

---

# **Centro Nacional de Investigación y Desarrollo Tecnológico**

**Subdirección Académica**

**Departamento de Ciencias Computacionales**

## **TESIS DE MAESTRÍA EN CIENCIAS**

**Caracterización de Expresiones Faciales con un Sensor RGBD**

presentada por  
**Ing. Rita Alejandra Morfin Villatoro**

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

Director de tesis  
**Dr. Raúl Pinto Elías**



Cuernavaca, Morelos a 26 de mayo del 2016  
OFICIO No. DCC/076/2016  
**Asunto:** Aceptación de documento de tesis

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

Por este conducto, los integrantes de Comité Tutorial de la **Ing. Rita Alejandra Morfin Villatoro**, con número de control M14CE016, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis profesional titulado "**Caracterización de expresiones faciales con un sensor RGBD**" 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.

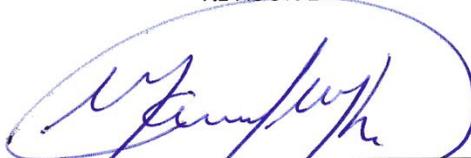
DIRECTOR DE TESIS



---

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

REVISOR 1



---

Dr. Manuel Mejía Lavalle  
Doctor en Ciencias Computacionales  
8342472

REVISOR 2



---

Dr. José Ruiz Ascencio  
Doctor en Ciencias  
5009035

REVISOR 3



---

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

C.p. Lic. Guadalupe Garrido Rivera - Jefa del Departamento de Servicios Escolares.  
Estudiante  
Expediente

AMR/lmz

Cuernavaca, Mor., 2 de junio de 2016  
OFICIO No. SAC/186/2016

**Asunto:** Autorización de impresión de tesis

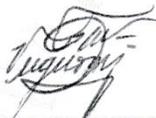
**C. RITA ALEJANDRA MORFÍN VILLATORO  
CANDIDATA AL GRADO DE MAESTRA EN CIENCIAS  
DE LA COMPUTACIÓN  
PRESENTE**

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado “**Caracterización de expresiones faciales con un sensor RGBD**”, 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 VICENTE GUERRERO RAMÍREZ  
SUBDIRECTOR ACADÉMICO**



**SEP TecNM  
CENTRO NACIONAL  
DE INVESTIGACIÓN  
Y DESARROLLO  
TECNOLÓGICO  
SUBDIRECCIÓN  
ACADÉMICA**

C.p. Lic. Guadalupe Garrido Rivera.- Jefa del Departamento de Servicios Escolares.  
Expediente

GVGR/mcr

## **Dedicatoria**

Le dedico este trabajo a:

*Mi familia y mi compañera Ali  
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 sus valiosos consejos y asesoría, los cuales contribuyeron a la terminación satisfactoria de este trabajo.

Agradezco a los revisores de mi trabajo: 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.

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 mentales, emociones y otros comportamientos que manifiestan las personas. La extracción y el seguimiento automático de componentes faciales en secuencias de imágenes son las principales tareas que deben resolver los sistemas de visión artificial enfocados al análisis del comportamiento humano, detección de expresiones faciales, interfaces hombre-máquina, entre otras áreas.

En este trabajo se propone utilizar una cámara RGBD para crear un sistema que localice automáticamente la cara y sus componentes en secuencias de imágenes en tiempo real. El rostro y los componentes faciales se detectaron mediante la utilización de la librería *facetracking*.

Se desarrolló un descriptor a partir de las unidades de acción de la librería *facetracking* para identificar las Unidades de Acción Facial (AUs - Action Unit) presentes con base en el Sistema de Codificación de Acciones Faciales (FACS) desarrollado por P. Ekman y W. Friesen.

Se utilizaron diferentes algoritmos de clasificación: árboles de decisión (J48 y RepTree), algoritmos que generan reglas (NNge y PART) y máquina de soporte vectorial (de *kernel* lineal, de base radial y sigmoide) para reconocer el descriptor de AU's, en clases correspondientes a expresiones faciales. Finalmente los resultados de los clasificadores con el vector fueron analizados y comparados.

En el estado del arte, los trabajos que obtuvieron el mejor resultado en reconocimiento, clasificando todas las expresiones faciales básicas con un sensor RGBD, fueron los siguientes: Li, 2011 [17] con un 88%, Stocchi, 2014 [8] con un 87% y Mao, 2015 [13] con un 83%; es importante mencionar que todos ellos utilizaron la Máquina de Soporte Vectorial. En esta tesis la Máquina de Soporte Vectorial de base radial obtuvo el mejor resultado con un 88.86% en la etapa de reconocimiento.

## **Abstract**

Facial expressions give nonverbal information about mental and emotional states, and other obvious human behavior. The automatic extraction and tracking of facial components in image sequences are the main tasks that Computer Vision Systems focused on analysis of human behavior, detection of facial expression, interfaces and human computer, among other fields, must solve.

This work intends to use a RGBD camera to create a system that automatically detect the face and its components in image sequences. The face and the facial components, were located by using the *facetracking* library.

A new descriptor was created to identify Facial Actions Units (AU's) based on the Unit of Actions of the *Facetracking* Library, according to the Facial Action Coding System (FACS) created by P. Ekman and W. Friesen.

A set of different algorithms of classification was used: Decision trees (J48 and RepTree), algorithms to generate rules (NNge and partial) and Support Vector Machine (linear, radial and sigmoid kernels) to assign the AU's descriptor to the corresponding facial expressions classes. Finally, the results of classifiers were analyzed and compared their performance.

In the state of art, the works that obtained the best result recognition and classifying all basic facial expressions with a RGBD sensor are: Li, 2011 [17] with 88%, Stocchi, 2014 [8] with 87% and Mao, 2015 [13] with 83%; it is important to mention that all of them used the Support Vector Machine. In this thesis Support Vector Machine radial basis it obtained the best result with 88.86% at the stage of recognition.

# Contenido

Lista de Figuras.....	iii
Lista de Tablas.....	v
Glosario.....	vi
<b>1 Introducción.....</b>	<b>1</b>
1.1 Descripción del problema.....	2
1.1.1 Descripción del problema general.....	3
1.1.2 Problema específico.....	4
1.2 Propuesta de solución.....	4
1.2.1 Objetivo general.....	4
1.2.2 Objetivos específicos.....	4
1.2.3 Alcances.....	4
1.2.4 Limitaciones.....	5
1.2.5 Justificación.....	5
1.3 Organización del documento.....	6
<b>2 Análisis del problema y propuesta de solución.....</b>	<b>7</b>
2.1 Localización y seguimiento del rostro.....	7
2.1.1 Detección del rostro.....	8
2.1.2 Extracción de características (componentes del rostro).....	10
2.2 Sistema de Codificación de Acciones Faciales (FACS).....	13
2.2.1 Unidades de Acción.....	15
2.3 Guión.....	17
2.3.1 Modelo del guión.....	17
2.4 Reconocimiento de expresiones faciales (clasificación).....	18
2.4.1 Algoritmos de clasificación.....	19
2.4.2 Características de los algoritmos de clasificación.....	19
2.4.3 Lista de clasificadores candidatos.....	23
2.5 Propuesta de solución.....	23
2.6 Discusión.....	24
<b>3 Análisis, diseño e implementación de la solución.....</b>	<b>25</b>
3.1 Análisis del sistema.....	25
3.2 Esquema del sistema.....	26
3.2.1 Etapa 1: Recopilación de los datos.....	26
3.2.1.1 Recopilar datos a partir de imágenes 3D.....	27
3.2.1.2 Recopilar datos a partir de imágenes 2D.....	28
3.2.2 Etapa 2: Librería facetracking.....	33
3.2.2.1 Detección de rostro y extracción de características.....	33
3.2.2.2 Unidades de acción.....	34
3.2.2.3 Descriptor-EF.....	35
3.2.3 Etapa 3: Algoritmos de Clasificación para las expresiones faciales.....	39
3.2.3.1 Algoritmos de clasificación implementados.....	39
3.2.4 Etapa 4: Historial de las expresiones faciales.....	41
3.2.4.1 Video (guión) con las expresiones faciales.....	41
3.2.4.2 Histórico con las expresiones faciales.....	42
3.3 Diseño e implementación del sistema.....	43
3.3.1 Descripción del ambiente de implementación.....	43
3.3.2 Diseño del sistema.....	44

3.3.3	<i>Implementación del sistema</i> .....	44
3.4	Discusión.....	45
<b>4</b>	<b>Experimentación y resultados</b> .....	<b>47</b>
4.1	Ambiente de pruebas .....	47
4.2	Plan de pruebas.....	48
4.2.1	<i>Prueba de iluminación con cámara RGBD</i> .....	48
4.2.2	<i>Prueba con imágenes 2D</i> .....	50
4.2.3	<i>Prueba a distintas distancia con respecto al sujeto de estudio</i> .....	52
4.2.4	<i>Prueba para obtener los valores de las 6 AU's de facetracking</i> .....	56
4.2.5	<i>Prueba para obtener las 15 AU's a partir de las 6 AU's de facetracking</i> .....	58
4.2.6	<i>Prueba de clasificación</i> .....	61
4.2.6.1	Prueba con MSV Lineal.....	62
4.2.6.2	Prueba con MSV Sigmoide .....	62
4.2.6.3	Prueba con MSV de base radial.....	63
4.2.6.4	Prueba con árbol de decisión J48.....	64
4.2.6.5	Prueba con árbol de decisión REPTree.....	65
4.2.6.6	Prueba con algoritmos de reglas NNge .....	66
4.2.6.7	Prueba con algoritmos de reglas PART.....	68
4.3	Tabla de matriz cruzada con las pruebas y objetivos.....	69
4.4	Discusión.....	70
<b>5</b>	<b>Conclusiones generales y trabajos futuros</b> .....	<b>74</b>
5.1	Objetivos alcanzados .....	74
5.2	Alcances.....	75
5.3	Aportaciones .....	76
5.4	Conclusiones.....	76
5.5	Trabajos futuros .....	77
	<b>Referencias</b> .....	<b>78</b>
	<b>Anexos</b> .....	<b>82</b>
	A. FACS: Unidades de acción .....	82
	B. Ejemplo de guión (estructura) .....	84
	C. Ejemplos de imágenes finales de pruebas .....	85
	D. Herramientas de desarrollo .....	88

## Lista de Figuras

Figura 1.1 Factores que influyen en las expresiones faciales [5] .....	1
Figura 1.2 Arquitectura general de un sistema de reconocimiento de expresiones faciales [5] .....	3
Figura 2.1 Músculos relacionados con AU's de la parte superior de la cara [26] .....	13
Figura 2.2 Ejemplos de AU's y combinaciones de la parte superior de la cara [26] ..	14
Figura 2.3 Ejemplos de AUs y combinaciones de la parte inferior de la cara [26].....	14
Figura 2.4 Secuencia de video con las escenas para generar expresiones [12].....	18
Figura 2.5 Secuencia de video con las escenas para generar todas expresiones básicas .....	18
Figura 3.1 Elementos que intervienen en el análisis del sistema .....	25
Figura 3.2 Arquitectura del sistema de reconocimiento de expresiones faciales (SREF) .....	26
Figura 3.3 Ejemplo del almacenamiento de las imágenes en la base de datos .....	27
Figura 3.4 Expresión de sorpresa en dos personas distintas .....	30
Figura 3.5 Expresión de enojo en dos personas distintas .....	31
Figura 3.6 Expresión de miedo en dos personas distintas .....	31
Figura 3.7 Expresión de tristeza en dos personas distintas .....	32
Figura 3.8 Expresión de alegría en dos formas distintas.....	32
Figura 3.9 Expresión de disgusto en dos personas distintas .....	33
Figura 3.10 Ejemplo del rango para la boca.....	38
Figura 3.11 Ejemplo del rango para las cejas .....	39
Figura 3.12 Pasos de la etapa 3.....	40
Figura 3.13 Aplicación Kinect Studio v1.8 .....	41
Figura 3.14 Ejemplo del video con el archivo historial.....	42
Figura 3.15 Diseño general del sistema .....	44
Figura 3.16 Diagrama de flujo del sistema SREF.....	45
Figura 4.1 Distintos focos de iluminación .....	48
Figura 4.2 Cuarto oscuro.....	49
Figura 4.3 Ejemplo de una iluminación estable .....	49
Figura 4.4 Ejemplo de un rostro con lentes .....	50

Figura 4.5 Imagen 2D en escala de grises con la malla <i>Candide</i> sobre el rostro.....	51
Figura 4.6 Imagen 2D a color con la malla <i>Candide</i> sobre el rostro.....	51
Figura 4.7 Imagen animada de una Figura humana.....	51
Figura 4.8 Captura del sujeto con el sensor a una distancia de: a) Distancia 60 cm b) Distancia 120 cm c) Distancia 200 cm .....	52
Figura 4.9 Promedio de las unidades de acción de la boca (AU0, AU1, AU2 y AU4) para las tres distancias.....	54
Figura 4.10 Valores máximos de las unidades de acción de la boca (AU0, AU1, AU2 y AU4) para las tres distancias.....	54
Figura 4.11 Valores mínimos de las unidades de acción de la boca (AU0, AU1, AU2 y AU4) para las tres distancias.....	54
Figura 4.12 Promedio de las unidades de acción de las cejas (AU3 y AU5) para las tres distancias .....	55
Figura 4.13 Valores máximos de las unidades de acción de las cejas (AU3 y AU5) para las tres distancias.....	55
Figura 4.14 Valores mínimos de las unidades de acción de las cejas (AU3 y AU5) para las tres distancias.....	55
Figura 4.15 Gráfica del promedio de todas las AU's a la misma distancia (50 cm) ...	56
Figura 4.16 Esta imagen representa las AU's 1, 2, 4 y 6. Las 15 primeras columnas representan las AU's y la última columna la expresión detectada.....	59
Figura 4.17 Esta imagen representa las AU's 10, 13, 15 y 16.....	59
Figura 4.18 Esta imagen representa las AUs 17, 20, 23 y 24 .....	60
Figura 4.19 Esta imagen representa las AUs 25, 26 y 27 .....	61
Figura 4.20 Modelo entrenado del árbol J48.....	64
Figura 4.21 Distribución del árbol REPTree .....	65
Figura 4.22 Reglas generadas por el algoritmo NNge .....	67
Figura 4.23 Reglas generadas con el algoritmo PART .....	68
Figura 4.24 Comparación entre los algoritmos de clasificación.....	70
Figura 4.25 MSV de base radial con los resultados de las etapas de entrenamiento y de reconocimiento .....	71
Figura 4.26 Ejemplo de un patrón incorrecto.....	71

## Lista de Tablas

Tabla 2.1 Funciones para la detección y extracción de características [23].....	7
Tabla 2.2 Técnicas para localizar el rostro .....	9
Tabla 2.3 Técnicas para localizar componentes del rostro .....	10
Tabla 2.4 Trabajos relacionados sobre unidades de acción .....	15
Tabla 2.5 AU's asociadas a las expresiones faciales utilizadas [24] .....	19
Tabla 2.6 Trabajos relacionados sobre los algoritmos de clasificación .....	20
Tabla 3.1 Lista de patrones (reglas) de Pantic & Rothkrantz [25] .....	29
Tabla 3.2 Lista de patrones para el SREF desarrollado en esta tesis .....	29
Tabla 3.3 Total de imágenes para cada expresión.....	30
Tabla 3.4 Funciones para la detección y extracción de características [23].....	33
Tabla 3.5 Rango de valores para las 6 AUs .....	35
Tabla 3.6 Distribución de las AUs de <i>facetracking</i> con respecto a las AUs de la literatura .....	36
Tabla 3.7 Un ejemplo del archivo con el historial de un sujeto de estudio .....	42
Tabla 3.8 Funciones del sistema SREF .....	45
Tabla 4.1 Ejemplo de los valores obtenidos de las 6 AU's .....	53
Tabla 4.2 Cómo funcionan las unidades de acción con la librería <i>facetracking</i> .....	57
Tabla 4.3 Matriz de confusión para MVS Lineal (etapa de reconocimiento) .....	62
Tabla 4.4 Matriz de confusión para la MVS Sigmoide (etapa de reconocimiento) ....	63
Tabla 4.5 Matriz de confusión para la MVS de base radial (etapa de reconocimiento) .....	63
Tabla 4.6 Matriz de confusión para el árbol J48 (etapa de reconocimiento) .....	64
Tabla 4.7 Matriz de confusión para el árbol REPTree (etapa de reconocimiento) ....	66
Tabla 4.8 Matriz de confusión del algoritmo NNge (etapa de reconocimiento) .....	68
Tabla 4.9 Matriz de confusión del algoritmo PART .....	69
Tabla 4.10 Matriz cruzada entre objetivos y pruebas realizadas.....	69
Tabla 4.11 Datos de Kinect del sujeto de la Figura 4.21 .....	72
Tabla 4.12 Datos erróneos para cada una de las expresiones faciales .....	72
Tabla 5.1 Resultados de los objetivos específicos .....	74
Tabla 5.2 Alcances de la tesis .....	75

## Glosario

+	Carácter para expresar que el rango de medida es un poco mayor
++	Carácter para expresar que el rango de medida es mayor que +
+++	Carácter para expresar que el rango de medida es mayor que ++
2D	2 dimensiones
3D	3 dimensiones
AU	Action Unit (Unidad de Acción Facial)
AU's	Action Units (Unidades de Acción Facial)
<i>Candide</i>	Malla parametrizada desarrollada específicamente para la codificación basada en el modelo de rostros humanos
CENIDET	Centro Nacional de Investigación y Desarrollo Tecnológico
Descriptor-EF	Descriptor de Expresiones Faciales con 15 unidades de acción
<i>Facetracking</i>	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)
<i>FPS</i>	<i>Frames Per Second</i> (Fotogramas por segundo)
Frame	Fotograma
<i>GPU</i>	<i>Graphics Processing Unit</i> (Unidad de Procesamiento Gráfico)
IR	Infra Rojo
<i>Kernel</i>	Núcleo (función discriminante) de la máquina de soporte vectorial
Kinect	Dispositivo de juego desarrollado por Microsoft
MSV	Máquina de Soporte Vectorial
OpenCV	Librería libre de visión artificial
RGB	<i>Red, Green, Blue</i> (Rojo, Verde, Azul)
RGBD	<i>Red, Green, Blue, Depth</i> (Rojo, Verde, Azul, Profundidad)
SDK	<i>Software Development Kit</i> (Kit de desarrollo de <i>software</i> )
SREF	Sistema de Reconocimiento de Expresiones Faciales
Toolkit	Kit de herramientas
Weka	Plataforma de <i>software</i> para el aprendizaje automático y la minería de datos escrito en Java y desarrollado en la Universidad de Waikato

# Capítulo 1

## 1 Introducción

Las expresiones faciales proporcionan información no verbal a cerca del comportamiento de las personas y son muchos los factores que influyen en su generación (ver Figura 1.1).

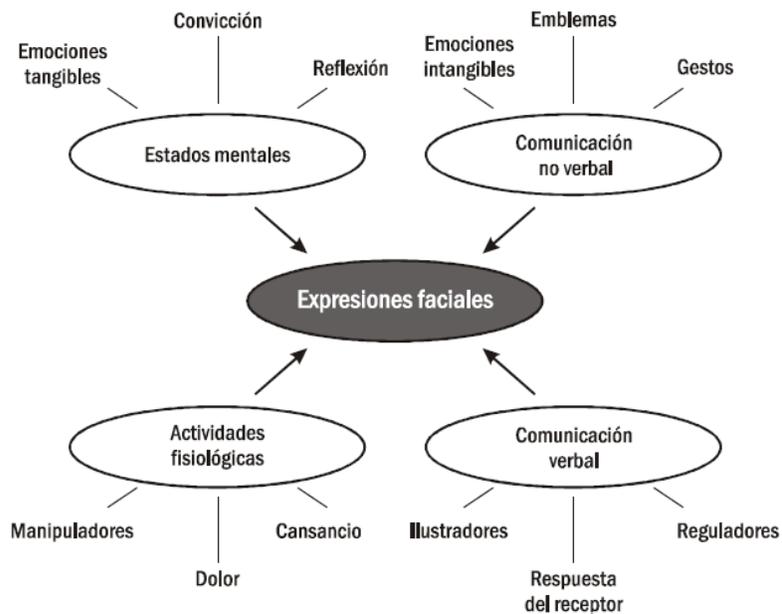


Figura 1.1 Factores que influyen en las expresiones faciales [5]

El reconocimiento automático mediante aplicaciones de computadora se ha empleado en la enseñanza del lenguaje de señas, en el desarrollo de interfaces hombre-máquina, en el estudio del comportamiento humano, ayudan a transmitir señales de comunicación no verbal, detección de enfermedades, facilitar instrumentos a los agentes para detectar expresiones faciales, como el miedo o la ira, inferir la conducta y así poder anticiparse a ella, por citar algunas áreas de aplicación.

Las primeras técnicas de reconocimiento facial se implementaron sobre imágenes 2D ya que anteriormente no se contaba con la tecnología suficiente para generar modelos 3D o cámaras RGBD; estas técnicas necesitan de condiciones controladas (iluminación, pose frontal y fondo que contraste con el rostro) para tener resultados [19].

Un problema con las técnicas 2D es que éstas son propensas a cambios de iluminación y plantean variaciones que afectan a la geometría percibida por la cámara y por lo tanto afecta al aspecto de los rasgos faciales. Para hacer frente a los problemas causados por las variaciones de pose, algunos investigadores propusieron el uso de múltiples vistas de la cara [1], crearon modelos 3D deformables montados en imágenes 2D [2] o en imágenes 3D [3], para así tener un ambiente menos controlado sobre el sujeto de estudio, porque en estos casos se obtiene más información sobre el rostro, debido a que se captura el rostro desde varios ángulos.

Con la llegada de la nueva tecnología de detección RGBD, ahora se puede conseguir la localización del rostro y sus componentes con la imagen de color 2D (RGB), junto con el mapa de profundidad (D) por medio de sensores de costo aceptable, como el Kinect de Microsoft®, softkinectik® y Leap Motion Controller®, entre otros. Una imagen RGBD proporciona información de profundidad y de color sobre la imagen capturada, a diferencia de una imagen RGB que sólo proporciona información de color. El mapa de profundidad proporciona información adicional por píxel, utilizando un proyector de láser infrarrojo combinado con una cámara.

En esta tesis se utilizaron los elementos que el sensor RGBD ofrece en conjunto con el conocimiento de las FACS (*Face Action Coding System*) desarrolladas por el psicólogo Paul Ekman [4], para reconocer las expresiones faciales básicas: felicidad, tristeza, miedo, enojo, sorpresa, desagrado (disgusto).

Se utiliza el sensor RGBD para capturar las imágenes del rostro; las unidades de acción se obtienen con la librería *facetracking*, ya que dicha librería obtiene información del rostro y sus componentes faciales; posteriormente es necesario utilizar una técnica de clasificación para finalmente obtener el resultado (la expresión facial). El sistema que se propone en este tema de tesis se podría utilizar como complemento para una entrevista.

El estímulo se le presenta al sujeto de estudio como un guión (secuencia de segmentos de video) acorde con los estímulos que se le pretenden hacer al sujeto de estudio; cada segmento de video está separado por una escena de transición neutral para que el sujeto de estudio vuelva a un estado neutral.

## **1.1 Descripción del problema**

En esta sección se describe el problema general y el problema específico de esta tesis.

En la Figura 1.2 se muestra la arquitectura general de un sistema para el reconocimiento de expresiones faciales, la cual comienza con la localización de la cara y termina en el reconocimiento de expresiones.

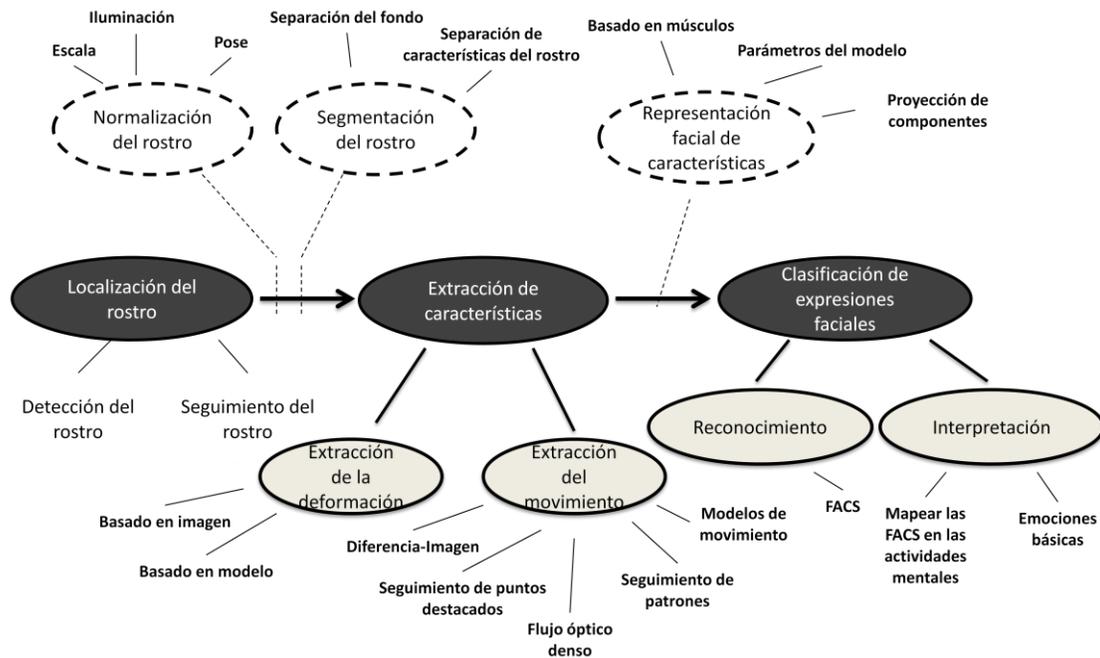


Figura 1.2 Arquitectura general de un sistema de reconocimiento de expresiones faciales [5]

### 1.1.1 Descripción del problema general

El reconocimiento de expresiones faciales tiene diferentes aplicaciones [6], [7] debido a que éstas transmiten emociones de forma visual. Para obtener el reconocimiento de expresiones faciales por medio de una máquina, son necesarias la localización de componentes en el rostro en la imagen de entrada, una base de conocimiento (Unidades de acción - FACS) y finalmente asociar (clasificar) la imagen de entrada a una expresión facial.

