



**DEPARTAMENTO DE ELÉCTRICA, ELECTRÓNICA Y
TELECOMUNICACIONES**

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y
TELECOMUNICACIONES**

**TRABAJO DE TITULACIÓN, PREVIO A LA OBTENCIÓN DEL
TÍTULO DE INGENIERO ELECTRÓNICO Y TELECOMUNICACIONES**

**TEMA: SISTEMA DE ENTRENAMIENTO PARA
RECONOCIMIENTO DE EMOCIONES A TRAVÉS DE GESTOS
FACIALES**

AUTORA: SANTANA LLUMIQUINGA, GEOVANNA ALEXANDRA

DIRECTOR: PhD. OLMEDO CIFUENTES, GONZALO FERNANDO

SANGOLQUÍ

2019



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA ELECTRÓNICA EN TELECOMUNICACIONES

CERTIFICACIÓN

Certifico que el trabajo de titulación, "Sistema de entrenamiento para reconocimiento de emociones a través de gestos faciales" fue realizado por la señorita Santana Llumíquina, Geovanna Alexandra el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 12 de Julio, 2019

Firma:

Ing. Gimedo Cifuentes Gonzalo Fernando Mgs

C. C: 1711696342



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA ELECTRÓNICA EN TELECOMUNICACIONES

AUTORÍA DE RESPONSABILIDAD

Yo, **Santana Llumiquinga, Geovanna Alexandra**, declaro que el contenido, ideas y criterios del trabajo de titulación: **“Sistema de entrenamiento para reconocimiento de emociones a través de gestos faciales”** es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Consecuentemente el contenido de la investigación mencionada es veraz.

Sangolquí 12 de Julio 2019

Firma

Geovanna Alexandra, Santana Llumiquinga

C.C.: 1723461032



DEPARTAMENTO DE ELÉCTRICA Y ELECTRÓNICA

CARRERA DE INGENIERÍA ELECTRÓNICA EN TELECOMUNICACIONES

AUTORIZACIÓN

Yo, **Santana Llumiquinga, Geovanna Alexandra** autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Sistema de entrenamiento para reconocimiento de emociones a través de gestos faciales”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí 12 de Julio 2019

Firma

Santana Llumiquinga, Geovanna Alexandra

C.C.: 1723461032

DEDICATORIA

El presente proyecto de investigación va dedicado principalmente a Dios, por regalarme la dicha de poner culminar esta meta, por regalarme a mi familia.

A mis padres Carlos y María, por creer en mí, apoyarme en todo lo que me propongo por ser mi pilar fundamental en el ámbito estudiantil en el ámbito laboral por ser mi apoyo incondicional en cada paso que doy, por ser lo que me impulsaron a ser cada día mejor en todo lo que me proponía realizar.

A mis hermanas Carla y Karol, por compartir conmigo tantas noches de desvelo en la realización de este proyecto de investigación y sobre todo por ser mi compañía desde el primer día que llegaron al mundo llegaron a ser mi motivo para seguir superándome.

A mi amor Diego, por ser ese apoyo que necesitaba cada vez que quería desmayar, por ser la persona que me impulsaba para lograr cumplir esta meta.

AGRADECIMIENTO

En primer lugar, quiero agradecer a todos mis maestros por compartirme sus conocimientos, principalmente quiero agradecer al Dr. Gonzalo Olmedo y a la Ing. Nancy Pareces por su paciencia y su dedicación durante el desarrollo de este proyecto de investigación.

Quiero agradecer a mis padres por su amor incondicional y por creer en mí, por estar conmigo en los días más difíciles. a mis hermanas por estar siempre motivándome durante la realización de este proyecto.

Quiero agradecer a Diego por estar ahí en los momentos más difíciles para mí, y por motivarme para culminar esta meta.

