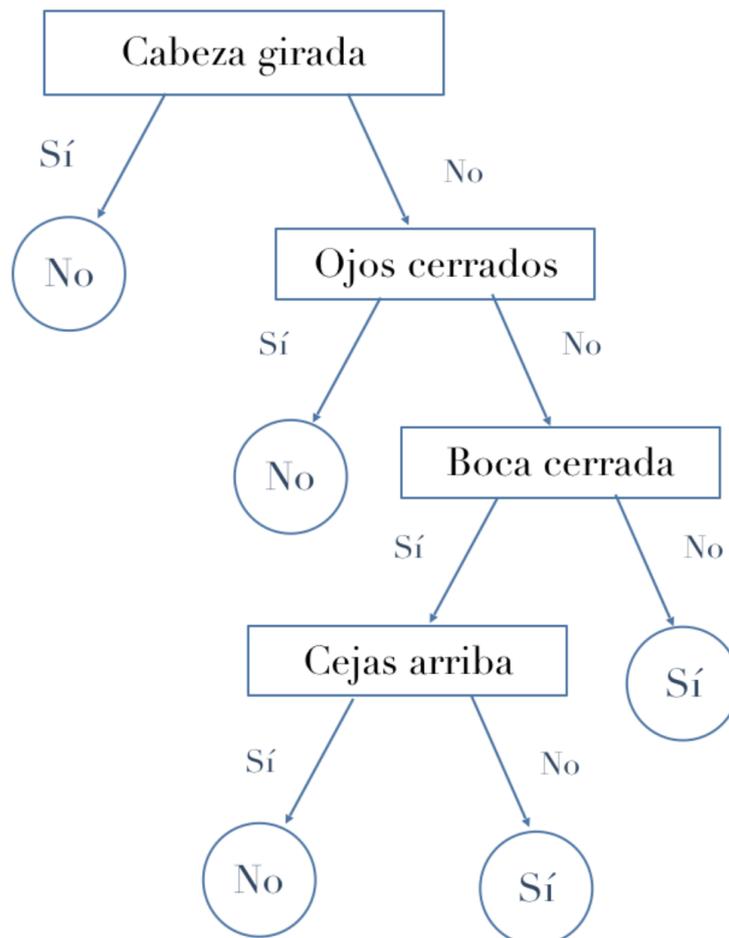


Figura 2.13. Árbol binario para evaluación de la motivación.

2.9 Videos a utilizar

Cuatro videos se utilizan en este trabajo. Para realizar la evaluación automática del interés empleando en las reglas de la Figura 2.13

En la Tabla 2.4 se muestra la información de los videos utilizados en este trabajo así como una breve descripción del mismo.

Tabla 2.4 Descripción de videos utilizados.

Número de video	Video	Descripción
1		La duración de este video es de 1 min. con 40 seg. trata sobre vuelos comerciales
2		la duración de este video es de 3 min. con 19 seg. trata sobre la evolución de la tierra
3		La duración de este video es de 3 min. con 03 seg. trata sobre la transformación de oruga a mariposa
4		La duración de este video es de 11 min. con 44. seg. trata sobre agujeros negros

2.10 Clasificación

El concepto de clasificación tiene tres enfoques:

1. Clasificación no supervisada también conocida como *clustering*: a diferencia de la supervisada no se cuenta con conocimiento *a priori*, por lo que no se tiene un banco de datos para entrenamiento del sistema. En este tipo de clasificación se cuenta con “objetos” o muestras, de las que no se sabe a qué clase o categoría pertenecen, entonces la finalidad es el descubrimiento de grupos de “objetos” cuyas características afines permiten separar las diferentes clases.

2. Clasificación supervisada: este tipo de clasificación cuenta con un conocimiento *a priori*, es decir para la tarea de clasificar un objeto dentro de una categoría o clase se cuenta con modelos ya clasificados es decir objetos agrupados que tienen características comunes. Se puede diferenciar dos fases dentro de este tipo de clasificación.

3. Clasificación semi-supervisados: Es una mezcla de los dos anteriores.

Existen diversos tipos de clasificadores tales como: métodos estadísticos clásicos, árboles de decisión, reglas, redes neuronales, máquinas de soporte vectorial, entre otros. Se debe tener en cuenta una serie de consideraciones con respecto a los clasificadores.

1. Exactitud – proporción de clasificaciones correctas.
2. Rapidez – tiempo que toma hacer la clasificación.
3. Claridad – qué tan comprensible es para los humanos.
4. Tiempo de aprendizaje – para obtener o ajustar el clasificador a partir de datos.

2.10.1 Algoritmos de clasificación utilizados en el estado del arte

En esta sección se mencionan los trabajos relacionados con los algoritmos de clasificación que se utilizaron para reconocer las expresiones faciales. Los trabajos revisados involucran el uso de una cámara RGBD. En la Tabla 2.5 se muestran algunos algoritmos del estado del arte utilizados para la clasificación en trabajos similares.

Tabla 2.5 Clasificadores del estado del arte.

Algoritmo de clasificación	Detalles del algoritmo usado	Porcentaje de reconocimiento
Máquina de Soporte Vectorial [18]	Algoritmo de clasificación.	75%
Máquina de Soporte Vectorial [19]	Kernel lineal, de base radial y sigmoide	Lineal (81%) Base radial (88%) Sigmoide (84%)
Máquina de Soporte Vectorial [21]	No especifica cuál fue el kernel utilizado	73%
Condicional (umbrales) [12]	El autor crea una condición para cada expresión facial	89%
EigenFaces [22]	El paradigma Eigenface supone que cualquier rostro deseado puede ser reconstruido a partir de la imagen media, mediante la adición de una serie de detalles.	64%
Máquina de Soporte Vectorial [23]	Sin registro	87%

Continuación de Tabla 2.5 Clasificadores del estado del arte.

Algoritmo de clasificación	Detalles del algoritmo usado	Porcentaje de reconocimiento
Árboles de decisión [27]	Sin registro	Trabajo en desarrollo
Máquina de Soporte Vectorial [28]	Sin registro	83%

2.11 Propuesta de solución

La propuesta para la solución al problema planteado (ver 1.1.3) se muestra en la figura 2.14 y se traduce en determinar el problema de detección de expresiones faciales para la interpretación de la atención de una persona utilizando una RGBD

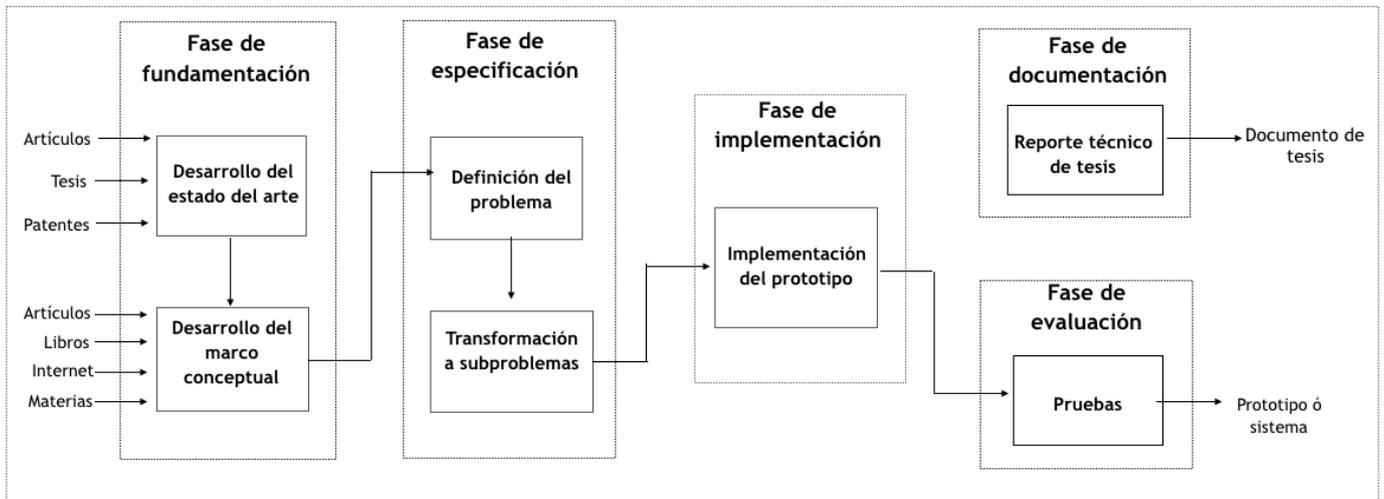


Figura 2.14 Propuesta de solución.

La fase de fundamentación es la parte donde es necesario leer artículos y tesis de trabajos relacionados para saber con qué algoritmos y de qué manera se está solucionando la problemática de la detección de rostro y elementos faciales.

En la fase de especificación se pretende definir la problemática de este trabajo para posteriormente dividirla en subproblemas y resolverlos uno a uno. Es decir identificar con qué elementos faciales se trabajará y por qué, así como los parámetros que se manejarán, en este trabajo se seleccionaron los elementos: ojos, ceja, boca, y cabeza; donde la cabeza es evaluada con base a dos parámetros: el giro arriba/abajo y el giro a los costados.

En la etapa de implementación se desarrollará el código necesario para hacer extracción, interpretación y clasificación de los elementos seleccionados para dar un veredicto sobre si la persona está interesada, desinteresada o en un estado neutral. Esto mediante la medición de la atención a partir de unos videos de prueba que el sujeto visualizará.

2.12 Discusión

Dentro del estado del arte utilizaron el sensor Kinect así como la librería *facetracking*. El sistema que se implementó en esta tesis puede detectar la motivación y desmotivación en los individuos. El sistema puede detectar expresiones sin embargo cuando el rostro está cubierto por accesorios presenta confusión ya que los elementos podrían no ser totalmente visibles. Se detectó que el elemento cabeza es el más importante por ello su prioridad a evaluar y posteriormente los demás elementos.

CAPÍTULO 3 Análisis, diseño e implementación de la solución

En este capítulo se describe el análisis y diseño del proceso necesario para la implementación del Sistema de Reconocimiento de Expresiones Faciales para la detección automática del interés; este proceso incluye: la recopilación de los datos (imágenes de rostros), la implementación de las funciones de *FaceTracking* para el reconocimiento y seguimiento del rostro y sus componentes, para reconocer las expresiones faciales de manera automática y la implementación del clasificador.

3.1 Análisis del sistema

El análisis del sistema se define con base en los requerimientos necesarios para cumplir con el objetivo planteado en este trabajo. El análisis tiene la finalidad de generar un modelo de los procesos y etapas necesarias para reconocer las expresiones faciales básicas y la definición de los elementos que integran el sistema. En el sistema intervienen los siguientes elementos: imágenes de rostros, librería *facetracking*, interpretación de los datos y algoritmo de clasificación, estos elemento se relacionan entre sí para concluir el sistema de reconocimiento, como se muestra en la Figura 3.1, con la finalidad de desarrollar una arquitectura, el diseño y la implementación del sistema.

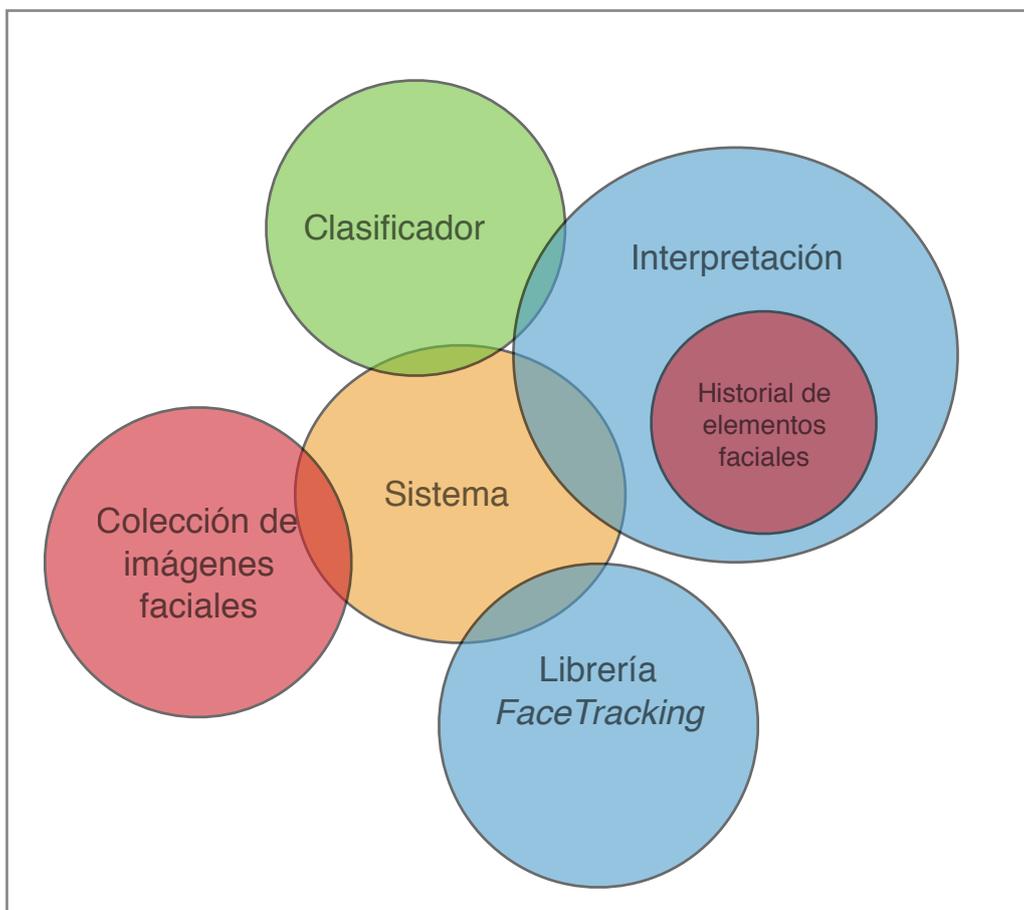


Figura 3.1 Elementos de la solución.

3.2 Esquema del sistema

La Figura 3.2 muestra la descripción general del sistema, en éste se definen sus componentes, los procesos y las relaciones entre éstos. Cada una de las etapas señaladas en el diagrama se encarga de actividades específicas para en conjunto dar el resultado de la interpretación clasificación de la motivación.

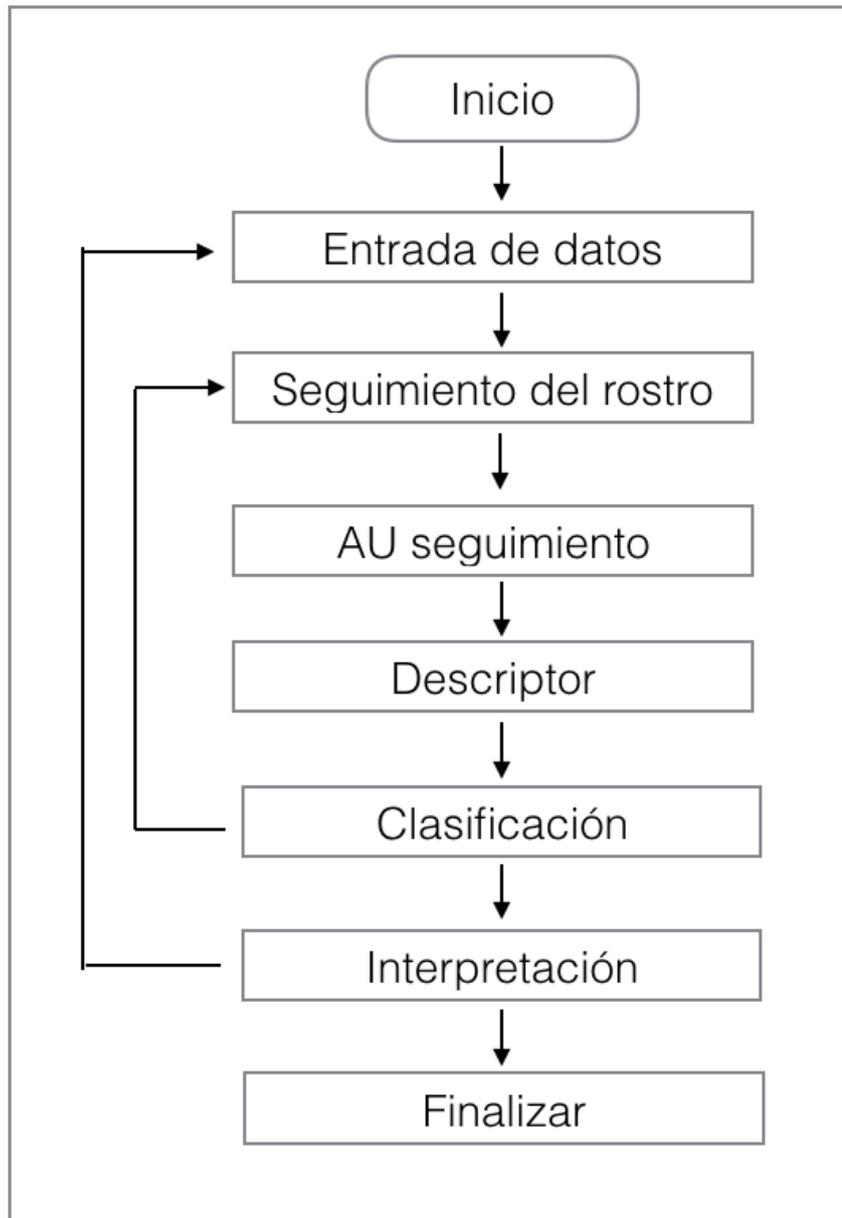


Figura 3.2 Esquema del sistema.

Entrada de datos se encarga de recibir la información del exterior para pasar a la etapa del seguimiento de rostro en la cual se realiza la detección del rostro para darle seguimiento y poder extraer los elementos faciales. AU se encarga de identificar los elementos faciales con los que se está trabajando y extraerlos.

Descriptor es la descripción y se enfoca en trabajar con elementos faciales y cabeza obtenidos, los cuales tienen valores numéricos entre 1 y -1 y dependerá de su posición el valor que se rescate. Interpretación trabaja con los valores obtenidos y hace una interpretación de ellos individualmente para saber en qué estado se encuentran y poder hacer una clasificación de los elementos en conjunto. Clasificar se encarga de trabajar con los elementos individuales de: ojos, boca, cabeza, ceja y con ello dar un veredicto de la motivación basándose en el árbol de combinaciones (Figura 2.12).

Es importante mencionar que la descripción varía entre una persona y otra por que este valor depende de la morfología del rostro de cada persona. En el anexo A se muestran rostros con variaciones, debido a que existen rostro alargados, rostros redondos o más pequeños incluso existe una diferencia entre rostros de mujeres y hombres por lo cual se entrenó el sistema con rostros de diversas personas de ambos sexos.

3.3 Sistema de reconocimiento de expresiones faciales

En esta sección se define el diseño de las cuatro etapas, en conjunto, el proceso, la entrada y salida del sistema de reconocimiento de la motivación.

En la Figura 3.3 se muestra la arquitectura del sistema.

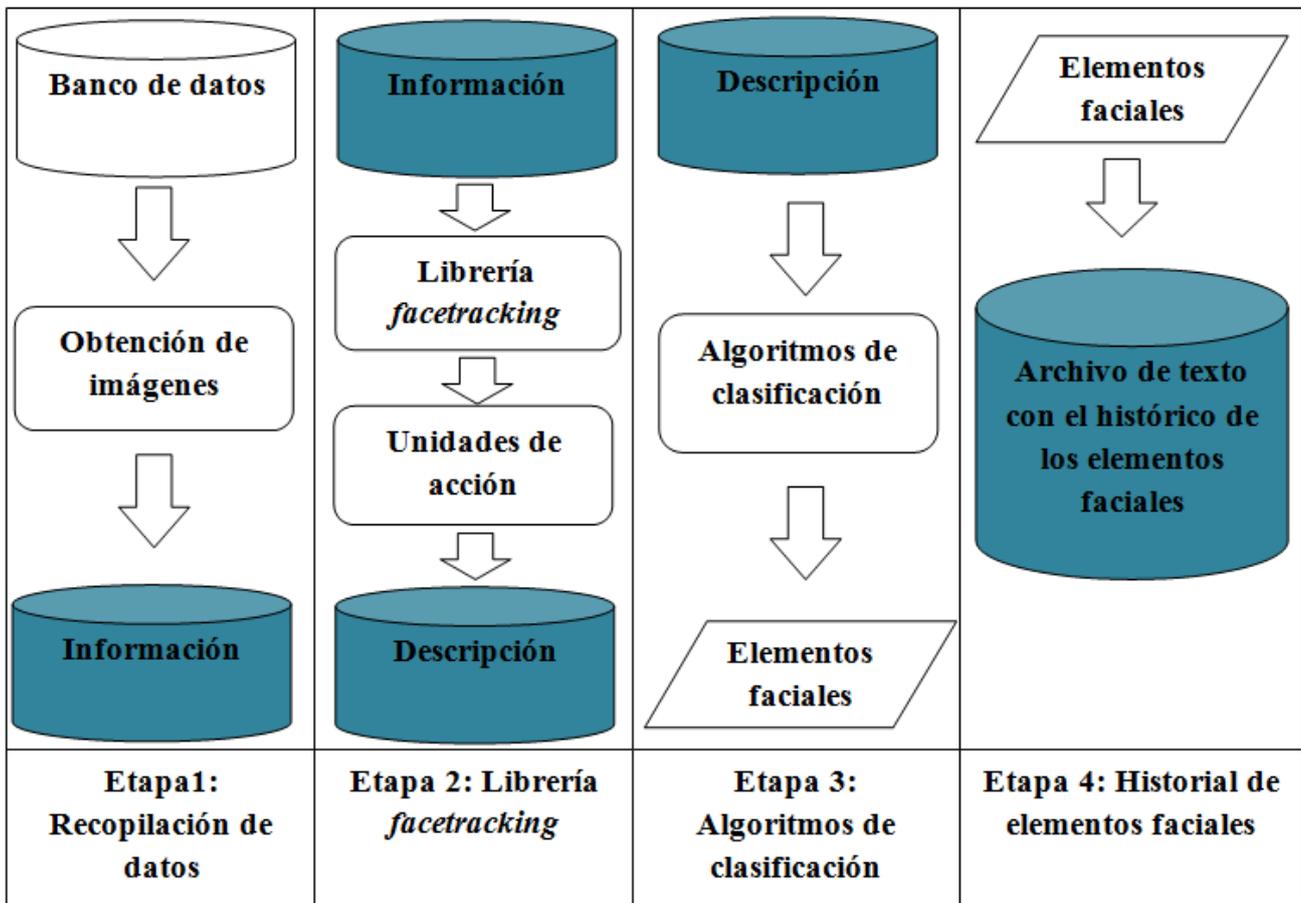


Figura 3.3 Arquitectura del sistema.