Para lograr la identificación de las expresiones faciales algunas soluciones encontradas en la literatura se basan en la utilización de un sensor RGBD, usan el SDK de Kinect (*Facetracking*) o la implementación de algoritmos externos, con condiciones controladas de iluminación, de pose y de gestos en el rostro (expresiones faciales exageradas), lo que permite la correcta o parcial detección de las expresiones faciales.

Los problemas detectados en los trabajos relacionados son:

- El porcentaje de clasificación de algunas expresiones (enojo, tristeza, miedo, sorpresa) están apenas por encima del 60% [7], [8].
- Si se usa la herramienta SDK de Kinect es necesario exagerar la expresión para obtener un resultado favorable [6], [8].
- La expresión disgusto no se pudo clasificar en los trabajos de [6], [8], los cuales usaron el SDK de Kinect.

Por lo tanto se propuso estudiar, evaluar, seleccionar e implementar las técnicas que mejor desempeño tengan según el estado del arte, bajo las siguientes restricciones:

- Utilizar Sensor Kinect RGBD.
- Usar el SDK de Kinect y el toolkit de Kinect.
- Obtener un reconocimiento de expresiones faciales básicas con resultados cercanos a tiempo real <sup>1</sup>aceptable.

### 1.1.2 Problema específico

Se detectaron varios problemas dentro del estado del arte mencionados en la sección anterior, en esta tesis se propone resolver el siguiente problema: identificar las seis expresiones faciales básicas, sin que el sujeto bajo estudio tenga que exagerar sus expresiones.

## 1.2 Propuesta de solución

Para el desarrollo de esta tesis se propuso alcanzar los siguientes objetivos:

### 1.2.1 Objetivo general

El objetivo primordial de esta propuesta es identificar las expresiones faciales básicas con un sensor RGBD, usando como base de conocimiento las Unidades de Acción de las FACS sin que el sujeto de estudio tenga la necesidad de exagerar la expresión.

### 1.2.2 Objetivos específicos

- Estudiar el SDK de Kinect para la generación de los datos asociados a las unidades de acción.
- Estudiar las FACS, para seleccionar las unidades de acción necesarias para cada expresión facial.
- Estudiar las técnicas para el reconocimiento de expresiones faciales.
- Seleccionar las técnicas que mejor desempeño tengan, según el estado del arte, para reconocer las expresiones faciales.
- Implementar las técnicas seleccionadas, probarlas, validarlas y afinarlas para este problema de investigación.

### 1.2.3 Alcances

- Realizar un procesamiento de secuencia de imágenes RGBD con las expresiones faciales básicas.
- Identificar en las imágenes las unidades de acción (AU26, AU20, AU15, AU10, AU4, AU2), estas son proporcionadas por la librería *Facetracking*.

---

<sup>1</sup> Se considera aceptable para este caso, cuando el proceso se realiza a 30 fps.

- Identificar además 15 unidades de acción para formar el descriptor-EF, con este descriptor se entrenan y prueban los algoritmos de clasificación para obtener las expresiones faciales.
- Se seleccionó una técnica para clasificar las expresiones faciales básicas con sus respectivas unidades de acción.
- Crear una base de datos del CENIDET (10 personas), también se utilizó una base de datos llamada Cohn-Kanade Extend [27] de la cual se seleccionaron 30 personas para complementar la base de datos.
- Detección de las expresiones faciales en tiempo real.

#### 1.2.4 Limitaciones

- Debe existir sólo un rostro en cada secuencia de video.
- No debe existir oclusión del rostro por accesorios lentes, barba, bigote y/o cabello.
- No se compromete que los algoritmos de clasificación den respuesta en tiempo real.
- Se utilizan las seis AU's (AU26, AU20, AU15, AU10, AU4, AU2) que proporciona Kinect.
- No se realiza una interpretación del estado de ánimo de las personas entrevistadas.

#### 1.2.5 Justificación

Las expresiones corporales, en particular las expresiones faciales proporcionan información no verbal útil en la comunicación humana, se pretende que esta información se obtenga con un sistema que utilice a Kinect y su respectivo SDK, además se complementa con algoritmos externos al toolkit de Kinect. De esta forma se tendría esa información no verbal asociada a la estímulo/respuesta establecida en el guión, ya que como se menciona antes, el sistema que se desarrolló en esta tesis se pensó en utilizarlo como complemento para una entrevista.

La identificación de expresiones faciales tiene otras aplicaciones las cuales abarcan generalmente el estudio del comportamiento humano, aplicaciones para estudio de mercado (herramienta CARA [9]), así como también aplicaciones para crear avatares, modelos 3D del rostro con expresiones, etc.

Kinect es una buena herramienta para el reconocimiento de las expresiones faciales dado que este sistema tiene una librería llamada *facetracking*; esta librería tiene las siguientes funcionalidades:

- Rastreo y seguimiento del rostro, por medio de una malla de puntos con información sobre los componentes principales del rostro (ojos, nariz, boca, cejas y mandíbula inferior).

- Unidades de acción, estas AU's están relacionadas con las cejas, la boca y la mandíbula inferior, para que posteriormente con una técnica de clasificación sea posible reconocer las expresiones faciales.

### **1.3 Organización del documento**

Capítulo 1. Comprenden los temas introductorios de esta tesis.

Capítulo 2. Se desarrolla un análisis del problema de tesis 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: FACS (unidades de acción), ejemplos del guión, imágenes finales, herramientas de desarrollo (*hardware* y *software*).

## Capítulo 2

### 2 Análisis del problema y propuesta de solución

En este capítulo se realiza un breve repaso a algunos conceptos utilizados en la implementación de los algoritmos seleccionados. Los conceptos están divididos en grupos: localización y seguimiento del rostro, FACS, guión 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 (ver Figura 1.2).

#### 2.1 Localización y seguimiento del rostro

En esta sección se realiza un breve repaso de la librería *facetracking* encargada de la localización y seguimiento del rostro, también se describen las técnicas utilizadas en el estado del arte para localizar y seguir el rostro.

En la Tabla 2.1 se describen las funciones básicas de la librería *facetracking* para la detección, seguimiento del rostro y sus componentes.

Tabla 2.1 Funciones para la detección y extracción de características [23]

Función	Descripción
<b><i>IFTFaceTracker</i></b> (tipo de dato devuelto: booleano)	Se encarga de la interfaz principal para inicializar la detección del rostro en forma síncrona mediante el paso de <i>frames</i> de color y profundidad. Verifica que tanto los <i>frames</i> de color como los de profundidad llegan desde el sensor Kinect.
<b><i>IFTImage</i></b> (tipo de dato devuelto: 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 a cerca de la imagen: el formato, la altura, la anchura, el tamaño de la imagen, y los bytes por píxel. Además, <i>IFTImage</i> 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.

Función	Descripción
<b><i>IFTResult</i></b> (tipo de dato, decimal)	Esta función está encargada de verificar que se haya detectado un rostro de forma exitosa, está encargada del seguimiento del rostro. <i>IFTResult</i> proporciona acceso a la siguiente información: - <i>GetFaceRect</i> - en el fotograma de vídeo ofrece información de las coordenadas del cuadro delimitador alrededor del rostro en seguimiento. - <i>Get2DShapePoints</i> - 2D (x, y) Coordenadas de los puntos clave en el rostro del fotograma de vídeo. Realiza un seguimiento de los 87 puntos en 2D.
<b><i>IFTModel</i></b> (tipo de dato, decimal)	Esta función crea la malla <i>Candide</i> sobre el rostro. La interfaz <i>IFTModel</i> 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: <i>GetSUCount</i> , <i>GetAUCount</i> - devuelven respectivamente el número de unidades de forma (SU) o unidades de acción (AU), utilizados en el modelo lineal 3D (malla <i>Candide</i> ). <i>GetTriangles</i> - devuelve triángulos del modelo de la malla 3D (índices de vértices). Cada triángulo tiene 3 índices de vértices que Figuran en el sentido horario. <i>GetVertexCount</i> - devuelve el número de vértices del modelo de la malla 3D

En la Tabla 2.1 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++.

### 2.1.1 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 2D para la localización del rostro y los que utilizaron una cámara RGBD. La Tabla 2.2 muestra de forma resumida las técnicas revisadas.

Tabla 2.2 Técnicas para localizar el rostro

Tipo de imagen o dispositivo de captura	Autor/ Referencia	Procesamiento para detectar el rostro	Pose del sujeto de estudio	Clasifica expresiones faciales
2D	Cho, 2009 [14]	Utiliza el algoritmo clasificador de características Haar	Vista frontal	No
	Paredes, 2009 [26]	Algoritmo que detecta la región más grande piel (rostro)	Vista frontal	Sí
Cámara RGBD	Ajmal, 2013 [9]	Nube de Puntos Procesamiento canónico (Frontal) Forma y Textura de Rostro	Variable	No
	Seddik, 2013 [7]	Nube de Puntos	Vista frontal	Sí
	Zhang, 2013 [10]	Obtiene manualmente el rostro a partir de la eliminación del fondo de la escena	Vista frontal	No
	Zhang, 2013 [11]	Obtiene automáticamente el rostro a partir de la eliminación del fondo de la escena	Variable	No
	Kakarla, 2014 [6]	Herramientas de SDK de Kinect: <i>FaceTracking</i>	Vista frontal	Sí
	Stocchi, 2014 [8]	Herramientas de SDK de Kinect: <i>FaceTracking</i>	Vista frontal	Sí
	Stocchi, 2014 [8]	Localización de la nariz. Esfera alrededor de ella para obtener rostro Escala de grises Momentos de Zernick	Vista frontal	Sí

Tipo de imagen o dispositivo de captura	Autor/ Referencia	Procesamiento para detectar el rostro	Pose del sujeto de estudio	Clasifica expresiones faciales
Cámara RGBD	Ruiz, 2014 [12]	Herramientas de SDK de Kinect: <i>FaceTracking</i>	Vista frontal	Sí
	Mao, 2015 [13]	Herramientas de SDK de Kinect: <i>FaceTracking</i>	Variable	Sí

En la Tabla 2.2 se describen los trabajos relacionados referentes a las técnicas para detectar el rostro, en la mayoría de los trabajos utilizaron una cámara RGBD como herramienta para obtener las imágenes, solo el autor Cho [14] utilizó imágenes 2D como imagen de entrada; es importante mencionar que existen otros trabajos cuya finalidad es reconocer expresiones faciales, pero en estos trabajos no fue necesario detectar el rostro, ya que usaron bases de datos con imágenes donde el rostro ya está segmentado, usaron estos datos solamente para detectar los componentes del rostro.

El proceso para detectar el rostro en casi todos los trabajos de la Tabla 2.2 se realiza fuera de línea con excepción de [6], [8], [11], [12] y [13], ellos realizan el proceso en tiempo real.

### 2.1.2 Extracción de características (componentes del rostro)

Esta es la etapa posterior a la detección del rostro y es muy importante detectar los componentes del rostro con precisión, para detectar en ellos las unidades de acción correspondientes. Los trabajos revisados se clasificaron en tres categorías: los que procesaron imágenes 2D para la localización del rostro, procesamiento con imágenes 3D y los que utilizaron una cámara RGBD. La Tabla 2.3 muestra de forma resumida las técnicas revisadas.

Tabla 2.3 Técnicas para localizar componentes del rostro

Tipo de imagen o dispositivo de captura	Autor / Referencia	Procesamiento para detectar los componentes del rostro	Componentes del rostro detectados	Comentarios
2D	Pantic, 2000 [22]	Detectar el contorno de los componentes faciales	Cejas, ojos, nariz y boca, mentón	Vista frontal y lateral del rostro

Tipo de imagen o dispositivo de captura	Autor / Referencia	Procesamiento para detectar los componentes del rostro	Componentes del rostro detectados	Comentarios
	Paredes, 2009 [26]	Utiliza detección de bordes, umbralizado y operadores morfológicos para detectar cejas y boca	Cejas, ojos y boca	Realiza un proceso extra para ajustar el contraste de la imagen y obtener un mejor resultado
3D	Lemaire, 2011 [16]	Puntos destacados (ICP - <i>Iterative Closest Points</i> )	Ojos, cejas, espacio entre las cejas, surcos alrededor de la boca y boca	Para generar el modelo entrenado, colocan manualmente los puntos destacados, posteriormente se localizan los puntos destacados dentro de cada región de del rostro
	Azazi, 2015 [19]	Puntos destacados (SURF)	Boca (marcas alrededor de la boca), cejas y ojos	A partir de puntos preestablecidos (boca y ojos) al generar movimiento en estas partes obtienen otros puntos en la zona alrededor de la boca y en las cejas
RGBD	Gong, 2009 [15]	Algoritmo de reconocimiento de forma facial básica (BFSC)	Cejas, ojos y boca	Extraen puntos característicos del rostro. Sólo usan la imagen de profundidad

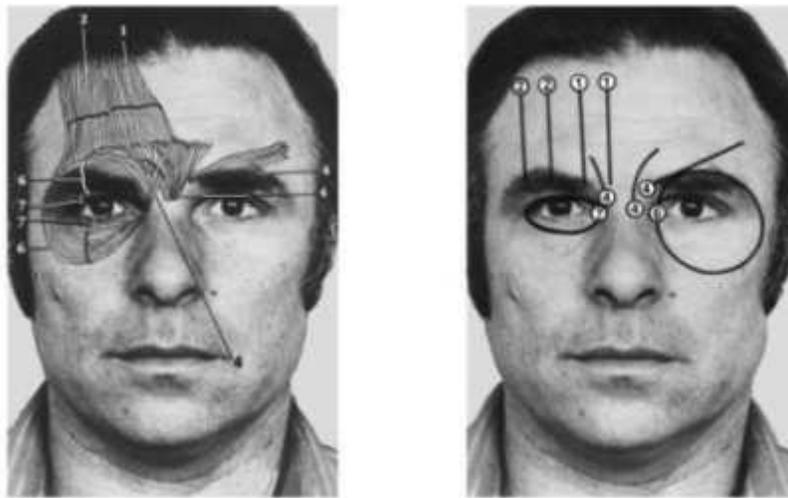
Tipo de imagen o dispositivo de captura	Autor / Referencia	Procesamiento para detectar los componentes del rostro	Componentes del rostro detectados	Comentarios
	Li, 2011 [17]	Algoritmo de detección de regiones (detecta la región de los ojos y la boca)	Cejas, ojos y boca	Utiliza la imagen de profundidad para detectar las regiones (cejas, ojos y boca)
	Vetros, 2011 [18]	Momentos de Zernike	Cejas, ojos y boca	Utiliza la imagen de profundidad Aplican un pre-proceso a la imagen de profundidad (ecualización del histograma adaptativo)
	Seddik, 2013 [7]	Algoritmos EigenFaces	Frente, cejas, ojos, nariz, boca y mentón	El algoritmo se aplica sobre la imagen de profundidad
	Kakarla, 2014 [6]	Herramientas de SDK de Kinect: <i>FaceTracking</i>	Cejas, boca y mandíbula inferior	Proceso en tiempo real
	Stocchi, 2014 [8]	Herramientas de SDK de Kinect: <i>FaceTracking</i>	Cejas, boca y mentón	Proceso en tiempo real
	Stocchi, 2014 [8]	Momentos de Zernike	Cejas, ojos y boca	Sólo utiliza la imagen de profundidad
	Ruiz, 2014 [12]	Herramientas de SDK de Kinect: <i>FaceTracking</i>	Cejas, boca y mentón	Proceso en tiempo real
	Mao, 2015 [13]	Herramientas de SDK de Kinect: <i>FaceTracking</i>	Cejas, ojos, nariz, boca y mentón	Proceso en tiempo real

Todos los trabajos presentados en la Tabla 2.3 tienen como finalidad encontrar expresiones faciales. También en la Tabla 2.3 las aplicaciones que utilizaron la librería *facetracking* realizan el proceso en tiempo real.

El sistema implementado utiliza el SDK de Kinect, específicamente la librería *facetracking* para detección y seguimiento del rostro.

## 2.2 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 [20]. 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. En la imagen etiquetada como “anatomía muscular” se observa la localización de los músculos y en la imagen “acción muscular” se indica la posición y dirección de acción de cada músculo [21]. De esta manera se etiquetan los movimientos en AU's.



**Anatomía muscular**

**Acción muscular**

Figura 2.1 Músculos relacionados con AU's de la parte superior de la cara [26]

El sistema FACS está compuesto por 46 AU's, de las cuales 30 están directamente relacionadas con la contracción de músculos específicos, otras corresponden a la aparición de líneas de expresión en regiones, por ejemplo en la parte superior de la nariz y en la región central de la frente. De estas AU's, 12 corresponden a la parte superior del rostro y 18 a la parte inferior. Las AU's pueden ocurrir de manera aislada o en combinación. En la Figura 2.2 se muestran algunas AU's y combinaciones de AU's relacionadas con la parte superior del rostro y en la Figura 2.3 AU's de la parte inferior [21].

<p><b>AU1</b></p>  <p>Elevación de parte interna de las cejas.</p>	<p><b>AU2</b></p>  <p>Elevación de parte externa de las cejas.</p>	<p><b>AU4</b></p>  <p>Disminución de altura y acercamiento de las cejas.</p>
<p><b>AU5</b></p>  <p>Elevación de párpado superior.</p>	<p><b>AU6</b></p>  <p>Elevación de mejillas.</p>	<p><b>AU7</b></p>  <p>Elevación de párpado inferior.</p>
<p><b>AU1+4</b></p>  <p>Elevación de parte interna de la cejas + generación de esfuerzo.</p>	<p><b>AU4+5</b></p>  <p>Disminución de altura y acercamiento de las cejas + elevación de párpado superior.</p>	<p><b>AU1+2</b></p>  <p>Elevación de la parte interna y externa de las cejas.</p>
<p><b>AU1+2+4</b></p>  <p>Las cejas son empujadas hacia arriba con esfuerzo.</p>	<p><b>AU1+2+5+6+7</b></p>  <p>Elevación de cejas, párpados y mejillas.</p>	<p><b>Neutral</b></p>  <p>Relajación de ojos cejas y mejillas.</p>

Figura 2.2 Ejemplos de AU's y combinaciones de la parte superior de la cara [26]

<p><b>AU9</b></p>  <p>Elevación del centro de labio superior. Presencia de arrugas en la parte superior de la nariz.</p>	<p><b>AU10</b></p>  <p>Elevación de labio superior. No hay presencia de arrugas en la parte superior de la nariz.</p>	<p><b>AU20</b></p>  <p>Movimiento hacia abajo y en forma lateral de los labios y la parte baja de la línea trazada entre la parte externa de nariz y comisuras. Aumento de la elongación de la boca.</p>
<p><b>AU15</b></p>  <p>Movimiento de las esquinas de la boca hacia abajo.</p>	<p><b>AU17</b></p>  <p>Elevación de región de la barbilla.</p>	<p><b>AU12</b></p>  <p>Separación de las esquinas de los labios en forma oblicua.</p>

Figura 2.3 Ejemplos de AUs y combinaciones de la parte inferior de la cara [26]

La librería *Facetracking* sólo maneja seis AUs relacionadas con las cejas, la boca y el mentón (la descripción de cada una de las seis AUs se encuentra en la tabla 3.5), a partir de estas seis AUs se generaron 15 AUs para crear un descriptor (la descripción de cada una de las 15 AUs se encuentra en la tabla 3.6). El descriptor con las 15 AUs sirve para clasificar las expresiones faciales.

### 2.2.1 Unidades de Acción

En esta sección se mencionan el estado del arte de los trabajos relacionados con las unidades de acción que se utilizaron para clasificar las expresiones faciales. Los trabajos revisados se clasificaron en tres categorías: los que procesaron imágenes 2D para la localización del rostro, procesamiento con imágenes 3D y los que utilizaron una cámara RGBD. La Tabla 2.4 muestra de forma resumida las técnicas revisadas.

Tabla 2.4 Trabajos relacionados sobre unidades de acción

Tipo de imagen o dispositivo de captura	Autor / Referencia	Unidades de acción utilizadas	Componentes del rostro detectados	Expresiones faciales	Detección en tiempo real
2D	Pantic, 2000 [22]	17 AU's: 1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25 y 26	Ojos, cejas, nariz, mentón y boca	Todas las expresiones básicas	No
	Paredes, 2009 [26]	17 AU's: 1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25 y 26	Ojos, cejas y boca	Todas las expresiones básicas	No
	Zhang, 2003 [24]	18 AU's : 1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25, 26 y 27	Ojos, cejas, nariz, mentón y boca	Todas las expresiones básicas	No
	Kotsia, 2005 [25]	17 AU's: 1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25 y 26	Ojos, cejas, nariz, mentón y boca	Todas las expresiones básicas	Sí

Tipo de imagen o dispositivo de captura	Autor / Referencia	Unidades de acción utilizadas	Componentes del rostro detectados	Expresiones faciales	Detección en tiempo real
3D	Lemaire, 2011 [16]	No hay información	Ojos, cejas, espacio entre las cejas, surcos alrededor de la boca y boca	Todas las expresiones básicas	No hay información
	Azazi, 2015 [19]	No hay información	Boca (marcas alrededor de la boca), cejas y ojos	Todas las expresiones básicas + neutral.	No
RGBD	Gong, 2009 [15]	No hay información	Cejas, ojos y boca	Todas las expresiones básicas	No hay información
	Li, 2011 [17]	No hay información	Cejas, ojos y boca	Todas las expresiones básicas	No
	Vetros, 2011 [18]	No hay información	Cejas, ojos y boca	Todas las expresiones básicas	No
	Seddik, 2013 [7]	No hay información	Frente, cejas, ojos, nariz, boca y mandíbula inferior	Todas las expresiones básicas + neutral	No
	Kakarla, 2014 [6]	6 AU's de la librería <i>facetracking</i>	Cejas, boca y mandíbula inferior	Expresiones básicas excepto disgusto	Sí
	Stocchi, 2014 [8]	6 AU's de la librería <i>facetracking</i>	Cejas, boca y mandíbula inferior	Expresiones básicas excepto disgusto	Sí
	Stocchi, 2014 [8]	No hay información	Cejas, ojos y boca	Expresiones básicas excepto disgusto	No

Tipo de imagen o dispositivo de captura	Autor / Referencia	Unidades de acción utilizadas	Componentes del rostro detectados	Expresiones faciales	Detección en tiempo real
	Ruiz, 2014 [12]	No hay información	Cejas, boca y mandíbula inferior	Alegría, tristeza, disgusto, miedo y neutral	Sí
	Mao, 2015 [13]	6 AU's de la librería <i>facetracking</i>	Cejas, ojos, nariz, boca y mandíbula inferior	Todas las expresiones básicas + neutral	Sí

En la Tabla 2.4 los autores que utilizaron un dispositivo RGBD utilizaron las unidades de acción de *facetracking*, hay autores que no especifican si usaron Kinect y/o *facetracking*. Algunos autores no mencionan qué unidades de acción utilizaron para clasificar las expresiones faciales.

Para el desarrollo del sistema de esta tesis se utiliza la librería *facetracking* para obtener las unidades de acción.

## 2.3 Guión

El guión se diseñó con la idea de recabar información facial de la forma más completa y fiable posible. Para ello se decidió crear una secuencia de vídeo especialmente diseñada para generar emociones en las personas sometidas a estudio de una forma controlada y diferenciada.

### 2.3.1 Modelo del guión

El vídeo (guión) proyectado a las personas grabadas está compuesto por seis secuencias diseñadas para generar diferentes emociones. Algunas escenas (escenas que generan desagrado y miedo) han sido tomadas del trabajo realizado por el Departamento de Psicología de la Universidad Pontificia [12] de Comillas por haber obtenido resultados positivos en estudios realizados anteriormente. Las otras escenas se descargaron de la web en base a sentido común (la descripción de las escenas se encuentran en el anexo B).

El video está basado en el modelo que utilizó Ruiz [12] para el desarrollo de su tesis, el cual consta de las siguientes partes (Figura 2.4):

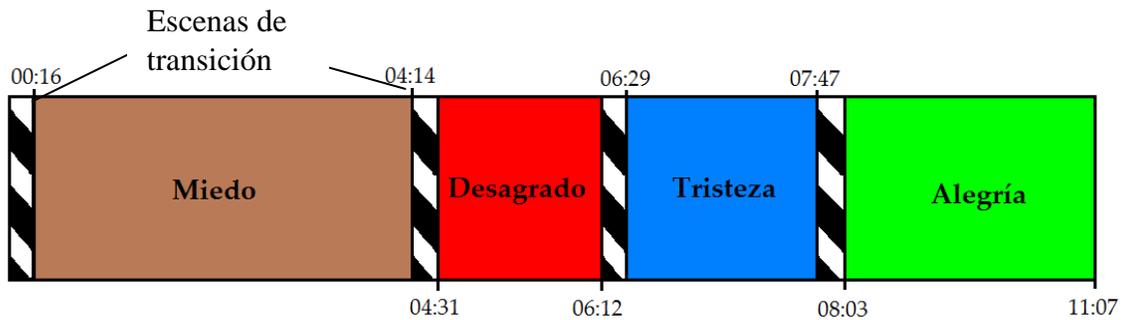


Figura 2.4 Secuencia de video con las escenas para generar expresiones [12]

El autor sólo estudió cuatro expresiones faciales (miedo, desagrado, tristeza y alegría), para las otras expresiones faltantes (enojo y sorpresa) se buscó videos en internet para propiciar esa reacción.

Las secuencias se encuentran separadas mediante breves escenas de transición para establecer una clara diferenciación entre ellas y evitar que las emociones se solapen. La duración total del vídeo es de 6 minutos y 55 segundos, tiempo suficiente para poder captar una gran cantidad de información sobre el rostro de una persona, pero lo suficientemente breve para mantener su atención.

El video para el desarrollo de esta tesis, está distribuido de la siguiente manera (Figura 2.5):



Figura 2.5 Secuencia de video con las escenas para generar todas expresiones básicas

Cada secuencia de video está separada por breves escenas de transición de 20 segundos. El video consta de estímulos audiovisuales.

## 2.4 Reconocimiento de expresiones faciales (clasificación)

La clasificación es la última parte de la metodología de solución; en esta sección se describe la función de los algoritmos de clasificación.