ÍNDICE DE CONTENIDOS

| | |
|---|-----|
| DEDICATORIA | v |
| AGRADECIMIENTO | vi |
| ÍNDICE DE CONTENIDOS | vii |
| ÍNDICE DE TABLAS | ix |
| ÍNDICE DE FIGURAS | x |
| RESUMEN | xi |
| ABSTRACT | xii |
| CAPÍTULO 1 | 1 |
| 1. INTRODUCCIÓN | 1 |
| 1.1. Antecedentes y justificación..... | 1 |
| 1.2. Objetivos | 2 |
| 1.2.1. General | 2 |
| 1.2.2. Específico | 2 |
| CAPÍTULO 2 | 3 |
| 2. ESTADO DEL ARTE..... | 3 |
| 2.1. Emoción | 3 |
| 2.2. Lenguaje corporal..... | 4 |
| 2.2.1. Gestos del Rostro Significado | 5 |
| 2.3. Tipos de expresiones | 7 |
| 2.3.1. Expresiones faciales emocionales | 7 |
| 2.3.2. Microexpresiones | 10 |
| 2.4. Técnicas de detección de rostro | 11 |
| 2.4.1. Análisis de componentes principales (PCA)..... | 11 |
| 2.4.2. LDA (Linear DiscriminantAnalysis)..... | 11 |
| 2.4.3. Viola-Jones..... | 11 |
| 2.5. Unidades de Acción (AU)..... | 12 |
| 2.6. Machine Learning | 15 |
| 2.7. Clasificadores | 16 |
| CAPÍTULO 3 | 19 |
| 3. MÉTODOS Y MATERIALES | 19 |

| | | |
|-------------------|---|----|
| 3.1. | Descripción del modelo de investigación..... | 19 |
| 3.2. | Base de datos | 20 |
| 3.3. | Viola-Jones..... | 21 |
| 3.3.1. | Detección y extracción de la cara..... | 21 |
| 3.3.2. | Detección y extracción de la boca..... | 22 |
| 3.4. | Machine Learning (aprendizaje automático)..... | 24 |
| 3.4.1. | Classificationlearner | 24 |
| 3.4.2. | Extracción de características básicas del rostro-Bag of words (banco de palabras) 25 | |
| 3.4.3. | Entrenamiento del sistema | 27 |
| 3.4.4. | Comparación | 29 |
| CAPÍTULO 4 | | 34 |
| 4. | ANALISIS DE RESULTADO OBTENIDOS | 34 |
| 4.1. | Resultado Obtenidos con clasificador KNN | 38 |
| 4.2. | Resultado Obtenidos con clasificador SVM | 39 |
| 4.3. | Resultados de Comparación | 41 |
| CAPÍTULO 5 | | 43 |
| 5. | CONCLUSIONES Y RECOMENDACIONES..... | 43 |
| 5.1. | Conclusiones | 43 |
| 5.2. | Recomendaciones..... | 44 |
| Bibliografía..... | | 45 |

ÍNDICE DE TABLAS

| | |
|--|----|
| Tabla 1. <i>Unidades de acción según los FACS</i> | 14 |
| Tabla 2. <i>Correspondencia de AU para cada expresión.</i> | 15 |
| Tabla 3. <i>Unidades de acción utilizadas</i> | 31 |
| Tabla 4. <i>Porcentajes de Respuesta</i> | 35 |
| Tabla 5. <i>Características de computadora portátil Sony Vaio</i> | 36 |
| Tabla 6. <i>Porcentajes de comparación entre clasificadores utilizados.</i> | 37 |
| Tabla 7. <i>Matriz de confusión de clasificador KNN.</i> | 38 |
| Tabla 8. <i>Matriz de confusión de clasificador SVM</i> | 39 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1. -Las seis emociones básicas y sus expresiones universales..... | 7 |
| Figura 2. Ejemplo SVM | 17 |
| Figura 3. Ejemplo de clasificador KNN | 18 |
| Figura 4. Diagrama de bloques del entrenamiento | 19 |
| Figura 5. Diagrama de bloques de la predicción | 20 |
| Figura 6. Método de Viola Jones de rostro..... | 22 |
| Figura 7. Método de Viola Jones de boca | 23 |
| Figura 8. Método de Viola Jones de boca | 24 |
| Figura 9. Imágenes organizadas por categorías..... | 25 |
| Figura 10. Creación de banco de características | 26 |
| Figura 11. Extracción de características | 27 |
| Figura 12. Pasos a seguir antes de entrenar un modelo..... | 28 |
| Figura 13. Diagrama de bloques de comparación | 30 |
| Figura 14. Puntos de interés con <i>OpenFace</i> | 31 |
| Figura 15. Activación de puntos de interés..... | 31 |
| Figura 16. Extracción de ojos y boca..... | 32 |
| Figura 17. Comparación de clasificadores | 41 |
| Figura 18. Comparación de Emociones..... | 42 |