En la Figura 3.4 se muestra la estructura del sistema en sus diferentes etapas.

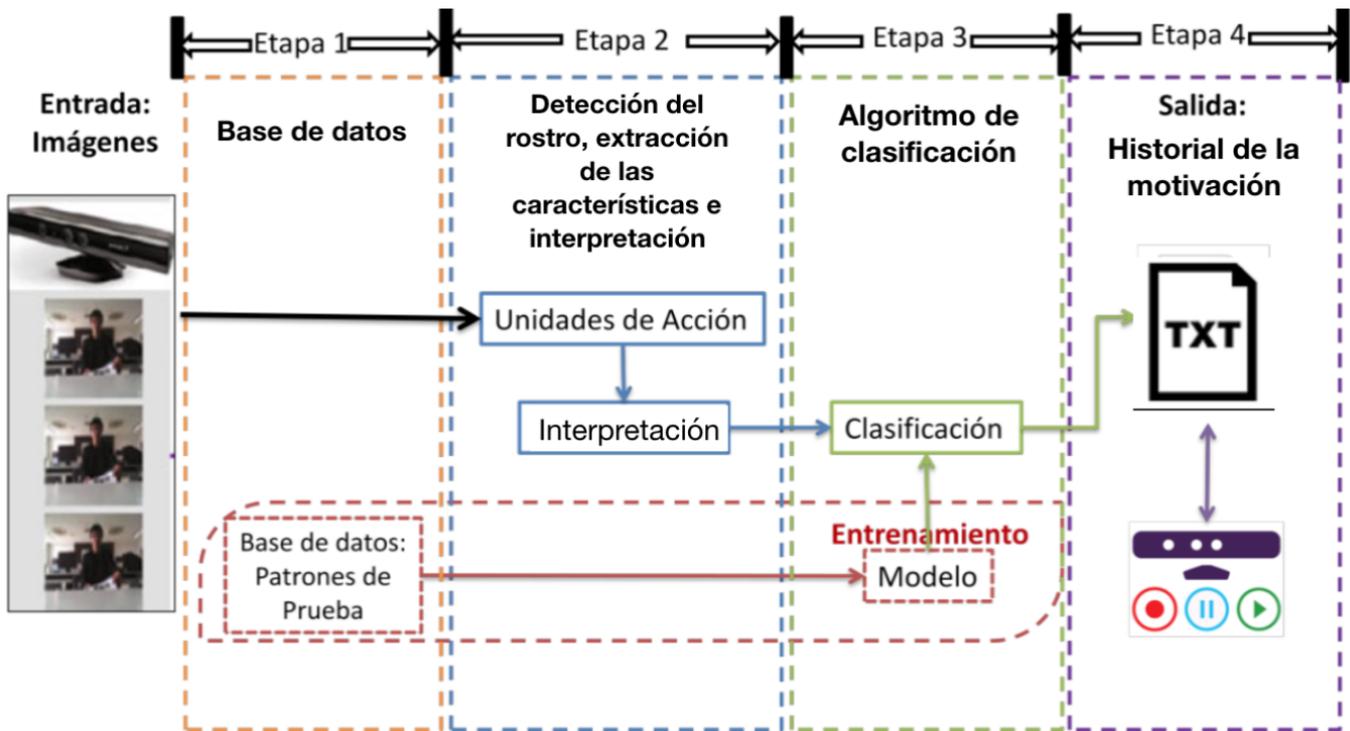


Figura 3.4 Funcionamiento del sistema por etapas.

3.3.1 Etapa 1: Recopilación de los datos

En esta etapa se realizó una recopilación de datos (imágenes del rostro) para entrenar el algoritmo de clasificación. Se realizó el banco de datos con personas de CENIDET, amigos y familiares y cuenta con 25 rostros de distintos sujetos. En la Figura 3.5 se muestran los ficheros de los individuos que contienen la información de los elementos: boca, ojos, ceja, cabeza de cada uno.

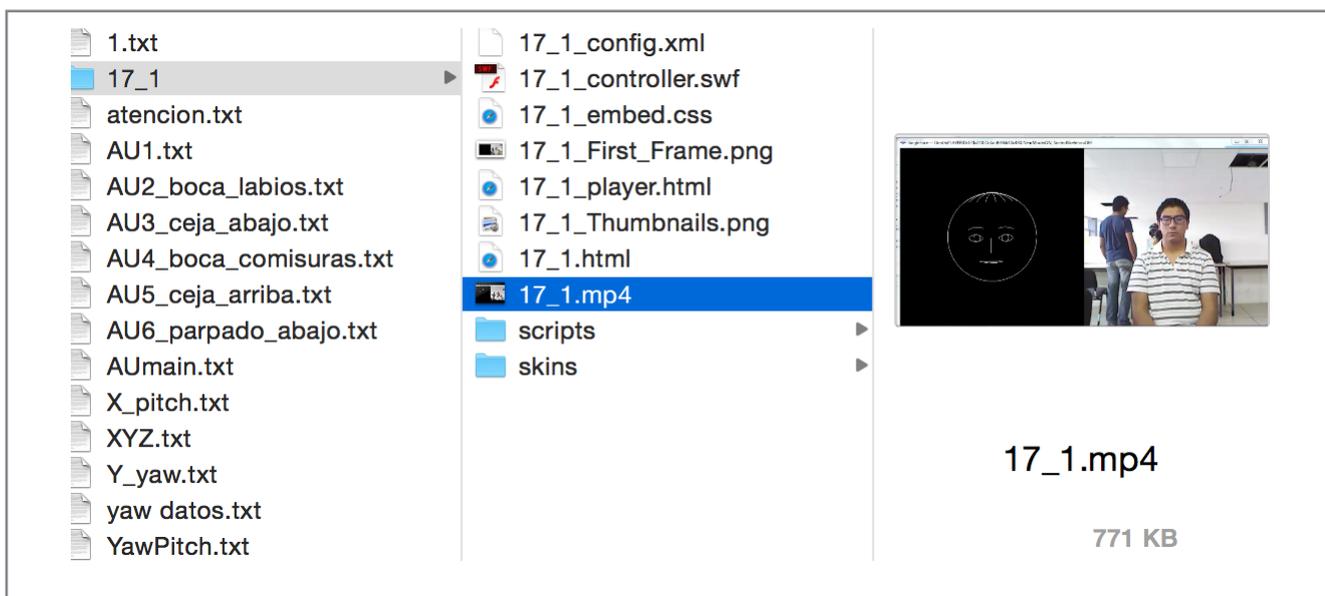


Figura 3.5 Contenido del banco de datos.

Cada uno de estos elementos fue almacenado por separado para poder medir de manera individual cada uno de ellos para posteriormente evaluarlos con base a las reglas definidas en el árbol y la frecuencia de ocurrencia en el bloque de tiempo, y de esta manera, llegar a una conclusión sobre el sujeto y su motivación ante la actividad a realizar.

Es importante mencionar que para capturar las imágenes no se usó un ambiente controlado, las condiciones para obtener las imágenes fueron las siguientes: En interior con una iluminación blanca, iluminación natural. La distancia fue variable para cada sujeto, la persona tuvo que estar de frente a la cámara RGBD, sin retirarle ninguna accesorio como lentes, flequillo, etc.

3.3.2 Etapa 2: Detección elementos faciales e interpretación [29]

En esta etapa se realiza la detección del rostro, extracción de los elementos faciales y la interpretación de los mismos esto, para saber si la cabeza está de frente o girada, ya sea en *pitch* o *yaw*, y los ojos están abiertos o cerrados, la ceja levantada o fruncida, y finalmente la boca cerrada o abierta y poder dar una interpretación de la posición de cada uno de estos elementos.

3.3.2.1 *FaceTracking*

Esta etapa se realiza a partir de la información generada en la etapa uno. Con base en la información se ejecutan las funciones de la librería *FaceTracking*, para la detección de rostro y extracción de características.

La librería *FaceTracking* realiza la detección, seguimiento de rostro y extracción de características de los componentes del rostro, la documentación oficial de Kinect no ofrece información detallada de las técnicas que utiliza para realizar el proceso, la información que ofrece Kinect se limita sólo a las funciones para su empleo por el usuario final.

3.3.2.1.1 Función *IFTFaceTracker*

La función *IFTFaceTracker* devuelve datos de tipo booleano. Se encarga de la interfaz principal para inicializar la detección del rostro en forma sincrónica mediante el paso de *frames* de color y profundidad. Verifica qué tanto los frames de color y profundidad llegan desde el sensor Kinect.

3.3.2.1.2 Función *IFTImage*

La función *IFTImage* devuelve datos de tipo booleano. Esta función se encarga del almacenamiento en memoria del proceso de seguimiento del rostro, esta función carga y libera memoria para el buen funcionamiento del proceso de reconocimiento de rostro. Dentro de esta función se encuentran métodos encargados del acceso a la información acerca de la imagen: el formato, la altura, la

anchura, el tamaño de la imagen y los bytes por pixel. Además, *IFTImage* proporciona métodos de ayuda para la copia de imágenes y dibujo de líneas de depuración son también parte de esta interfaz.

3.3.2.1.3 Función *IFTResult*

La función *IFTResult* es de tipo decimal. Esta función está encargada de verificar que se haya detectado un rostro de forma exitosa y del seguimiento del rostro.

IFTResult proporciona acceso a la siguiente información: *GetFaceRect* en el fotograma de video ofrece información de las coordenadas del cuadro delimitador alrededor del rostro en seguimiento. *Get2DShapePoints* 2D (x, y) coordenadas de los puntos clave en el rostro del fotograma de video. Realiza un seguimiento de los 87 puntos en 2D.

3.3.2.1.4 Función *IFTModel*

La función *IFTModel* es de tipo decimal. Esta función crea la malla *Candide* sobre el rostro. La interfaz *IFTModel* proporciona una manera de convertir los resultados de rastreo a una malla de vértices en 3D en el espacio de la cámara. La interfaz proporciona varios métodos para obtener varias propiedades del modelo: *GetSUCount*, *GetAUCount* devuelven respectivamente el número de unidades de forma (SU) o unidades de acción (AU), utilizados en el modelo lineal 3D (malla *Candide*). *GetTriangles* devuelve triángulos del modelo de la malla 3D (índices de vértices). Cada triángulo tiene 3 índices de vértices que figuran en el sentido horario. *GetVertexCount* devuelve el número de vértices del modelo de la malla 3D.

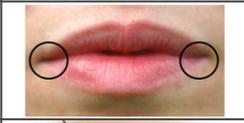
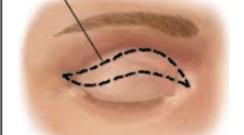
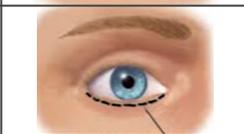
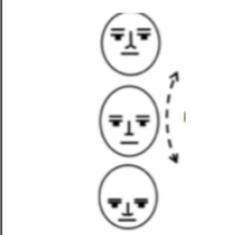
3.3.2.2 Interpretación de las unidades de acción

En este caso se usan las unidades de acción de *facetracking*, las cuales son 4 unidades que comprenden las cejas, la boca y la mandíbula inferior. En la Tabla 3.1 se observa que todas las AU's generan valores desde -1 hasta 1 con incrementos de 0.10. Cada AU en la Tabla 3.1 tiene una descripción que indica el movimiento que realiza cada uno dependiendo si el valor es negativo o positivo.

En la Tabla 3.1 la primera columna indica el nombre de la AU, la columna de valores está dividida en tres secciones: un valor negativo hasta -1 que indica un movimiento, un valor positivo (hasta 1) significa un movimiento contrario al negativo y un valor de cero significa ausencia de movimiento y esto quiere decir que, es un estado neutral, para la cabeza se evaluar los grados de giro, con valores positivos y negativos.

En la tabla 3.1 se muestran los valores que pueden tomar cada uno de los elementos con los que se está trabajando y mediante este valor se dará una interpretación dependiendo del valor numérico de cada uno de estos elementos. Además es necesario tomar en cuenta un historial del valor de los elementos ya que con un valor instantáneo no es correcto dar un veredicto, y de esta manera es posible descartar valores atípicos que pudieran haber sido un error.

Tabla 3.1 Rango de valores para las AU's y elementos.

AU Kinect	Valores negativos	Valor neutro	Valores positivos	Imagen
<i>BrowLower</i>	-1 a 0 Ceja levantada	0	1 a 0 Ceja abajo	
<i>MouthCornerLift</i>	-1 a 0 Boca cerrada	0	1 a 0 Boca abierta	
<i>UpperEyelid</i>	-1 a 0 Ojos cerrados	0	1 a 0 Ojos abiertos	
<i>LowerEyelid</i>	-1 a 0 Ojos cerrados	0	1 a 0 Ojos abiertos	
<i>Pitch</i>	Grados -45 a 0 Cabeza agachada	0	Grados 0 a 45 Cabeza agachada	
<i>Yaw</i>	Grados -45 a 0 Cabeza girada a la izquierda	0	Grados 0 a 45 Cabeza girada a la derecha	

Dado que el valor numérico de los elementos definirá el estado de los mismos, en la Figura 3.6 se muestra el elemento cejas con el rango de valores y la posición que le corresponde.

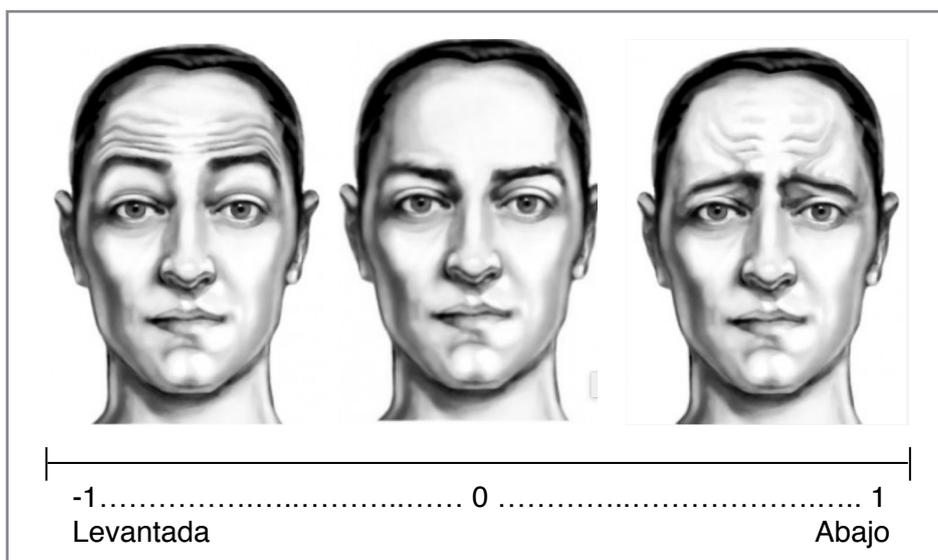


Figura 3.6 Valores que toma el elemento cejas [33].

En la Figura 3.6 se muestra las cejas elevadas y en la parte inferior un valor numérico negativo, esto es por que para el análisis de cejas se utilizó la AU *BrowLower* por tanto al tener una elevación se consigue un valor negativo. En el centro de la imagen se tiene un rostro con cejas en posición neutral y en la parte inferior valor 0 esto por que las cejas están en reposo. Finalmente en la parte derecha se tienen las cejas fruncidas hacia abajo y por ello la AU *BrowLower* da valor positivo. Los valores se tornarán decimales en los casos intermedios; esto dependerá de la intensidad o exageración de este gesto, una ceja ligeramente levantada podría tomar un valor de -0.3 por ejemplo.

En la Figura 3.7 se muestra el elemento ojos con el rango de valores y la posición que le corresponde según la apertura de ellos, es importante mencionar que se utilizan dos AU para la evaluación de los ojos en este trabajo: *UpperEyelid* y *LowerEyelid*, de esta manera se evalúa el ojo de la parte inferior y superior.

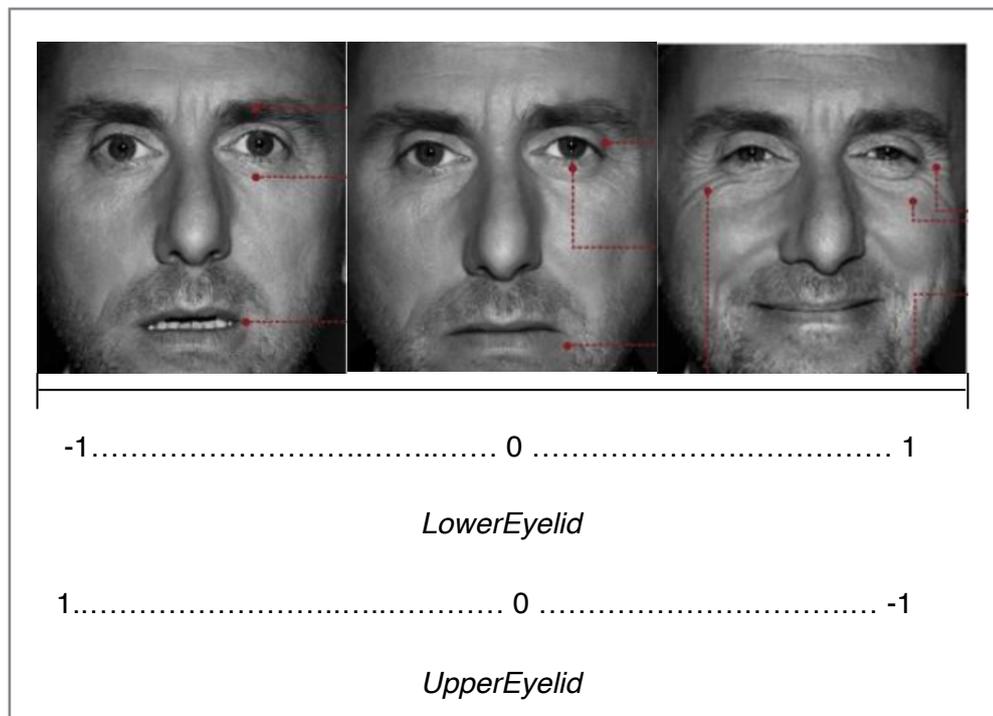


Figura 3.7 Valores que toma el elemento ojos [33].

En la Figura 3.7 se muestran dos líneas de valores, en la primera a la izquierda se tiene valores negativos, debido a que corresponde a la AU *LowerEyelid* la cual evalúa la parte de abajo del ojo (párpado inferior) por tanto da un valor negativo cuando el ojo está muy abierto y la cuenca ocular es estirada hacia abajo. En la segunda parte a la izquierda se tiene valores positivos, debido a que corresponde a la AU *UpperEyelid* la cual evalúa la parte de arriba del ojo (párpado superior) por tanto da un valor positivo cuando el ojo está muy abierto y la cuenca ocular es estirada hacia arriba.

En la Figura 3.8 se muestra el elemento boca el cual es evaluado acorde a su valor numérico.

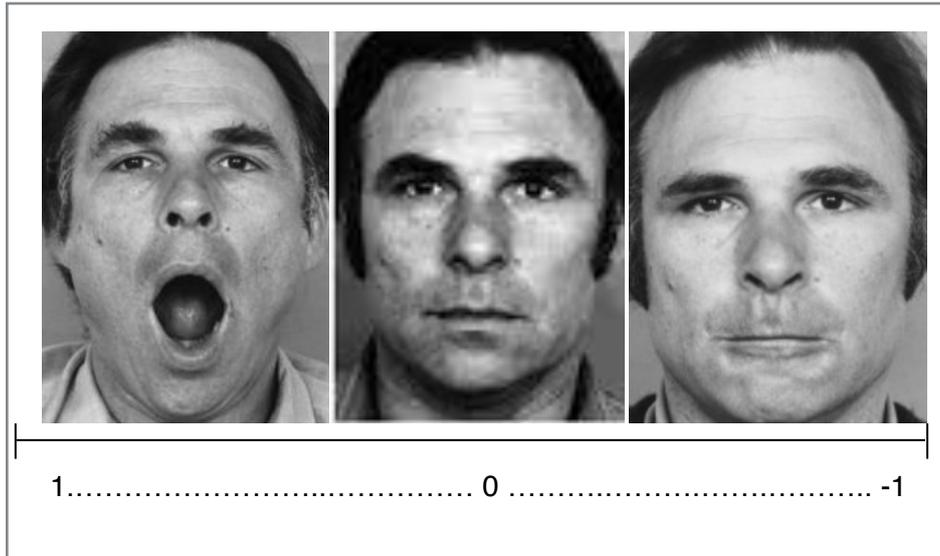


Figura 3.8 Valores que toma el elemento boca[33] .