### 2.4.1 Algoritmos de clasificación

Los algoritmos de clasificación se encuentran en la última parte del sistema para reconocer expresiones faciales. Para clasificar las imágenes en expresiones faciales, es necesario entrenar algún clasificador. El modelo obtenido con el entrenamiento recibirá un vector de características arbitrario y regresará la clase a la que pertenece.

La relación que existe entre las AU's y las expresiones se presenta en la Tabla 2.5 basada en [24].

Tabla 2.5 AU's asociadas a las expresiones faciales utilizadas [24]

Expresión facial	Características primarias					Características auxiliares				
	AU(s)	AU(s)	AU(s)	AU(s)	AU(s)	AU(s)	AU(s)	AU(s)	AU(s)	AU(s)
Alegría	12					25	26	16		
Tristeza	1	15				4	7	25	26	
Enojo	2	4	7	23	24	16	25	26		
Miedo	20	1+5	5+7			4	5	7	25	26
Sorpresa	5	26	27	1+2						

La Tabla 2.5 de relaciones entre AU's y expresiones faciales no se puede aplicar como un conjunto de reglas para clasificar las imágenes de acuerdo a los vectores basados en FACS por la siguiente razón: el sistema puede generar como salida una combinación de AU's distinta a las expuestas en la Tabla 2.5, por lo que debe existir alguna solución para clasificar la imagen en la clase con la que tenga mayor similitud. Los algoritmos de aprendizaje automático proporcionan esta característica, asociar o clasificar cierto patrón a la clase que más se le parezca, creando una línea de frontera a partir de una función discriminante para separar dichas clases.

### 2.4.2 Características de los algoritmos de clasificación

En esta sección se mencionan los trabajos relacionados con los algoritmos de clasificación que se utilizaron para reconocer las expresiones faciales. Los trabajos revisados se organizaron en tres categorías: los que procesaron imágenes 2D para la localización del rostro, procesamiento con imágenes 3D y los que utilizaron una cámara RGBD. La Tabla 2.6 muestran de forma resumida las técnicas revisadas.

Tabla 2.6 Trabajos relacionados sobre los algoritmos de clasificación

<b>Tipo de imagen o dispositivo de captura</b>	<b>Autor / Referencia</b>	<b>Algoritmo de clasificación</b>	<b>Características</b>	<b>Porcentaje de reconocimiento</b>
2D	Pantic, 2000 [22]	Sistema Experto	-Sistema experto basado en reglas (Tabla de reglas) -Elimina información ambigua	90%
	Paredes, 2009 [26]	C5.0	-Se basa en árboles de decisión para crear reglas tipo IF-THEN que evalúan los parámetros o variables del vector característico para clasificar el vector en alguna clase	90%
	Zhang, 2003 [24]	Red Bayesiana Dinámica	- El reconocimiento de expresiones faciales se logra mediante la fusión no sólo de las observaciones visuales actuales, sino también de las evidencias visuales anteriores -El reconocimiento se hace más robusto y preciso a través de la modelización del comportamiento temporal de las expresiones faciales -Elimina información ambigua	Sin información
	Kotsia, 2005 [25]	Máquina de Soporte Vectorial	-Algoritmo de aprendizaje -Utilizan kernel de base radial y polinomial	Base radial (93%) Polinomial (97%)

<b>Tipo de imagen o dispositivo de captura</b>	<b>Autor / Referencia</b>	<b>Algoritmo de clasificación</b>	<b>Características</b>	<b>Porcentaje de reconocimiento</b>
3D	Lemaire, 2011 [16]	Máquina de Soporte Vectorial	-Algoritmo de aprendizaje -Utilizan kernel lineal	75%
	Azazi, 2015 [19]	Máquina de Soporte Vectorial	-Se introduce la integración las AUs con las estimaciones de probabilidad con la MSV para maximizar la variación entre-clase	84%
RGBD	Gong, 2009 [15]	Máquina de Soporte Vectorial	-Algoritmo de aprendizaje -No hay más información	75%
	Li, 2011 [17]	Máquina de Soporte Vectorial	-Kernel lineal, de base radial y sigmoide	Lineal (81%) Base radial (88%) Sigmoide (84%)
	Vetros, 2011 [18]	Máquina de Soporte Vectorial	-No especifica cuál fue el kernel utilizado	73%
	Kakarla, 2014 [6]	Condicionales (umbrales)	-El autor crea una condición para cada expresión facial	89% (no incluye disgusto)

Tipo de imagen o dispositivo de captura	Autor / Referencia	Algoritmo de clasificación	Características	Porcentaje de reconocimiento
RGBD	Seddik, 2013 [7]	<i>EigenFaces</i> .	El paradigma Eigenface supone que cualquier rostro deseado $x$ puede ser reconstruido a partir de la imagen media $m$ mediante la adición de una serie de detalles. Se crea un conjunto de vectores A continuación, los vectores propios se obtienen utilizando la matriz de covarianza y se aplica un análisis de componentes principales con el fin de producir la imagen-media similar. Cualquier nuevo rostro se compara entonces con él utilizando el cálculo de distancias	64%
	Stocchi, 2014 [8]	Máquina de Soporte Vectorial	No hay información	87%
	Ruiz, 2014 [12]	Arboles de decisión	No especifica	Solo existe información de alegría (97%) y disgusto (96%)
	Mao, 2015 [13]	Máquina de Soporte Vectorial	No hay información	83%

En la Tabla 2.6 algunos de los autores [6], [8] y [12] no clasificaron todas las expresiones faciales básicas y utilizaron la librería *facetracking* (6 AU's). Los otros autores sí detectaron todas las expresiones faciales básicas.

También se observa en Tabla 2.6 que la mayoría de los autores usaron la máquina de soporte vectorial (MSV) para clasificar las expresiones faciales. También se observa que los autores que no utilizaron MSV usaron algoritmos de decisión o algoritmos que generan reglas. Es por ello que se utilizaron estos tres tipos de algoritmos para implementarlos en el sistema propuesto en esta tesis.

### 2.4.3 Lista de clasificadores candidatos

Se identificó en el estado del arte que los algoritmos utilizados para clasificar las expresiones faciales son: algoritmos de decisión, algoritmos que generan reglas (Sistemas expertos y reglas tipos IF-THEN) y la mayoría de los trabajos utilizaron máquina de soporte vectorial con diferentes *kernels*. Por lo tanto en este trabajo se implementaron los siguientes algoritmos de clasificación:

- Máquina de Soporte Vectorial (MSV) [33]:
  - Lineal
  - Sigmoide
  - De base radial
- Árboles de decisión [34]:
  - J48
  - REPTree
- Algoritmos de reglas [34]:
  - NNge
  - Part

Cabe mencionar que se utilizaron esos árboles de decisión y algoritmos de reglas ya que obtuvieron un valor mayor al 98% de instancias correctamente clasificadas en la aplicación Weka [34].

La máquina de soporte vectorial con los diferentes *kernels* se implementaron usando Opencv [33].

## 2.5 Propuesta de solución

La propuesta de solución para resolver el problema de detección de expresiones faciales con una cámara RGBD es la siguiente:

- Utilizar la librería *facetracking* para detección y seguimiento del rostro y sus componentes.
- Utilizar las 6 unidades de acción de *facetracking* para generar un descriptor (descriptor-EF) con 15 unidades de acción que se activan (1) siempre y cuando estén presentes en el rostro del sujeto de estudio, de lo contrario permanecerán inactivas (0).

- Compilar una base de datos con las expresiones faciales básicas más la expresión neutral; el objetivo de esta base de datos es crear un conjunto de patrones para entrenar y validar los algoritmos de clasificación implementados, esta base de datos consiste en dos partes:
  1. Base de datos a partir del guión (10 sujetos).
  2. Base de datos Kohn-Canade Extend (30 sujetos).
- Implementar los algoritmos de clasificación para obtener las expresiones faciales.
- Finalmente el sistema genera un histórico con la captura (video) de la sesión de grabación y un archivo que contiene las expresiones faciales obtenidas de cada *frame* de la grabación.

## 2.6 Discusión

Dentro del estado del arte se detectaron algunos trabajos Kakarla [6], Stocchi [8] y Mao [13] en donde es necesario exagerar la expresión, esto quiere decir que los valores deben estar lo más cercano de -1 o 1 para detectarse en cada AU y ellos utilizaron la librería *facetracking*. También se encontró que los autores [6] y [8] no lograron clasificar disgusto.

El sistema que se implementó en esta tesis puede detectar las seis expresiones faciales sin exagerar, es decir, no es necesario llegar a valores cercanos a -1 o 1 para cada AU y reconocer una expresión facial. Finalmente el sistema puede detectar expresiones con o sin exagerar, ya que se crearon más rangos y por lo tanto más unidades de acción (la Tabla 3.6 se muestra la organización de dichos rangos).

## Capítulo 3

### 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. Este proceso incluye la recopilación de los datos (imágenes de rostros), la implementación de las funciones de *facetracking* para el reconocimiento y seguimiento de rostro y sus componentes, el desarrollo de un descriptor nombrado descriptor-EF (descriptor de expresiones faciales) para reconocer las expresiones faciales y la implementación de algoritmos de clasificación.

#### 3.1 Análisis del sistema

El análisis del sistema se define con base en los requerimientos necesarios para cumplir con el objetivo planteado en esta tesis. 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 análisis intervienen los siguientes elementos: la imágenes de rostros, librería *facetracking* y los algoritmos de clasificación, estos elemento se relacionan entre sí 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 que intervienen en el análisis del sistema

### 3.2 Esquema del sistema

El esquema que se presenta en la Figura 3.2 muestra la descripción general del sistema, en éste se definen sus componentes, los procesos y las relaciones entre éstos. El esquema propuesto se divide en cuatro etapas: etapa uno, recopilación de los datos (imágenes del rostro); etapa dos, implementación de la librería *facetracking*, etapa tres, implementación de algoritmos de clasificación y etapa cuatro, historial de expresiones faciales.

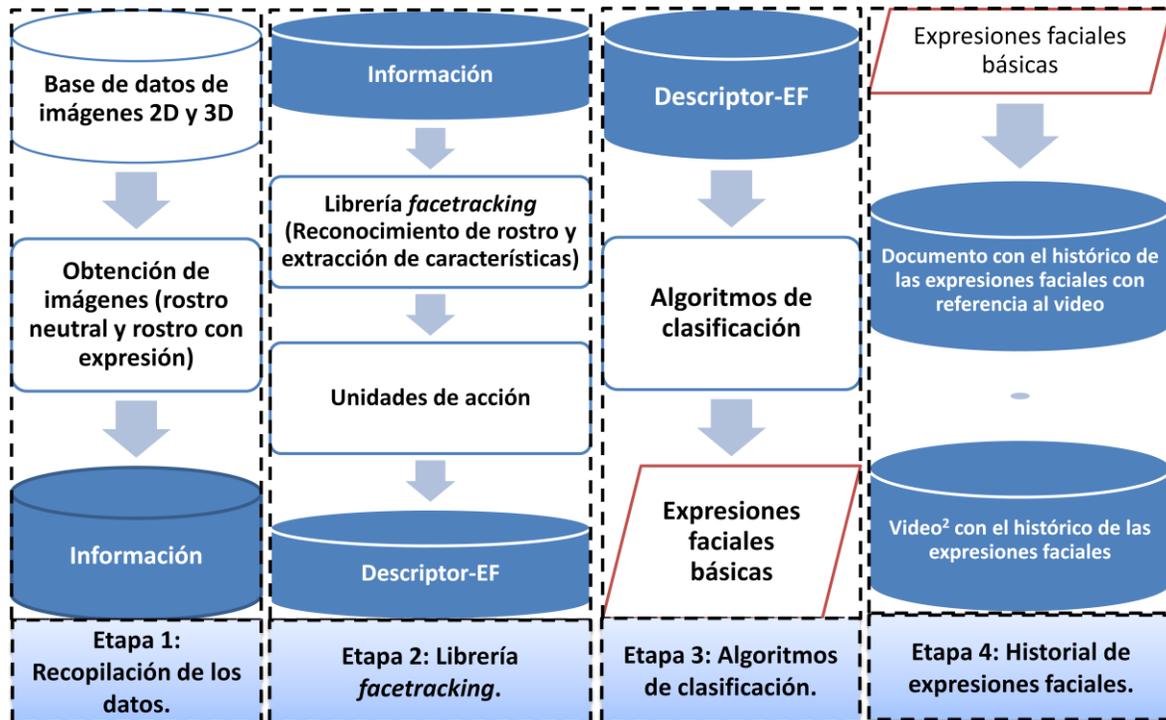


Figura 3.2 Arquitectura del sistema de reconocimiento de expresiones faciales (SREF)

#### 3.2.1 Etapa 1: Recopilación de los datos

En esta etapa se realizó una recopilación de datos (imágenes del rostro) para entrenar los algoritmos de clasificación y poner a prueba el SREF. La recopilación de los datos constó de 2 partes:

- Parte 1: Recopilar datos a partir de imágenes 3D
- Parte 2: Recopilar datos a partir de imágenes 2D.

<sup>2</sup> El video se genera con Kinect Studio y se graba en sincronía con el sistema (SREF).

### 3.2.1.1 Recopilar datos a partir de imágenes 3D

Para obtener los datos, se realizó un procedimiento; éste consistió en la recopilación de imágenes de expresiones básicas con 10 personas, el método para obtener las imágenes fue el siguiente:

- El sujeto de estudio se coloca frente a la cámara RGBD a aproximadamente un metro de distancia, éste observa un video (guión) que lo estimula a realizar las distintas expresiones.
- Después se analiza el video para obtener fotogramas que contengan dichas expresiones faciales básicas.

En este procedimiento se obtiene el fotograma de la expresión. Este archivo se guarda dentro de una carpeta con el nombre del sujeto (ejemplo, sujeto\_1), dentro de ella, se agrega una carpeta con el nombre de la expresión detectada y dentro se encuentra el fotograma de la expresión Figura 3.3.

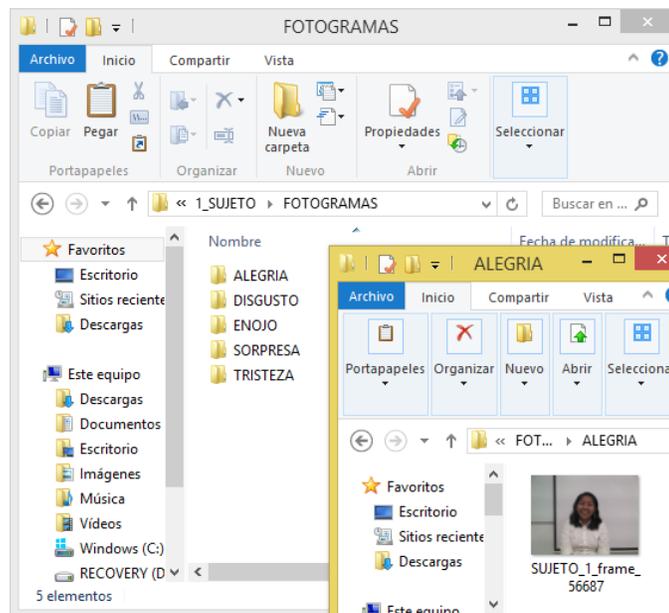


Figura 3.3 Ejemplo del almacenamiento de las imágenes en la base de datos

Es importante mencionar que para capturar las imágenes fue necesario tener un ambiente controlado, las condiciones para obtener las imágenes fueron las siguientes:

- En interior con una iluminación blanca estable sin otros tipos de fuentes de iluminación.
- La distancia adecuada para captar a la persona fue de 1 metro. La razón de esto se debe a que en esta distancia el sujeto tiene una perfecta visión del guión y las facciones del rostro se observan adecuadamente.

- La persona tuvo que estar de frente a la cámara RGBD, sin ningún objeto que obstruya la visión del rostro.

El experimento anterior no logró obtener todas las expresiones faciales en algunos sujetos de estudio (la distribución de los patrones de cada expresión facial se encuentra en la Tabla 3.3, en la segunda columna), por lo cual se tuvo que acceder a bases de datos de secuencias de imágenes con las distintas expresiones faciales básicas [27].

### 3.2.1.2 Recopilar datos a partir de imágenes 2D

Para obtener más ejemplos de expresiones faciales básicas se accedió a la base de datos que desarrolló Cohn Kanade llamada Cohn Kanade Extend [27]. Las imágenes de Cohn Kanade fueron tomadas con una cámara RGB, y además el autor agrega un archivo Excel en el cual da información de las unidades de acción que están presentes en cada una de las imágenes.

La base de datos Cohn Kanade Extend contiene 597 secuencias de imágenes de un total de 123 sujetos [35], sujetos de edad media (hombres y mujeres), la pose es frontal, la secuencia de imágenes de cada sujeto inicia con una expresión neutral y termina con la expresión resultado. Dentro de la base de datos cada sujeto de estudio puede o no puede tener todas las expresiones faciales básicas.

La recopilación de datos a partir de imágenes 2D consistió en obtener los datos de las unidades de acción de la base de datos Cohn Kanade Extend con la cámara RGBD, el experimento se realizó de la siguiente manera:

- Se proyectaron las imágenes (secuencia de imágenes) de la base de datos Cohn Kanade Extend en una pantalla de 32"; se colocó la cámara RGBD (sensor Kinect) frente a la pantalla a 1 metro de distancia
- Se grabó la imagen (imagen de la base de datos Cohn Kanade Extend) de cada expresión, para después analizar el archivo con los resultados y guardar la imagen de la expresión con su patrón correspondiente.

De esta base de datos se analizaron las imágenes de 30 sujetos (hombres y mujeres de entre 20 y 30 años). Los archivos que se obtuvieron con este procedimiento, se guardaron de la misma manera que en la parte 1 (recopilar datos a partir de imágenes 3D).

Se agregaron 47 reglas a partir de los trabajos de Kotsia & Pitas [25], Pantic & Rothkrantz [22] y de Zhang & Ji [24], estos autores utilizaron las mismas AU's (se excluyen las AU's de la nariz y de los ojos; ya que las AU's de *facetracking* sólo utilizan las cejas, la boca y la mandíbula inferior). Especialmente se obtuvieron los patrones de Pantic & Rothkrantz [25], ya que estos autores generaron unas reglas para cada expresión facial (Tabla 3.1).

Tabla 3.1 Lista de patrones (reglas) de Pantic &amp; Rothkrantz [25]

AUs	Emoción	AUs	Emoción	AUs	Emoción	AUs	Emoción
1+2	<b>Sorpresa</b>	1	<b>Tristeza</b>	23+17	<b>Enojo</b>	10+17	<b>Disgusto</b>
2	<b>Enojo</b>	4	<b>Enojo</b>	23+(25/26)	<b>Enojo</b>	10+(25/26)	<b>Disgusto</b>
6	<b>Felicidad</b>	5	<b>Sorpresa</b>	23	<b>Enojo</b>	10	<b>Disgusto</b>
1+4+5+7	<b>Miedo</b>	7	<b>Enojo</b>	24+17+26	<b>Enojo</b>	9+(25/26)	<b>Disgusto</b>
1+4+5	<b>Miedo</b>			24+17	<b>Enojo</b>	9+17	<b>Disgusto</b>
1+4+7	<b>Tristeza</b>	27	<b>Sorpresa</b>	24+26	<b>Enojo</b>	9	<b>Disgusto</b>
1+5+7	<b>Miedo</b>	20+(25/26)	<b>Miedo</b>	24	<b>Enojo</b>	12+(25/26)	<b>Felicidad</b>
1+4	<b>Tristeza</b>	20	<b>Miedo</b>	10+16+(25/26)	<b>Enojo</b>	12	<b>Felicidad</b>
1+5	<b>Miedo</b>	15+(25/26)	<b>Tristeza</b>	10+17+(25/26)	<b>Disgusto</b>	16+(25/26)	<b>Enojo</b>
1+7	<b>Tristeza</b>	15	<b>Tristeza</b>	9+17+(25/26)	<b>Disgusto</b>	17	<b>Tristeza</b>
5+7	<b>Miedo</b>	23+17+(25/26)	<b>Enojo</b>	12+16+(25/26)	<b>Felicidad</b>	26	<b>Sorpresa</b>

En la Tabla 3.1 se muestra la lista de patrones (reglas) de Pantic & Rothkrantz [25], el símbolo + significa conjunción ("y"), las AU's que se encuentran entre paréntesis significa que solo una de ellas estará presente, por lo tanto se toman como dos patrones, por ejemplo: 20+(25/26) = 20+25 ó 20+26. Para esta tesis se tomaron los siguientes patrones de la Tabla 3.1 (ver Tabla 3.2):

Tabla 3.2 Lista de patrones para el SREF desarrollado en esta tesis

AUs	Emoción	AUs	Emoción	AUs	Emoción	AUs	Emoción
1+2	<b>Sorpresa</b>	1	<b>Tristeza</b>	23+17	<b>Enojo</b>	10+17	<b>Disgusto</b>
2	<b>Enojo</b>	4	<b>Enojo</b>	23+(25/26)	<b>Enojo</b>	10+(25/26)	<b>Disgusto</b>
6	<b>Felicidad</b>			23	<b>Enojo</b>	10	<b>Disgusto</b>
6+12	<b>Felicidad</b>			24+17+26	<b>Enojo</b>		
		1+2+(26/27)	<b>Sorpresa</b>	24+17	<b>Enojo</b>		
		27	<b>Sorpresa</b>	24+26	<b>Enojo</b>		
		20+(25/26)	<b>Miedo</b>	24	<b>Enojo</b>	12+(25/26)	<b>Felicidad</b>
1+4	<b>Tristeza</b>	20	<b>Miedo</b>	10+16+(25/26)	<b>Enojo</b>	12	<b>Felicidad</b>
		15+(25/26)	<b>Tristeza</b>	10+17+(25/26)	<b>Disgusto</b>	16+(25/26)	<b>Enojo</b>
16	<b>Neutral</b>	15	<b>Tristeza</b>	17+(25/26)	<b>Disgusto</b>	17	<b>Tristeza</b>
Todas inactivas	<b>Neutral</b>	23+17+(25/26)	<b>Enojo</b>	12+16+(25/26)	<b>Felicidad</b>	26	<b>Sorpresa</b>

En la Tabla 3.2 se muestra la lista de reglas obtenidos a partir de la Tabla 3.1, en donde se agregaron cuatro patrones: dos para la expresión neutral, uno para alegría y otro para sorpresa. El total de los reglas de entrenamiento fue de 47 (las 43 reglas restantes se obtuvieron de la Tabla 3.1), estas reglas se adaptaron al descriptor-EF (descripción en la sección de etapa 2: librería *facetracking*) de forma binaria. En la Tabla 3.3 se observa la distribución de estas reglas para cada expresión facial.

El total de patrones fue de 162. La Tabla 3.3 tiene el total de imágenes para cada expresión facial.

Tabla 3.3 Total de imágenes para cada expresión

EXPRESIONES	GUIÓN	IMÁGENES	REGLAS (Tabla 3.2)	TOTAL
<b>MIEDO</b>	2	11	3	<b>16</b>
<b>TRISTEZA</b>	1	19	6	<b>26</b>
<b>ALEGRÍA</b>	11	23	7	<b>41</b>
<b>SORPRESA</b>	1	16	5	<b>22</b>
<b>ENOJO</b>	5	14	16	<b>35</b>
<b>DISGUSTO</b>	2	10	8	<b>20</b>
<b>NEUTRAL</b>			2	<b>2</b>
<b>TOTAL:</b>				<b>162</b>

La Tabla 3.3 está definida de la siguiente forma: la primera columna tiene el nombre de cada una de las expresiones, la segunda columna muestra el total de imágenes para cada expresión de la primera parte (recopilación de datos a partir de imágenes 3D), la tercera columna es el total de imágenes obtenidas de la segunda parte (recopilación de datos a partir de imágenes 2D), la cuarta columna es el total de reglas obtenidas de la Tabla 3.2 (se tomaron las reglas como imágenes) y la última columna muestra el total de imágenes de cada expresión facial. El total de las imágenes recolectadas fue de 162.

Para cumplir con el objetivo de identificar expresiones faciales sin que los sujetos bajo estudio tengan que exagerar, se agregaron algunas imágenes donde el sujeto no tiene que abrir exageradamente la boca o subir demasiado las cejas, etc., en la Figuras 3.4 - 3.9 se muestra algunos ejemplos de imágenes de este tipo:



Figura 3.4 Expresión de sorpresa en dos personas distintas

En la Figura 3.4 el sujeto del lado izquierdo tiene una expresión exagerada mientras que en el sujeto de lado derecho no la tiene.



Figura 3.5 Expresión de enojo en dos personas distintas

En la Figura 3.5 se muestran diferentes niveles para prensar los labios y el nivel de la distancia horizontal de la boca reducida.



Figura 3.6 Expresión de miedo en dos personas distintas

En la Figura 3.6 se puede observar que la diferencia radica en la boca, el sujeto de la izquierda abre un poco la boca, mientras que el sujeto de la derecha tiene la boca cerrada, sin embargo ambos conservan la distancia horizontal de la boca al máximo nivel.

Para la Figura 3.7 se puede observar que el sujeto de la izquierda levanta un poco la parte interna de las cejas y baja un poco las comisuras de los labios, mientras que el sujeto de la derecha sólo tiene un valor reducido para la distancia de la boca horizontal y las cejas están en posición neutral.



Figura 3.7 Expresión de tristeza en dos personas distintas

En la Figura 3.8 se presenta un caso especial, ya que la misma persona realiza dos tipos de sonrisa y el algoritmo de clasificación detecta ambos casos como alegría. Podría decirse que el sujeto del lado izquierdo tiene una sonrisa más relajada, en cambio el sujeto del lado derecho tiene una sonrisa más exagerada.



Figura 3.8 Expresión de alegría en dos formas distintas

La Figura 3.9 muestra ejemplos de dos sujetos con expresión de disgusto, la cual es muy diferente una de la otra; la expresión de disgusto es muy variada, pero las unidades que son estrictamente necesarias para la identificación de disgusto en este trabajo son las siguientes: AU17 (levantar la mandíbula) sujeto de la izquierda y la AU10 (levantar labio superior) sujeto de la derecha. Para el caso de este trabajo sí es necesario tener activas estas unidades de acción para que la identificación sea correcta, por esta razón se buscaron imágenes de este tipo.



Figura 3.9 Expresión de disgusto en dos personas distintas

### 3.2.2 Etapa 2: Librería *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* [23], para la detección de rostro y extracción de características.

#### 3.2.2.1 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. Las principales funciones se describen en la Tabla 3.4.

Tabla 3.4 Funciones para la detección y extracción de características [23]

Función	Descripción
<b><i>IFTFaceTracker</i></b> (tipo de dato devuelto: booleano)	Se encarga de la interfaz principal para inicializar la detección del rostro en forma sincrónica mediante el paso de <i>frames</i> de color y profundidad. Verifica que tanto los <i>frames</i> de color y profundidad lleguen desde el sensor Kinect.
<b><i>IFTImage</i></b> (tipo de dato devuelto: 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 píxel. Además, <i>IFTImage</i> proporciona métodos de ayuda para la copia de imágenes y dibujo de líneas de depuración son también parte de esta interfaz.

Función	Descripción
<b><i>IFTResult</i></b> (tipo de dato, decimal)	Esta función está encargada de verificar que se haya detectado un rostro de forma exitosa y del seguimiento del rostro. <i>IFTResult</i> proporciona acceso a la siguiente información: - <i>GetFaceRect</i> - en el fotograma de vídeo ofrece información de las coordenadas del cuadro delimitador alrededor del rostro en seguimiento. - <i>Get2DShapePoints</i> - 2D (x, y) Coordenadas de los puntos clave en el rostro del fotograma de vídeo. Realiza un seguimiento de los 87 puntos en 2D.
<b><i>IFTModel</i></b> (tipo de dato, decimal)	Esta función crea la malla <i>Candide</i> sobre el rostro. La interfaz <i>IFTModel</i> 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: <i>GetSUCount</i> , <i>GetAUCount</i> - devuelven respectivamente el número de unidades de forma (SU) o unidades de acción (AU), utilizados en el modelo lineal 3D (malla <i>Candide</i> ). <i>GetTriangles</i> - Devuelve triángulos del modelo de la malla 3D (índices de vértices). Cada triángulo tiene 3 índices de vértices que Figuran en el sentido horario. <i>GetVertexCount</i> - devuelve el número de vértices del modelo de la malla 3D.

En la Tabla 3.4 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.

### 3.2.2.2 Unidades de acción

En éste caso se usan las unidades de acción de *facetracking*, las cuales son 6 unidades que comprenden las cejas, la boca y la mandíbula inferior. En la Tabla 3.5 se observa que todas las AU's generan valores desde -1 a 1 con incrementos de 0.10. Cada AU en la Tabla 3.6 tiene una descripción que indica el movimiento que realiza cada uno dependiendo si el valor es negativo o positivo.

En la Tabla 3.5 la primera columna indica el nombre de la AU, la columna de valores está dividida en tres secciones: un valor negativo  $< 0$  hasta -1 que indica un movimiento, un valor positivo  $> 0$  hasta 1 significa un movimiento contrario al negativo y un valor de cero significa ausencia de movimiento y ésto quiere decir que, es un estado neutral.

Tabla 3.5 Rango de valores para las 6 AUs

AU de Kinect	VALORES		
	Negativo (-1)	0	Positivo (1)
<b>AU[0] Labio superior</b>	Prensar Labios 	Neutral	Levantar el labio superior 
<b>AU[1] Mandíbula</b>	Boca cerrada - subir al máximo la mandíbula inferior 	Neutral	Boca totalmente abierta 
<b>AU[2] Distancia Horizontal de la Boca</b>	Mínima distancia horizontal 	Neutral	Máxima distancia horizontal 
<b>AU[3] Cejas interior</b>	Arriba 	Neutral	Abajo (fruncir ceño) 
<b>AU[4] Comisura de los labios</b>	Comisura de labios arriba 	Neutral	Comisura de labios abajo 
<b>AU[5] Cejas</b>	Abajo 	Neutral	Arriba 

### 3.2.2.3 Descriptor-EF

A partir de estas 6 unidades de acción (Tabla 3.6), se generan 15 unidades de acción que se activan, siempre y cuando estén presentes en el rostro, de lo contrario permanecen en 0 (inactivas).

En los trabajos de Kotsia & Pitas [25], Pantic & Rothkrantz [22] y de Zhang & Ji [24] se identificó que todos ellos utilizan las mismas unidades de acción para clasificar cada expresión facial. *Facetracking* utiliza 6 de estas AUs según la documentación oficial; pero *facetracking* da valores con incrementos de 0.10, esto significa que se pueden crear más rangos para cada AU de *facetracking*, en la Tabla 3.6 se puede observar cómo quedan estas agrupaciones para cada AU, se excluyen 2 AUs relacionadas con los ojos y una relacionada con la nariz, ya que *facetracking* no incluye estos elementos del rostro dentro de sus unidades de acción.

En la Tabla 3.6 primera columna contiene la AU que utiliza *facetracking* y cada AU tiene la condición para crear cada una de las AUs de la literatura (AU real).

Se crearon dos variables (*ceint* y *cej*)<sup>3</sup> respectivamente para la AU[3] y AU[5], porque es necesario obtener un valor neutral para las cejas de cada sujeto al iniciar con la grabación del video, ya que Kinect obtiene valores neutrales distintos para cada sujeto. Las otras unidades de acción no tienen un valor especial. Continuando con la Tabla 3.6 la segunda columna indica el número de la AU real con su respectiva descripción y la última columna muestra la imagen de cada AU.

Tabla 3.6 Distribución de las AUs de *facetracking* con respecto a las AUs de la literatura

Unidad de Acción de <i>facetracking</i> . Condición.	Unidad de Acción, Descripción	Imagen
AU[3] <code>if(AU[3] &lt; ceint )</code>	1, levantar cejas parte interior	
AU[5] <code>if(AU[5] &gt; cej )</code>	2, levantar cejas parte externa	
AU[3] <code>if(AU[3] &gt; ceint )</code>	4, fruncir el ceño	
No utiliza	5, abrir los ojos completamente	
AU[5] <code>if(AU[5] &lt; cej )</code>	6, bajar las cejas a la altura de los ojos	
No utiliza	9, arrugar la nariz	
No utiliza	7, cerrar un poco los ojos	

<sup>3</sup> Estos valores *ceint* y *cej* se obtienen al iniciar la sesión de grabación, a cada sujeto de estudio se le pide estar en un estado neutral sin realizar ningún movimiento en el rostro, para capturar estas dos variables y así tener un valor neutral exacto para cada sujeto de estudio. Kinect no obtiene un valor cero (neutral) para las cejas, es distinto en cada sujeto.

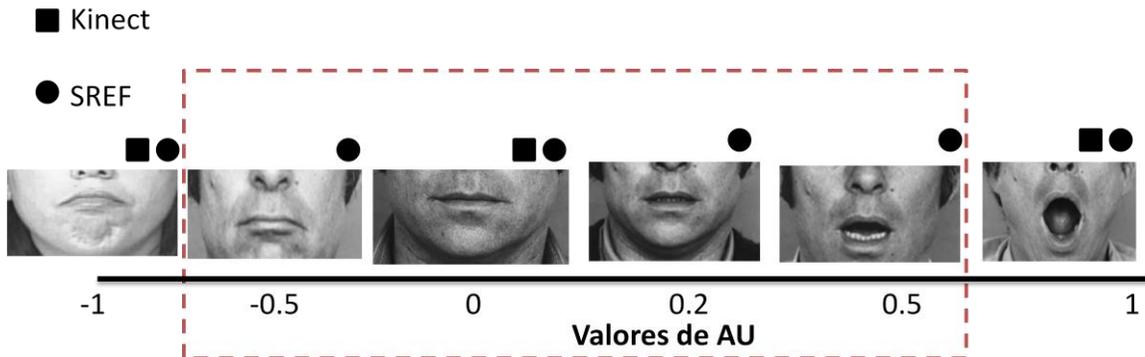
Unidad de Acción de <i>facetracking</i> . Condición.	Unidad de Acción, Descripción	Imagen
AU[0] <code>if(AU[0] &gt; 0.6 )</code>	10, levantar labio superior	
AU[4] <code>if(AU[4] &lt; 0 )</code>	12/13, comisura de los labios hacia arriba	
AU[4] <code>if(AU[4] &gt; 0 )</code>	15, comisura de los labios hacia abajo	
AU[2] y AU[1] <code>if(AU[2] &lt;= -0.2 &amp;&amp; AU[2] &gt; -0.4 &amp;&amp; AU[1] &lt;=0)</code>	16, boca cerrada normal prensando muy poco los labios	
AU[1] <code>if(AU[1] &lt;= -0.2 )</code>	17, mandíbula inferior hacia arriba	
AU[2] y AU[1] <code>if(AU[2] &gt; 0.4 &amp;&amp; AU[1] &lt;=0)</code>	20, distancia horizontal de la boca máxima y boca cerrada	
AU[2] y AU[1] <code>if(AU[2] &lt;= 0.4 &amp;&amp; AU[2] &gt; 0.2 &amp;&amp; AU[1] &lt;=0)</code>	23, pensar poco los labios y subir mandíbula inferior	
AU[2] y AU[1] <code>if(AU[2] &lt; -0.4 &amp;&amp; AU[1] &lt;= 0)</code>	24, distancia horizontal de la boca mínima y boca cerrada	
AU[1] <code>if(AU[1] &gt; 0 &amp;&amp; AU[1] &lt;= 0.3)</code>	25, abrir muy poco la boca.	
AU[1] <code>if(AU[1] &gt; 0.3 &amp;&amp; AU[1] &lt;= 0.5)</code>	26, abrir la boca.	
AU[1] <code>if(AU[1] &gt; 0.5)</code>	27, abrir completamente la boca	

En la primera columna de la Tabla 3.6 se muestran los rangos para cada unidad de acción, estos rangos representan cada una de las imágenes de la última columna de la misma tabla; según Paul Ekman [21], las unidades de acción pueden estar presentes o ausentes en cada sujeto, no existe un valor intermedio para cada AU, es por ello que se crearon patrones con valores 1 (activo) y 0 (inactivo) para cada AU. Cada AU se activa siempre y cuando cumpla con su respectiva condición, de lo

contrario permanecerá inactiva. El descriptor-EF está formado con estas 15 unidades de acción. Este descriptor sirve para continuar con la etapa 3 del proyecto.

Como se presenta en la Tabla 3.5 Kinect sólo reconoce valores extremos (-1, 0 y 1) para reconocer cada AU, mientras que en la Tabla 3.6 se observa que el sistema SREF reconoce más variables de movimiento sin tener que llegar a los extremos (sin tener que exagerar la expresión).

La Figura 3.10 muestra un ejemplo entre los valores de Kinect y los valores del SREF para la boca; esta figura tiene una leyenda en la parte superior izquierda que indica sí la imagen es reconocida por Kinect (cuadro relleno) y/o por el sistema SREF (circulo relleno); cada una de las imágenes tiene un cuadro relleno o un circulo relleno dependiendo de qué sistema reconoce dicha imagen y en la parte de abajo de cada imagen se encuentra el valor numérico para cada una.



Como se observa en la Figura 3.10 Kinect sólo reconoce los extremos y el valor neutro, el sistema SREF reconoce más variables del movimiento de la boca. En la Figura 3.10 lo que se encuentra dentro del rectángulo de línea punteada se considera como una expresión natural (expresión sin exagerar) y lo que se encuentra fuera de este rectángulo es considerado como una expresión exagerada.

La Figura 3.11 presenta un ejemplo entre los valores de Kinect y los valores de SREF para las cejas; esta figura tiene una leyenda en la parte superior izquierda que indica sí la imagen es reconocida por Kinect (cuadrado relleno) y/o por el sistema SREF (circulo relleno); cada una de las imágenes tiene un circulo relleno o un cuadrado relleno dependiendo de qué sistema reconoce dicha imagen.

En la Figura 3.11 la línea horizontal de los valores tiene la variable *cej*, porque el valor neutral de las cejas para cada sujeto es único, entonces esto significa que sí el valor aumenta, las cejas suben y sí el valor disminuye, las cejas bajan.

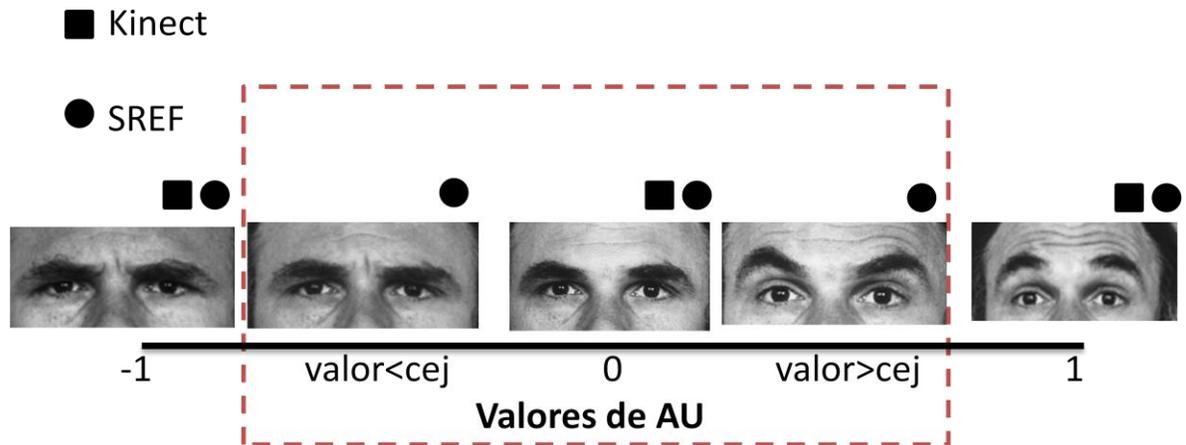


Figura 3.11 Ejemplo del rango para las cejas

Como se observa en la Figura 3.11 Kinect sólo reconoce los extremos y el valor neutro, el sistema SREF reconoce más variables del movimiento de las cejas. En la Figura 3.11 lo que se encuentra dentro del rectángulo de línea punteada se considera como una expresión natural (expresión sin exagerar) y lo que se encuentra fuera de este rectángulo es considerado como una expresión exagerada.

### 3.2.3 Etapa 3: Algoritmos de Clasificación para las expresiones faciales

Esta etapa se realiza a partir de la generación del descriptor-EF. Con base en esta información se efectúa la implementación de los algoritmos de clasificación, para obtener finalmente las expresiones faciales básicas.

#### 3.2.3.1 Algoritmos de clasificación implementados

Los algoritmos de clasificación que se implementaron fueron los siguientes:

- Máquina de Soporte Vectorial (Se utilizó la librería OpenCV):
  - Lineal
  - Sigmoide
  - De base radial
- Árboles de decisión (Se utilizó Weka):
  - J48
  - REPTree
- Algoritmos de reglas (Se utilizó Weka):
  - NNge
  - Part

Estos algoritmos de reglas y árboles de decisión se implementaron porque con ellos se logró el 100% de los patrones correctamente clasificados al generar los modelos

entrenados. También se identificó que en algunos trabajos [12][26] usaron árboles de decisión para clasificar las expresiones faciales.

La máquina de soporte vectorial (MSV) se implementó porque en algunos trabajos relacionados [17][19][25] se obtuvieron buenos resultados.

Para la implementación de los algoritmos de clasificación se realizó una lista de todas las imágenes con su respectivo descriptor-EF, de esta lista se utilizó el 80% para entrenar los algoritmos de clasificación y 20% para probarlos. En la Figura 3.10 se muestran los pasos de esta etapa.

La etapa 3: algoritmos de clasificación para las expresiones faciales, está compuesta por una serie de pasos (Figura 3.12).

- En el paso 1, se generó una lista con los patrones de entrenamiento (80% del total de patrones), cada patrón contiene el descriptor-EF y su expresión facial correspondiente.
- En el paso 2, se entrena cada algoritmo de clasificación con los patrones de entrenamiento, cada algoritmo genera un modelo entrenado con reglas o pesos que servirán para probarlos con los patrones de prueba (paso 3).
- En el paso 3, finalmente se obtendrá la expresión facial para cada uno de los patrones de prueba.

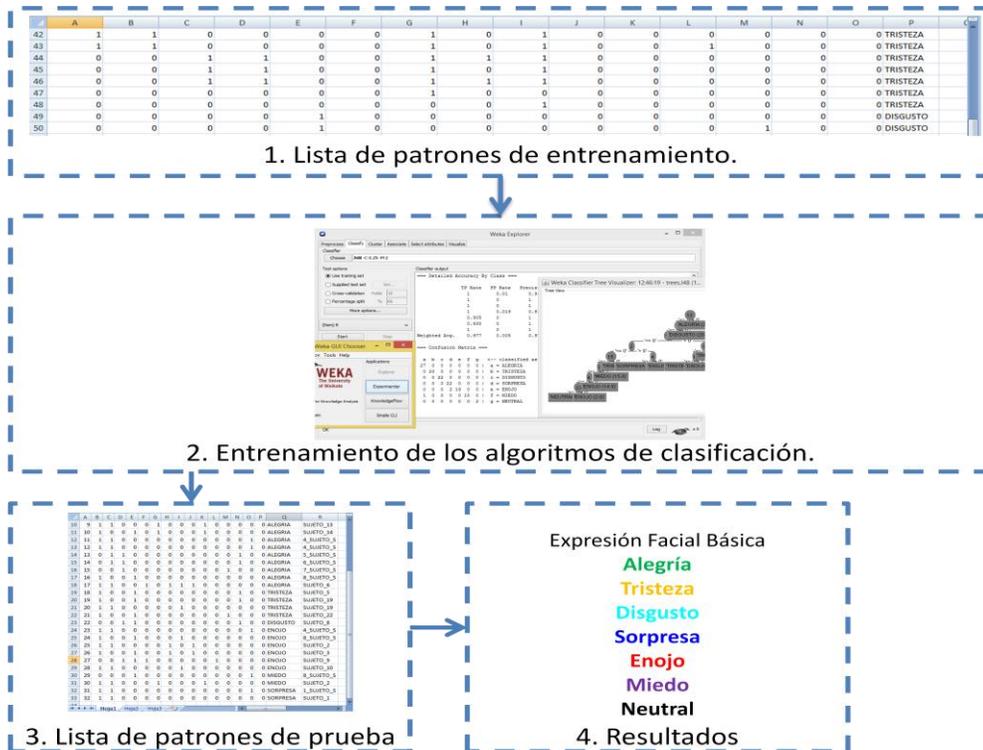


Figura 3.12 Pasos de la etapa 3

### 3.2.4 Etapa 4: Historial de las expresiones faciales

Finalmente se obtiene un historial con las expresiones faciales básicas, estas expresiones se obtienen al realizar la grabación de la sesión del sujeto de estudio; se obtiene el video de la respuesta obtenida a partir del estímulo (guión) y un archivo que contiene el descriptor-EF, la expresión facial y el milisegundo en el cual se tomó esa captura.

#### 3.2.4.1 Video (guión) con las expresiones faciales

El video de cada sujeto se genera a partir de la herramienta Kinect Studio v1.8, esta herramienta captura la imagen RGB, la imagen de profundidad y una imagen en 3D generada a partir de las imagen RGB y profundidad. El video se guarda en un formato .xed (formato de Kinect Studio), se reproduce con esta herramienta (Kinect Studio v.1.8 es una aplicación incluida en el SDK de Kinect v1.8). Se guarda el video de la sesión, la Figura 3.13 es un ejemplo de la aplicación Kinect Studio v1.8.

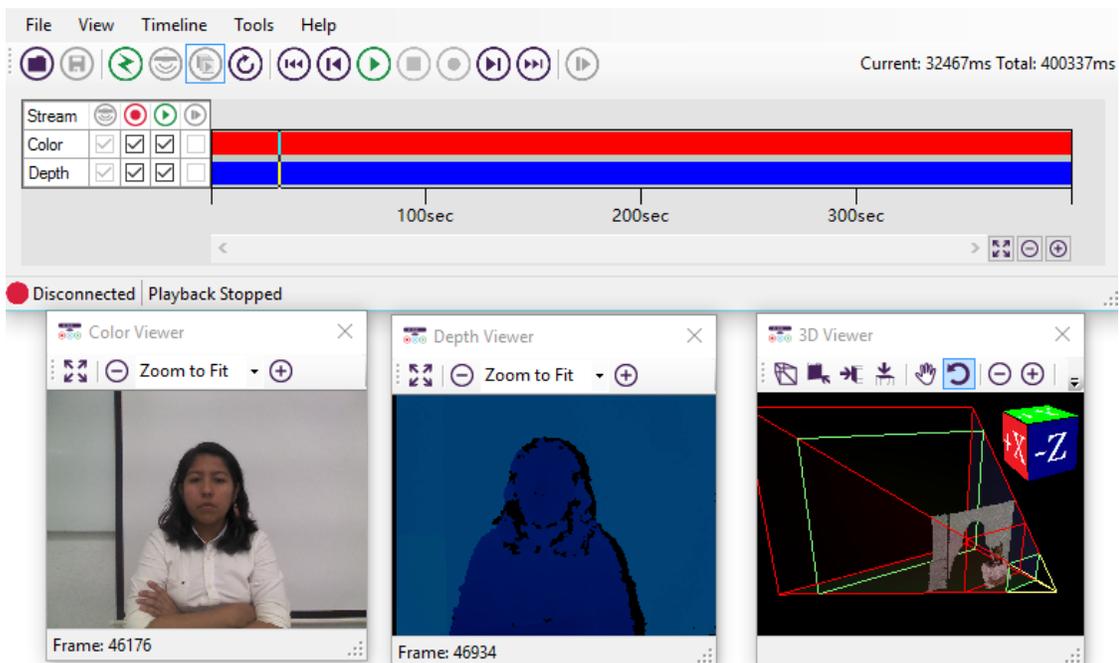


Figura 3.13 Aplicación Kinect Studio v1.8

La aplicación Kinect Studio v1.8 es una herramienta que permite guardar un video del sensor Kinect; esta aplicación guarda la imagen RGB, profundidad y genera una vista 3D adicional (ver Figura 3.13). Esta aplicación tiene una línea de tiempo para reproducir el video en el milisegundo deseado, esto es de gran importancia ya que se guarda el historial en un archivo que contiene la captura de cada *frame* y con esto se accede a la imagen en ese mismo milisegundo para verificar que los datos sean congruentes con el archivo del historial de las expresiones faciales.

### 3.2.4.2 Histórico con las expresiones faciales

Esta sección se refiere al archivo que contiene los datos obtenidos a partir del SREF, este archivo de excel contiene el descriptor-EF, las unidades de acción de *facetracking* en bruto, la expresión facial y el tiempo en el que se capturó cada *frame* (Tabla 3.7).

Se accede al video para verificar la captura de los datos y ver cuál fue la respuesta obtenida a partir del SREF dentro del archivo excel con el historial. En la Figura 3.14 se muestra un ejemplo del video en el milisegundo 387,638 y el archivo excel en ese mismo milisegundo con el resultado de ese *frame*.

Tabla 3.7 Un ejemplo del archivo con el historial de un sujeto de estudio

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	2	4	6	10	13	15	16	17	20	23	24	25	26	27	LAB SUP	MANDIBULA	DIS HZTL	CEJA INT	COMISURA	CEJA EXT	EXPRESION	TIEMPO	milisegundc
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.857863	-0.033301	0.914877	0.235219	-0.419689	-0.130398	,ALEGRIA	393.281	386798
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.836319	-0.03595	0.913359	0.215726	-0.378315	-0.159817	,ALEGRIA	393.333	386850
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.817913	-0.037916	0.835706	0.212704	-0.369206	-0.149667	,ALEGRIA	393.395	386912
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.813259	-0.026866	0.855337	0.207891	-0.352235	-0.163412	,ALEGRIA	393.429	386946
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.813259	-0.026866	0.855337	0.207891	-0.352235	-0.163412	,ALEGRIA	393.475	386992
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.758655	-0.029755	0.807617	0.235406	-0.370288	-0.119231	,ALEGRIA	393.527	387044
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.577	387094
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.633	387150
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.686	387203
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.732	387249
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.778	387295
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.825	387342
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.866	387383
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.921	387438
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	393.97	387487
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	394.023	387540
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	394.066	387583
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	394.121	387638
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	394.166	387683
0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	0.723846	-0.031092	0.840498	0.232226	-0.418643	-0.134157	,ALEGRIA	394.221	387738

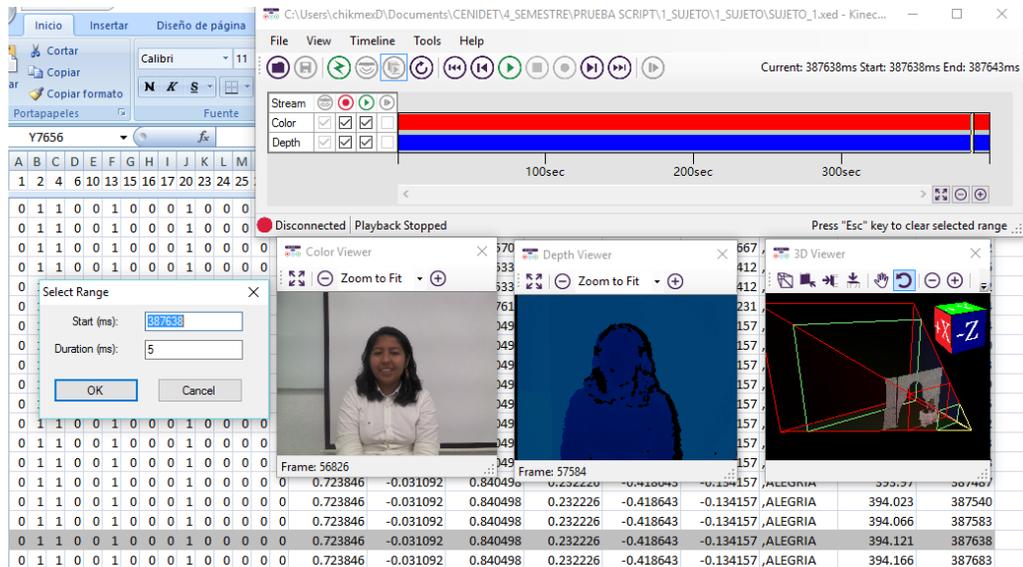


Figura 3.14 Ejemplo del video con el archivo historial

En la Figura 3.14 se observa que el resultado de la expresión facial en el archivo de excel es: Alegría y la imagen concuerda con este resultado, porque el sujeto está sonriendo. En esta Figura 3.14 se puede observar en la parte izquierda de la figura, que hay un recuadro llamado *Select Range*, con el cual se puede ingresar directamente el milisegundo que se desea verificar dentro del video.