En la Figura 3.9 se muestra el elemento cabeza con moviendo en *pitch*, en el costado derecho de la imagen se muestra una barra con valor numérico en grados, en la parte superior se tiene 45° y corresponde a la cabeza girada hacia arriba, la imagen de en medio corresponde a la cabeza en posición de frente 0°, en la parte inferior se tiene la cabeza girada hacia abajo con valor numérico negativo de -45°. El grado máximo de giro definido es de -45° a 45° ya que fuera de ese rango es posible perder la detección del rostro.

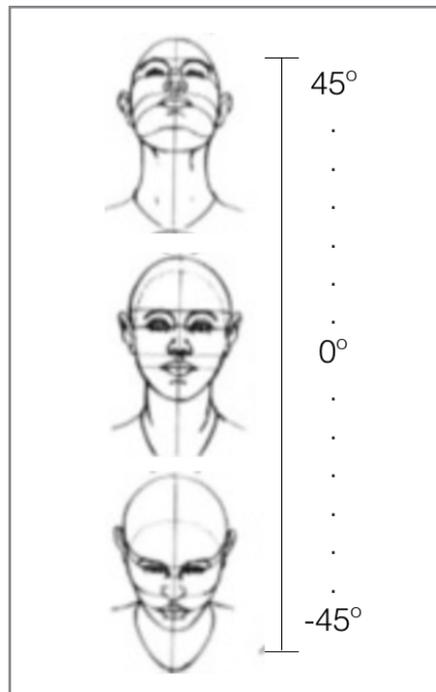


Figura 3.9 Valores que toma cabeza en movimiento *pitch* [42].

En la Figura 3.10 se muestra el elemento cabeza con moviendo en *yaw*, en la parte inferior de la imagen se muestra una barra con valor numérico en grados, del lado izquierdo se tiene 45° y corresponde a la cabeza girada hacia la derecha, la imagen de en medio corresponde a la cabeza en posición de frente y con valor numérico de 0, del lado derecho se tiene la cabeza girada a la izquierda con valor numérico negativo. En los extremos fue definido que el grado máximo de giro fuera de -45° a 45° ya que fuera de ese rango es posible perder la detección del rostro.

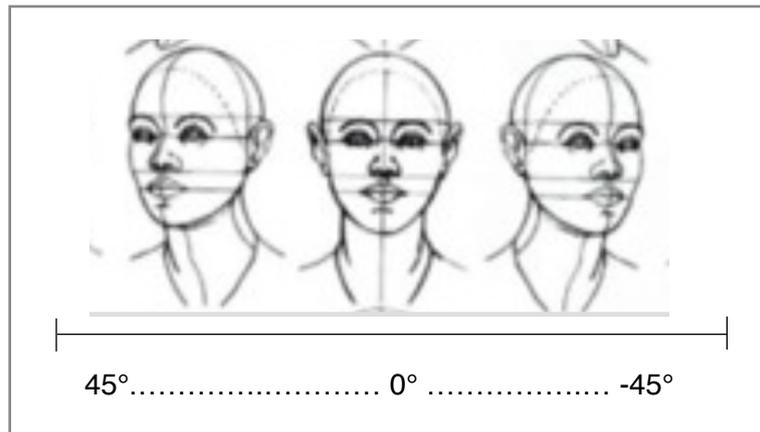


Figura 3.10 Valores que toma cabeza en movimiento *yaw* [34].

3.3.2.2.1 Histórico

Mediante el histórico se evalúa el comportamiento de los elementos faciales, es decir si la boca está abierta por un lapso de tiempo o los ojos cerrados y así para cada uno de los elementos con los que se está trabajando.

Se tomaron bloques de dos segundos aproximadamente con 31 datos cada bloque, (un bloque por cada AU). Dichas AU's tiene valores de -1 a 1 y, dependiendo del elemento, se le asigna un sí o un no; y para la cabeza, el ángulo. Finalmente se calcula el porcentaje de atención prestada mediante las condiciones del árbol.

El cálculo del historico se divide en 6 etapas:

Etapas 1.- Captura de instancias. Se capturan 31 instancias donde se almacenan los datos, los cuales corresponden a cada uno de los elementos con los que se está trabajando y en cada registro está contenido el valor numérico de cada elemento. El registro de éstos 31 datos en el bloque toma un tiempo aproximado de dos segundos (esto varía dependiendo de la carga que tenga el procesador).

Etapa 2.- Los datos entran a una función discriminante, la cual se encarga de asignar a una clase, cada uno de los registros del bloque de 31 datos. Las clases son: interesado, desinteresado y neutral. La clase predominante de los registros se asigna a todo el bloque, es decir si de los 31 registros 25 fueron asignados a la clase interesado, el bloque entero pertenecerá a la clase interesado. La evaluación del bloque será enviada a un árbol de decisión.

Etapa 3.- El árbol de decisión recibe la clase predominante de cada bloque, entonces son evaluados acorde a las reglas establecidas en dicho árbol.

Etapa 4.- Se decide si el bloque pertenecerá a la clase, interesado, desinteresado o neutral.

Etapa 5.- Los bloques van siendo recibidos para crear el histórico, se trabaja con todos los bloques generados durante la prueba. Una vez que el video termina, se calcula con todos los bloques un porcentaje de cómo avanza la atención en el transcurso del tiempo de la prueba.

Etapa 6.- Finalmente con todas las evaluaciones, se calcula un porcentaje de atención de la siguiente forma. (ver ecuación 1).

$$x = \frac{b}{a} * 100 \quad (1)$$

Donde a es la cantidad total de evaluaciones realizadas por el sistema durante el video y b es la cantidad de veces que el sistema dio como resultado interesado. Donde x es el porcentaje de atención prestada en la actividad durante el total de tiempo del video.

El tiempo es un factor importante dentro del historial, no solamente se trata de la frecuencia de ocurrencia sino de un proceso en el cual una acción o pose de los elementos se repite por un periodo prolongado y en diversos bloques, ya que si constantemente el sujeto de prueba esta con la boca abierta, o la cabeza girada se entenderá que no se está prestando atención porque en un tiempo determinado se repitió esta acción.

En la Figura 3.11 se muestra el proceso con el cual se lleva a cabo el histórico y con base en él se llega a una conclusión de motivado, desmotivado o neutral.

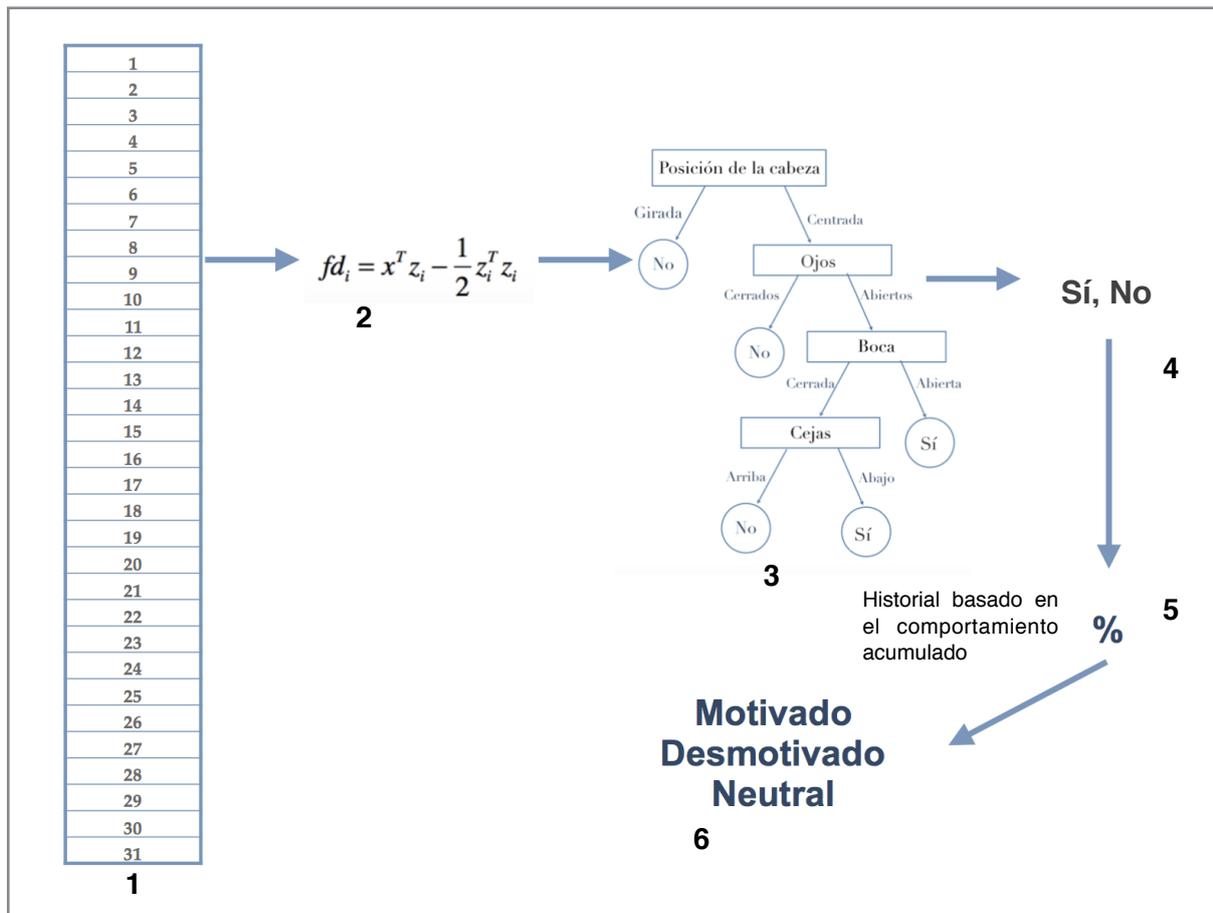


Figura 3.11 Cálculo del histórico del comportamiento acumulado.

3.3.3 Etapa 3: Clasificación Función discriminante por distancia

En esta tesis se utilizó un clasificador de tipo supervisado. El objetivo de un análisis discriminante es encontrar la combinación de las variables independientes que mejor permitan diferenciar o discriminar a los grupos, una vez encontrada esa combinación (la función discriminante) podría ser utilizada para clasificar nuevos casos, ya que se trata de una técnica de análisis multivariable que es capaz de aprovechar las relaciones existentes entre una gran cantidad de variables, para maximizar la capacidad de discriminación.

En esta tesis se trabajó con función discriminante por distancia [30]. En la Ecuación 2 se calcula la media de clase utilizando todos los datos del conjunto de entrenamiento, esto para conseguir una referencia de la media de todos los datos perteneciente a cada una de las clases.

$$Z_i = \frac{1}{p} \sum_{j=1}^p x_{ij} \quad (2)$$

En la Ecuación 3 se calcula el dato a evaluar menos las medias de clase, esto genera un resultado para cada clase que se tenga.

$$d(x, z_i) = \sqrt{(x - z_i)^T (x - z_i)} \quad (3)$$

A partir de la Ecuación 4 hasta la Ecuación 8 se pasa de la ecuación de la distancia euclidiana a una forma donde se obtienen factores, que pretenden maximizar el factor discriminante entre los datos para una clasificación más precisa.

$$(d(x, z_i))^2 = \left(\sqrt{(x - z_i)^T (x - z_i)} \right)^2 \quad (4)$$

$$(d(x, z_i))^2 = (x - z_i)^T (x - z_i) = \|x - z_i\|^2 \quad (5)$$

$$(d(x, z_i))^2 = x^T x - 2x^T z_i + z_i^T z_i \quad (6)$$

$$d(x, z_i) = x^T z_i - \frac{1}{2} z_i^T z_i \quad (7)$$

$$fd_i = x^T \cdot z_i - \frac{1}{2} z_i^T \cdot z_i \quad (8)$$

Finalmente con la Ecuación 9 se determina la pertenencia de clase del nuevo dato. El valor más alto dado mediante la función discriminante indicará la pertenencia.

$$x \in a_i \text{ sii } fd_i(x) > fd_j(x) \quad \forall i, j, i \neq j \quad (9)$$

Donde:

Z media de clase

p cantidad de datos por clase

x dato a evaluar

3.3.4 Etapa 4: Salida de resultado

En la etapa 4 se obtiene una salida numérica, estos valores son utilizados para interpretar cada uno de los elementos faciales trabajados.

En la Figura 3.12 se muestra los valores de salida del resultado del individuo uno dicho archivo está constituido con cinco columnas correspondientes a los elementos utilizados en este trabajo. La columna uno pertenece a la boca. La columna dos pertenece a las cejas. La columna tres pertenece a los ojos. Las columnas cuatro y cinco pertenecen a la cabeza, en dichas columnas se evalúa si la cabeza está girada en el eje de las x y de las y .

En la etapa 4 también se obtiene un archivo .txt en cual contiene el resultado de la atención y el elemento que lo detonó, así como el porcentaje de atención.

Boca	Cejas	Ojos	Pitch	Yaw
-0.165868,	-0.763584,	0.301822,	7.65376,	0.853391
-0.139053,	-0.691069,	0.316516,	7.65376,	0.853391
-0.139053,	-0.691069,	0.316516,	8.09205,	1.02296
-0.139053,	-0.691069,	0.316516,	8.09205,	1.02296
-0.128762,	-0.619216,	0.324486,	8.09205,	1.02296
-0.128762,	-0.619216,	0.324486,	8.37194,	0.818902
-0.119333,	-0.574434,	0.320037,	8.37194,	0.818902
-0.125838,	-0.556226,	0.358182,	8.37194,	0.818902
-0.121804,	-0.474208,	0.324653,	8.77252,	0.977755
-0.120601,	-0.415957,	0.322494,	9.01695,	0.584132
-0.114646,	-0.446667,	0.362835,	9.01695,	0.584132
-0.114646,	-0.446667,	0.362835,	9.01695,	0.584132
-0.106673,	-0.348365,	0.316137,	9.01695,	0.584132
-0.107203,	-0.397813,	0.271117,	8.01447,	-0.66638
-0.11742,	-0.41282,	0.225502,	7.87379,	-1.05254
-0.128499,	-0.477193,	0.189751,	7.87379,	-1.05254
-0.128499,	-0.526257,	0.161068,	8.73107,	-1.28019
-0.128499,	-0.526257,	0.161068,	9.69076,	-0.9682
-0.12996,	-0.562705,	0.146882,	9.69076,	-0.9682
-0.12996,	-0.562705,	0.146882,	9.69076,	-0.9682
-0.112841,	-0.557705,	0.140585,	10.0261,	0.0174773
-0.112841,	-0.557705,	0.140585,	9.42762,	-0.197841
-0.112841,	-0.557705,	0.140585,	9.42762,	-0.197841
-0.0171124,	-0.627539,	0.176204,	8.9345,	-0.81641
-0.0171124,	-0.627539,	0.176204,	8.9345,	-0.81641
-0.0171124,	-0.627539,	0.176204,	9.59627,	-1.06552
-0.0174506,	-0.576267,	0.138227,	7.60607,	-0.21748
-0.0096067,	-0.542692,	0.201126,	8.29538,	-0.59167
-0.0219465,	-0.544191,	0.263231,	8.29538,	-0.59167
-0.0367487,	-0.541917,	0.294882,	6.10186,	-0.890724
-0.055846,	-0.546551,	0.291081,	6.10186,	-0.890724
-0.0755789,	-0.544604,	0.331103,	7.57621,	-1.65437
-0.0755789,	-0.544604,	0.331103,	7.83439,	-1.04553
-0.0755789,	-0.544604,	0.331103,	7.83439,	-1.04553
-0.0755789,	-0.544604,	0.331103,	7.36884,	-0.678774
-0.0755789,	-0.544604,	0.331103,	7.36884,	-0.678774
-0.0897061,	-0.508313,	0.388127,	7.36884,	-0.678774
-0.0982786,	-0.547226,	0.40049,	6.76167,	-1.2102
-0.0982786,	-0.547226,	0.40049,	7.14396,	-1.59249
-0.0982786,	-0.547226,	0.40049,	7.14396,	-1.59249
-0.0982786,	-0.547226,	0.40049,	7.77758,	-1.87191
-0.0982786,	-0.547226,	0.40049,	7.14396,	-1.59249
-0.0982786,	-0.547226,	0.40049,	7.77758,	-1.87191
-0.0982786,	-0.547226,	0.40049,	7.61506,	-1.40056
-0.0982786,	-0.547226,	0.40049,	7.61506,	-1.40056
-0.104391,	-0.530863,	0.396479,	7.73109,	-1.89686
-0.104391,	-0.530863,	0.396479,	8.56926,	-1.68377
-0.104391,	-0.530863,	0.396479,	8.56926,	-1.68377
-0.104391,	-0.530863,	0.396479,	7.96086,	-1.1947

Figura 3.12 Estructura del archivo .txt.

En la Tabla 3.2 se muestra la columna de atención y elemento detonante se muestra si se prestó atención o no y qué elemento detonó esta clasificación. En las colinas porcentaje de atención se tiene el porcentaje acumulado hasta llegar al último registro donde se muestra la conclusión y en qué porcentaje se prestó atención a la actividad.

Tabla 3.2 Atención y porcentaje.

Atención y elemento detonate	Porcentaje de atención	Atención y elemento detonate	Porcentaje de atención
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:92%, No:8%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:92.3077%, No:7.69231%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:92.5926%, No:7.40741%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:92.8571%, No:7.14286%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:93.1034%, No:6.89655%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:94.4444%, No:5.55556%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:94.5946%, No:5.40541%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:94.7368%, No:5.26316%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:94.8718%, No:5.12821%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:95%, No:5%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:95.122%, No:4.87805%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:95.2381%, No:4.7619%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:95.3488%, No:4.65116%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:95.4545%, No:4.54545%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:95.5556%, No:4.44444%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:95.6522%, No:4.34783%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: No -> ceja -1	Si:93.617%, No:6.38298%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:93.75%, No:6.25%
Atencion: Si -> boca 1	Si:100%, No:0%	Atencion: Si -> boca 1	Si:93.8776%, No:6.12245%
Atencion: No -> ceja -1	Si:95%, No:5%	Atencion: Si -> boca 1	Si:94%, No:6%
Atencion: No -> ceja -1	Si:90.4762%, No:9.52381%	Atencion: Si -> boca 1	Si:94.1176%, No:5.88235%
Atencion: Si -> boca 1	Si:90.9091%, No:9.09091%	Atencion: Si -> boca 1	Si:94.2308%, No:5.76923%
Atencion: Si -> boca 1	Si:91.3043%, No:8.69565%	Atencion: Si -> boca 1	Si:94.3396%, No:5.66038%
Atencion: Si -> boca 1	Si:91.6667%, No:8.33333%	Atencion: Si -> boca 1	Si:94.4444%, No:5.55556%

3.4 Discusión

El análisis y diseño del sistema determinaron los elementos que intervienen para la solución del problema y las etapas necesarias para la implementación del sistema, éstas son: la recopilación de banco de datos (el cual fue creado para este trabajo), la implementación de la librería *facetracking* para detección del rostro y extracción de las características, la interpretación de las características extraídas y finalmente la implementación de un algoritmo para la clasificación de la motivación de los sujetos en motivado, desmotivado o neutral. Es importante recordar que no se puede dar un veredicto por el comportamiento de un instante por lo cual se creó un histórico del comportamiento y con base en ello clasificar.

CAPÍTULO 4 Experimentación y resultados

En este capítulo se presentan las pruebas realizadas al experimentar con la librería *facetracking*, las pruebas realizadas y los resultados obtenidos con el sistema implementado de visión para reconocimiento automática de la motivación. Estas pruebas tienen por objetivo observar lo obtenido en diversas etapas debido a que se realizaron pruebas individuales a los elementos y posteriormente en conjunto, también se evalúa el cumplimiento de los objetivos y alcances definidos para este trabajo.

4.1 Ambiente de pruebas

Para realizar cada una de las pruebas se requirió de elementos de hardware, software y un banco de datos de imágenes, las características de cada uno se describen a continuación.

Hardware-Laptop

Procesador: Intel(R) Core(TM) i3 CPU 2.60 GHz

Memoria RAM: 8.00 GB.

Sistema Operativo de 64 bits: Windows 7.

Kinect xbox 360, Kinect Windows

Software

SDK de Kinect: Esta herramienta contiene librerías para el desarrollo de aplicaciones con Kinect. Especialmente se utilizó la librería Facetracking para detección de rostros y extracción de características

Kinect Studio: Herramienta que permite guardar videos con información RGB y profundidad.

Visual Studio: Visual Studio express 2012, en la versión 12.0.

Lenguaje C++.