### **3.3 Diseño e implementación del sistema**

En esta sección se describe el ambiente, el diseño y la implementación de las etapas de recopilación de los datos (imágenes de rostro) y el SREF descritas en el apartado 3.2.

#### **3.3.1 Descripción del ambiente de implementación**

La etapa 1: recopilación de los datos se generó a partir de las imágenes de 10 personas, dichas personas observaron un guión para estimular y generar las expresiones faciales básicas. Para complementar esta base de imágenes se accedió a la base de datos Cohn-Kanade Extend [27] de la cual se obtuvieron imágenes de las expresiones faciales básicas de 30 personas. Estas imágenes se almacenaron para generar la base de datos con la cual se entrenaron los algoritmos de clasificación (80% de los datos) y se probaron con el Sistema de Reconocimiento de Expresiones Faciales.

La etapa 2: librería *facetracking* está implementada en el lenguaje de programación C++ para Windows [23] que realiza el reconocimiento y seguimiento del rostro y sus componentes, en esta etapa también se obtienen las 6 unidades de acción de *facetracking* y finalmente se genera el descriptor-EF con 15 unidades de acción.

En la etapa 3: se implementa el algoritmo de clasificación (previamente entrenado) para generar finalmente la expresión facial (una de las seis expresiones básicas). Todos los algoritmos de clasificación se implementaron en lenguaje de programación C++.

La etapa 4: muestra el histórico de las expresiones faciales detectadas, se genera un video con las expresiones, este video se guarda con el formato de Kinect Studio v1.8; esta herramienta sirve para guardar la sesiones de los sujetos de estudio (Kinect Studio v1.8 es parte del SDK de Kinect [28]). En esta etapa también se guarda un archivo con el historial, el cual contiene el descriptor-EF, las expresiones faciales detectadas y el segundo en el que se realizó dicha captura para toda la sesión de grabación.

### 3.3.2 Diseño del sistema

En esta sección se define el diseño de las cuatro etapas, en conjunto en la Figura 3.15 se observa en detalle el proceso, la entrada y salida del SREF (Sistema de reconocimiento de expresiones faciales).

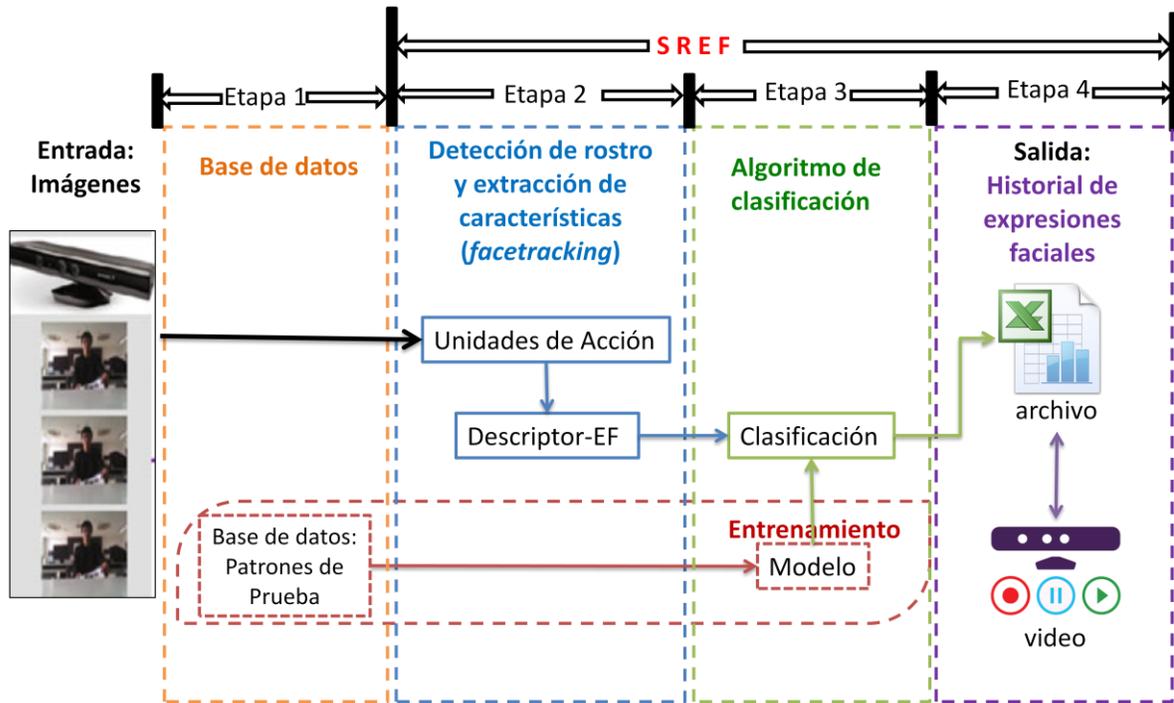


Figura 3.15 Diseño general del sistema

En la Figura 3.15 la etapa 1 está encargada de la recopilación de los datos para el funcionamiento del SREF. Se observa (Figura 3.14) que la etapa 2, 3 y 4 conforman el SREF entrenado.

### 3.3.3 Implementación del sistema

En esta sección se define la implementación del sistema en la Figura 3.15 se observa en detalle el proceso, la entrada y salida del SREF (Sistema de reconocimiento de expresiones faciales).

En la Figura 3.16 los datos de entrada (Entrada\_datos) son las imágenes capturadas por la cámara RGBD, después el proceso pasa por todas las funciones hasta llegar a la parte en donde se imprime el resultado (salida), los datos que se imprimen son: el descriptor-EF (contiene las 15 AUs), la expresión facial y el tiempo en milisegundo de esa captura. Finalmente el proceso continua hasta que el usuario cierra el sistema.

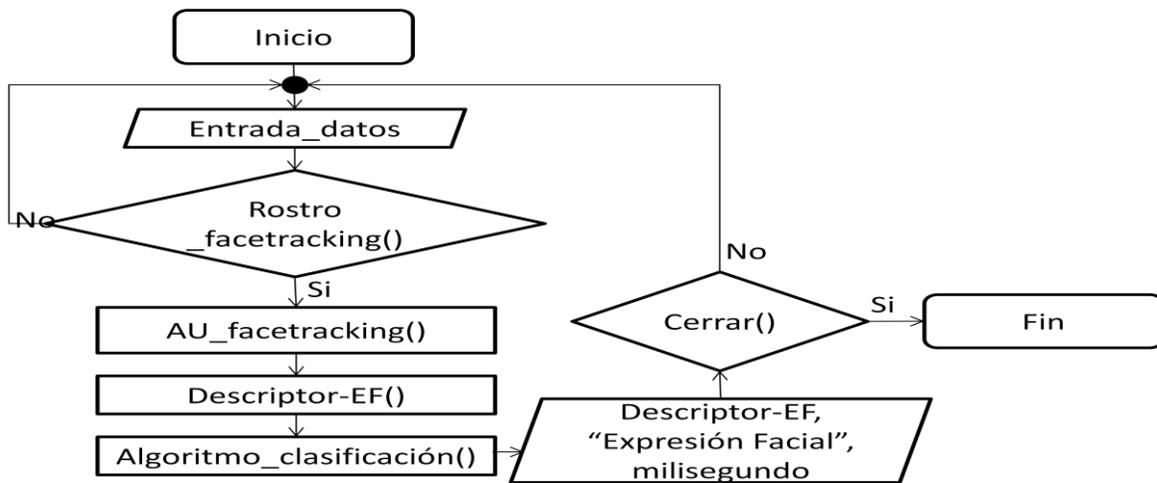


Figura 3.16 Diagrama de flujo del sistema SREF

Las funciones de la Figura 3.16 se describen en la Tabla 3.8.

Tabla 3.8 Funciones del sistema SREF

Función	Descripción
<i>Rostro_facetracking()</i>	Esta función se encarga de reconocer el rostro, colocando la malla <i>Candide</i> sobre el rostro. También se reconocen los componentes del rostro, es parte de la librería <i>facetracking</i>
<i>AU_facetracking()</i>	Esta función obtiene los valores de cada una de las seis unidades de acción, esta función es parte de la librería <i>facetracking</i>
<i>Descriptor-EF()</i>	Esta función se encarga de asignar cada una de las seis AUs, obtenidas de la función <i>AU_facetracking()</i> a su respectiva AU de las 15 AUs, de esta forma se crea un descriptor con las 15 AUs.
<i>Algoritmo_clasificación()</i>	Esta función realiza el trabajo de clasificar el descriptor-EF en una de las seis expresiones faciales
<i>Cerrar()</i>	Esta función realiza la instrucción de cerrar el sistema

### 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 los datos (base de datos de imágenes del rostro), la implementación de la librería *facetracking* para detección del rostro y extracción de las características, finalmente la implementación de algoritmos que permitan clasificar las expresiones faciales básicas. El sistema (SREF) se desarrolló en C++.

La forma en la que se obtienen los valores con el sistema SREF, permite una mayor cantidad de variabilidad en las expresiones faciales. Las Figuras 3.10 y 3.11 explican cuando un movimiento del rostro se considera natural (sin exagerar) y cuando un movimiento se considera exagerado.

Es importante mencionar que no se puede definir un valor fijo sin exagerar, ya que este valor depende de la morfología del rostro de cada persona, esto quiere decir que, existen rostros alargados, rostros redondos o más pequeños y este valor se adapta a la morfología de cada uno de ellos.

## Capítulo 4

### 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 al aplicar cada uno de los algoritmos de clasificación. Estas pruebas tienen por objetivo medir el desempeño del sistema.

#### 4.1 Ambiente de pruebas

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

##### Hardware

- Laptop  
Procesador: Intel(R) Core(TM) i7-45100 CPU 2.00Ghz 2.60 GHz  
Memoria RAM: 8.00 GB  
Sistema Operativo de 64 bits: Windows 8.1, 10.  
GPU: NVIDIA GeForce 840M

##### Software

- Kinect Studio: Herramienta que permite guardar videos con información RGB y profundidad [28]
- 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.

##### Base de datos

Se utilizaron dos bases de datos de distintas fuentes. La primera es Cohn-Kanade Extend desarrollada por Cohn Kanade y sus colaboradores en la universidad de Pittsburgh en el año 2010 [27]. La segunda se creó en CENIDET para esta tesis. El objetivo de crear una base de datos fue para usar el guión como estímulo para generar las expresiones faciales.

En la base de datos Cohn-Kanade Extend participaron personas de distintas razas. En las imágenes la iluminación del escenario es blanca. La cara permanece en posición frontal respecto al dispositivo de captura, la distancia de captura es de un metro y el área del rostro representa en promedio el 30% del total de la imagen. Con estas características se simplifica el trabajo de la localización de la cara. Esta base de datos también tiene un documento con las unidades de acción presentes en cada imagen.

La base de datos creada se capturó a partir de la forma natural en que los participantes generan las expresiones faciales. Los videos se capturaron en el interior de un edificio con iluminación blanca de 100 watts. La cara permanece en posición frontal respecto al dispositivo de captura, la distancia de captura es de un metro y su área representa en promedio el 18% del total de la imagen. Las personas que participaron son de nacionalidad mexicana. En estos videos el fondo del escenario es inmóvil.

Con estas bases de datos se tiene material diverso para probar la aplicación desarrollada, por ejemplo: 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 las dos bases de datos. En la base de datos de CENIDET se obtuvieron los fotogramas de las 10 personas donde se hacían presentes las expresiones faciales.

### 4.2.1 Prueba de iluminación con cámara RGBD

#### Objetivo

Realizar pruebas para verificar el correcto funcionamiento del dispositivo en diferentes condiciones de iluminación.

#### Desarrollo

Se realizaron capturas del rostro en distinto tipos de iluminación, ya que la cámara RGBD tuvo problemas para obtener los datos.

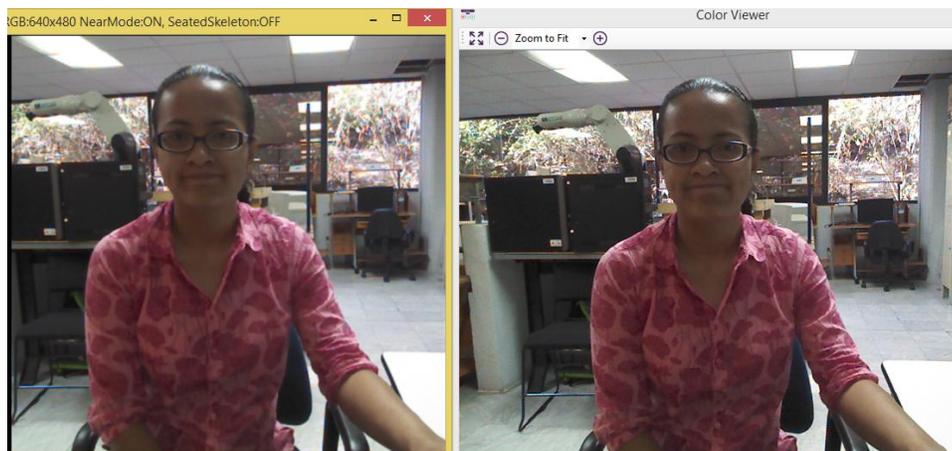


Figura 4.1 Distintos focos de iluminación

Como se observa en la Figura 4.1 este cuarto tiene una iluminación variable con lo cual se generan problemas; en el fondo está la iluminación natural del sol, arriba del sujeto de estudio hay dos lámparas (100 watts) que crean sombras en el rostro, por lo tanto el sensor no reconoce el rostro (no se logra colocar la malla sobre el rostro).

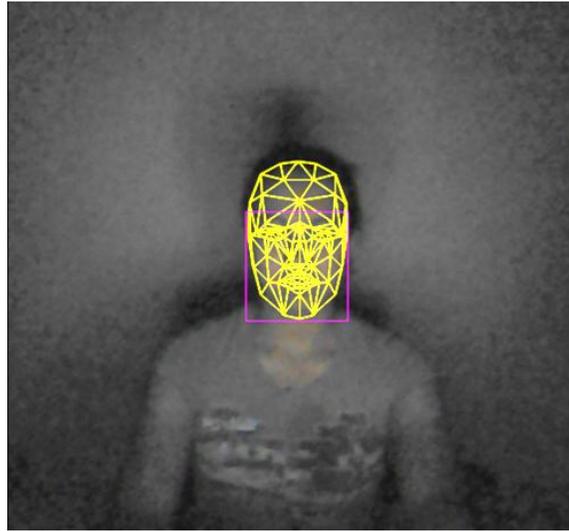


Figura 4.2 Cuarto oscuro

En la Figura 4.2 se observa el sujeto de estudio en un cuarto oscuro, la única iluminación se obtiene a partir de la pantalla de la computadora y el sensor es capaz de reconocer el rostro, la imagen se observa en escala de grises.



Figura 4.3 Ejemplo de una iluminación estable

En la Figura 4.3 se observa que el cuarto tiene una iluminación blanca estable que no crea sombras importantes en el rostro.



Figura 4.4 Ejemplo de un rostro con lentes

En la Figura 4.4 se observa que el sensor no reconoce adecuadamente las cejas, ya que los lentes causan confusión; la malla coloca la línea de las cejas un poco más abajo de donde se encuentran en realidad. También se observa en esta figura que el cuarto tiene poca iluminación y sin embargo detecta el rostro.

### Observaciones

- El dispositivo no es capaz de reconocer el rostro en un ambiente con muchas fuentes de iluminación, sobre todo cuando se presentan sombras en el rostro.
- Se encontró que el dispositivo es capaz de reconocer rostros en un cuarto oscuro, siempre y cuando exista algún tipo de iluminación sobre el rostro (Figura 4.2).
- Para que el dispositivo logre capturar el rostro es necesario un cuarto con iluminación estable o uniforme sobre el rostro (Figura 4.3).
- También se observó que los lentes pueden causar error en la captura de las cejas, por lo tanto se tomaron capturas de sujetos sin lentes (Figura 4.4).

### 4.2.2 Prueba con imágenes 2D

#### Objetivo

Verificar que el sensor reconoce imágenes 2D, evaluar el nivel de reconocimiento y los escenarios en los que cuales es posible reconocer una imagen 2D con el sensor.

#### Desarrollo

Se colocó el sensor frente a una pantalla de 32" donde se proyectaron las imágenes en 2D.

En la Figura 4.5 se muestra una imagen 2D en escala de grises con la malla se adaptó bien, ya que el tamaño del rostro es proporcional a la distancia en que se realizó la captura, esto quiere decir que, el rostro proyectado debe tener el tamaño de un rostro real, a cualquier distancia que se ponga la pantalla con respecto al sensor de captura.



Figura 4.5 Imagen 2D en escala de grises con la malla *Candide* sobre el rostro

Se logró observar que el sensor puede reconocer imágenes 2D en escala de grises (Figura 4.5) y a color (Figura 4.6).

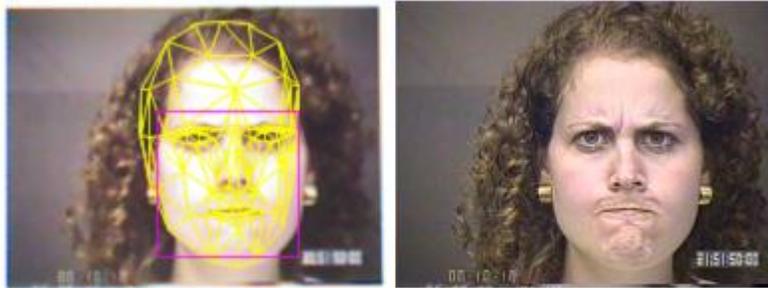


Figura 4.6 Imagen 2D a color con la malla *Candide* sobre el rostro

Se capturaron imágenes animadas con la forma de un rostro humano (Figura 4.7) y la malla se colocó sobre el rostro. El sensor sólo reconoce rostros humanos.

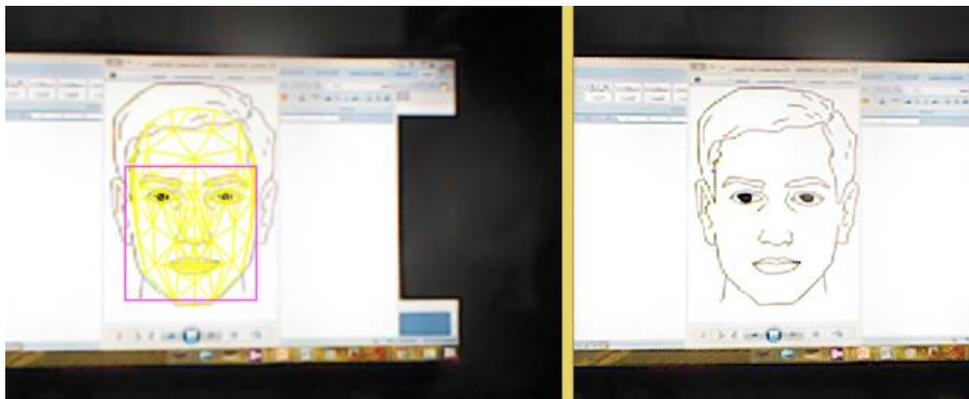


Figura 4.7 Imagen animada de una Figura humana

## Observaciones

Es importante mencionar que se realizó una prueba adicional para comprobar que el sensor Kinect utiliza el sensor infrarrojo. Al bloquear el sensor infrarrojo la malla *Candide* desaparece, de igual forma si sólo se bloquea la lente de la cámara RGB la malla desaparece.

Con esto se puede concluir que el sensor Kinect utiliza ambos mecanismos (sensor infrarrojo y cámara RGB) para que la malla se posicione sobre el rostro y así realizar el reconocimiento de las expresiones faciales.

También es importante que el rostro de la imagen 2D esté proyectado de un tamaño real a la distancia de captura para que la toma de los datos sean correctos, esto quiere decir, que si se coloca la imagen 2D a un metro de distancia y el rostro es demasiado pequeño, la malla *Candide* no se colocará bien sobre el rostro, es necesario colocar la imagen 2D de un tamaño real para que sea reconocido correctamente (Figura 4.5 y 4.6).

### 4.2.3 Prueba a distintas distancia con respecto al sujeto de estudio

#### Objetivo

Se realizó esta prueba para verificar los datos de salida de las AU's del Kinect colocando al sujeto de estudio a diferentes distancias (60 cm, 120 cm y 200 cm), esto con el fin de comprobar que los datos no cambian según la distancia a la que se encuentre el sujeto y así tener valores confiables para que posteriormente sean usados para la etapa de clasificación de las expresiones.

#### Desarrollo

La prueba consiste en colocar al sujeto en cada una de las distancias predefinidas (Figura 4.8, a) 60 cm, b) 120 cm y c) 200 cm).

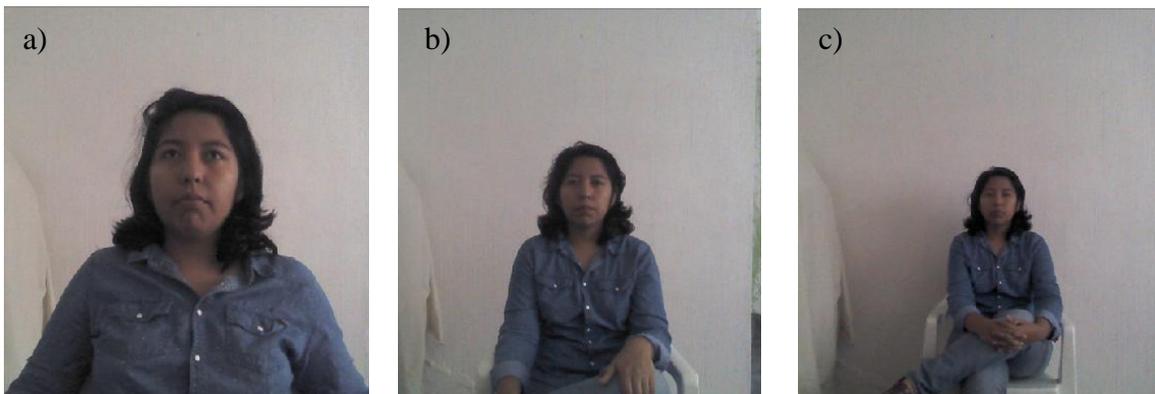


Figura 4.8 Captura del sujeto con el sensor a una distancia de: a) Distancia 60 cm b) Distancia 120 cm c) Distancia 200 cm

Para medir los cambios en el rostro, el sujeto mueve las cejas, la boca y la mandíbula inferior, en esta secuencia: levantar las cejas, fruncir el ceño (bajar las cejas de la parte interna), abrir completamente la boca, prensar los labios y sonreír.

Realizar los movimientos, arriba indicados, se hace con el objetivo de observar los valores de cada una de las AU's para que lleguen a su máximo nivel y así se puede observar la escala con sus respectivos valores que van desde valores negativos hasta positivos, teniendo como valor neutral 0 para todas las AU's. Este procedimiento se realiza para cada una de las distancias.

Para cada posición (distancia - 60 cm, 120 cm y 200 cm) donde se colocó al sujeto en estudio, se generaron 200 patrones (200 capturas de datos) con los valores de cada una de las AU's, como se muestra en la Tabla 4.1.

Tabla 4.1 Ejemplo de los valores obtenidos de las 6 AU's

	AU0	AU1	AU2	AU3	AU4	AU5
NUM	Labio Superior	Mandíbula inferior	Distancia de la Boca horizontal	Cejas	Comisura de los labios	Cejas parte interior
1	0.238338	-0.007295	-0.618133	0.347778	0.048803	-0.170859
2	0.031912	0.036582	-0.622207	0.372464	0.108486	-0.147597
...	...	...	...	...	...	...
200	0.000259	0.034004	-0.599044	0.348165	0.079166	-0.165779

Se tomaron los primeros 100 patrones para las unidades de acción relacionadas con las cejas (AU3 Y AU5) y los otros 100 para la unidades relacionadas con la boca (AU0, AU1, AU2 y AU4), esto se debe a que la grabación inicia con los movimientos de las cejas y posteriormente la boca, así de esta manera las gráficas serán más reducidas y facilitan la comprensión.

En las gráficas (Figuras 4.9 - 4.11) se presentan las unidades de acción relacionadas con la boca (AU0, AU1, AU2, AU4) para las 3 distancias:

En la Figura 4.9 la gráfica representa el promedio de cada una de las AUs relacionadas con la boca para las 3 distancias, se puede observar que la diferencia de los valores varían en 0.1. En la Figura 4.10 la gráfica representa el valor máximo de cada AU relacionada con la boca para las 3 distancias, de igual forma tiene valores cercanos. En la Figura 4.11 la gráfica representa el valor mínimo para cada AU y los valores también se parecen (la diferencia entre ellos es de 0.1 en algunos casos).

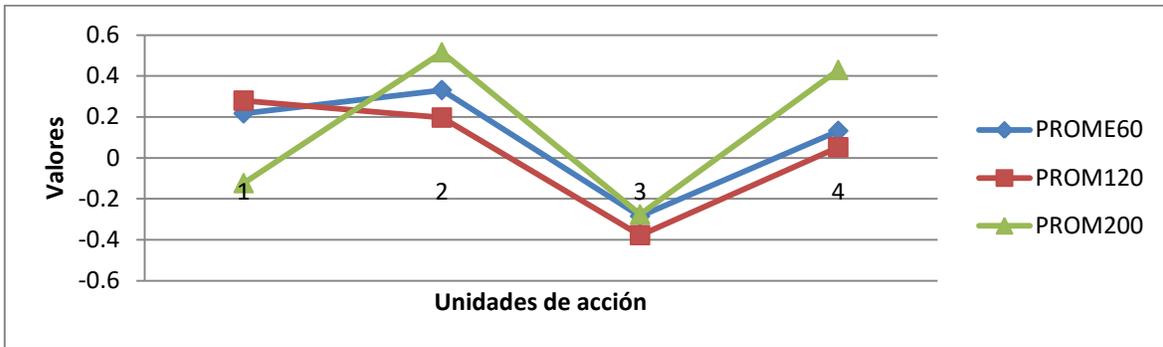


Figura 4.9 Promedio de las unidades de acción de la boca (AU0, AU1, AU2 y AU4) para las tres distancias

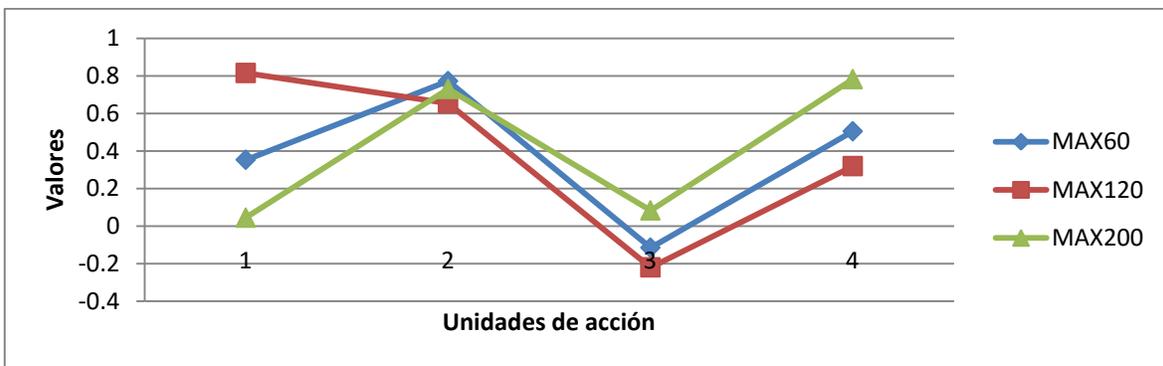


Figura 4.10 Valores máximos de las unidades de acción de la boca (AU0, AU1, AU2 y AU4) para las tres distancias

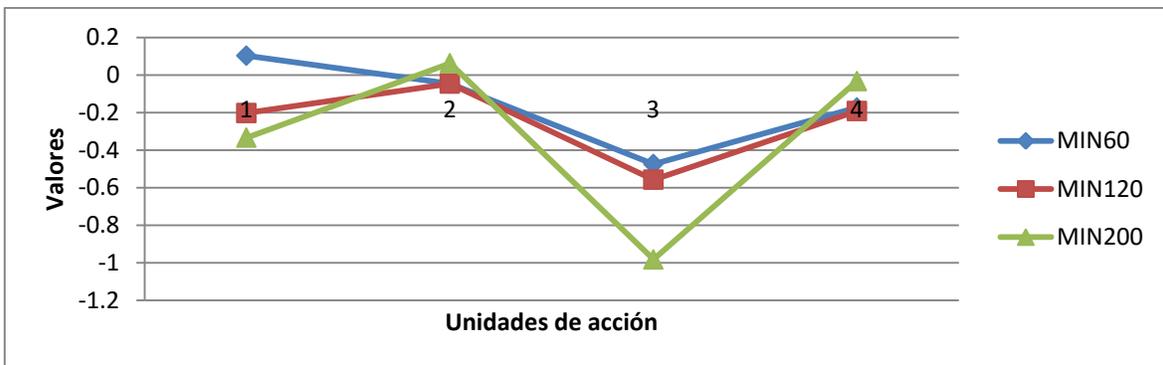


Figura 4.11 Valores mínimos de las unidades de acción de la boca (AU0, AU1, AU2 y AU4) para las tres distancias

A continuación se muestran las gráficas (Figuras 4.12 - 4.14) para las unidades de acción relacionadas con las cejas (AU3, AU5) para las 3 distancias:

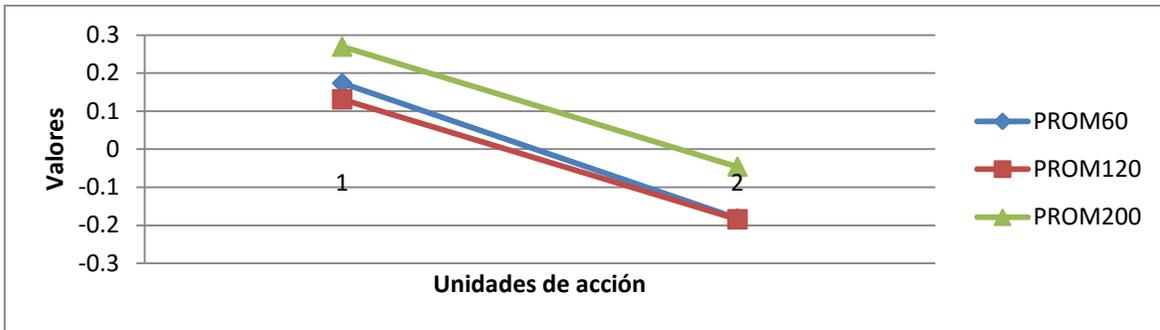


Figura 4.12 Promedio de las unidades de acción de las cejas (AU3 y AU5) para las tres distancias

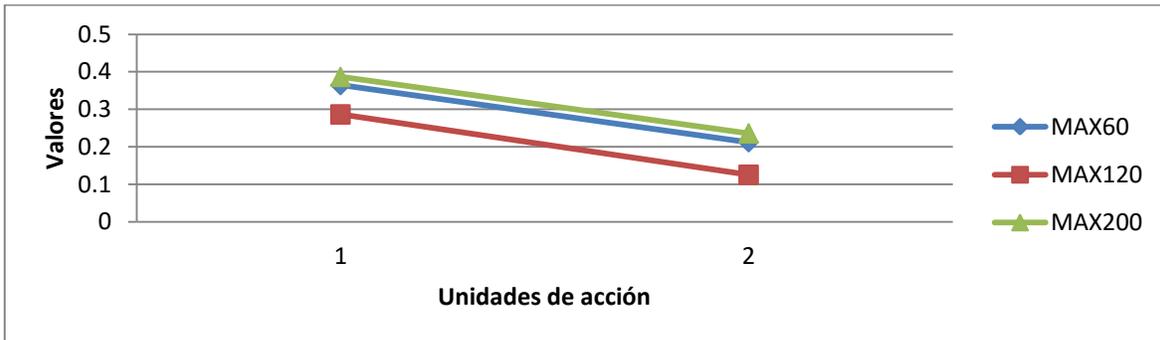


Figura 4.13 Valores máximos de las unidades de acción de las cejas (AU3 y AU5) para las tres distancias

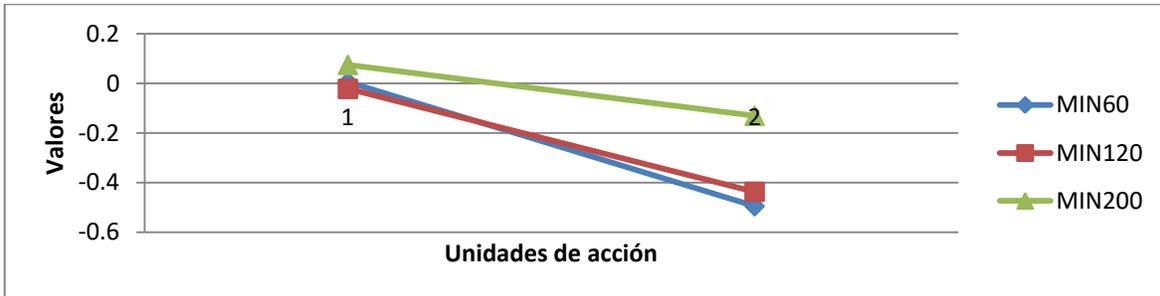


Figura 4.14 Valores mínimos de las unidades de acción de las cejas (AU3 y AU5) para las tres distancias

En la Figura 4.12 la gráfica representa el promedio de cada una de las AUs relacionadas con las cejas para las 3 distancias, se puede observar que en los valores de las distancias de 60 cm y 120 cm existe una variación de 0.05, la distancia de 200 cm varía en 0.1 con respecto de las de 60 cm y 120 cm.

En la Figura 4.13 la gráfica representa el valor máximo de cada AU relacionada con las cejas para las 3 distancias, de igual forma tiene valores cercanos, como en la Figura 4.12.

En la Figura 4.14 la gráfica representa el valor mínimo para cada AU y los valores también se parecen, la diferencia entre la distancia 60 cm y 120 cm es de 0.05 y éstas con respecto a la distancia 200 cm es mayor 0.2.

La siguiente gráfica (Figura 4.15) muestra el resultado que se obtuvo al realizar la misma prueba a la misma distancia.

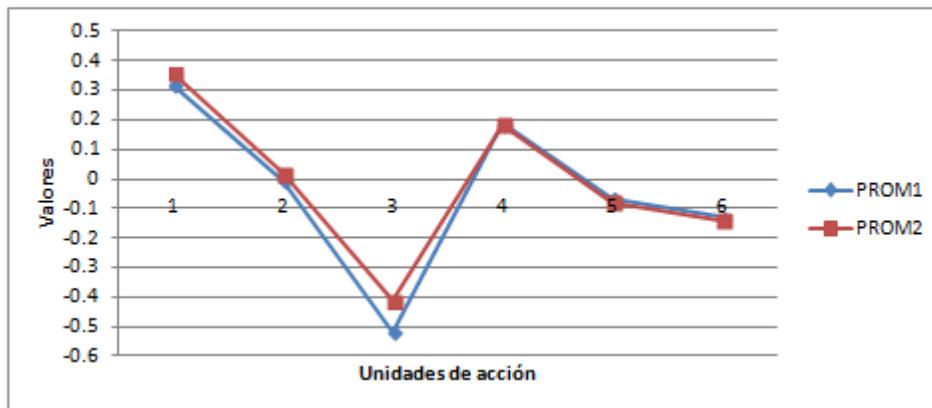


Figura 4.15 Gráfica del promedio de todas las AU's a la misma distancia (50 cm)

En la Figura 4.15. se puede observar que existen valores parecidos en las 6 AU (la diferencia es mínima de 0.05). El sujeto realizó el mismo movimiento a la misma distancia dos veces.

### Observaciones

Con esto se concluye que la distancia no afecta demasiado el resultado siempre y cuando el sujeto esté ubicado entre 60 - 120 cm, ya que como se observó en las gráficas (Figura 4.9 -4.15) entre más alejado está el sujeto respecto de la cámara el resultado es más inestable (la diferencia crece de 0.1 - 0.3, para cada valor de AU). La distancia adecuada para la captura es de 60 a 120 cm, es por ello que el sujeto se colocó a aproximadamente 1 m de distancia de la cámara para realizar las capturas.

#### 4.2.4 Prueba para obtener los valores de las 6 AU's de *facetracking*

##### Objetivo

Conocer cómo funcionan las unidades de acción de la librería *facetracking*.

##### Desarrollo

La librería *facetracking* reconoce las unidades de acción en el rostro de la persona (genera un valor para cada unidad de acción), las unidades de acción tienen valores de 0 (neutral), negativos y positivos, para cada AU cada valor tiene un significado

distinto en el movimiento de cada componente del rostro, los resultados se muestran en la Tabla 4.2.

Tabla 4.2 Cómo funcionan las unidades de acción con la librería *facetracking*

AU de Kinect	VALORES		
	Negativo (-1)	0	Positivo (1)
<b>AU[0] Labio superior</b>	Prensar Labios 	Neutral	Levantar el labio superior 
<b>AU[1] Mandíbula</b>	Boca cerrada - subir al máximo la mandíbula inferior 	Neutral	Boca totalmente abierta 
<b>AU[2] Distancia Horizontal de la Boca</b>	Mínima distancia horizontal 	Neutral	Máxima distancia horizontal 
<b>AU[3] Cejas de la parte interior</b>	Arriba 	Neutral	Abajo (fruncir ceño) 
<b>AU[4] Comisura de los labios</b>	Comisura de labios arriba 	Neutral	Comisura de labios abajo 
<b>AU[5] Cejas</b>	Abajo 	Neutral	Arriba 

Como se puede observar en la Tabla 4.2 las AU's tienen desde valores negativos hasta positivos, pero según la teoría de Paul Ekman [21] las unidades de acción están presentes o no, en el rostro, es decir, cada unidad de acción representa el movimiento de un músculo; por ejemplo, para AU[3] que representa las cejas, un músculo está encargado de elevar las cejas, mientras que otro músculo se encarga de bajar las cejas, de esta forma se tendrían dos unidades de acción (según Paul Ekman) y una unidad de acción (según *facetracking*). Esto significa que con las unidades de *facetracking* se pueden obtener otras unidades de acción.

### Observaciones

Las seis unidades de acción de *facetracking* representan valores desde -1 hasta 1, con incrementos entre ellos de 0.001, con estos valores se pueden obtener unidades de acción que se asemejen mejor a la teoría de Paul Ekman, utilizando unidades de acción que se activen si están presentes y en ausencia se mantengan inactivas.

### 4.2.5 Prueba para obtener las 15 AU's a partir de las 6 AU's de *facetracking*

#### Objetivo

Verificar que las 15 unidades de acción se activaban correctamente al realizar el movimiento correspondiente para cada unidad de acción.

#### Desarrollo

Es importante mencionar que las 15 unidades de acción que se presentan a continuación son parte de las 44 unidades de acción de Paul Eckman; se seleccionaron sólo 15 AU's porque están relacionadas con las cejas, la boca y la mandíbula inferior, ya que la librería de *facetracking* utiliza estos componentes del rostro para representar las expresiones faciales. Los trabajos de Kotsia & Pitas [25], Pantic & Rothkrantz [22] y de Zhang & Ji [24] utilizaron estas mismas AU's para representar las expresiones faciales básicas.

Se verificó que cada una de las AU esté activa, sin que otras se activen, para validar que el Kinect obtuviera los valores correctamente.

La Figura 4.16 está formada por las siguientes imágenes:

- La imagen de la esquina superior izquierda representa la AU1 (levantar las cejas de la parte interna) se puede observar que está activada, también está activa la AU16 porque la boca está neutral.
- La imagen de la esquina superior derecha representa la AU2 (levantar las cejas de la parte externa) de igual forma está activa.
- La imagen de la esquina inferior izquierda representa la AU4 (fruncir el ceño) está activa y también la AU6 (bajar las cejas), porque es muy complicado fruncir el ceño sin bajar las cejas de la parte externa.

- La imagen inferior derecha representa la AU6 (bajar las cejas de la parte externa) y está activada.

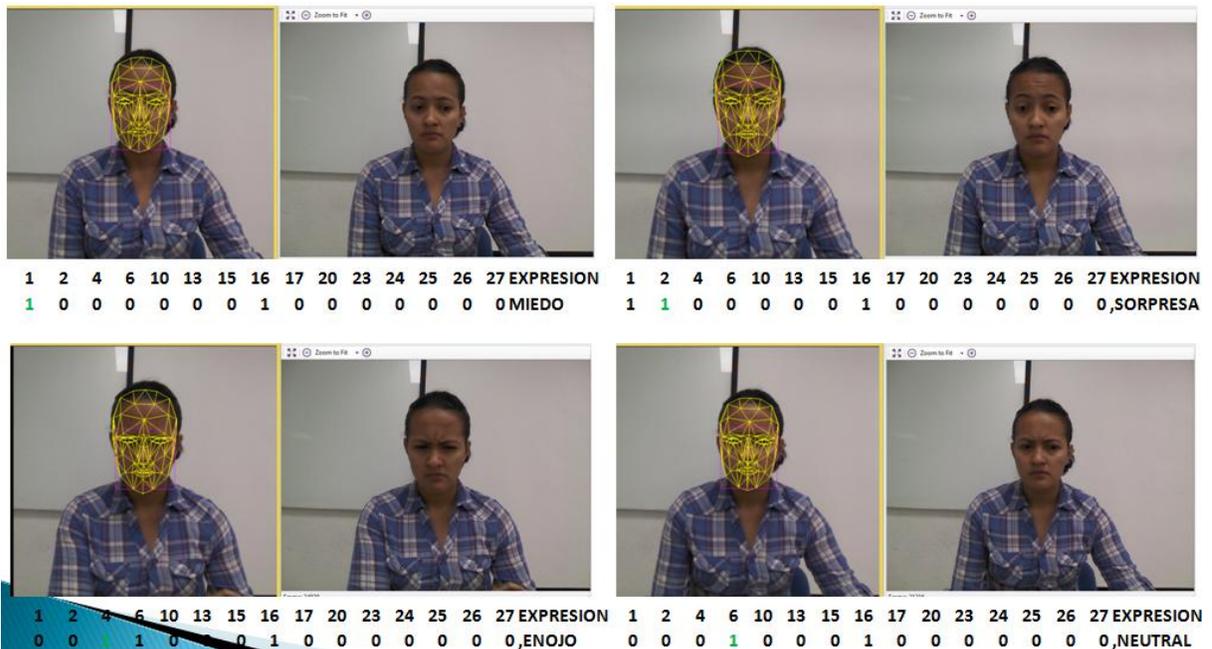


Figura 4.16 Esta imagen representa las AU's 1, 2, 4 y 6. Las 15 primeras columnas representan las AU's y la última columna la expresión detectada

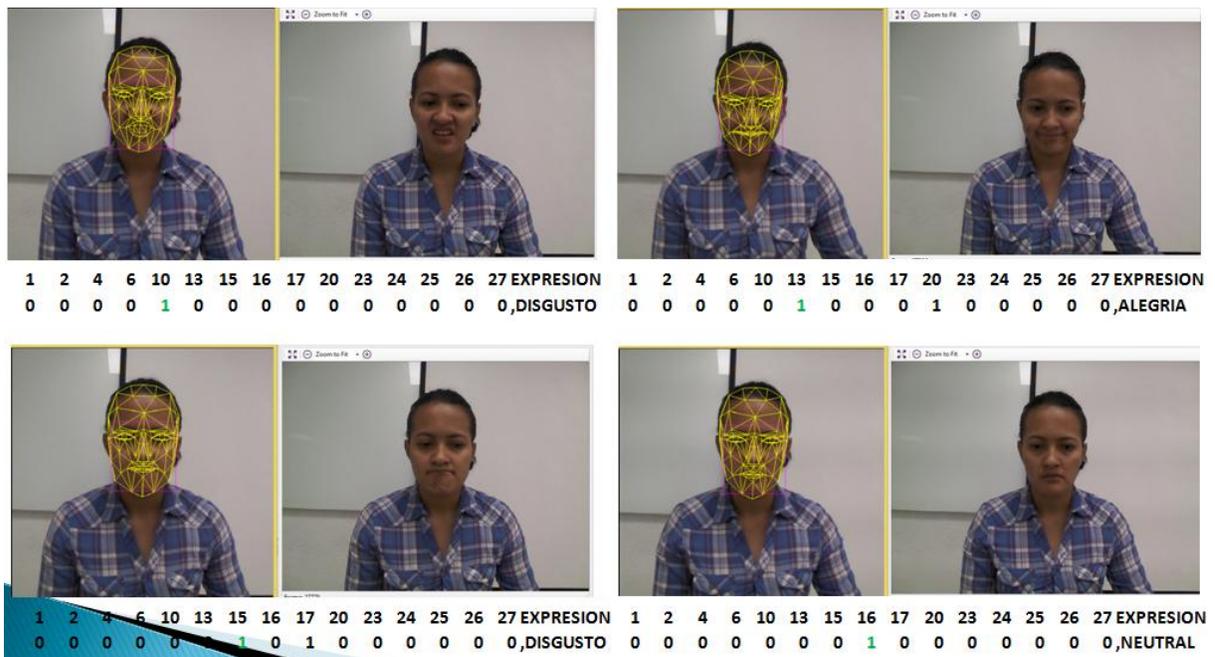


Figura 4.17 Esta imagen representa las AU's 10, 13, 15 y 16

La Figura 4.17 está conformada por las siguientes imágenes:

- La imagen superior izquierda representa la AU10 (levantar el labio superior) está activa.
- La imagen superior derecha representa la AU13 (comisura del labio hacia arriba) también está activa, se observa en esa misma imagen que la AU20 está activa esto se debe a que la distancia horizontal de la boca supera el rango y por lo tanto se activa (es muy complicado realizar este movimiento sin que se vea afectada la distancia horizontal de la boca).
- La imagen inferior izquierda representa la AU15 (comisura de los labios hacia abajo) activada, se puede observar que la AU17 está activa porque al realizar ese movimiento también se subió la mandíbula inferior.
- La imagen inferior derecha representa la AU16 (distancia normal de la boca) y está activa.

La Figura 4.18 está compuesta por la siguientes imágenes:

- La imagen superior izquierda representa la AU17 (mandíbula inferior hacia arriba) está activa, también está activada la AU15 por las comisuras de los labios están hacia abajo.
- La imagen superior derecha representa la AU20 (máxima distancia horizontal de la boca) la cual está activa.
- La imagen inferior izquierda representa la AU23 (prensar los labios sin llegar a tener la máxima distancia horizontal de la boca) está activa.
- La imagen inferior derecha representa la AU24 (distancia horizontal mínima de la boca) está activa.



Figura 4.18 Esta imagen representa las AUs 17, 20, 23 y 24

La Figura 4.19 está compuesta por las siguientes imágenes:

- La imagen superior izquierda representa la AU25 (bajar un poco la mandíbula inferior +) está activa.
- La imagen superior derecha representa la AU26 (baja la mandíbula inferior ++) está activa.
- La imagen inferior izquierda representa la AU27 (bajar completamente la mandíbula inferior +++) está activa.

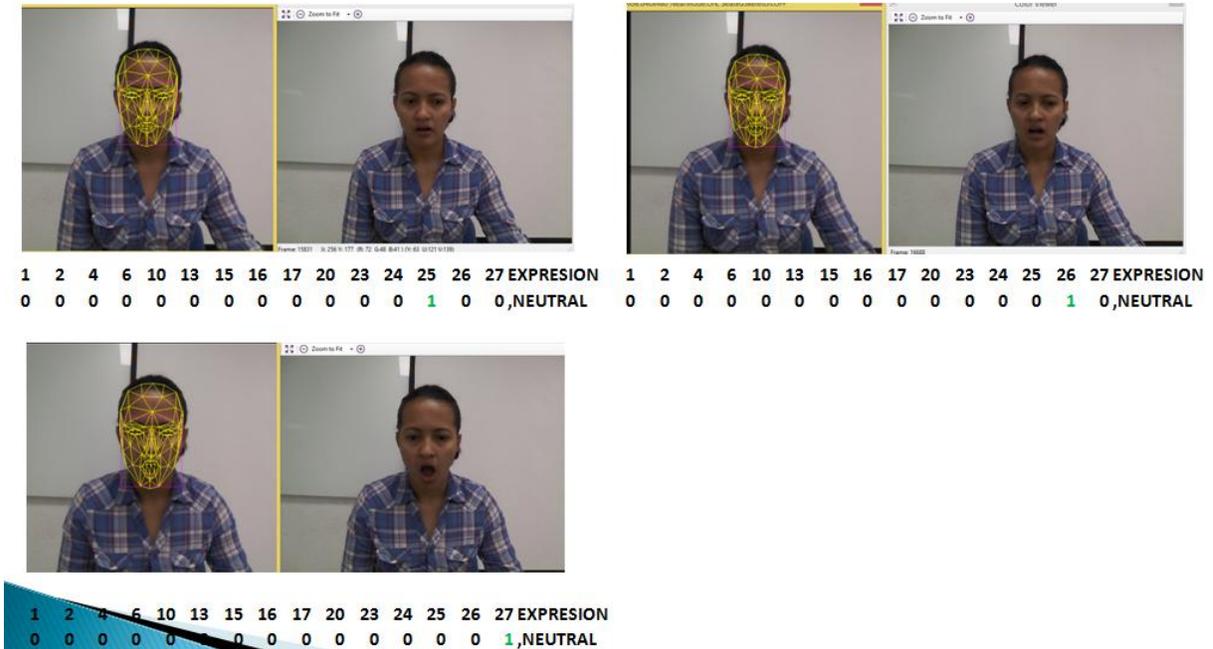


Figura 4.19 Esta imagen representa las AUs 25, 26 y 27

### Observaciones

Se detectó que las unidades de acción se activan adecuadamente realizando el movimiento respectivo para cada unidad de acción.

### 4.2.6 Prueba de clasificación

Se usaron distintos algoritmos de clasificación, para evaluar el resultado de cada uno de ellos y para verificar qué algoritmo de clasificación es el que mejor resultados obtiene para identificar las expresiones faciales básicas.

El modelo entrenado de cada uno de los algoritmos de clasificación se generó con el 80% de los patrones totales y se validó con el 20%.

#### 4.2.6.1 Prueba con MSV Lineal

##### Objetivo

Validar la eficacia de la máquina de soporte vectorial de kernel lineal.

##### Desarrollo

La máquina de soporte vectorial de kernel lineal realiza la clasificación de la forma más sencilla, donde la frontera de separación entre clases está dada por líneas rectas. El resultado de este algoritmo se muestra en la Tabla 4.3.

Tabla 4.3 Matriz de confusión para MVS Lineal (etapa de reconocimiento)

	ALE.	TRIS.	DIS.	SOR.	ENO.	MIE.	NEU.	TOTAL
ALEGRIA	44	<b>81.82</b>	4.55	0.00	4.55	6.82	2.27	0.00
TRISTEZA	24	0.00	<b>83.33</b>	0.00	4.17	0.00	12.50	0.00
DISGUSTO	23	0.00	0.00	<b>95.65</b>	0.00	4.35	0.00	0.00
SORPRESA	24	0.00	0.00	0.00	<b>100.00</b>	0.00	0.00	0.00
ENOJO	27	0.00	7.41	3.70	7.41	<b>77.78</b>	3.70	0.00
MIEDO	18	11.11	0.00	0.00	0.00	5.56	<b>83.33</b>	0.00
NEUTRAL	2	0.00	0.00	0.00	0.00	100.00	0.00	<b>0.00</b>
TOTAL								<b>74.56</b>

##### Observaciones

El porcentaje promedio para la MVS lineal fue de 74.56%, no logró clasificar la expresión neutral y para la expresión enojo el porcentaje fue muy bajo. La expresión de sorpresa fue del 100% y las otras expresiones tuvieron un resultado promedio de 80%.

#### 4.2.6.2 Prueba con MSV Sigmoide

##### Objetivo

Validar la eficacia de la máquina de soporte vectorial de kernel sigmoide.