Banco de datos

Se creó en CENIDET para esta tesis. La base de datos creada se capturó a partir de la forma natural en que los participantes generan las expresiones faciales. Los tomas se capturaron en el interior de un edificio con iluminación artificial. La cara permanece en posición frontal respecto al dispositivo de captura, la distancia de captura es variable y las personas que participaron son de nacionalidad mexicana y china. En estos videos el fondo del escenario es variable.

Con estas bases de datos se tiene material diverso para probar la aplicación desarrollada, por ejemplo: sexo masculino y femenino, distintos tonos de piel, rasgos faciales y tamaño de la cara.

4.2 Plan de pruebas

El objetivo de las pruebas es validar el correcto funcionamiento de los módulos del sistema, así como comprobar que se alcanzaron los objetivos planteados. Se realizaron las pruebas con el banco de datos a 20 personas; donde se evalúa individualmente cada elemento para corroborar el correcto funcionamiento de la detección y los valores numéricos capturados y la interpretación y clasificación el conjunto.

4.2.1 Prueba de cambios de iluminación

Se realizaron pruebas con cambios en la iluminación para corroborar el desempeño del Kinect con diferente iluminación. En la Figura 4.1 se muestra una captura con iluminación natural y se puede apreciar que la malla de detección de rostro de Kinect está correctamente posicionada sobre el rostro.

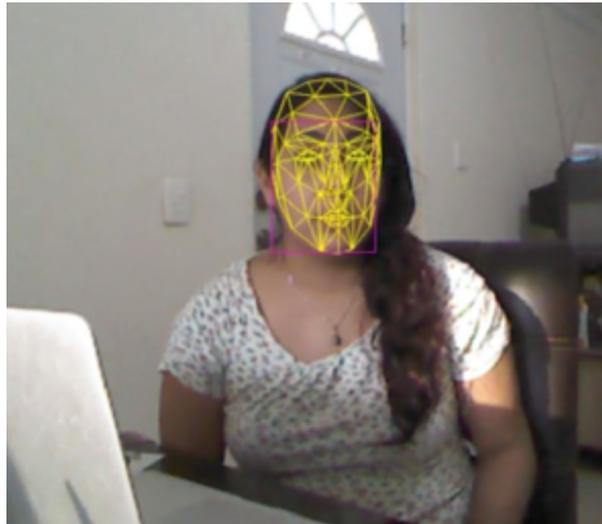


Figura 4.1 Captura con iluminación natural.

En la Figura 4.2 se muestra una captura con iluminación artificial y se puede apreciar que la malla de detección de rostro de Kinect está correctamente posesionada sobre el rostro, sin embargo al realizar repetitivamente esta prueba y con diferente intensidad de luz artificial se logró diferenciar que usando la luz artificial se detectan con mayor precisión los elementos trabajados.

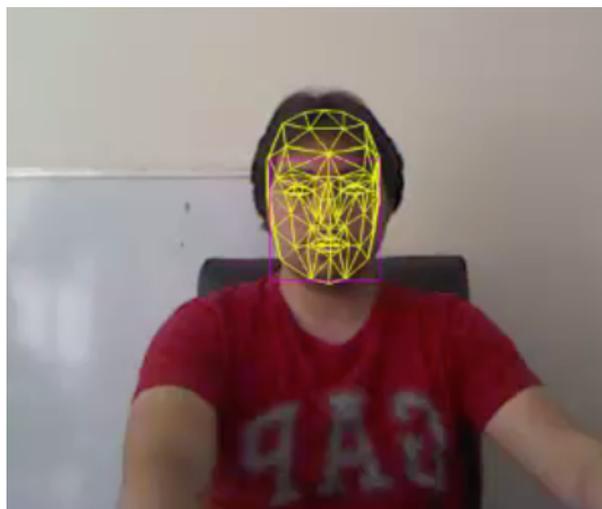


Figura 4.2 Captura con iluminación artificial.

4.2.2 Prueba a distintas distancia con respecto al sujeto de estudio

Se realizó la prueba de la distancia para verificar los datos colocando al sujeto de prueba a diversas distancias, esto con el fin de comprobar que los datos no cambian con la distancia, además que el sistema no siempre podrá ser utilizado en un ambiente controlado. En la Figura 4.3 se muestra el sujeto de prueba a diferente distancia del sensor RGBD; en el lado izquierdo el sujeto se encuentra a 120 cm. y del lado derecho a 70 cm.

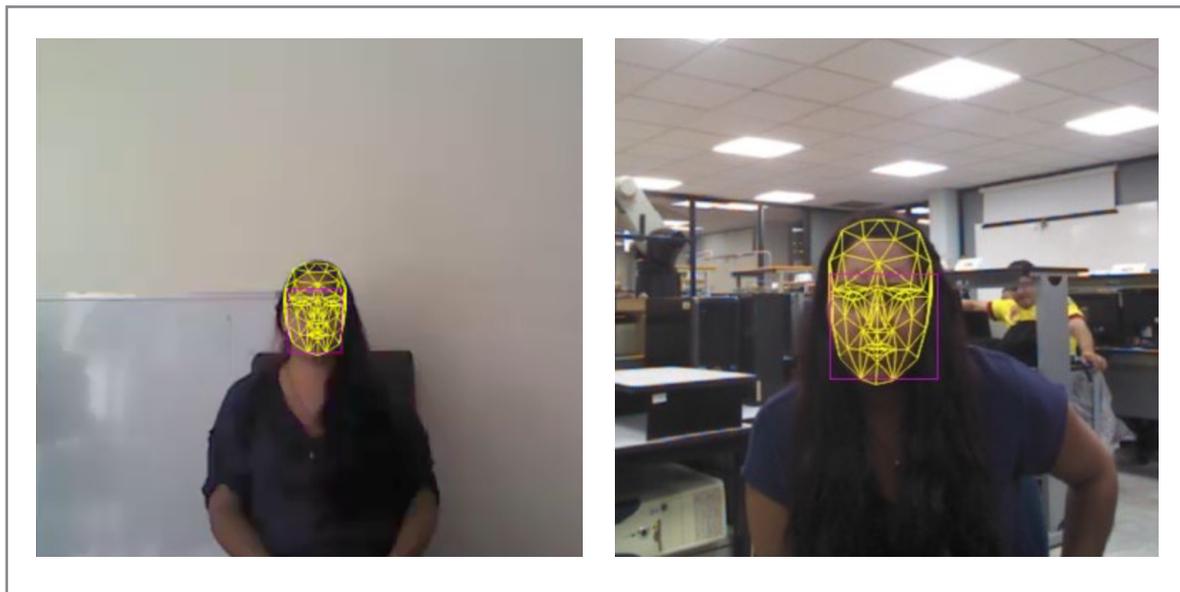


Figura 4.3 Captura en diferente distancia.

Como se puede apreciar en la Figura 4.3 la distancia no influye en la detección del rostro, sin embargo debe considerarse que la distancia máxima recomendable es de 250 cm para evitar la pérdida de información de los elementos trabajados.

4.3.3 Prueba de identificación de los elementos individualmente

Para medir los cambios en el rostro se realizó la prueba donde se le pide al individuo que mueva las cejas, la boca, los ojos y la cabeza para corroborar que la detección de los movimientos correspondan a lo establecido en la documentación de las AU's del Toolkit de Kinect.

En la Figura 4.4 se muestra al sujeto de prueba con las cejas hacia abajo, para corroborar que el valor obtenido se encuentra en el rango de 0 a 1 debido a lo establecido en la Figura 3.5, donde se especifica que la ceja fruncida al utilizar la AU *LowerEyeLid* obtendrá valores positivos.

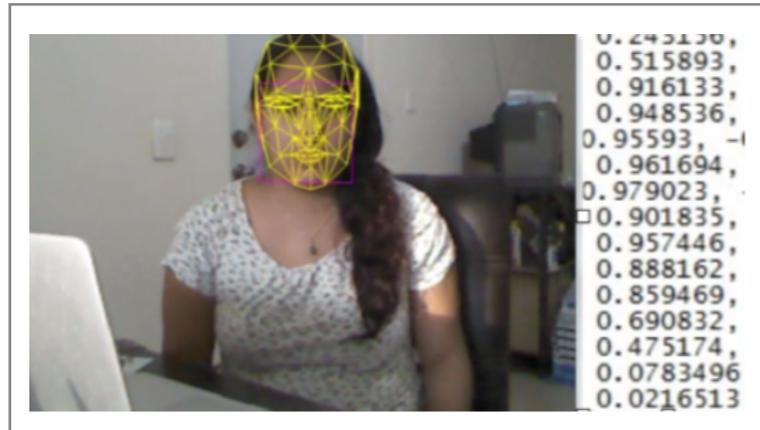


Figura 4.4 Elemento cejas hacia abajo.

En la Figura 4.5 se muestra al sujeto de prueba con la boca abierta, para corroborar que el valor obtenido se encuentra en el rango 0 a 1 debido a la establecido en la Figura 3.7 de la sección anterior, donde se especifica que la boca abierta al utilizar la AU *MouthCornerLift* obtendrá valores positivos.

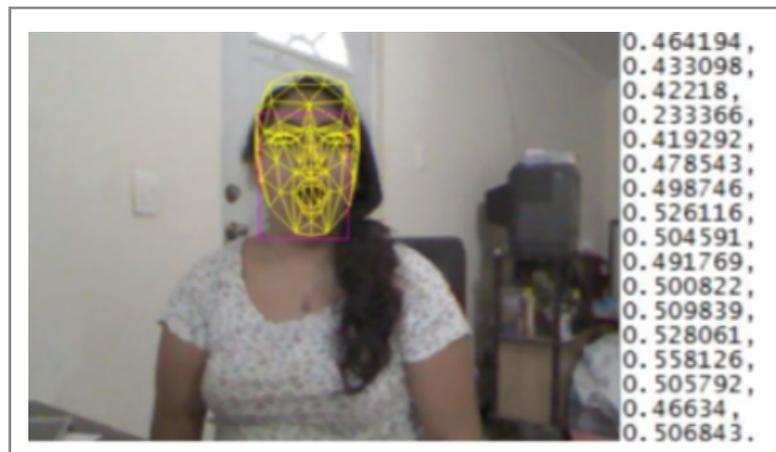


Figura 4.5 Elemento boca abierta.

En la Figura 4.6 se muestra al sujeto de prueba con los ojos abiertos, para corroborar que el valor obtenido se encuentra en el rango establecido en la Figura 3.6, donde se especifican los valores que pertenecen a los ojos abiertos al utilizar las AU's *LowerEyeLid* y *UpperEyeLid*.

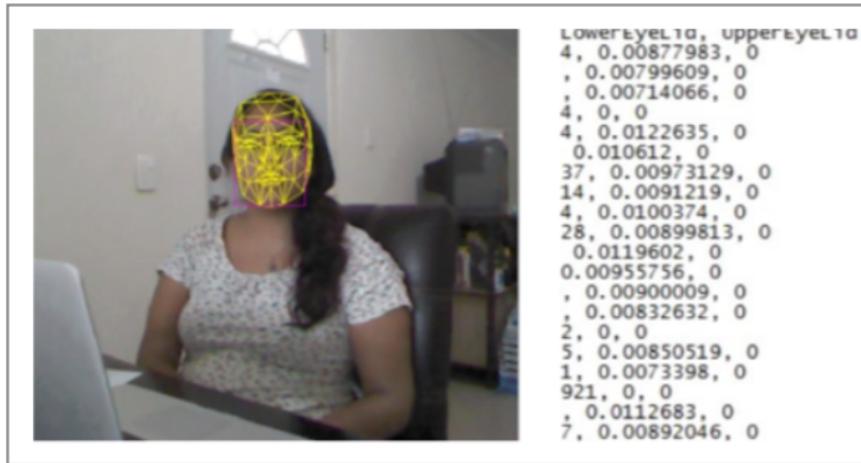


Figura 4.6 Elemento ojos abiertos.

En la Figura 4.7 se muestra al sujeto de prueba con la cabeza girada, para corroborar que el valor obtenido se encuentra entre el rango establecido para el movimiento de *yaw* en la Figura 3.9, donde se especifica los valores que pertenecen a la cabeza con este movimiento.

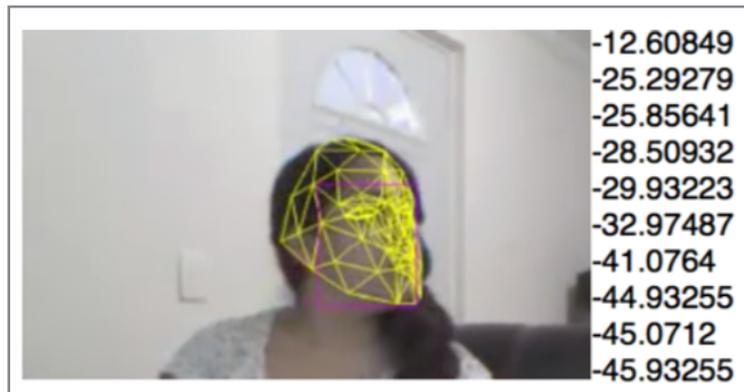


Figura 4.7 Elemento cabeza girada en *yaw*.

4.3.4 Pruebas de las combinaciones de los elementos faciales

Una vez verificado que los valores de los elementos eran capturados correctamente, se le pidió a los sujetos de prueba que realizaran movimientos específicos para verificar la correcta identificación al ser combinados. En la Figura 4.8 se muestra al sujeto de prueba para la rama uno.

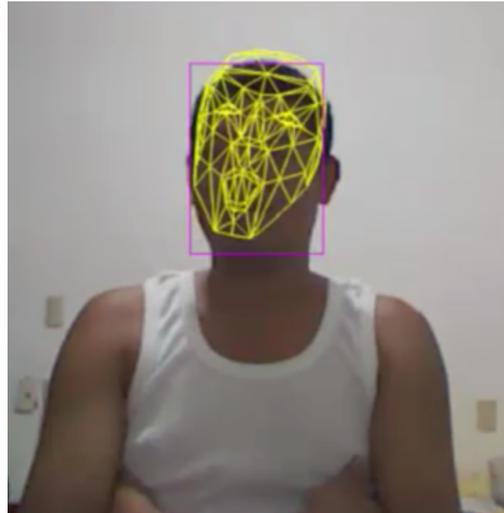


Figura 4.8 Sujeto de prueba para la rama 1 del árbol.

Figura 4.9. prueba de la combinación de elementos para la rama 1 del árbol “posición de la cabeza → girada → no”.

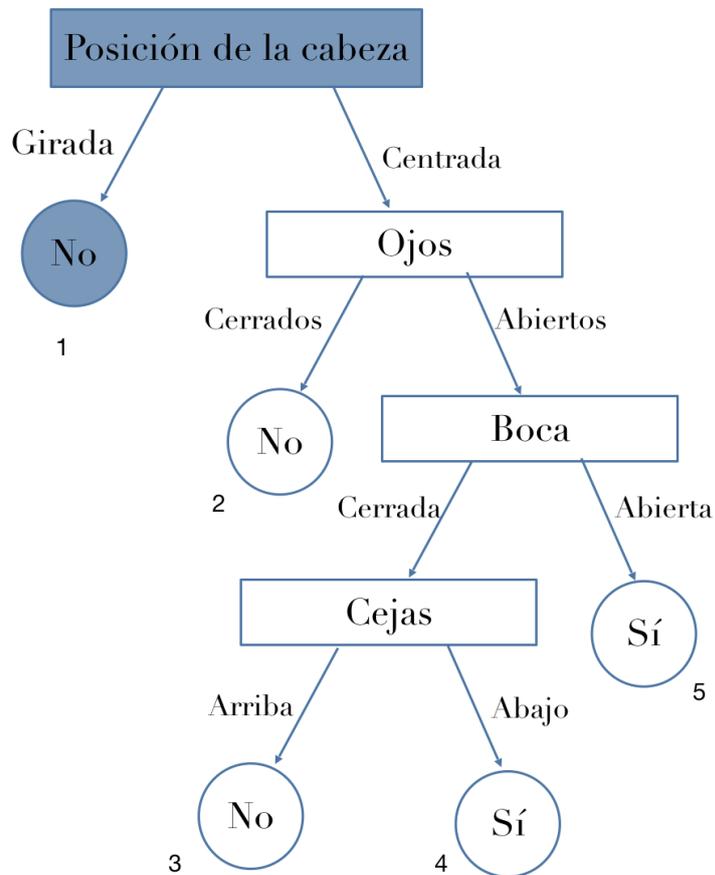


Figura 4.9 Combinaciones para rama uno.

En la Figura 4.10 se muestra al sujeto de prueba para la rama dos.

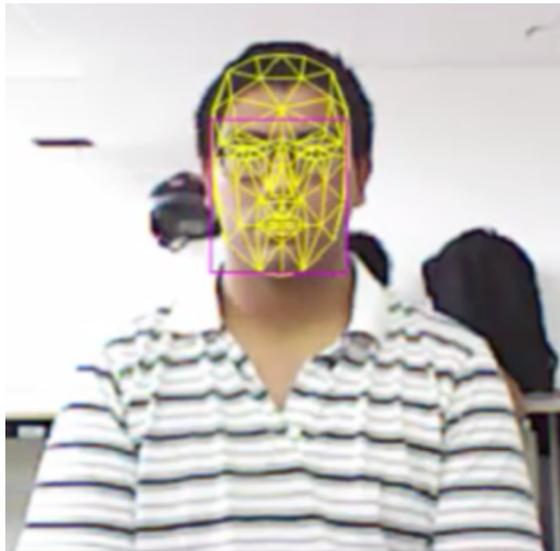


Figura 4.10 Combinaciones para rama uno.

Figura 4.11 prueba de la combinación de elementos para la rama 2 “posición de la cabeza → centrada → ojos → cerrados → no

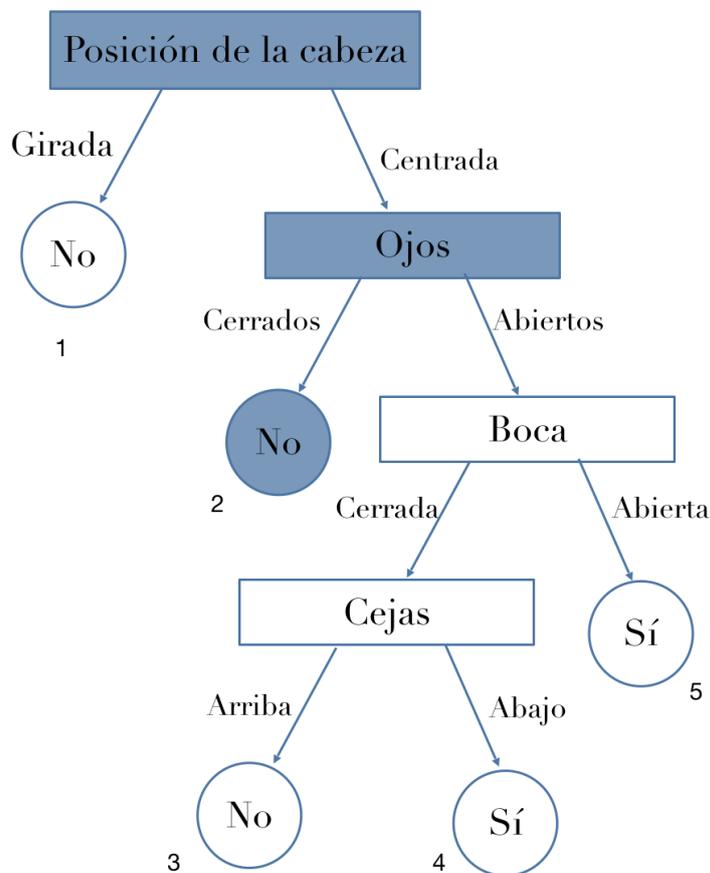


Figura 4.11 Combinaciones para rama dos.

En la Figura 4.12 se muestra al sujeto de prueba para la rama tres.

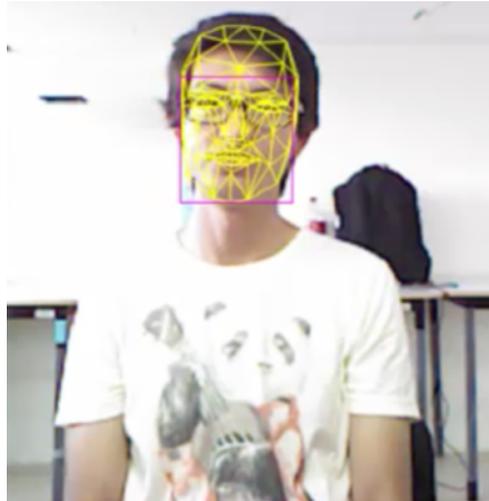


Figura 4.12 Combinaciones para la rama tres.