##### Desarrollo

La máquina de soporte vectorial de kernel sigmoide utiliza una función sigmoidea para realizar la clasificación. Esta función se modifica y se afina dependiendo del problema a resolver, los valores iniciales que se usaron fueron:

Para el  $\gamma = 1$ .

El valor de coeficiente  $c_0 = 1$  (este valor se modificó con incrementos de 0.1, pero el resultado no mejoró).

Iteraciones = 10'000,000.

El resultado de este algoritmo se muestra en la Tabla 4.4.

Tabla 4.4 Matriz de confusión para la MVS Sigmoide (etapa de reconocimiento)

	ALE.	TRIS.	DIS.	SOR.	ENO.	MIE.	NEU.	TOTAL
ALEGRIA	44	<b>47.73</b>	0.00	27.27	25.00	0.00	0.00	0.00
TRISTEZA	24	100.00	<b>0.00</b>	0.00	0.00	0.00	0.00	0.00
DISGUSTO	23	65.22	0.00	<b>0.00</b>	34.78	0.00	0.00	0.00
SORPRESA	24	100.00	0.00	0.00	<b>0.00</b>	0.00	0.00	0.00
ENOJO	27	96.30	0.00	0.00	3.70	<b>0.00</b>	0.00	0.00
MIEDO	18	88.89	0.00	11.11	0.00	0.00	<b>0.00</b>	0.00
NEUTRAL	2	100.00	0.00	0.00	0.00	0.00	0.00	<b>0.00</b>
TOTAL								<b>6.82</b>

### Observaciones

El resultado de este algoritmo (MVS sigmoide) fue el más bajo de todos los algoritmos de clasificación implementados en esta tesis, no se encontró un valor inicial adecuado para llegar a un mejor resultado.

### 4.2.6.3 Prueba con MSV de base radial

#### Objetivo

Validar la eficacia de la máquina de soporte vectorial de kernel de base radial.

#### Desarrollo

La máquina de soporte vectorial de kernel de base radial clasifica problemas no lineales. Este kernel utiliza el valor gamma (gamma = 1) dentro de su ecuación. Iteraciones = 10'000,000.

El resultado de este algoritmo se muestra en la Tabla 4.5.

Tabla 4.5 Matriz de confusión para la MVS de base radial (etapa de reconocimiento)

	ALE.	TRIS.	DIS.	SOR.	ENO.	MIE.	NEU.	TOTAL
ALEGRIA	44	<b>81.82</b>	4.55	0.00	4.55	6.82	2.27	0.00
TRISTEZA	24	0.00	<b>83.33</b>	0.00	4.17	0.00	12.50	0.00
DISGUSTO	23	0.00	0.00	<b>95.65</b>	0.00	4.35	0.00	0.00
SORPRESA	24	0.00	0.00	0.00	<b>100.00</b>	0.00	0.00	0.00
ENOJO	27	0.00	7.41	3.70	7.41	<b>77.78</b>	3.70	0.00
MIEDO	18	11.11	0.00	0.00	0.00	5.56	<b>83.33</b>	0.00
NEUTRAL	2	0.00	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>
TOTAL								<b>88.84</b>

### Observaciones

La MVS de base radial obtuvo el mejor resultado de todos los algoritmos que se implementaron. El porcentaje más bajo es para la expresión facial enojo con 77.78%.



El resultado mostrado en la Tabla 4.6 es la validación del algoritmo J48 con los patrones de prueba.

### Observaciones

El árbol J48 fue el segundo mejor algoritmo de clasificación dentro de los algoritmos que se implementaron para clasificar las expresiones faciales. El porcentaje de clasificación más bajo fue para la expresión facial enojo con 70.37%. Las otras expresiones faciales tienen un porcentaje mayor a 81%. La diferencia del árbol J48 con la MSV de base radial (este algoritmo es el que obtuvo el mejor desempeño) se debe a que el árbol J48 tiene un menor porcentaje de clasificación para la expresión facial enojo.

#### 4.2.6.5 Prueba con árbol de decisión REPTree

##### Objetivo

Validar la eficacia del algoritmo de clasificación REPTree.

##### Desarrollo

El algoritmo REPTree (por sus siglas en inglés Reduced Error Pruning Tree - Árbol con poda de error reducida), construye un árbol de decisión/regresión utilizando información de ganancia/varianza para la reducción de errores en la poda del árbol (con *backfitting*).

El modelo entrenado del algoritmo REPTree se obtuvo con el programa Weka (versión 3.6), el árbol REPTree está distribuido de la siguiente manera (Figura 4.21).

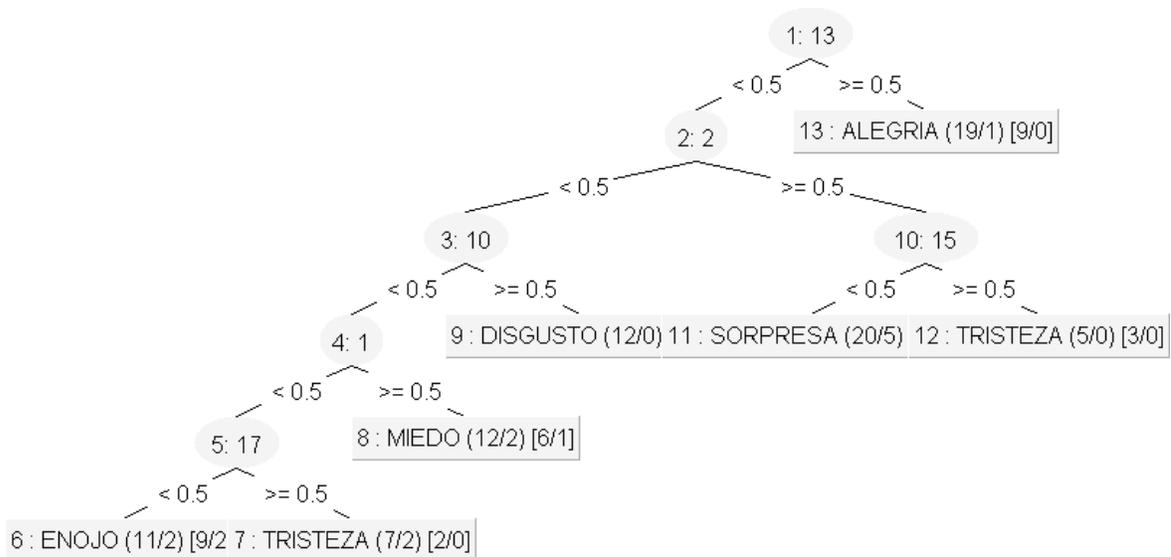


Figura 4.21 Distribución del árbol REPTree

Se utilizó el algoritmo REPTree, ya que durante el entrenamiento para generar el modelo se obtuvo el 100% de instancias correctamente clasificadas en promedio general. El resultado de este algoritmo al validarlo con los patrones de prueba se muestra en la Tabla 4.7:

Tabla 4.7 Matriz de confusión para el árbol REPTree (etapa de reconocimiento)

	ALE.	TRIS.	DIS.	SOR.	ENO.	MIE.	NEU.	TOTAL
ALEGRIA	44	<b>81.82</b>	4.55	0.00	9.09	2.27	2.27	0.00
TRISTEZA	24	0.00	<b>62.50</b>	0.00	4.17	8.33	25.00	0.00
DISGUSTO	23	0.00	8.70	<b>82.61</b>	4.35	4.35	0.00	0.00
SORPRESA	24	0.00	0.00	0.00	<b>100.00</b>	0.00	0.00	0.00
ENOJO	27	0.00	3.70	3.70	25.93	<b>59.26</b>	7.41	0.00
MIEDO	18	11.11	0.00	0.00	0.00	5.56	<b>83.33</b>	0.00
NEUTRAL	2	0.00	0.00	0.00	0.00	100.00	0.00	<b>0.00</b>
TOTAL								<b>67.07</b>

### Observaciones

Este algoritmo REPTree obtuvo un porcentaje de 67% de promedio general, dentro de los algoritmos de clasificación implementados para clasificar las expresiones faciales. No reconoció la expresión neutral y reconoció totalmente los patrones para la expresión de sorpresa.

#### 4.2.6.6 Prueba con algoritmos de reglas NNge

##### Objetivo

Validar la eficacia del algoritmo de clasificación de reglas NNge.

##### Desarrollo

El algoritmo NNge (Non-Nested With Generalization Exemplars - Ejemplares no anidados con generalización) realiza la generalización mediante la fusión de ejemplares, formando hiper-rectángulos<sup>4</sup> en el espacio de características que representan reglas de combinación con conjunción interna. NNge forma una generalización cada vez que se añade un nuevo ejemplar para la base de datos, mediante la unión a su vecino más cercano de la misma clase [36].

El modelo entrenado del algoritmo NNge se obtuvo con el programa Weka (versión 3.6), el algoritmo NNge generó las siguientes reglas (Figura 4.22):

<sup>4</sup> Un hiper-rectángulo también llamado caja, para este algoritmo NNge es un almacén de objetos de n-espacios euclidianos [36].

```

class NEUTRAL IF : 1=0.0 ^ 2=0.0 ^ 4=0.0 ^ 6=0.0 ^ 10=0.0 ^ 13=0.0 ^ 15=0.0 ^
0.0<=16<=1.0 ^ 17=0.0 ^ 20=0.0 ^ 23=0.0 ^ 24=0.0 ^ 25=0.0 ^ 26=0.0 ^ 27=0.0 (2)
class MIEDO IF : 1=1.0 ^ 2=0.0 ^ 4=0.0 ^ 0.0<=6<=1.0 ^ 10=0.0 ^ 13=0.0 ^ 15=0.0 ^
0.0<=16<=1.0 ^ 17=0.0 ^ 0.0<=20<=1.0 ^ 0.0<=23<=1.0 ^ 24=0.0 ^ 0.0<=25<=1.0 ^
0.0<=26<=1.0 ^ 27=0.0 (15)
class MIEDO IF : 1=1.0 ^ 2=1.0 ^ 4=0.0 ^ 6=0.0 ^ 10=0.0 ^ 13=1.0 ^ 15=0.0 ^ 16=0.0
^ 17=0.0 ^ 20=1.0 ^ 23=0.0 ^ 24=0.0 ^ 25=0.0 ^ 26=0.0 ^ 27=0.0 (1)
class ALEGRIA IF : 0.0<=1<=1.0 ^ 2=0.0 ^ 4=0.0 ^ 0.0<=6<=1.0 ^ 0.0<=10<=1.0 ^
13=1.0 ^ 15=0.0 ^ 16=0.0 ^ 0.0<=17<=1.0 ^ 0.0<=20<=1.0 ^ 23=0.0 ^ 24=0.0 ^
0.0<=25<=1.0 ^ 0.0<=26<=1.0 ^ 27=0.0 (17)
class ALEGRIA IF : 1=1.0 ^ 2=1.0 ^ 4=0.0 ^ 6=0.0 ^ 10=1.0 ^ 13=1.0 ^ 15=0.0 ^
16=0.0 ^ 17=0.0 ^ 20=1.0 ^ 23=0.0 ^ 24=0.0 ^ 25=0.0 ^ 26=0.0 ^ 27=0.0 (3)
class ENOJO IF : 1=0.0 ^ 0.0<=2<=1.0 ^ 4=0.0 ^ 6=0.0 ^ 10=0.0 ^ 13=0.0 ^ 15=0.0 ^
16=0.0 ^ 17=0.0 ^ 20=0.0 ^ 23=1.0 ^ 24=0.0 ^ 25=0.0 ^ 26=0.0 ^ 27=0.0 (3)
class ENOJO IF : 1=0.0 ^ 2=0.0 ^ 4=1.0 ^ 0.0<=6<=1.0 ^ 10=0.0 ^ 13=0.0 ^ 15=0.0 ^
0.0<=16<=1.0 ^ 17=0.0 ^ 20=0.0 ^ 0.0<=23<=1.0 ^ 0.0<=24<=1.0 ^ 0.0<=25<=1.0 ^
0.0<=26<=1.0 ^ 27=0.0 (14)
class ENOJO IF : 1=0.0 ^ 2=1.0 ^ 4=1.0 ^ 6=0.0 ^ 10=0.0 ^ 13=0.0 ^ 15=0.0 ^ 16=0.0
^ 17=0.0 ^ 20=0.0 ^ 23=0.0 ^ 24=0.0 ^ 25=0.0 ^ 26=0.0 ^ 27=0.0 (2)
class ENOJO IF : 1=0.0 ^ 2=1.0 ^ 0.0<=4<=1.0 ^ 6=0.0 ^ 10=0.0 ^ 13=0.0 ^ 15=0.0 ^
16=0.0 ^ 17=0.0 ^ 20=0.0 ^ 23=0.0 ^ 24=1.0 ^ 25=0.0 ^ 26=0.0 ^ 27=0.0 (2)
class SORPRESA IF : 0.0<=1<=1.0 ^ 2=1.0 ^ 4=0.0 ^ 6=0.0 ^ 10=0.0 ^ 13=0.0 ^ 15=0.0
^ 0.0<=16<=1.0 ^ 17=0.0 ^ 20=0.0 ^ 23=0.0 ^ 24=0.0 ^ 0.0<=25<=1.0 ^ 0.0<=26<=1.0 ^
0.0<=27<=1.0 (22)
class DISGUSTO IF : 0.0<=1<=1.0 ^ 0.0<=2<=1.0 ^ 0.0<=4<=1.0 ^ 0.0<=6<=1.0 ^ 10=1.0
^ 13=0.0 ^ 15=0.0 ^ 16=0.0 ^ 0.0<=17<=1.0 ^ 20=0.0 ^ 0.0<=23<=1.0 ^ 24=0.0 ^
0.0<=25<=1.0 ^ 0.0<=26<=1.0 ^ 27=0.0 (20)
class DISGUSTO IF : 1=0.0 ^ 2=0.0 ^ 4=1.0 ^ 6=1.0 ^ 10=0.0 ^ 13=0.0 ^ 15=0.0 ^
16=0.0 ^ 17=1.0 ^ 20=0.0 ^ 23=0.0 ^ 24=0.0 ^ 25=0.0 ^ 26=0.0 ^ 27=0.0 (2)
class TRISTEZA IF : 0.0<=1<=1.0 ^ 2=0.0 ^ 0.0<=4<=1.0 ^ 6=0.0 ^ 10=0.0 ^ 13=0.0 ^
15=0.0 ^ 16=0.0 ^ 17=1.0 ^ 20=0.0 ^ 23=0.0 ^ 24=0.0 ^ 25=0.0 ^ 26=0.0 ^ 27=0.0
class TRISTEZA IF : 0.0<=1<=1.0 ^ 0.0<=2<=1.0 ^ 0.0<=4<=1.0 ^ 0.0<=6<=1.0 ^ 10=0.0
^ 13=0.0 ^ 15=1.0 ^ 0.0<=16<=1.0 ^ 0.0<=17<=1.0 ^ 20=0.0 ^ 23=0.0 ^ 0.0<=24<=1.0 ^
25=0.0 ^ 26=0.0 ^ 27=0.0 (17)

```

Figura 4.22 Reglas generadas por el algoritmo NNge

Se utilizó el algoritmo NNge, ya que durante el entrenamiento para generar el modelo se obtuvo el 97% en promedio general de instancias correctamente clasificadas. El resultado de este algoritmo al validarlo con los patrones de prueba se muestra en la Tabla 4.8.

### Observaciones

Este algoritmo obtuvo un porcentaje de 86.54% en promedio general, el porcentaje más bajo de clasificación fue para la expresión alegría con 54.55% y el más alto fue de 100% para las expresiones de sorpresa y neutral.

Es importante mencionar que este algoritmo genera reglas, si algún patrón no se encuentra dentro de estas reglas, no se obtiene ninguna respuesta. En la lista de patrones de prueba, el algoritmo NNge no dio resultados para algunos patrones.

Tabla 4.8 Matriz de confusión del algoritmo NNge (etapa de reconocimiento)

	ALE.	TRIS.	DIS.	SOR.	ENO.	MIE.	NEU.	TOTAL
ALEGRIA	44	<b>54.55</b>	0.00	0.00	4.55	2.27	29.55	0.00
TRISTEZA	24	0.00	<b>83.33</b>	0.00	4.17	0.00	8.33	0.00
DISGUSTO	23	0.00	0.00	<b>95.65</b>	0.00	4.35	0.00	0.00
SORPRESA	24	0.00	0.00	0.00	<b>100.00</b>	0.00	0.00	0.00
ENOJO	27	0.00	7.41	3.70	7.41	<b>77.78</b>	3.70	0.00
MIEDO	18	0.00	0.00	0.00	0.00	0.00	<b>94.44</b>	0.00
NEUTRAL	2	0.00	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>
TOTAL								<b>86.54</b>

#### 4.2.6.7 Prueba con algoritmos de reglas PART

##### Objetivo

Validar la eficacia del algoritmo de clasificación de reglas PART.

##### Desarrollo

El algoritmo PART (PART significa parcial por sus siglas en inglés partial) genera una lista de decisiones de cada clase. Utiliza el algoritmo "separa y vencerás". Construye un árbol de decisión C4.5 parcial en cada iteración y hace que el mejor nodo de cada clase del árbol sea una regla [37].

El modelo entrenado con el algoritmo PART se obtuvo con el programa Weka (versión 3.6), el algoritmo PART generó las siguientes reglas (Figura 4.23):

PART decision list	
13 > 0: ALEGRIA (28.0/1.0)	17 <= 0 AND 4 > 0: ENOJO (17.0)
10 > 0: DISGUSTO (20.0)	15 <= 0 AND 1 > 0: MIEDO (15.0)
17 > 0 AND 15 > 0: TRISTEZA (14.0)	6 <= 0 AND 15 > 0: TRISTEZA (3.0)
17 > 0 AND 6 <= 0: TRISTEZA (3.0)	4 <= 0 AND 23 <= 0: NEUTRAL (2.0)
2 > 0 AND 4 <= 0: SORPRESA (24.0/2.0)	4 <= 0: ENOJO (2.0)
15 <= 0 AND 1 <= 0 AND	

Figura 4.23 Reglas generadas con el algoritmo PART

Se utilizó el algoritmo PART, ya que durante el entrenamiento para generar el modelo se obtuvo el 98% de instancias correctamente clasificadas en promedio

general. El resultado de este algoritmo con los patrones de prueba se muestra en la Tabla 4.9.

Tabla 4.9 Matriz de confusión del algoritmo PART

	ALE.	TRIS.	DIS.	SOR.	ENO.	MIE.	NEU.	TOTAL
ALEGRIA	44	<b>81.82</b>	2.27	2.27	4.55	6.82	2.27	0.00
TRISTEZA	24	0.00	<b>83.33</b>	0.00	4.17	0.00	12.50	0.00
DISGUSTO	23	0.00	0.00	<b>95.65</b>	0.00	4.35	0.00	0.00
SORPRESA	24	0.00	0.00	0.00	<b>100.00</b>	0.00	0.00	0.00
ENOJO	27	0.00	7.41	3.70	14.81	<b>70.37</b>	3.70	0.00
MIEDO	18	11.11	0.00	0.00	0.00	0.00	<b>83.33</b>	5.56
NEUTRAL	2	0.00	0.00	0.00	0.00	0.00	0.00	<b>100.00</b>
TOTAL								<b>87.79</b>

### Observaciones

Este algoritmo obtuvo un porcentaje de 87.79% en promedio general, el porcentaje más bajo de clasificación fue para la expresión enojo con 70.37% y el más alto fue de 100% para las expresiones de sorpresa y neutral.

El algoritmo PART generó respuesta para todos los patrones a diferencia del algoritmo NNge, que para algunos patrones no generó respuesta; este algoritmo tuvo mejor desempeño que el algoritmo NNge.

### 4.3 Tabla de matriz cruzada con las pruebas y objetivos

La Tabla 4.10 muestra una matriz cruzada entre las pruebas que se realizaron y los objetivos que se cumplieron con dichas pruebas, las X marcan las pruebas que cumplen con dichos objetivos.

Tabla 4.10 Matriz cruzada entre objetivos y pruebas realizadas

Objetivos	Pruebas					
	4.2.1	4.2.2	4.2.3	4.2.4	4.2.5	4.2.6
Estudio del SDK de Kinect: generación de los datos asociados a las unidades de acción	X	X	X	X	X	
Estudiar las FACS, para seleccionar las unidades de acción necesarias para cada expresión facial			X	X	X	
Estudiar las técnicas para el reconocimiento de expresiones faciales			X	X	X	X
Seleccionar las técnicas que mejor desempeño tengan, según el estado del arte, para reconocer las expresiones faciales				X	X	X
Implementación de las técnicas seleccionadas, probarlas, validarlas y afinarlas para este problema de investigación				X	X	X

En la Tabla 4.10 se observa la relación entre las pruebas con los objetivos planteados en esta tesis. Se muestran actividades que no se tenían contempladas, el sensor presento problemas para capturar los datos, fue por eso que se realizaron las pruebas de iluminación, para tener una condición adecuada para capturar los datos.

La prueba con imágenes 2D se realizó porque el sensor puede capturar datos de estas imágenes, cabe mencionar que gracias a esto se logró complementar la base de datos con los patrones de pruebas para validar el buen funcionamiento del sistema.

#### 4.4 Discusión

Como se puede observar en la gráfica (Figura 4.24) el mejor resultado es para la máquina de soporte vectorial de Kernel de base radial. El algoritmo J48 tiene resultados cercanos a 88.84%. El porcentaje más bajo es el que se obtuvo con la máquina de soporte vectorial de Kernel Sigmoide.

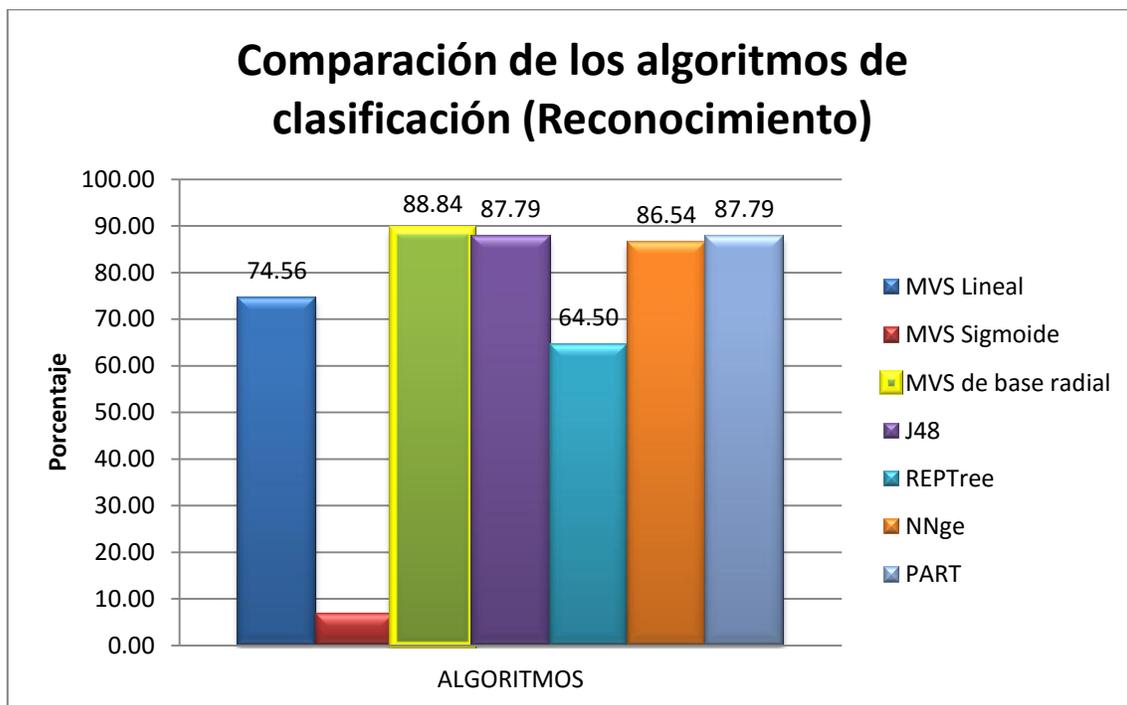


Figura 4.24 Comparación entre los algoritmos de clasificación

En conclusión, el algoritmo que mejor porcentaje obtuvo en la clasificación para los datos de prueba para el desarrollo de este proyecto, fue la máquina de soporte vectorial de Kernel de base radial.

La Figura 4.25 tiene el resultado de la MSV de base radial de las etapas de entrenamiento y reconocimiento, ya que obtuvo el mejor resultado. En la gráfica se

observa que sólo la clase MIEDO obtuvo un error (93.75%) de clasificación en la etapa de entrenamiento, las otras clases tienen un 100% de instancias correctamente clasificadas para la etapa de entrenamiento.

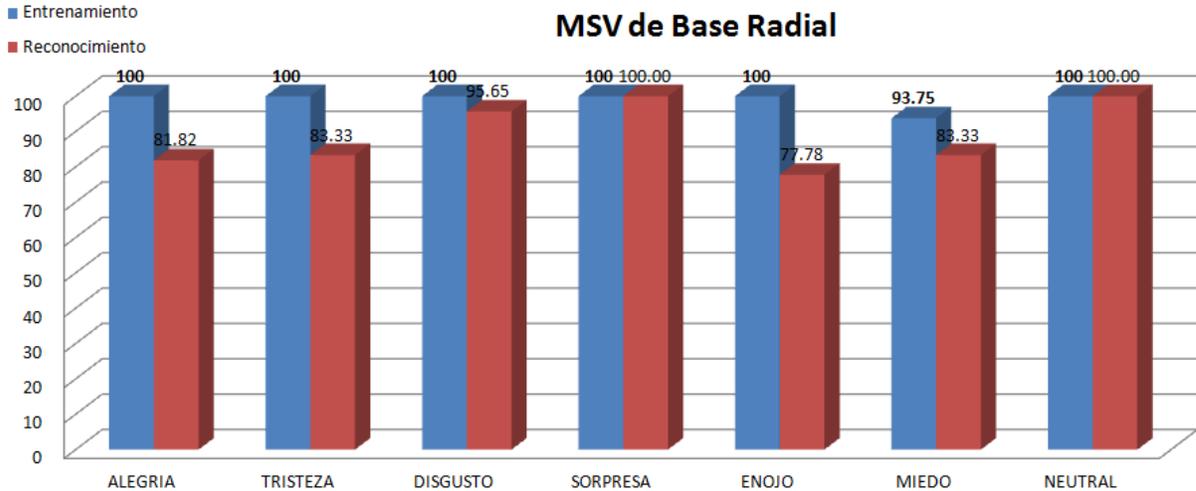


Figura 4.25 MSV de base radial con los resultados de las etapas de entrenamiento y de reconocimiento

También es importante mencionar que el sistema de la cámara RGBD a veces obtiene falsos positivos, esto quiere decir que, el sistema de la cámara RGBD erróneamente activa unidades de acción que en realidad no están presentes en las imágenes y con ello se crean patrones incorrectos para cierta expresión facial. Se identificó que el problema de detección del rostro se debe a la presencia de sombras en el rostro.