Figura 4.13 se muestra la prueba de la combinación de elementos para la rama 3 “posición de la cabeza → centrada → ojos → abiertos → boca → cerrada → cejas → arriba → no”

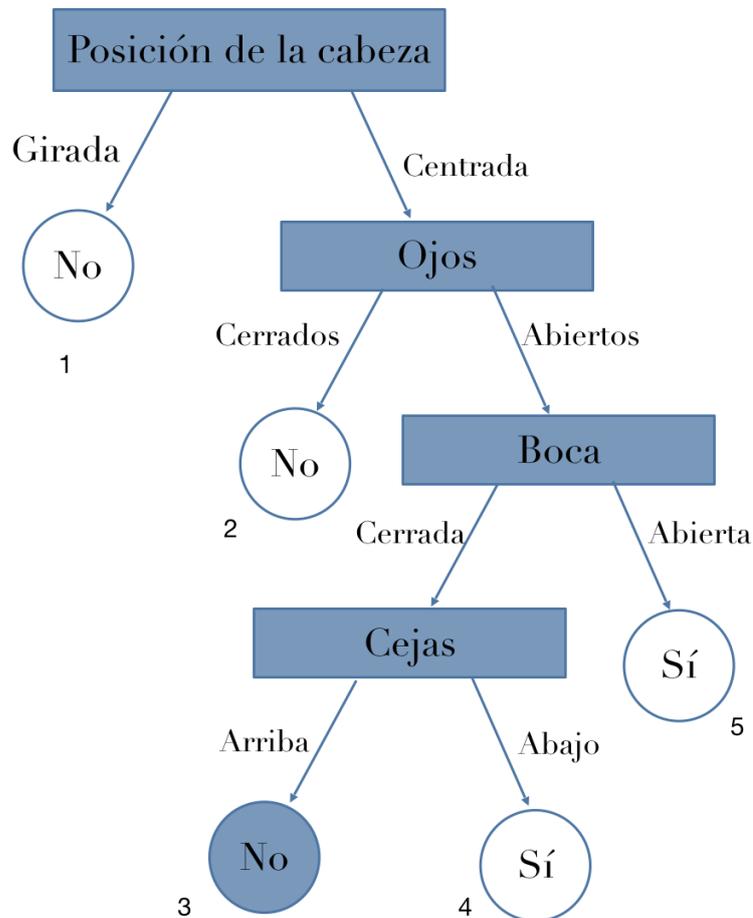


Figura 4.13 Combinaciones para la rama tres.

En la Figura 4.14 se muestra al sujeto de prueba para la rama cuatro.



Figura 4.14 Combinaciones para la rama cuatro.

En la Figura 15 se muestra la prueba de la combinación de elementos para la rama 4 “posición de la cabeza → centrada → ojos → abiertos → boca → cerrada → cejas → abajo → sí”.

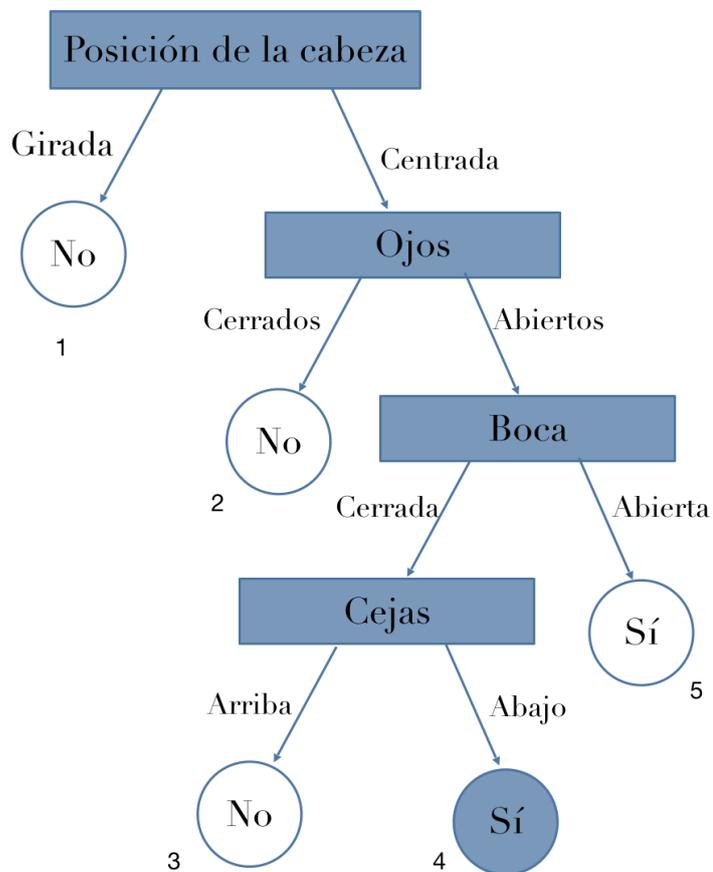


Figura 4.15 Combinaciones para la rama cuatro.

En la Figura 4.16 se muestra al sujeto de prueba para la rama cinco.



Figura 4.16 Combinaciones para la rama seis.

En la Figura 4.17 se muestra la prueba de la combinación de elementos para la rama 5 “posición de la cabeza → centrada → ojos → abiertos → boca → abierta → sí”.

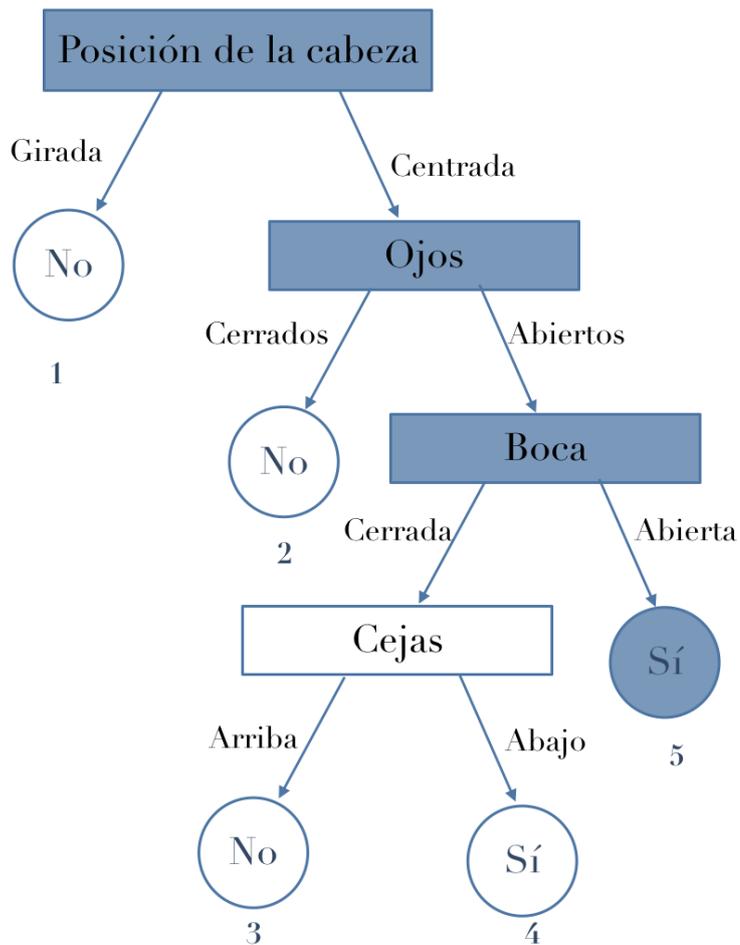


Figura 4.17 Combinaciones para la rama cuatro.

4.3.5 Resultados de la valoración de la atención de los sujetos

Para evaluar la atención de los sujetos se utilizaron cuatro videos de prueba, los cuales eran reproducidos para los participantes, no todos los sujetos observaban los mismos videos debido a la duración de los mismos. Algunos videos eran más interesantes que otros. En la Tabla 2.4 se muestra la portada de los videos utilizados en estas pruebas, uno trata sobre los vuelos en avión, otro sobre la evolución de la tierra los agujeros negros, un tercero sobre la metamorfosis de la oruga a mariposa y finalmente un video que habla sobre agujeros negros.

En la Figura 4.18 se muestra el sujeto de prueba uno, el cual fue grabado mientras visualizaba el video número dos mencionado en la Tabla 2.4 Como se puede apreciar en la Figura 4.18 el sujeto de prueba en ocasiones prestaba atención mientras realizaba la actividad encomendada.



Figura 4.18 Sujeto de prueba 1.

En la Figura 4.19 se muestran la gráfica de la atención prestada y cómo se fue modificando en bloques de aproximadamente dos segundos hasta concluir la actividad. Mediante el sistema se concluye que estuvo interesado durante el 42.55% de la actividad.

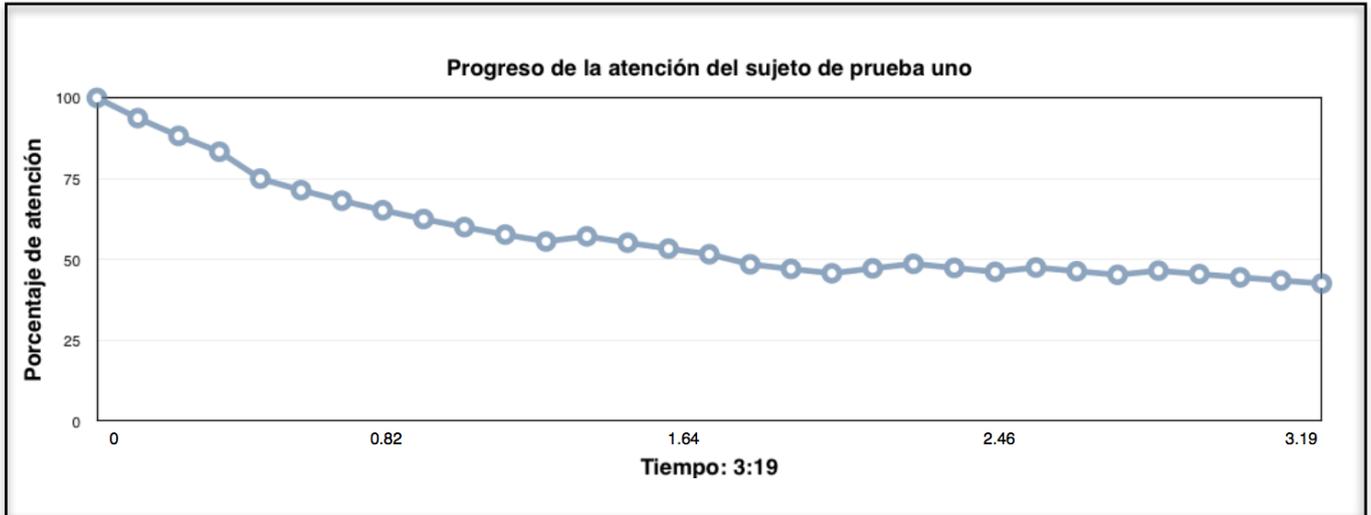


Figura 4.19. Histórico de la atención del sujeto de prueba uno.

El histórico es obtenido mediante el cálculo de la ecuación 1, siempre toma en cuenta el valor anterior obtenido, para realizar el cálculo. En la Figura 4.20 se muestra la gráfica de los resultados del individuo uno con el video dos, donde cada barra corresponde al bloque que contiene 31 datos y el valor de porcentaje de atención del individuo en la parte izquierda de la gráfica se presenta el porcentaje de veces que se clasificó como sí; es decir para la primera barra esto corresponde a 23 datos de un total de 31 dando un porcentaje de 78% para el primer bloque de 31 datos se concluyó que sí se estaba prestando atención.

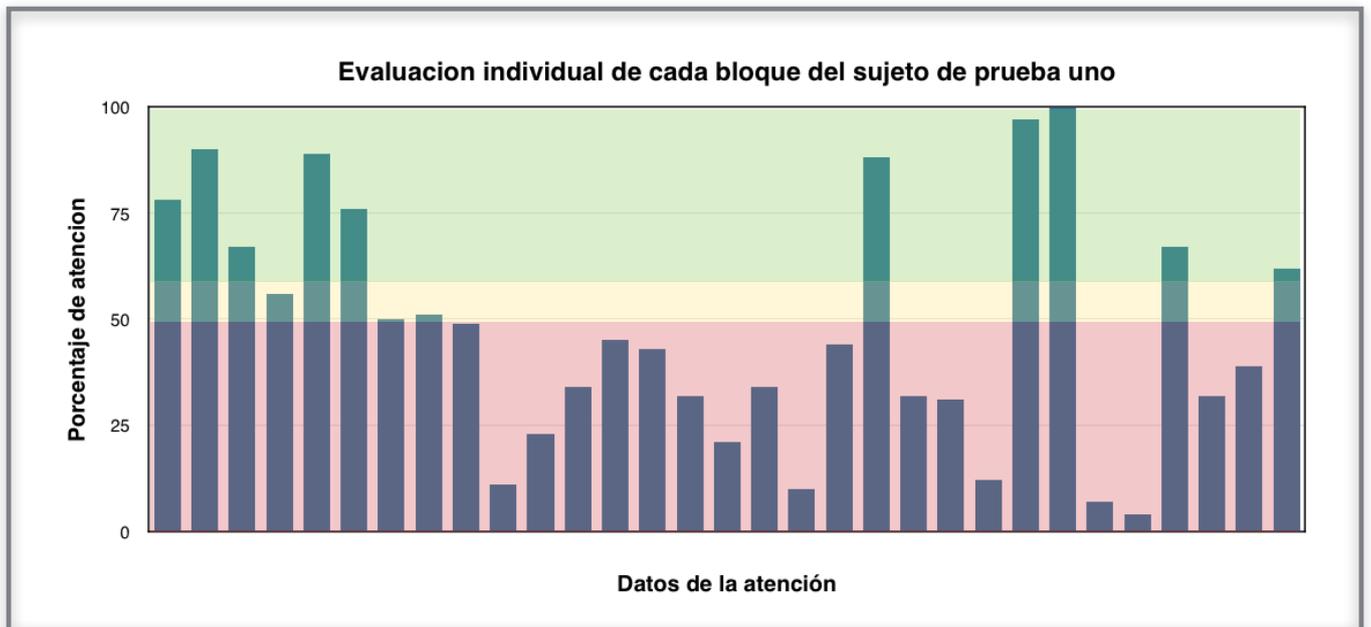


Figura 4.20 Evaluación de la atención sin histórico para el sujeto de prueba 1.

En la Figura 4.21 se tienen tres colores sobre la gráfica que corresponden a la atención. La sección

que cubre el color rojo corresponde al conjunto de bloques donde no se prestó atención durante la actividad. La sección que cubre el color amarillo corresponde a un estado neutral y finalmente el color verde para los bloques donde no se mostró interés.

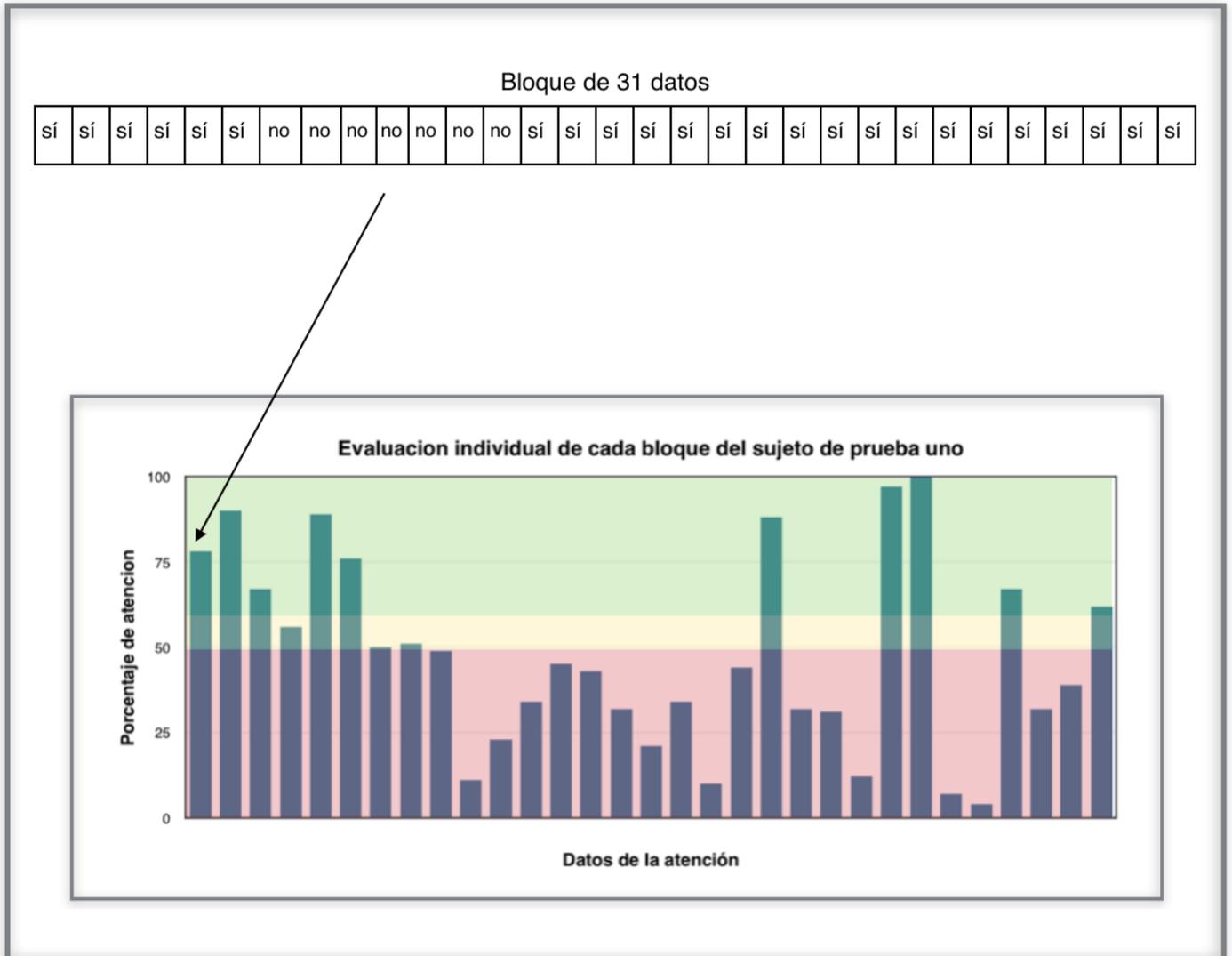


Figura 4.21 Evaluación de los elementos del un bloque.

En la Figura 4.21 se tiene la descripción de cómo está formada cada una de las barras de la Figura 4.20. Cada barra contiene la información de un bloque, en la parte superior de la imagen se muestra un vector, el cual corresponde al primer bloque de resultados del sujeto de prueba uno; dicha barra tiene un 78% de valores con resultado “ sí ”, por tanto del vector de 31 datos, 24 corresponden a *sí* y 7 a *no*. Como se puede apreciar en el bloque de 31 datos la mayoría son valores “sí”.

En la Figura 4.22 se muestra al sujeto de prueba dos el cual fue grabado mientras visualizaba el video número tres. Como se puede apreciar el sujeto de prueba dos tuvo algunos movimientos mientras realizaba la actividad encomendada, sin embargo en las imágenes parece estar atendiendo constantemente.



Figura 4.22 Sujeto de prueba 2.

En la Figura 4.23 se muestra la gráfica de la atención prestada y cómo se fue modificando hasta concluir la actividad. Mediante el sistema se obtiene que estuvo interesado durante el 90.89% de la actividad.

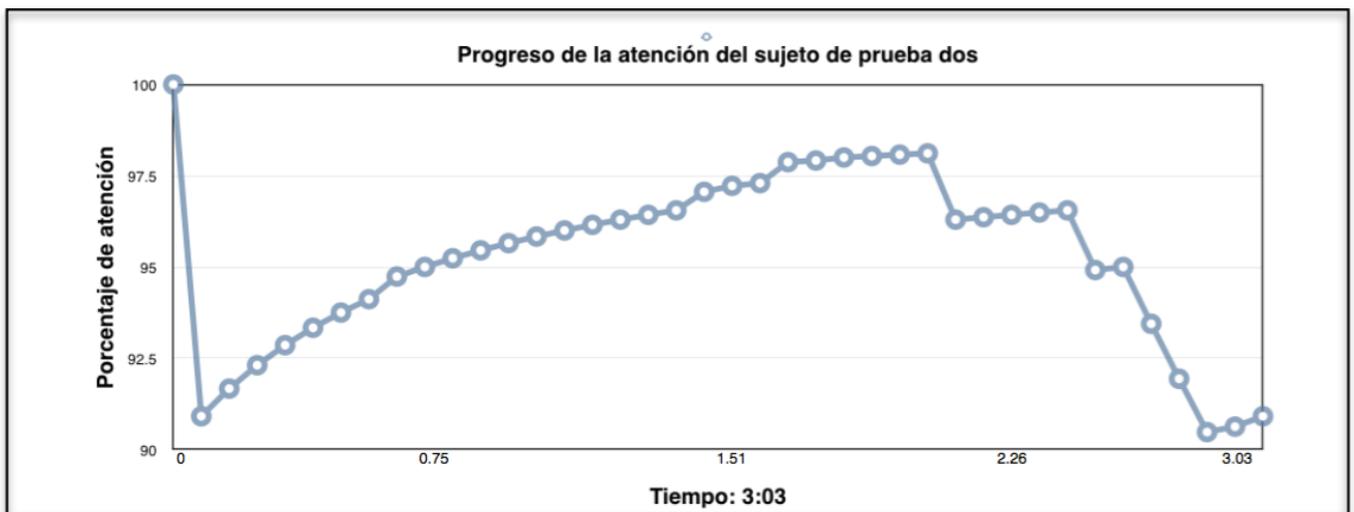


Figura 4.23 Histórico de la atención del sujeto de prueba uno.

En la Figura 4.24 se muestra la gráfica que corresponde a la evaluación de la atención sin tomar en cuenta el histórico para el sujeto de prueba 2 donde se puede apreciar que la mayoría de las barras se encuentran en la zona de color verde.

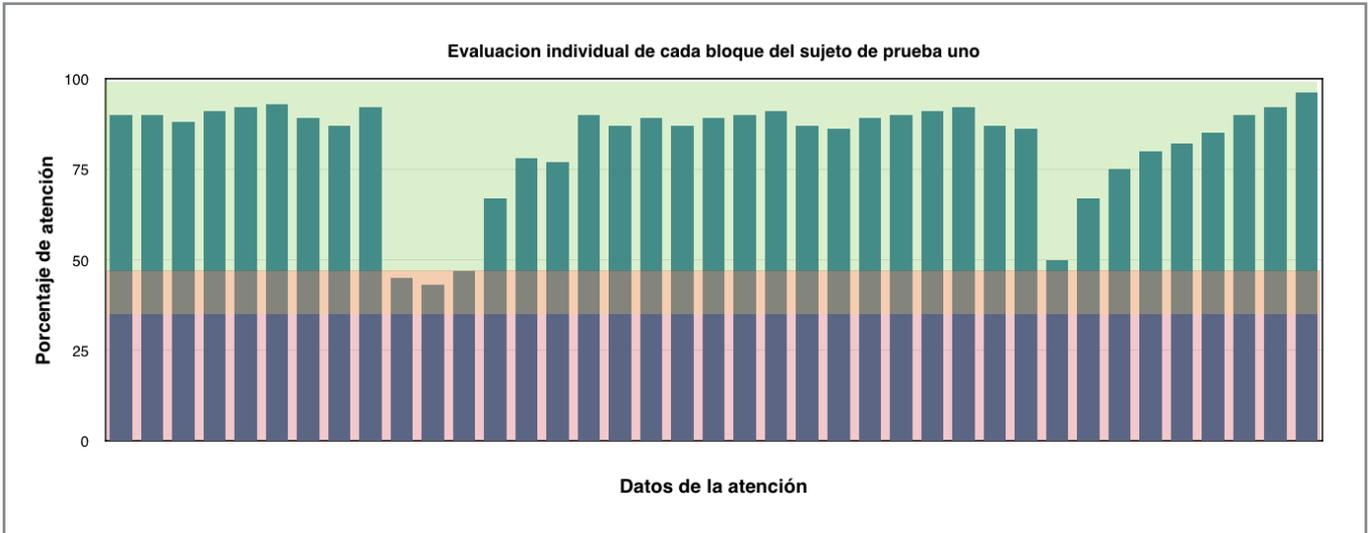


Figura 4.24 Evaluación de la atención sin histórico para el sujeto de prueba 2.

En la Figura 4.25 se muestra el sujeto de prueba tres, el cual fue grabado mientras visualizaba el video número tres y se puede apreciar que muestra un interés en la actividad de frente atendiendo.

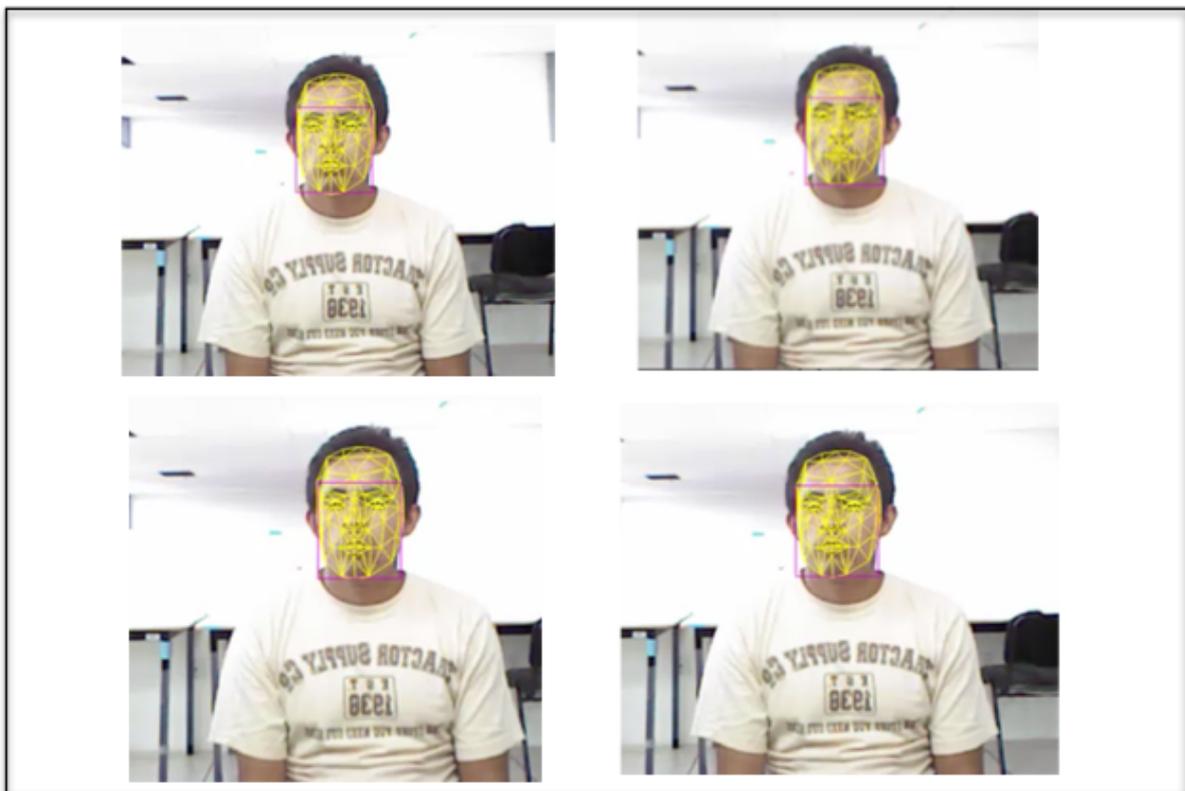


Figura 4.25 Sujeto de prueba 4.

En la Figura 2.26 se muestra la gráfica de la atención prestada del sujeto cuatro concluyendo que estuvo interesado durante el 96.55% de la actividad.

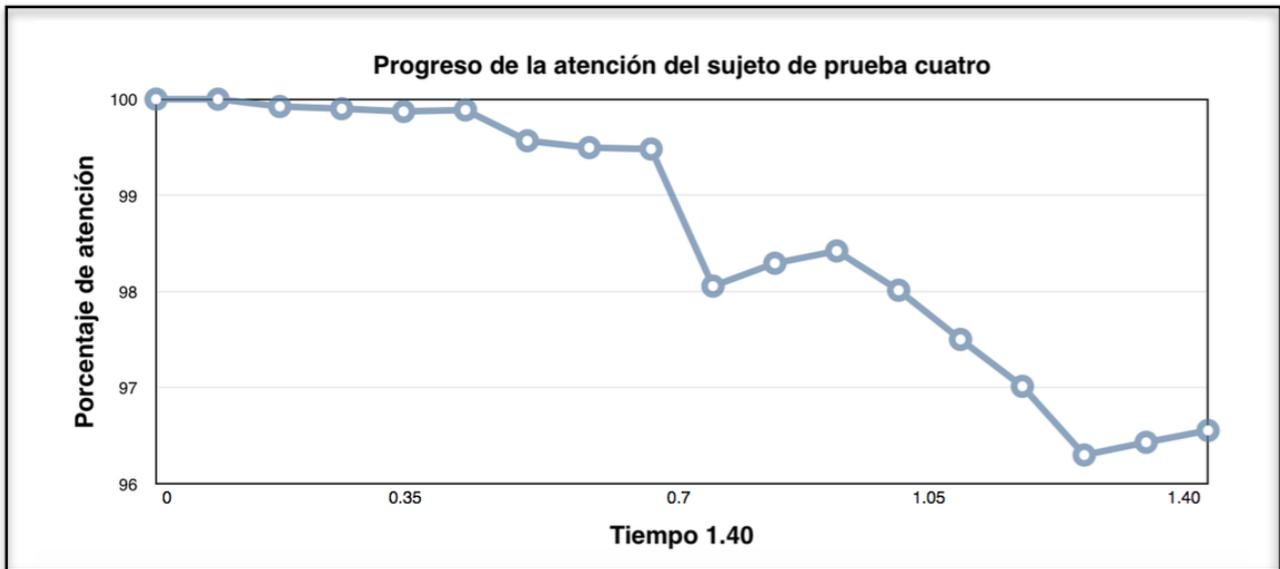


Figura 4.26 Histórico de la atención del sujeto de prueba tres.

En la Figura 4.27 se muestra la gráfica que corresponde a la evaluación de la atención sin tomar en cuenta el histórico para el sujeto de prueba 2 donde se puede apreciar que la mayoría de las barras se encuentran en la zona de color verde esto debido a que el sujeto de prueba 3 estuvo el 96.55% del tiempo atento.

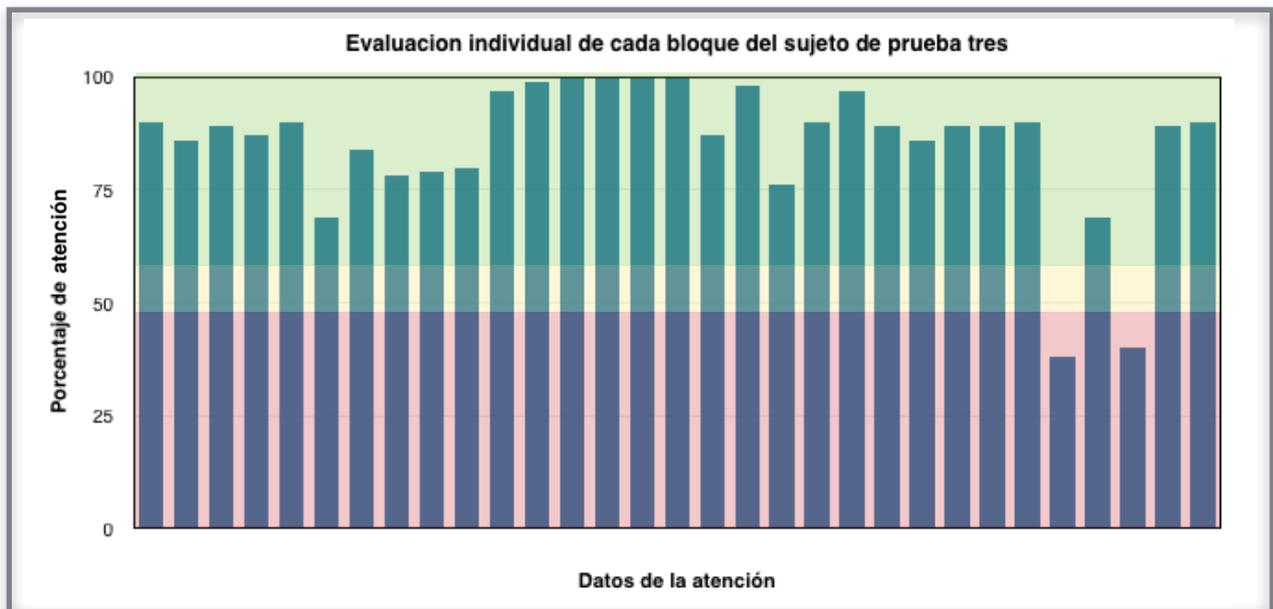


Figura 4.27 Evaluación de la atención sin histórico para el sujeto de prueba 3.

En la Figura 4.28 se muestra el sujeto de prueba cinco el cual fue grabado mientras visualizaba el video número uno y muestra un cierto interés en la actividad aunque se encuentra consumiendo caramelo.

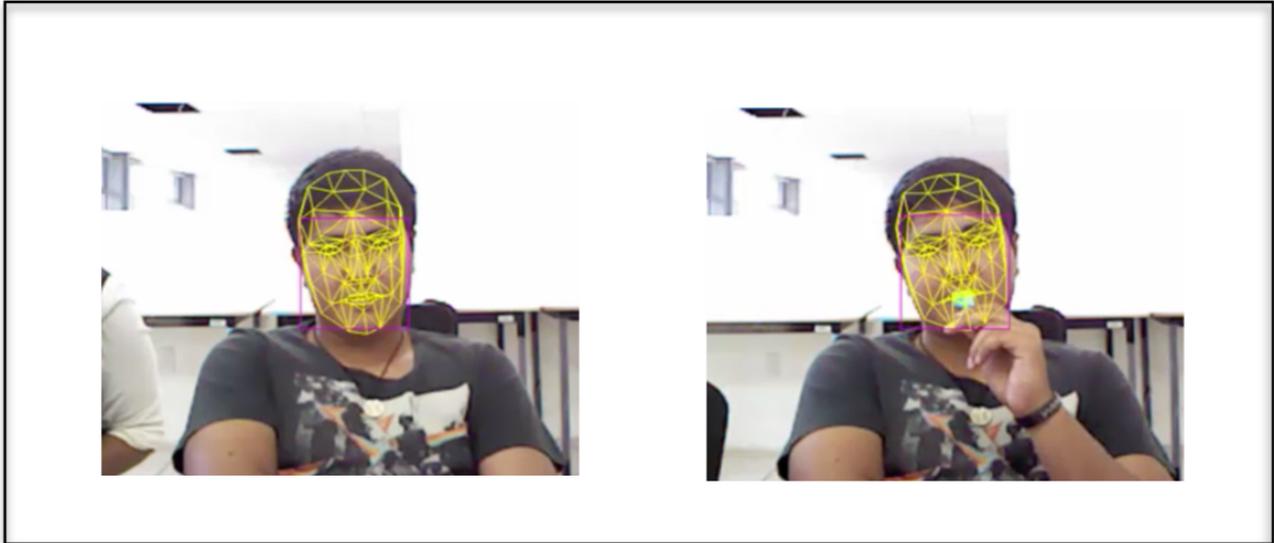


Figura 4.28 Sujeto de prueba 4.

En la Figura 4.29 se muestra la gráfica de la atención prestada del sujeto cinco concluyendo que estuvo interesado en un 52.17% de la actividad, esto por la oclusión del elemento boca, sin embargo el individuo estuvo atento todo el tiempo, este es un claro ejemplo de una interpretación errónea por oclusión.

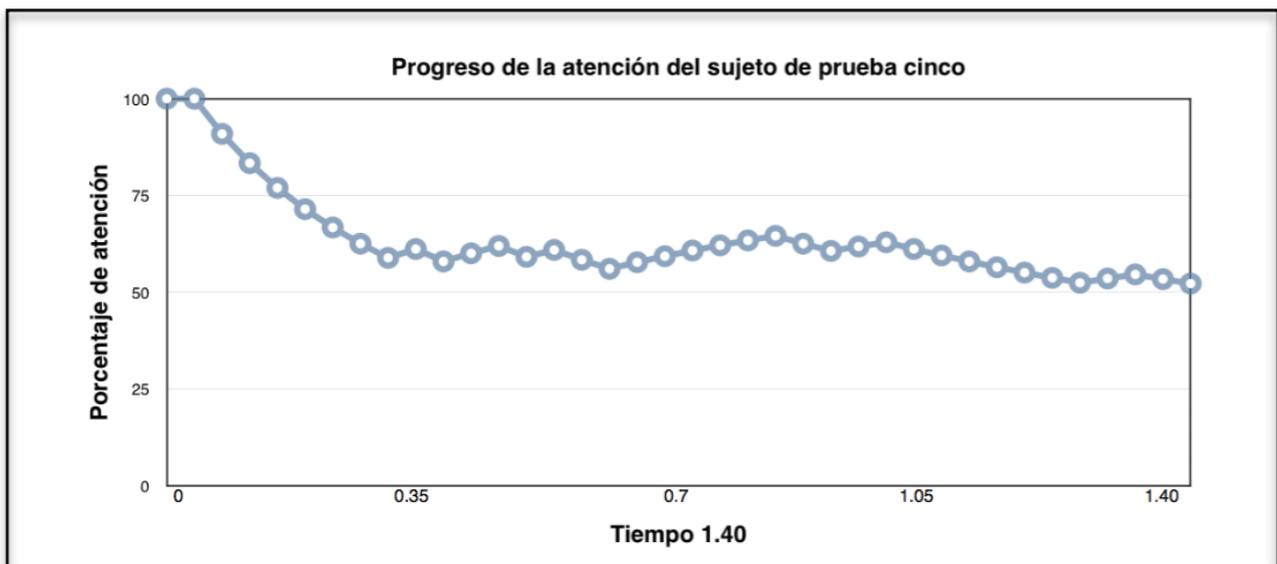


Figura 4.29 Histórico de la atención del sujeto de prueba cinco.

En la Figura 4.30 se muestra la gráfica que corresponde a la evaluación de la atención sin tomar en cuenta el histórico para el sujeto de prueba 4.

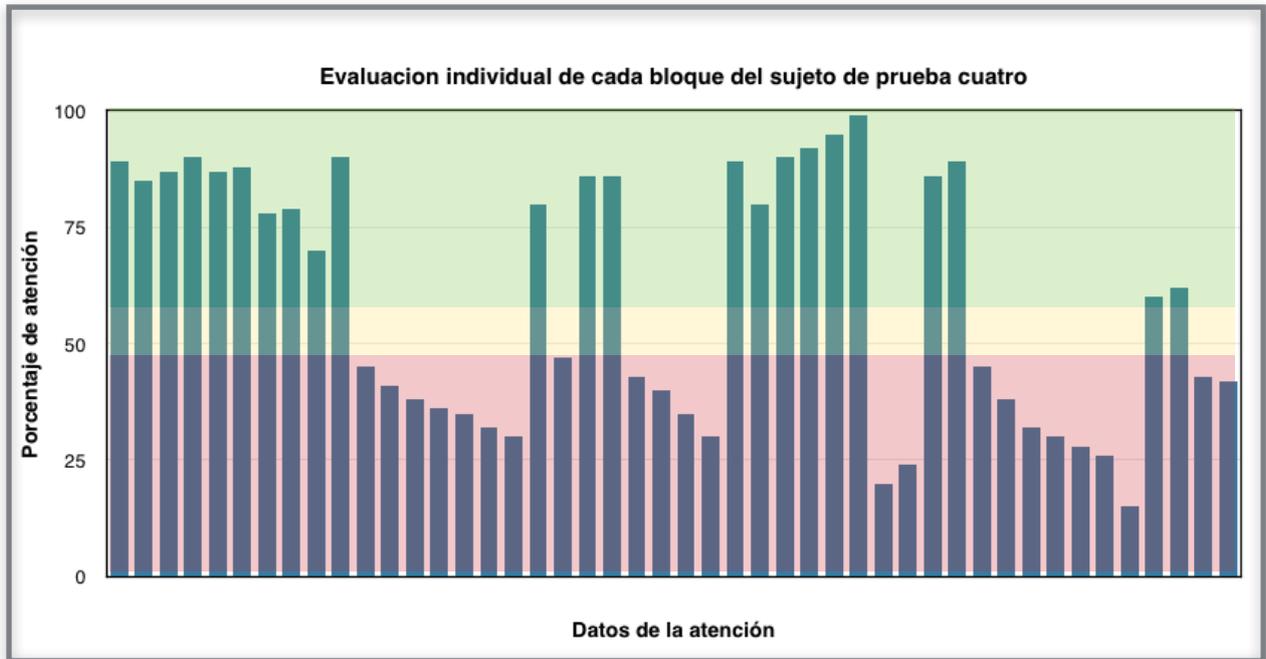


Figura 4.30 Evaluación de la atención sin histórico para el sujeto de prueba 4.

En la Figura 4.31 se muestra el sujeto de prueba tres, el cual fue grabado mientras visualizaba el video número tres y no prestó atención al video, sólo al principio después de un minuto el sujeto se durmió, como se puede apreciar el sujeto tiene aspecto somnoliento.

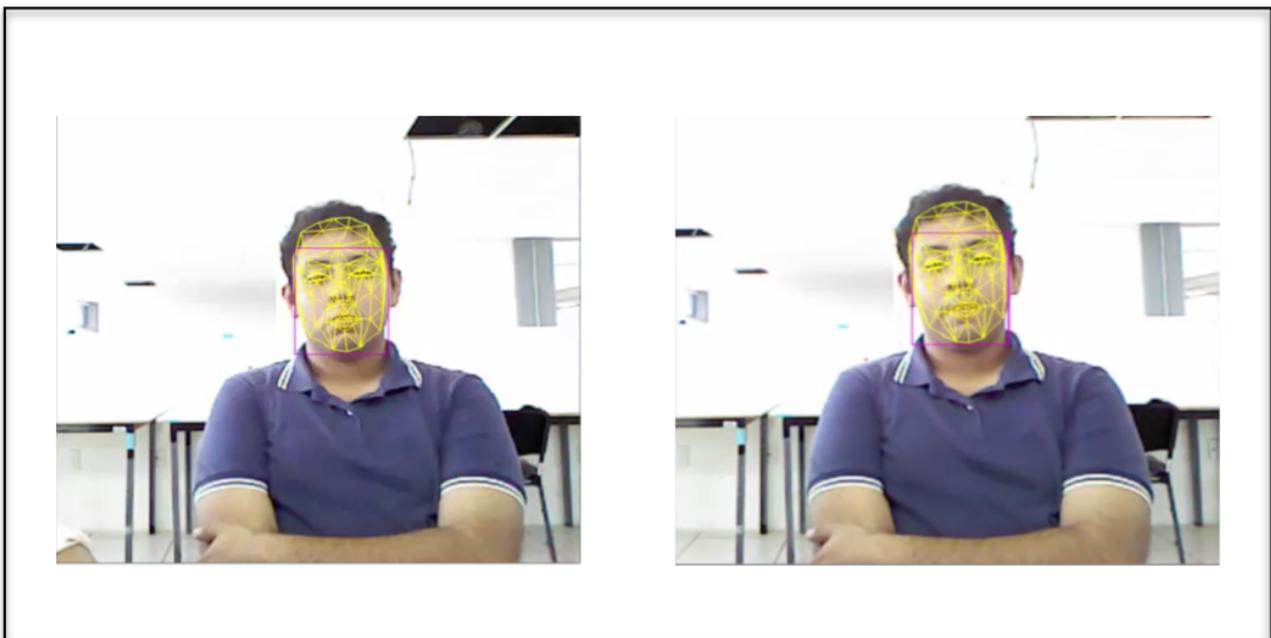


Figura 4.31 Sujeto de prueba 5.