La Figura 4.26 muestra el patrón con clasificación como una expresión facial de enojo, en la cual se puede observar que tienen activadas las AUs 2 (levantar las cejas), 4 (fruncir el ceño) y 25 (abrir levemente la boca); se observa que el sujeto no está frunciendo el ceño, pero sí levanta levemente las cejas, el sistema de la cámara RGBD está dando datos erróneos.



Figura 4.26 Ejemplo de un patrón incorrecto

Si se analiza sólo los datos de la Tabla 4.11 se observa por ejemplo que en la variable  $cej = -0.135552$  la cual indica el valor neutral de la parte exterior de las cejas; según en la imagen (Figura 4.26) el valor para AU[5] (parte exterior de las cejas) fue de 0.104801. En la Tabla 4.2, si el valor de AU[5] es mayor significa que la parte exterior de las cejas suben, este valor es correcto en la imagen (Figura 4.26), la parte exterior de las cejas se elevaron un poco y por lo tanto está activa la casilla 2 de la Figura 4.26, por lo tanto los **datos del sensor son correctos**.

Tabla 4.11 Datos de Kinect del sujeto de la Figura 4.21

AU[0]	AU[1]	AU[2]	AU[3]	AU[4]	AU[5]	ceint	cej
0.496249	0.007987	0.449958	0.242564	-0.442782	0.104801	0.240818	-0.135552

Ahora se analizará la AU[3] (parte interior de las cejas), la variable  $ceint = 0.240818$  la cual indica el valor neutral de la parte interior de las cejas; según la Tabla 4.11 el valor para AU[3] (parte interior de las cejas) fue de 0.242564. En la Tabla 4.2, si el valor de AU[3] es mayor significa que la parte interior de las cejas bajan, pero en la imagen la parte interior de las cejas no están abajo, por lo tanto los **datos del sensor son erróneos**.

Al analizar el conjunto de patrones de prueba se encontraron casos erróneos como en la Figura 4.26, ya que se verificó cada imagen con su respectivos datos del sensor y se detectaron más casos erróneos. En la Tabla 4.12 se observan los casos erróneos para cada expresión facial.

Tabla 4.12 Datos erróneos para cada una de las expresiones faciales

EXPRESIÓN	CASOS ERRONEOS
ALEGRÍA	8
TRISTEZA	4
DISGUSTO	1
SORPRESA	0
ENOJO	6
MIEDO	2

El total de casos erróneos de la Tabla 4.12 es de 21, lo cual significa que son el 12% del total de patrones (Tabla 3.3).

Se verificó cada una de las imágenes con sus respectivos datos del sensor para cada uno de los casos erróneos. Se identificó que la malla *Candide* no se colocó bien sobre el rostro, en estos casos la malla no captó correctamente el movimiento de las cejas o el abrir de la boca y es por ello que los datos no son los correctos.

Si se descartan esos casos erróneos (12%) donde el sensor proporcionó datos incongruentes con su respectiva imagen, podría significar que tener un porcentaje

de 88% (MVS de kernel de base radial) se puede deber a que el sensor captura datos erróneamente en algunos casos.

Para reducir este problema se podría utilizar otros puntos de la malla *candidate*, como las unidades de forma [23], utilizar el contorno del rostro y obtener distancias a partir de otros puntos característicos.

## Capítulo 5

### 5 Conclusiones generales y trabajos futuros

En este capítulo se presentan los objetivos alcanzados, aportaciones, las conclusiones finales y los trabajos futuros del presente trabajo.

#### 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 "identificar las expresiones faciales básicas capturadas por un sensor RGBD, usando como base de conocimiento las Unidades de Acción de las FACS sin que el sujeto en estudio tenga la necesidad de exagerar la expresión".

El objetivo general se cumplió en su totalidad al presentar pruebas de reconocimiento de las expresiones faciales antes mencionadas con base en un conjunto de características de los componentes faciales bajo estudio. Estas características se basan en la identificación de las unidades de acción.

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 de excel que contiene el descriptor-EF (15 unidades de acción), las unidades de acción facial (AU's) en bruto de *facetracking*, la expresión facial presentes de cada *frame* y el tiempo (milisegundo) de la captura de cada *frame*.

Los detalles de los objetivos específicos alcanzados se explican en la Tabla 5.1.

Tabla 5.1 Resultados de los objetivos específicos

Objetivo	Comentarios
Estudio del SDK de Kinect: generación de los datos asociados a las unidades de acción.	Se realizaron seis pruebas finales para verificar el funcionamiento del sensor RGBD. Se obtuvieron los datos de las unidades de acción a partir la librería <i>facetracking</i> .

Objetivo	Comentarios
Estudiar las FACS, para seleccionar las unidades de acción necesarias para cada expresión facial.	Se seleccionaron 15 unidades de acción relacionadas con las cejas, boca y mandíbula inferior, estas 15 unidades se obtuvieron a partir de las 6 unidades de acción de <i>facetracking</i> .
Estudiar las técnicas para el reconocimiento de expresiones faciales.	La librería <i>facetracking</i> tiene funciones para la detección de rostro y extracción de características (unidades de acción). Se utilizaron siete algoritmos de clasificación para clasificar las expresiones faciales.
Seleccionar las técnicas que mejor desempeño tengan, según el estado del arte, para reconocer las expresiones faciales.	Se implementaron siete algoritmos de clasificación para identificar las expresiones faciales básicas usando Opencv para los algoritmos de MSV y Weka para los algoritmos de reglas y árboles de decisión.
Implementación de las técnicas seleccionadas, probarlas, validarlas y afinarlas para este problema de investigación.	Se evaluaron los resultados de los siete algoritmos de clasificación. Se evaluó el desempeño del reconocimiento del rostro y sus componentes y de las unidades de acción.

## 5.2 Alcances

A continuación se presentan los alcances que se presentaron en la propuesta, los cuales se cumplieron en el desarrollo de esta tesis (Tabla 5.2).

Tabla 5.2 Alcances de la tesis

Alcances	Comentarios
Realizar un procesamiento de secuencia de imágenes RGBD con las expresiones faciales básicas	Se realizaron pruebas de iluminación, se detectó que el sensor reconoce imágenes planas
Identificar en las imágenes las unidades de acción (AU26, AU20, AU15, AU10, AU4, AU2), estas son proporcionadas por la librería <i>Facetracking</i> .	Se detectaron las seis unidades de acción por medio de <i>facetracking</i>
Identificar además 15 unidades de acción para formar el descriptor-EF, con este descriptor se entrenan y prueban los algoritmos de clasificación para obtener las expresiones faciales	Se identificaron las 15 unidades de acción a partir de <i>facetracking</i> y con estas 15 AUs se generó el descriptor-EF

Alcances	Comentarios
Se seleccionó una técnica para clasificar las expresiones faciales básicas con sus respectivas unidades de acción	Se usaron los algoritmos de clasificación: MVS (base radial, sigmoide y lineal), J48, REPTree, PART y NNge.
Crear una base de datos del CENIDET (10 personas), también se utilizó una base de datos llamada Cohn-Kanade Extend [27] de la cual se seleccionaron 30 personas para complementar la base de datos.	Base de datos compilada, con un total de 40 sujetos, la distribución de las expresiones faciales se encuentra en la Tabla 3.3
Detección de las expresiones faciales en tiempo real	Se considera tiempo real para este sistema el procesar a 30 fps

### 5.3 Aportaciones

Se desarrolló un sistema de reconocimiento de expresiones faciales que identifica las expresiones en tiempo real y también crea un archivo con el historial de las expresiones identificadas.

Se compiló una base de datos, con imágenes 2D y 3D, con 40 sujetos en total. La base de datos cuenta con el fotograma de la expresión, su respectivo patrón y una captura del rostro con la malla *Candide*.

Se realizó una implementación para convertir las 6 AUs de *facetracking* en 15 AUs, esto con el fin de tener un mejor desempeño en la clasificación de las expresiones faciales.

Se implementaron los algoritmos de clasificación con Weka: REPTree, J48, NNge, PART; la Máquina de soporte vectorial con kernels lineal, de base radial y sigmoide se implementó a partir de la librería OpenCV.

El sistema finalmente reconoce expresiones faciales sin que el usuario tenga que exagerar la expresión facial, ya que el sistema cuenta con más rangos y no es necesario llegar a tener valores cercanos a -1 o 1 para cada AU (ver Figuras 3.10 y 3.11).

### 5.4 Conclusiones

En este trabajo de investigación se estudiaron y analizaron las herramientas de desarrollo de Kinect para realizar la identificación y seguimiento del rostro, se estudiaron e implementaron los algoritmos de clasificación para identificar las expresiones faciales básicas. El sistema implementado puede capturar imágenes 2D y 3D del rostro usando la cámara RGBD.

El sistema de visión desarrollado puede identificar las expresiones faciales básicas sin que el sujeto en estudio tenga que exagerar las expresiones, esto quiere decir que, hablando en términos de los valores extremos que obtiene el sensor (-1 a 1), el sujeto en estudio no tiene que llegar a estos extremos para que el sistema pueda identificar una expresión facial (ver Figura 3.10 y Figura 3.11).

La base de datos compilada entre los sujetos que observaron el guión y los sujetos que se tomaron de la base de datos Cohn Kanade Extend, está compuesta por los datos de 40 personas, cada sujeto tiene: el fotograma de la expresión, una captura con la malla *Candide* sobre el rostro y un archivo con su respectivo patrón.

Se identificó que es necesario tener un ambiente controlado para capturar adecuadamente a los sujetos de estudio. Es necesario capturar al sujeto en interior con una iluminación donde no se generen sombras sobre el rostro. La distancia en que se recomienda colocar la cámara RGBD frente al sujeto de estudio debe estar comprendida entre 60 y 120 cm.

El algoritmo de clasificación que mejor resultados tuvo fue la máquina de soporte vectorial de Kernel de base radial con un 88.84%.

Finalmente se debe mencionar que a pesar de los problemas que se tuvieron con la correcta identificación del rostro con la cámara RGBD y con los problemas para generar las expresiones faciales con el guión, en este trabajo de investigación se logró cumplir con todos los objetivos y los alcances que se tenían especificados al comienzo.

## 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.

- Realizar el seguimiento del rostro ante la presencia de más de un rostro en la imagen.
- Utilizar la información de los ojos y la nariz para clasificar las expresiones faciales básicas.
- Utilizar las unidades de forma que tiene la librería *facetracking*, estas unidades de forma son puntos localizados en ciertas partes del rostro como: cejas, ojos, nariz, boca y mandíbula inferior.

## Referencias

- [1] Maja Pantic, Member, IEEE, and Leon J. M. Rothkrantz (2004). Facial Action Recognition for Facial Expression Analysis From Static Face Images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. Vol. 34, pp. 1449 - 1461.
- [2] Z. Wen, T. Huang (2003), Capturing subtle facial motions in 3D face tracking. the *IEEE International Conference on Computer Vision*. IEEE, Vol. 2, pp. 1343–1350.
- [3] Filareti Tsalakanidou, Sotiris Malassiotis (2010). Real-time 2D+3D facial action and expression recognition. *Pattern Recognition*. Elsevier, Vol. 43, pp. 1763–1775.
- [4] Human Face. Description Of Facial Action Coding System. Disponible en: <http://face-and-emotion.com/dataface/facs/description.jsp>, Fecha de consulta: 07/10/2015.
- [5] B. Fasel and J. Luetin (2003), "Automatic Facial Expression Analysis: A Survey". *Pattern Recognition*, vol. 36, pp. 259-275.
- [6] 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.
- [7] 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.
- [8] Luca Stocchi. (2014) 3d Facial Expressions Recognition Using The Microsoft Kinect. School of Computing, Dublin City University, Dublin, Ireland, artículo no publicado.
- [9] CARA Emotion Measurement. IMRSV Inc. Disponible en: <https://www.kairos.com/docs/>. Fecha de consulta: 27/04/16.
- [10] 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.

- [11] 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.
- [12] Rodrigo Ruiz Martin (2014). Reconocimiento De Expresiones Faciales Mediante Imágenes Adquiridas Con Cámara Kinect. Universidad Pontificia de Comillas, Madrid. Tesis de licenciatura.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] P. Ekman and W. Friesen (1978), Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press.

- [21] 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.
- [22] 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.
- [23] Face tracking. Consultado: 21/03/2016. <https://msdn.microsoft.com/en-us/library/jj130970.aspx>
- [24] 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.
- [25] 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.
- [26] Daniel Paredes (2009). Seguimiento y Caracterización de Componentes del Rostro para la Detección de Expresiones Faciales. Tesis de maestría CENIDET.
- [27] Base de datos Cohn-Kanade versión 2. Consultado: 04/04/2016. <http://www.pitt.edu/~emotion/ck-spread.htm>
- [28] SDK de Kinect. Download Kinect for Windows SDK v1.8. Consultado: 04/04/2016. <https://www.microsoft.com/en-us/download/details.aspx?id=40278>
- [29] Matthew Szymczyk. How Does The Kinect 2 Compare To The Kinect 1? Consultado: 04/04/2016. <http://zugara.com/how-does-the-kinect-2-compare-to-the-kinect-1>
- [30] Microsoft Framework 4. Consultado: 04/04/2016. <https://www.microsoft.com/es-mx/download/details.aspx?id=17851>
- [31] Kinect for Windows Developer Toolkit v1.8. Consultado: 04/04/2016. <https://www.microsoft.com/en-us/download/details.aspx?id=40276>
- [32] Microsoft Visual Studio Express 2012 para escritorio de Windows. Consultado: 04/04/2016. <https://www.microsoft.com/es-mx/download/details.aspx?id=34673>
- [33] Opencv versión 2.4.9. Consultado:04/04/2016. <http://opencv.org/downloads.html>.

- [34] Weka Data Mining Software. Consultado 27/04/2016.  
<http://www.cs.waikato.ac.nz/ml/weka/>
- [35] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. IEEE, 13-18 June 2010, pp. 94 - 101.
- [36] Brent Martin, (1995). INSTANCE-BASED LEARNING: Nearest Neighbour with Generalisation. Departamento de ciencias de la computación de la Universidad de Waikato Hamilton, New Zealand. Tesis de maestría.
- [37] Eibe Frank, Ian H. Witten (1998). Generating Accurate Rule Sets Without Global Optimization. ICML '98 Proceedings of the Fifteenth International Conference on Machine Learning. Pages 144-151.
- [38] @lucchocon09. Final del proyecto de la bruja de Blair. Consultado 19/05/2016.  
©Youtube: <https://www.youtube.com/watch?v=snsJxIKGcik>
- [39] @juan candia. trainspotting (escena del baño). Consultado 19/05/2016.  
©Youtube: [https://www.youtube.com/watch?v=Bn0cn\\_ayFUw](https://www.youtube.com/watch?v=Bn0cn_ayFUw)
- [40] @ Encargados IURDTV. Perro llora en la tumba de su dueña. Consultado 19/05/2016. ©Youtube: <https://www.youtube.com/watch?v=XiTM-UNemKE>
- [41] @yocris95. Auto que asusta. Consultado 19/05/2016. ©Youtube: [https://www.youtube.com/watch?v=h\\_chFJjSwWQ](https://www.youtube.com/watch?v=h_chFJjSwWQ)
- [42] @Detroit Iron. Vicious BULLY Learns The HARD Way About Picking On Small Kids. Consultado 19/05/2016. ©Live Leak: [http://www.liveleak.com/view?i=596\\_1437955898](http://www.liveleak.com/view?i=596_1437955898)
- [43] @Nelson Rivas. Sujeto haciendo ejercicio. Consultado 19/05/2016. ©Youtube: <https://www.youtube.com/watch?v=1VMAj3ktDCI>

## Anexos

### A. FACS: Unidades de acción

Unidades de acción utilizadas en el desarrollo de esta tesis (Tabla A.1).

Tabla A.1. 15 unidades de acción para el reconocimiento de expresiones faciales básicas

No.	No. de AU	Unidad de Acción, Descripción	Imagen
1	1	Levantar cejas parte interior.	
2	2	Levantar cejas parte externa.	
3	4	Fruncir el ceño.	
4	6	Bajar las cejas a la altura de los ojos.	
5	10	Levantar labio superior.	
6	12/13	Comisura de los labios hacia arriba.	
7	15	Comisura de los labios hacia abajo.	
8	16	Boca cerrada normal prensando muy poco los labios.	
9	17	Mandíbula inferior hacia arriba.	
10	20	Distancia horizontal de la boca máxima y boca cerrada.	
11	23	Prensar poco los labios y subir mandíbula inferior.	

No.	No. de AU	Unidad de Acción, Descripción	Imagen
12	24	Distancia horizontal de la boca mínima y boca cerrada.	
13	25	Abrir muy poco la boca.	
14	26	Abrir la boca.	
15	27	Abrir completamente la boca.	

## B. Ejemplo de guión (estructura)

El guión utilizado en esta tesis, está formado por un conjunto de videos que tiene por objetivo generar un estímulo en los sujetos de estudio, este tiene la siguiente estructura (Figura B.1).



Figura B.1. Estructura del guión para generar todas las expresiones básicas

Cada secuencia de video está separada por breves escenas de transición (20 segundos), las escenas de transición tienen instrucciones para que el sujeto que observe los videos vuelva a una postura relajada y tranquila. El video tiene una duración total de 6:30 minutos.

Composición del video:

- Miedo: Escenas de una casa desconocida de la película de la bruja de Blair [38].
- Desagrado: Escena del baño de la película *Trainspotting* [39].
- Tristeza: Video de un perro llorando sobre la tumba de su dueño [40].
- Sorpresa: Escena de un carro en un paisaje tranquilo y de repente sale un personaje gritando [41].
- Enojo: Escena de bullying a un niño propiciado por sus compañeros de clases [42].
- Alegría: Una persona realizando poses ridículas [43].

### C. Ejemplos de imágenes finales de pruebas

Las imágenes que se presentan a continuación son capturas del Sistema implementado en esta tesis. La Figura C.1 muestra una captura del SREF, en esta figura se observa la imagen con y sin la malla sobre rostro, abajo de estas imágenes se presenta el descriptor-EF y el resultado (ALEGRIA) para esa captura.



<b>1</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>	<b>13</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>20</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>RESULTADO</b>
1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	ALEGRIA

Figura C.1 Captura (imagen y descriptor-EF) del SREF, sujeto muestra un patrón de alegría

La Figura C.2 muestra una captura del SREF, en esta figura se observa la imagen con y sin la malla sobre rostro, abajo de estas imágenes se presenta el descriptor-EF y el resultado (SORPRESA) para esa captura.



<b>1</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>	<b>13</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>20</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>RESULTADO</b>
1	1	0	0	0	0	1	1	1	0	0	1	0	0	0	SORPRESA

Figura C.2 Captura (imagen y descriptor-EF) del SREF, sujeto muestra un patrón de sorpresa

La Figura C.3 muestra una captura del SREF, en esta figura se observa la imagen con y sin la malla sobre rostro, abajo de estas imágenes se presenta el descriptor-EF y el resultado (TRISTEZA) para esa captura.



<b>1</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>	<b>13</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>20</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>RESULTADO</b>
1	0	0	1	0	0	1	0	1	0	0	1	0	0	0	TRISTEZA

Figura C.3 Captura (imagen y descriptor-EF) del SREF, sujeto muestra un patrón de tristeza

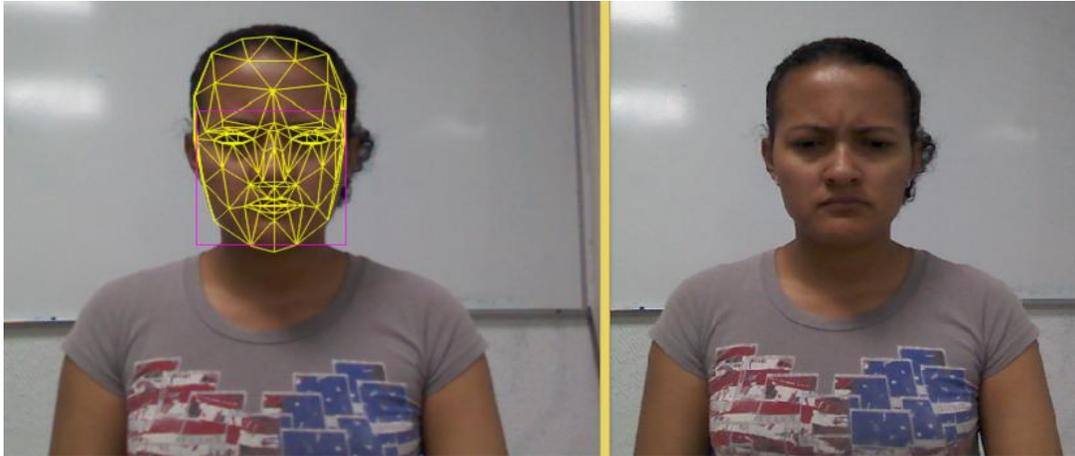
La Figura C.4 muestra una captura del SREF, en esta figura se observa la imagen con y sin la malla sobre rostro, abajo de estas imágenes se presenta el descriptor-EF y el resultado (DESAGRADO) para esa captura.



<b>1</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>	<b>13</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>20</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>RESULTADO</b>
0	0	0	0	1	0	0	1	1	0	0	1	0	0	0	DESAGRADO

Figura C.4 Captura (imagen y descriptor-EF) del SREF, sujeto muestra un patrón de desagrado

La Figura C.5 muestra una captura del SREF, en esta figura se observa la imagen con y sin la malla sobre rostro, abajo de estas imágenes se presenta el descriptor-EF y el resultado (ENOJO) para esa captura.



<b>1</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>	<b>13</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>20</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>RESULTADO</b>
0	0	1	1	0	0	0	1	0	0	0	1	0	0	0	ENOJO

Figura C.5 Captura (imagen y descriptor-EF) del SREF, sujeto muestra un patrón de enojo

La Figura C.6 muestra una captura del SREF, en esta figura se observa la imagen con y sin la malla sobre rostro, abajo de estas imágenes se presenta el descriptor-EF y el resultado (MIEDO) para esa captura.



<b>1</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>	<b>13</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>20</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>RESULTADO</b>
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	MIEDO

Figura C.6 Captura (imagen y descriptor-EF) del SREF, sujeto muestra un patrón de miedo

### D. Herramientas de desarrollo

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

#### Hardware

- Cámara RGBD:** La versión 1 de Kinect para Windows. Un sensor RGBD (Figura D.1) es un dispositivo capaz de crear un entorno visual 3D con los diferentes sensores ópticos, los cuales son:
  - Sensor infra rojo – emisor (Fuente de Luz IR)
  - Sensor infra rojo – receptor (Sensor CMOS IR).
  - Cámara RGB.

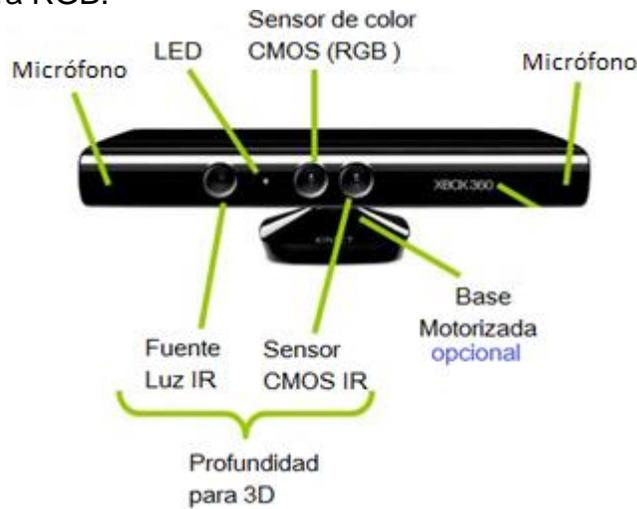


Figura D.1 Componentes de un Sensor RGBD.

Características del sensor RGBD (Tabla D.1):

Tabla D.1 Características del sensor RGBD versión 1 [29].

CARACTERÍSTICA	KINECT PARA WINDOWS V1
CÁMARA RGB	640 X 480 px a 32 bit @30 fps
CÁMARA DE PROFUNDIDAD	640 x 480 px a 16 bit @30 fps
DISTANCIA MÁXIMA DE PROFUNDIDAD	~4.5 m
DISTANCIA MÍNIMA DE PROFUNDIDAD	40 CM en modo cerca
CAMPO DE VISIÓN HORIZONTAL	57°
CAMPO DE VISIÓN VERTICAL	43°
MOTOR ROTATIVO	Sí
ARTICULACIONES DEL ESQUELETO	20
RASTREO MÁXIMO DE ESQUELETOS	2
USB	2.0
SISTEMA OPERATIVO SOPORTADO	Win 7, Win 8

- **Laptop:** Procesador dual 2.0 GHz - 2.6 GHz, 8 GB de RAM.

### **Software de desarrollo**

- **Sistema Operativo:** El sistema funciona en Windows 7, 8, 8.1 y 10.
- **Microsoft .NET Framework [30]:** Es necesario tener instalada la versión 4.0 o superior para el correcto funcionamiento del SDK de Kinect.
- **SDK de Kinect [28]:** Se utilizó la versión 1.8 (funciona para la v1 del sensor Kinect) para obtener las librerías necesarias para el funcionamiento del sensor.
- **Toolkit de Kinect [31]:** 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).
- **Kinect Studio [28]:** 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).
- **Visual Studio 2012 [32]:** En esta plataforma se desarrolló el sistema de reconocimiento de expresiones faciales.
- **OpenCV [33]:** Se utilizó la versión 2.9 para implementar el algoritmo de máquina de soporte vectorial.