En la Figura 4.32 se muestra la gráfica de la atención prestada, se puede apreciar cómo la atención disminuyó hasta llegar a un porcentaje de atención de 11.84%.

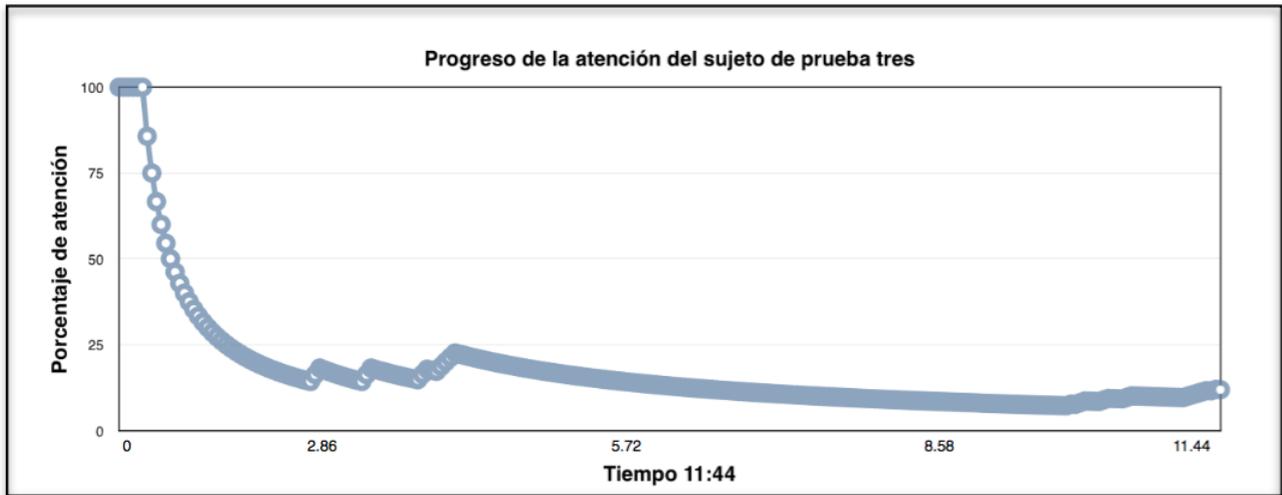


Figura 4.32 Histórico de la atención del sujeto de prueba cinco.

4.3.6 Matriz de confusión

Para la elaboración de la matriz cruzada se tomó en cuenta 61 pruebas correspondientes a los cuatro videos, pruebas realizadas a los 20 individuos.

En la tabla 4.1 se muestra la matriz de confusión donde el 89.79% corresponde a las veces que el sistema calculó que el sujeto de prueba estaba prestando atención y realmente lo estaba haciendo y el 91.66% corresponde a las veces que el sistema evaluó que el sujeto de prueba no estaba prestando atención y realmente no lo estaba haciendo. El 10.21% corresponde a las veces que el sistema determinó que el sujeto estaba prestando atención y fue erróneo ya que el sujeto no prestaba atención. Finalmente el 8.34% corresponde al porcentaje de los sujetos que el sistema dijo que no prestaron atención sin embargo sí estaban atendiendo.

	Sí	No
Sí	89.79%	10.21%
No	8.34%	91.66%

Tabla 4.1. Matriz de confusión para la atención.

Los datos para la matriz de confusión fueron tomados de los resultados de los sujetos de prueba y los videos que visualizaron así como de la identificación del interés. En la tabla 4.2 se muestra el desglose de los datos utilizados en la matriz de confusión.

Tabla 4.2. Desglose de los datos utilizados en la matriz de confusión.

Sujeto	Cantidad de videos vistos	Aciertos en la clasificación	Observaciones
1	4	3	El error en la interpretación fue debido a que el individuo está tomando agua de una botella
2	3	3	Correcta identificación en todos los videos
3	4	2	Demasiada iluminación directa lo cual causó errores en la detección de los elementos
4	4	3	La ceja del individuo es muy elevada por naturaleza
5	2	1	El individuo estaba comiendo una paleta de caramelo por lo cual el elemento boca no era correctamente interpretado
6	4	4	Correcta identificación en todos los videos
7	2	1	Demasiada iluminación directa lo cual causó errores en la detección de los elementos
8	2	1	El individuo usaba lentes de marco grueso, lo cual causaba confusión en el sistema
9	3	3	Correcta identificación en todos los videos
10	4	3	El individuo constantemente se tocaba el rostro por lo cual no había una correcta interpretación de los elementos
11	4	3	El individuo usaba lentes, lo cual causó confusión en el sistema
12	4	4	Demasiada iluminación directa lo cual causó errores en la detección de los elementos
13	3	3	Correcta identificación en todos los videos
14	3	1	El individuo tenía ojos muy pequeños, tanto que parecían estar cerrados
15	2	2	Correcta identificación en todos los videos
16	4	3	Demasiada iluminación directa lo cual causó errores en la detección de los elementos
17	2	2	Correcta identificación en todos los videos
18	3	2	Demasiada iluminación directa lo cual causó errores en la detección de los elementos
19	4	4	Correcta identificación en todos los videos
20	4	4	Correcta identificación en todos los videos

4.4 Discusión

Se estudiaron las combinaciones posibles de los elementos faciales trabajados para conseguir una interpretación automática del interés o desinterés mediante las expresiones faciales básicas. El sistema implementado puede analizar el comportamiento mientras un individuo realiza una prueba para la evaluación, en este caso se utilizaron cuatro videos para evaluar la atención de los individuos y de esta manera descubrir la motivación que tenían por la actividad. El banco de datos fue creada para esta tesis y los sujetos de prueba fueron veinte de los cuales se tiene grabado su comportamiento durante la prueba para corroborar los resultados obtenidos por el sistema.

Se identificó que es necesario tener un ambiente controlado para que el sistema trabaje mejor ya que la iluminación y accesorios utilizados por los individuos causan errores en el sistema al momento de la correcta detección de los elementos trabajados. La distancia en que se recomienda colocar el Kinect frente al sujeto de estudio debe estar comprendida entre 60 y 120 cm.

La evaluación de las pruebas de este sistema obtuvo que para casos donde del sistema determinó que los sujetos estaban prestando atención y realmente estaban prestando atención fue de 89.79% y para los casos donde el sistema valoró que los sujetos de prueba no prestaron atención y en realidad no lo hicieron fue de 90.72%.

Es importante mencionar que en el uso accesorios tales como lentes puede influir en la detección de los elementos: ojos y cejas, por ejemplo el cristal cubre los ojos y podrían estar reflejando el entorno, lo que dificultaría la observación de los mismos. En el caso de la ceja, se pudo apreciar que los lentes de marco grueso causan confusión con las cejas. En la Figura 4.31 se muestra un sujeto con lentes de marco grueso.

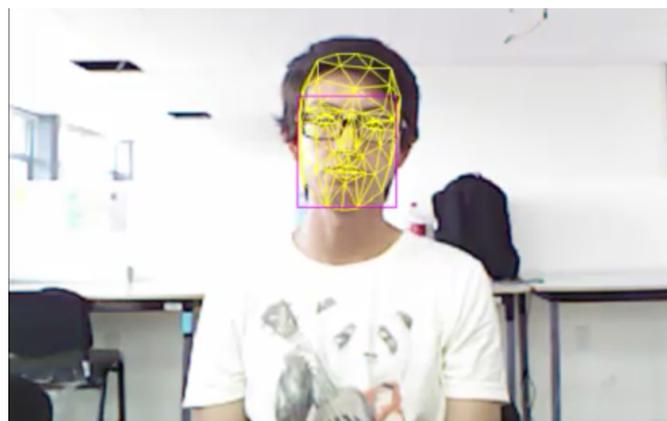


Figura 4.31 Sujeto con gafas de marco grueso.

En la Tabla 4.3 se muestra la información de la actividad del sujeto de la Figura 4.31 como se puede apreciar a partir de la segunda columna se tiene que el sujeto no se encuentra atento debido al

elemento ceja, pero en realidad este sujeto durante la actividad siempre estuvo atento pero cuando inclinó un poco la cabeza el marco de los lentes quedó posicionado sobre el área de las cejas, por tanto lo tomaba como parte del elemento ceja y clasifica como no interesado.

Tabla 4.3 Atención y elemento detonan para la clasificación.

Nivel de atención del sujeto de prueba y elemento detonante para la clasificación			
Atencion: Si -> boca 1	Atencion: Si -> boca 1	Si:66.6667%, No:33.3333%	Si:47.5%, No:52.5%
Si:100%, No:0%	Si:100%, No:0%	Atencion: No -> ceja -1	Atencion: No -> ceja -1
Atencion: Si -> boca 1	Atencion: Si -> boca 1	Si:64.2857%, No:35.7143%	Si:46.3415%, No:53.6585%
Si:100%, No:0%	Si:100%, No:0%	Atencion: Si -> boca 1	Atencion: No -> ceja -1
Atencion: Si -> boca 1	Atencion: Si -> boca 1	Si:65.5172%, No:34.4828%	Si:45.2381%, No:54.7619%
Si:100%, No:0%	Si:100%, No:0%	Atencion: No -> ceja -1	Atencion: No -> ceja -1
Atencion: Si -> boca 1	Atencion: Si -> boca 1	Si:63.3333%, No:36.6667%	Si:44.186%, No:55.814%
Si:100%, No:0%	Si:100%, No:0%	Atencion: No -> ceja -1	Atencion: No -> ceja -1
Atencion: Si -> boca 1	Atencion: Si -> boca 1	Si:61.2903%, No:38.7097%	Si:43.1818%, No:56.8182%
Si:100%, No:0%	Si:100%, No:0%	Atencion: No -> ceja -1	Atencion: No -> ceja -1
Atencion: Si -> boca 1	Atencion: No -> ceja -1	Si:59.375%, No:40.625%	Si:42.2222%, No:57.7778%
Si:100%, No:0%	Si:94.7368%, No:5.26316%	Atencion: No -> ceja -1	Atencion: No -> ceja -1
Atencion: Si -> boca 1	Atencion: No -> ceja -1	Si:57.5758%, No:42.4242%	Si:41.3043%, No:58.6957%
Si:100%, No:0%	Si:90%, No:10%	Atencion: No -> ceja -1	Atencion: No -> ceja -1
Atencion: Si -> boca 1	Atencion: No -> ceja -1	Si:55.8824%, No:44.1176%	Atencion: No -> ceja -1
Si:100%, No:0%	Si:85.7143%, No:14.2857%	Atencion: No -> ceja -1	Si:38.7755%, No:61.2245%
Atencion: Si -> boca 1	Atencion: No -> ceja -1	Si:54.2857%, No:45.7143%	Atencion: No -> ceja -1
Si:100%, No:0%	Si:81.8182%, No:18.1818%	Atencion: No -> ceja -1	Si:38%, No:62%
Atencion: Si -> boca 1	Atencion: No -> ceja -1	Si:52.7778%, No:47.2222%	Atencion: No -> ceja -1
Si:100%, No:0%	Si:78.2609%, No:21.7391%	Atencion: No -> ceja -1	Si:37.2549%, No:62.7451%
Atencion: Si -> boca 1	Atencion: No -> ceja -1	Si:51.3514%, No:48.6486%	Atencion: No -> ceja -1
Si:100%, No:0%	Si:75%, No:25%	Atencion: No -> ceja -1	Si:36.5385%, No:63.4615%
Atencion: Si -> boca 1	Atencion: No -> ceja -1	Si:50%, No:50%	Atencion: Si -> boca 1
Si:100%, No:0%	Si:72%, No:28%	Atencion: No -> ceja -1	Si:37.7358%, No:62.2642%
Atencion: Si -> boca 1	Atencion: No -> ceja -1	Si:48.7179%, No:51.2821%	Atencion: Si -> boca 1
Si:100%, No:0%	Si:69.2308%, No:30.7692%	Atencion: No -> ceja -1	Si:38.8889%, No:61.1111%

En estos experimentos los accesorios utilizados por los sujetos fueron lentes y cabello cubriendo el rostro, sin embargo es necesario mencionar que los tapabocas o bufandas podrían causar conflicto en el momento de la clasificación ya que también cubren parte del rostro y los elementos trabajados.

En la Figura 4.32 se tiene al sujeto de prueba seis, el cual es de los pocos que tiene un estado neutral.



Figura 4.32 Sujeto de prueba

En la Figura 4.33 se muestra la gráfica de la atención prestada, se puede apreciar cómo la atención disminuyó hasta llegar a un porcentaje de atención de 49.28%.

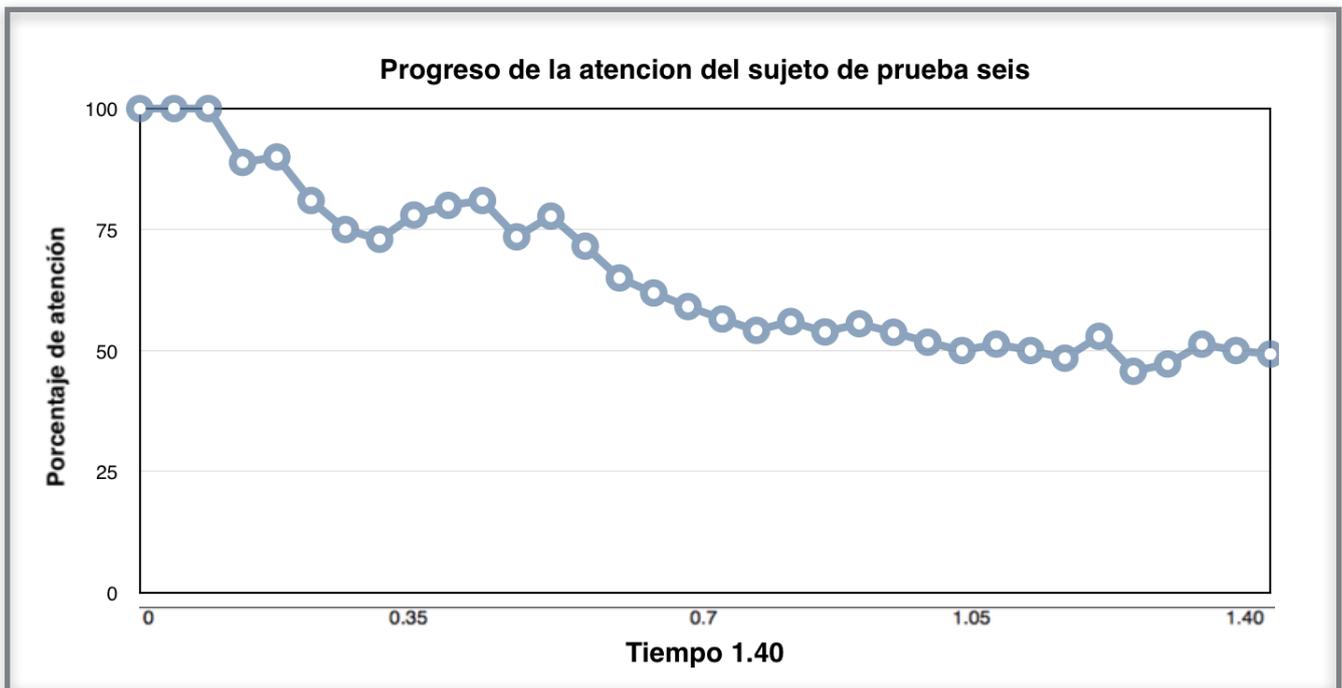


Figura 4.33 Histórico de la atención del sujeto de prueba seis.

En la Figura 4.33 se puede apreciar cómo el sujeto comenzó poniendo atención, sin embargo conforme pasó el tiempo quedó en un estado neutral valorándose cerca del 50% el tiempo que prestó atención. El estado neutral se consigue cuando el sujeto de prueba no está totalmente atento pero tampoco distraído, por ello su promedio de atención en la actividad, ronda el 50%.

CAPÍTULO 5 Conclusiones generales y trabajos futuros

En este capítulo se presentan los objetivos alcanzados, logros, las conclusiones finales y los trabajos futuros del presente trabajo así como las oportunidades de mejora.

5.1 Objetivos alcanzados

A continuación se presenta el objetivo general de esta tesis, el cual se presentó al inicio de la investigación y su cumplimiento se mostró en las secciones de análisis, diseño e implementación de la solución, experimentación y resultados.

El objetivo primordial de esta propuesta es “detectar e interpretar mediante las expresiones faciales el comportamiento de un sujeto para clasificarlo en motivado desmotivado o neutral con un sensor RGBD”.

El sistema de visión artificial desarrollado tiene como entrada una secuencia de imágenes que contienen un solo rostro, sobre el cual se realiza la localización automática de los componentes faciales mediante la librería *facetracking*. La salida de este sistema es un archivo de texto en formato txt que contiene las descripciones de cada uno de los elementos faciales trabajados y la combinación de las descripciones de esos elementos con una decisión final de a qué clase pertenecen, además incluyen a los elementos que causaron dicha clasificación.

Los detalles de los objetivos específicos alcanzados se explican a continuación y los alcanzados son señalados con una (✓) a la izquierda:

Objetivo General

✓ Estudiar e implementar técnicas para el reconocimiento de expresiones faciales (o corporales) y asociarlas a ciertos estados de ánimo que pudieran ser agrupados en tres categorías: motivado, neutral y desmotivado.

Objetivos Específicos

✓ 1. Se evaluará el sensor RGB del Kinect y su correspondiente SDK para el reconocimiento de expresiones faciales.

✓ 2. Seleccionar elementos del rostro que denoten motivación, desmotivación y el estado neutral.

✓ 3. Obtener bancos de datos de rostros.

✓ 4. Se evaluará el trabajo de Rita Alejandra Morfín Villatoro para ver qué pudiera ser útil en esta tesis.

5.2 Alcances

- ✓ 1. Se trabajarán de 2 a 5 elementos ya sean faciales o corporales.
- ✓ 2. De la imagen o video capturado, se realizará la extracción automática de los elementos faciales definidos.
- ✓ 3. De los elementos que fueron extraídos se realizará la clasificación de acuerdo a sus características en: motivado, desmotivado y neutral.
- ✓ 4. Se evaluó el trabajo de Rita Alejandra Morfín Villatoro para ver que pudiera ser útil en esta tesis.

5.3 Logros

Se desarrolló un sistema de reconocimiento de expresiones faciales que identifica la motivación y crea un archivo con el historial de las expresiones identificadas individualmente y en conjunto.

Se creó un banco de datos con imágenes 3D, con 25 sujetos en total. El banco de datos cuenta con los valores numéricos y con la captura del rostro con la malla de Kinect.

Se realizó una implementación para que mediante elementos faciales fuera posible reconocer la motivación.

Se implementó el algoritmo de clasificación mediante el uso de una Función Discriminante por Distancia sin necesidad de utilizar OpenCV, lo cual facilita la integración del proyecto ya que no ocasiona conflicto con las librerías del OpenCV y el Toolkit de Kinect.

El sistema reconoce las expresiones faciales sin que el usuario tenga que exagerar la expresión facial, para detectar la motivación.

5.4 Conclusiones

En esta investigación se estudiaron y analizaron las herramientas de desarrollo de Kinect para realizar la identificación y seguimiento del rostro, se estudiaron las combinaciones posibles de los elementos faciales trabajados para conseguir una interpretación automática del interés o desinterés mediante las expresiones faciales básicas. El sistema implementado puede analizar el comportamiento mientras un individuo realiza una prueba para la evaluación, en este caso se

utilizaron cuatro videos para determinar la atención de los individuos y de esta manera descubrir la motivación que tenían por la actividad.

El banco de datos fue creado para esta tesis y los sujetos de prueba fueron veinte de los cuales se tiene grabado su comportamiento durante la experimentación para corroborar los resultados obtenidos por el sistema.

Se identificó que es necesario tener un ambiente controlado para que el sistema trabaje mejor ya que la iluminación y accesorios utilizados por los individuos causan errores en el sistema al momento de la correcta detección de los elementos trabajados, en las pruebas realizadas con la iluminación se concluyó que la iluminación artificial causa menos problemas que la iluminación natural. La distancia a la que se recomienda colocar el Kinect frente al sujeto de estudio es entre 60 y 120 cm para evitar problemas con la detección del rostro.

En la experimentación se obtuvo que para casos donde el sistema determinó que los sujetos estaban prestando atención y realmente estaban prestando atención fue de 89.79% para los casos donde el sistema evaluó que los sujetos de prueba no prestaron atención y en realidad no lo hicieron fue de 90.72%.

5.5 Trabajos futuros

Durante el desarrollo de este proyecto se detectaron algunos puntos de interés que son parte de los trabajos futuros para extender y mejorar el actual sistema desarrollado. Es posible utilizar las unidades de acción para tomar en cuenta los gestos de la nariz.

El estado neutral evaluando la neutralidad de los elementos es un trabajo complejo debido a que cada persona tiene un rostro diferente y su estado neutral diferirá del de los demás por lo cual sería un tema de investigación completo, en el que se pudiera ahondar sobre el estado neutral.

Al igual que el estado neutral, las oclusiones podrían ser un tema a investigación, pues al detectar que un elemento no se puede analizar por oclusión sería posible descartarlo y evaluar con base a los demás elementos.

Referencias

- [1] R.S. Baker, A.T. Corbett, and K.R. Koedinger, Detecting Student Misuse of Intelligent Tutoring Systems, In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems*, MaceiÃs, , Alagoas, Brazil, (2004), 531–540.
- [2] R.S. Baker, A.T. Corbett, K.R. Koedinger, and I. Roll, Generalizing Detection of Gaming the System Across a Tutoring Curriculum, In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, Jhongli, Taiwan, (2006), 402–411.
- [3] R.S. Baker, A.T. Corbett, K.R. Koedinger, and A.Z. Wagner, Off-Task Behavior in the Cognitive Tutor Classroom: When Students “Game the System”, In *Proceedings of ACM CHI 2004: Computer-Human Interaction*, Vienna, Austria, (2004), 383–390.
- [4] R.S. Baker, J.A. Walonoski, N.T. Heffernan, I. Roll, A.T. Corbett, and K.R. Koedinger (In press), Why Students Engage in “Gaming the System” Behavior in Interactive Learning Environments, *Journal of Interactive Learning Research (JILR)*, Chesapeake, VA: AACE.
- [5] R.L. Bangert-Drowns and C. Pyke, A Taxonomy of Student Engagement with Educational Software: An Exploration of Literate Thinking with Electronic Text, *Journal of Educational Computing Research* 24(3) (2001), 213–234.
- [6] C.R. Beal, L. Qu, and H. Lee, Classifying Learner Engagement Through Integration of Multiple Data Sources, In *Proceedings of the 21st National Conference on Artificial Intelligence*, Menlo Park, California (2006), 2–8, AAAI Press.
- [7] Juan Jose Rosas Varela. Localización seguimiento y caracterización de expresiones faciales. Centro nacional de Investigación y Desarrollo Tecnológico. Tesis de maestría. Junio 2016.
- [8] Rita A. Morfin Villatoro. Caracterización de Expresiones Faciales con un sensor RGBD. Centro nacional de Investigación y Desarrollo Tecnológico. Reporte de avance de tesis de maestría. Diciembre 2015.
- [9] Face tracking, <https://msdn.microsoft.com/en-us/library/jj130970.aspx>, Consultado: 21/03/2016.
- [10] Junguk Cho, Shahnam Mirzaei, Jason Oberg, Ryan Kastner (2009). Fpga- based face detection system using Haar classifiers. FPGA '09 Proceedings of the ACM/SIGDA international symposium on Field programmable gate arrays, ACM, pp. 103-112.

- [11] Daniel Paredes (2009). Seguimiento y Caracterización de Componentes del Rostro para la Detección de Expresiones Faciales. Tesis de maestría CENIDET.
- [12] Bassem Seddik, Houda Maâmatou, Sami Gazzah, Thierry Chateau, Najoua Essoukri Ben Amara (2013). Unsupervised Facial Expressions Recognition and Avatar Reconstruction from Kinect. IEEE - 2013 10th International Multi- Conference on Systems, Signals & Devices. March 2013, pp. 1-6.
- [13] CARA Emotion Measurement. IMRSV Inc. Disponible en: <https://www.kairos.com/docs/>. Fecha de consulta: 27/04/16.
- [14] Murase H. and Nayar S. , (1995) Visual learning and recognition of 3D objects from appearance, International Journal of Computer Vision, no.14, pp.5-24.
- [15] Turk M. and Pentland A., (1991) Eigenfaces for Recognition, Journal of Cognitive Neuroscience, vol. 3, no. 1, pp.71-86, ISSN 2250-2459.
- [16] Bressan M., Guillamet D. and Vitrià J., (2003) Using an ICA representation of local color histograms for object recognition. Pattern Recognition 36(3), ISSN 2447-2454, pp. 691-701. 2003.
- [17] Boqing Gong, Yueming Wang, Jianzhuang Liu, Xiaoou Tang (2009). Automatic Facial Expression Recognition on a Single 3D Face by Exploring Shape Deformation. MM '09 Proceedings of the 17th ACM international conference on Multimedia, ACM, pp. 569-572.
- [18] Xiaoli Li, Qiuqi Ruan, Gaoyun An (2013). Analysis of Range Images Used in 3D Facial Expression Recognition. TENCON 2013 - 2013 IEEE Region 10 Conference (31194), IEEE. 22-25 Oct. 2013, pp. 1 - 4.
- [19] Nicholas Vretos, Nikos Nikolaidis, Ioannis Pitas (2011). 3d facial expression recognition using zernike moments on depth Images. 2011 18th IEEE International Conference on Image Processing, IEEE. 11-14 Sept. 2011, pp. 773 - 776.
- [20] Bassem Seddik, Houda Maâmatou, Sami Gazzah, Thierry Chateau, Najoua Essoukri Ben Amara (2013). Unsupervised Facial Expressions Recognition and Avatar Reconstruction from Kinect. IEEE - 2013 10th International Multi- Conference on Systems, Signals & Devices. March 2013, pp. 1-6. (es la 12)
- [21] Mounika Kakarla, Prof.G.Ram Mohana Reddy. (2014). A Real Time Facial Emotion Recognition Using Depth Sensor And Interfacing With Second Life Based Virtual 3D Avatar. International Conference on Recent Advances and Innovations in Engineering IEEE, May 2014, pp. 1-7.

- [22] Luca Stocchi. (2014) 3d Facial Expressions Recognition Using The Microsoft Kinect. School of Computing, Dublin City University, Dublin, Ireland, Image Processing (ICIP), 2011 18th IEEE International Conference: pp. 773-776: 2011
- [23] Rodrigo Ruiz Martin (2014). Reconocimiento De Expresiones Faciales Mediante Imágenes Adquiridas Con Cámara Kinect. Universidad Pontificia de Comillas, Madrid. Tesis de licenciatura.
- [24] Qi-rong MAO, Xin-yu PAN, Yong-zhao ZHAN, Xiang-jun SHEN (2015). Using Kinect for real-time emotion recognition via facial expressions. Frontiers of Information Technology & Electronic Engineering, Springer link, Vol.16, pp. 272-282.
- [25] P. Ekman and W. Friesen (1978), Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press.
- [26] Paul Ekman, Ph.D, Wallace V. Friesen, Ph.D, Joseph C. Hager, Ph.D. (2002). The Manual Facial Action Coding System. Nexus division of Network Information Research Corporation.
- [27] Qi-rong MAO, Xin-yu PAN, Yong-zhao ZHAN, Xiang-jun SHEN (2015). Using Kinect for real-time emotion recognition via facial expressions. Frontiers of Information Technology & Electronic Engineering, Springer link, Vol.16, pp. 272-282.
- [28] <https://msdn.microsoft.com/en-us/library/jj130970.aspx> Consultado: 21/03/2016.
- [29] http://omicron1.rssing.com/chan-7953211/all_p137.html. Consultado 20/11/2016.
- [30] Francisco Mario Hernandez Tejera, Jose Javier Lorenzo Navarro, Reconocimiento de formas clasificación y aprendizaje, Capitulo 3, Universidad de Las Palmas Gran Canaria, 2008.
- [31] http://lenguajedecuerpo.blogspot.mx/2015_07_01_archive.html. Consultado 09/05/2016.
- [32] <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0003628> Consultado 01/05/2016.
- [33] <http://sabia.tic.udc.es/gc/Contenidos%20adicionales/trabajos/3D/Realidad%20Virtual/web/dispositivos/trackers.html> Consultado 22/04/2016.
- [34] http://analistanoverbal.blogspot.mx/2013_08_01_archive.html Consultado 02/01/2016.

-
- [35] Jixia Zhang, Haibo Wang, Shaoguo Liu, Jiangyong Duan, Ying Wang, Chunhong Pan (2013). An integrated graph-based face segmentation approach from kinect videos. Image Processing (ICIP), 2013 20th IEEE International Conference on. 15-18 Sept. 2013, pp. 2733 - 2737.
- [36] Jixia Zhang, Haibo Wang, Shaoguo Liu, Franck Davoine, Chunhong Pan, Shiming Xiang (2013). Active learning based automatic face segmentation for kinect video. Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. 26-31 May 2013, pp. 1816 - 1820.
- [37]] Maja Pantic, Member, IEEE, and Leon J. M. Rothkrantz (2000). Expert system for automatic analysis of facial expressions. Image and Vision Computing. Elsevier, vol. 18, pp. 881–905.
- [38] Pierre Lemaire, Boulbaba Ben Amor, Mohsen Ardabilian (2011). Fully Automatic 3D Facial Expression Recognition using a Region-Based Approach. J-HGBU '11 Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding, ACM, pp. 53-58.
- [39] Amal Azazi, Syaheerah Lebai Lutfi, Ibrahim Venka, Fernando Fernández- Martínez (2015). Towards a robust affect recognition: Automatic facial expression recognition in 3D faces. Expert Systems with Applications. Elsevier. Vol. 42, pp. 3056–3066.
- [40] I. Kotsia and I. Pitas (2005). Real time facial expression recognition from image sequences using support vector machines. ICIP 2005. IEEE International Conference on Image Processing. Vol. 2, pp. II - 966-9.
- [41] Yongmian Zhang, Qiang Ji (2003). Facial Expression Understanding in Image Sequences Using Dynamic and Active Visual Information Fusion. IEEE International Conference on Computer Vision. Vol. 2, pp. 1297 - 1304.

Anexos

A. Variaciones del rostro con los que se realizó la experimentación

Las diferencias de forma y tamaño de los elementos faciales causa variaciones en el reconocimiento del rostro, por ello el sistema fue entrenado con sujetos de diversas características como: caras redondas, ovaladas; tamaños de ojos, ceja poblada y delgada, boca pequeña y grande. en la Figura A.1 se muestran algunos sujetos con características faciales diferentes, de diversas edades y sexos.

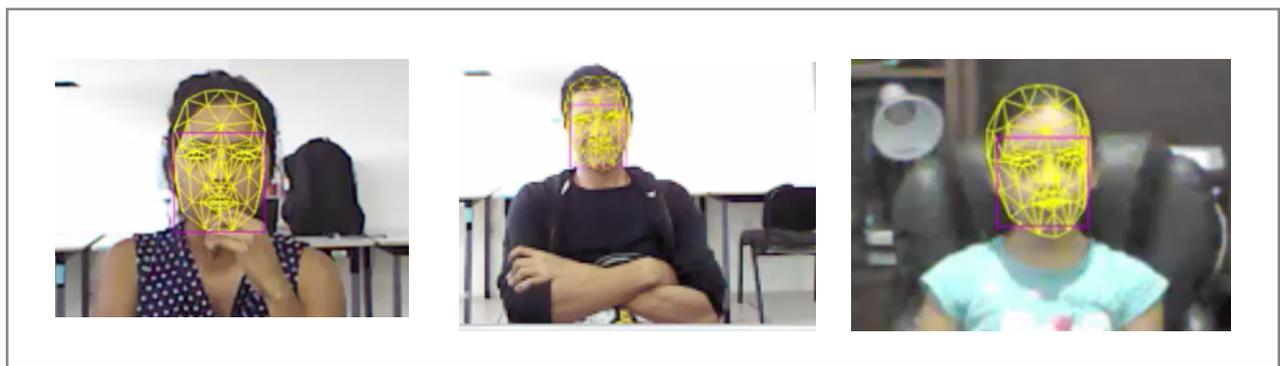


Figura A.1 Rostros para entrenamiento del sistema con variaciones de edades y sexo.

Las diferencias entre el tamaño de los elementos faciales también influye en la correcta clasificación e interpretación, la mayor diferencia entre los elementos se da entre personas de sexo opuesto, sin embargo, se puede apreciar en la Figura A.2 que estos sujetos tienen diferente tamaño de ojos, boca.

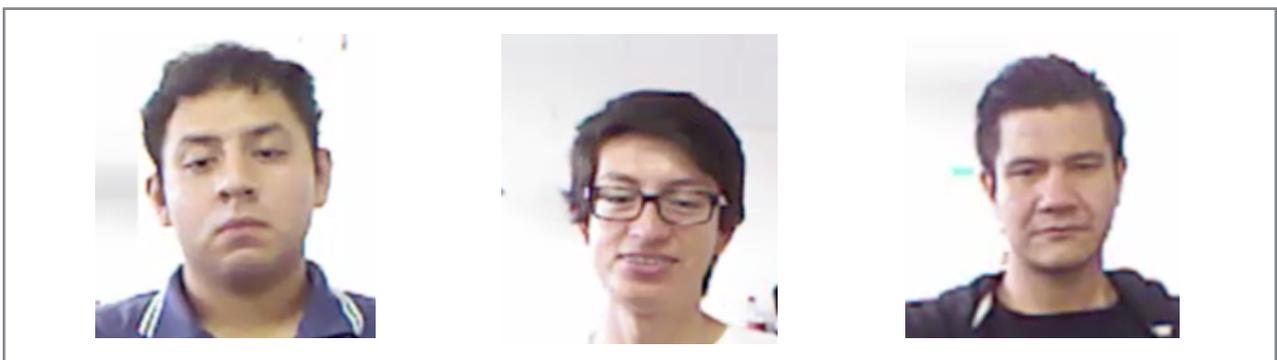


Figura A.2 Rostros con diversos tamaños de ojos y boca.

B. Video estímulo para accionar los elementos

Para verificar la correcta implementación de las reglas de decisión se utilizaron videos estímulo para accionar las ramas del árbol con base a los elementos faciales. En la Figura B.1 se muestran escenas del video utilizado para generar somnolencia y de esta manera probar la regla que establece los ojos cerrados como falta de atención. Se trata de un video aburrido con música de fondo tranquila y un tema poco atractivo ya que es conocido desde los estudios de primaria.

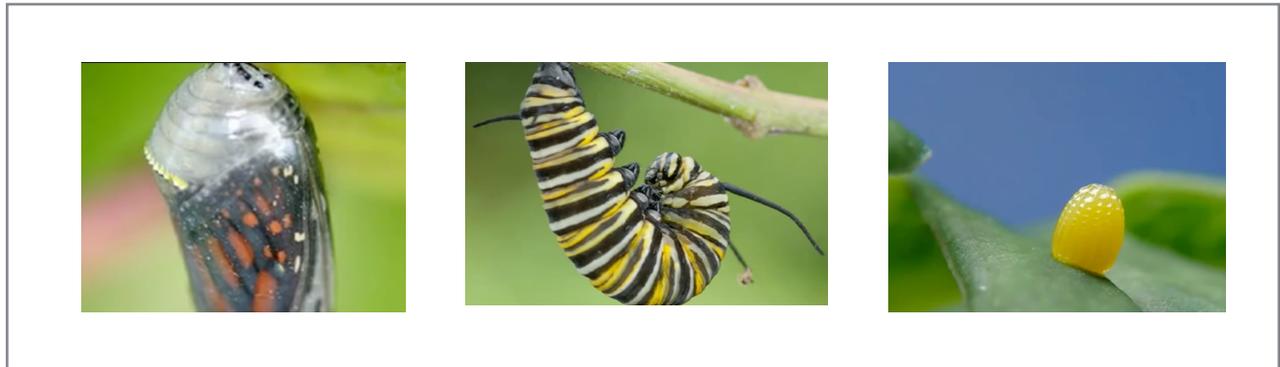


Figura B.1 Escenas del video utilizado para generar falta de motivación.

En la Figura B.2 se muestra el sujeto de prueba que visualizó el video mencionado en la Figura B.1. Como se puede apreciar el sujeto se a dormido por lo cual los ojos han sido evaluados cerrados.

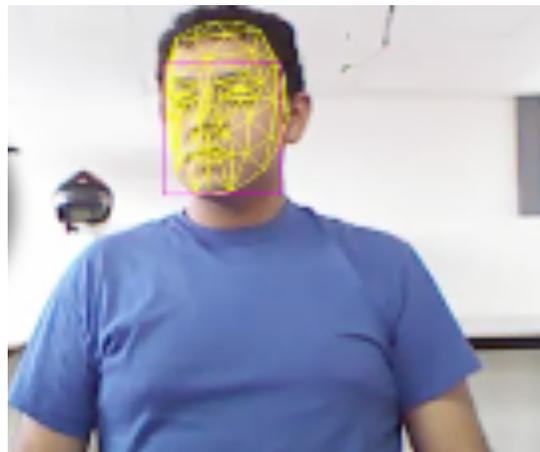


Figura B.2 Sujeto de prueba desmotivado.

En la Figura B31 se muestran escenas del video utilizado para generar atención y de esta manera probar la regla que establece que los elementos agudizados denotan motivación. Se trata de un video con un tema atractivo y de curiosidades sobre los vuelos comerciales.



Figura B.3 Escenas del video utilizado para generar motivación.

En la Figura B.4 se muestra el sujeto de prueba que visualizó el video mencionado en la Figura B.3. Como se puede apreciar el sujeto esta de frente, con ojos abiertos y demás sentidos agudizados.

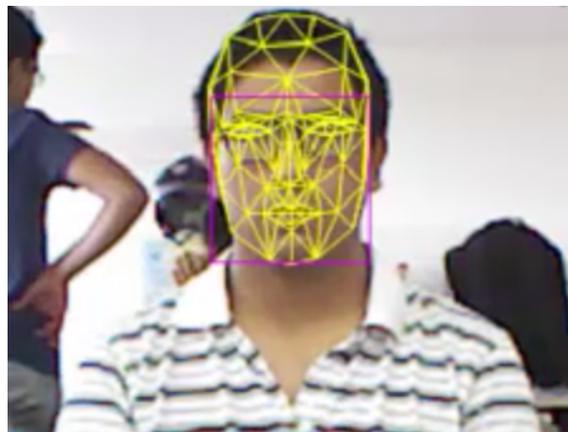
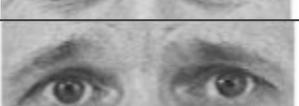
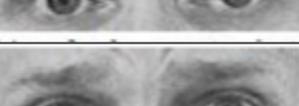


Figura B.4 Sujeto de prueba motivado.

C. AU's utilizadas

Tabla C.1. Unidades de acción utilizadas en el desarrollo de esta tesis.

Tabla C.1. 5 Unidades de acción para el reconocimiento de expresiones faciales básicas

AU	Descripción	Imagen
AU1	Elevación de la parte interna de las cejas	
AU2	Elevación de la parte externa de las cejas	
AU5	Elevación del párpado superior	
AU7	Elevación del párpado inferior	
AU1+4	Elevación de la parte interna de la ceja + generación de esfuerzo	
AU1+2	Elevación de la parte interna y externa de las cejas	
AU1+2+4	Las cejas son empujadas hacia arriba con esfuerzo	
Neutral	Relajación de ojos cejas y mejillas	
AU25	Abrir muy poco la boca.	
AU26	Abrir la boca.	
AU27	Abrir completamente la boca.	

D. Configuración

A continuación se describen brevemente las tecnologías hardware y software empleadas para el desarrollo de este trabajo.

Hardware

1. **Cámara RGBD:** Kinect el cual es un sensor RGBD capaz de crear un entorno visual 3D con los diferentes sensores ópticos, los cuales son:
 - Sensor infrarrojo – emisor (Fuente de Luz IR).
 - Sensor infrarrojo – receptor (Sensor CMOS).
 - Cámara RGB.
2. **Laptop:** Procesador dual 2.5 GHz - 2.6 GHz, 8 GB de RAM, procesador i3.

Software

1. **Sistema Operativo:** El sistema funciona en Windows 7, 8.
2. **Microsoft .NET Framework:** Es necesario tener instalada la versión 4.0 o superior para el correcto funcionamiento del SDK de Kinect.
3. **SDK de Kinect:** Se utilizó la versión 1.8 (funciona para la v1 del sensor Kinect) para obtener las librerías necesarias para el funcionamiento del sensor.
4. **Toolkit de Kinect:** Se utilizó la versión 1.8 para obtener ejemplos de aplicaciones que utilizan la librería *facetracking* para localización y seguimiento del rostro y sus componente (esta herramienta se encuentra dentro del paquete SDK de Kinect).
5. **Kinect Studio:** De igual forma se utilizó la versión 1.8, esta herramienta se utilizó para generar los videos, la aplicación captura la imagen RGB y la imagen de profundidad (esta herramienta se encuentra dentro del paquete SDK de Kinect).
6. **Visual Studio 2012:** En esta plataforma se desarrolló el sistema de reconocimiento de expresiones faciales.
7. **OpenCV:** Se utilizó la versión 2.9 para implementar el algoritmo de máquina de soporte vectorial.