RESUMEN

En el presente trabajo de investigación se desarrolló un algoritmo de análisis capaz de predecir emociones faciales universales básicas como son: miedo, feliz, triste, enojo, asco y sorpresa, a través de gestos faciales mediante un sistema previamente entrenado, utilizando aprendizaje automático supervisado en función de *Machine Learning*. Fueron obtenidas las características de 6 emociones básicas de una base de datos libre de emociones, misma que fue utilizada para este proyecto. Las características de cada una de estas imágenes fueron extraídas utilizando *Classification Learner* de Matlab, para posteriormente generar un banco de las mismas y generar un modelo de entrenamiento con las características previamente extraídas de cada emoción. Investigando y evaluando diferentes clasificadores. Dicho modelo previamente entrenado será utilizado en la predicción de las emociones con distintos rostros. Se realiza un algoritmo para encontrar emociones a través de una comparación de AUs extraídas de la imagen a predecir con una base de datos previamente realizada de las mismas para realizar una comparación entre resultados de la predicción previamente obtenida y el resultado del algoritmo de comparación realizado para obtener la emoción. Se muestra mediante resultados el clasificador SVM es eficiente en comparación al clasificador KNN, ya que dicho clasificador es específicamente para el uso de imágenes.

PALABRAS CLAVES:

- **PREDICCIÓN**
- **CLASSIFICATION LEARNER**
- **MACHINE LEARNING**
- **MODELO ENTRENADO**

ABSTRACT

In this research work is developed an algorithm of analysis can predict basic universal facial emotions such as: fear, happy, sad, anger, disgust and surprise, through facial gestures through the use of a trained system. Automatic learning function. They were to take into account the characteristics of 6 emotions. The characteristics of each one. Investigating and evaluating different classifiers. This model has been used in the prediction of emotions with faces. An algorithm is performed to find emotions through a comparison of AUs extracted from the image to be predicted with a database in which a relation between the results of the prediction has to be established. Get the emotion. The results show that the SVM classifier is efficient compared to the KNN classifier, and that the classifier is specifically for the use of images.

KEYWORDS:

- **PREDICTION**
- **CLASSIFICATION LEARNER**
- **MACHINE LEARNING**
- **TRAINED**

MODEL

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. Antecedentes y justificación

Las expresiones faciales son las encargadas de entregar la mayor información en cuanto a gestos y emociones se refiere. Las emociones son un aspecto increíblemente importante de la vida humana ya que investigaciones básicas sobre las mismas en las últimas décadas han producido varios descubrimientos que han llevado a importantes aplicaciones del mundo real. Como por ejemplo, aplicaciones basadas en el reconocimiento de emociones mediante gestos faciales, las cuales detectan no solamente macroexpresiones (gestos notables a la vista del ser humano), sino también microexpresiones (expresiones que no son visibles al ojo humano ya que son generadas en milésimas de segundo). Otro trabajo pionero en este campo es el de Ekman and Friesen, que sugieren que existe seis expresiones faciales prototípicas básicas reconocidas universalmente, como son: miedo, feliz, triste, enojo, asco, sorpresa (D.Matsumoto, 2011).

El reconocimiento de emociones es utilizado en una gran variedad de campos, como son en Psicología, Educación, Marketing, entre otros. Un sistema entrenado y basado en gestos faciales ayudaría en un sinnúmero de actividades, por ejemplo, se podría utilizar en la educación de personas con capacidades especiales al momento de interactuar con un computador midiendo el nivel de satisfacción mediante el reconocimiento de emociones, ya que el reconocimiento de los estados afectivos del usuario es un tema muy importante en el ámbito de la interacción humano-computador (R.W. Picard, 2014). Sería de mucha utilidad también en las cortes para la detección de la sinceridad de personas asociadas con los diversos casos.

En el 2014, *Milborrow and Nicolls* desarrollaron un método para posicionar puntos en rostros frontales usando descriptores. Existen aplicaciones desarrolladas estrictamente para la detección y clasificaciones de emociones como es *Emotion Api* lanzada por *Microsoft* en el año 2016 la misma que es totalmente gratuita.

1.2. Objetivos

1.2.1. General

Desarrollar un sistema de entrenamiento para reconocimiento de emociones a través de gestos faciales utilizando técnicas de *Machine Learning*.

1.2.2. Específico

- Investigar y utilizar una base de datos libre con gestos faciales básicos universales en el cual la temática sean las emociones.
- Investigar y seleccionar uno o varios clasificadores para entrenar el sistema y generar un modelo.
- Desarrollar un algoritmo de detección de emociones a través del modelo obtenido.
- Evaluar los modelos escogidos y utilizados mediante entrenamiento del sistema.
- Desarrollar un algoritmo para detección de emociones a través de AUs basado en PCA.
- Comparar entre la predicción de emociones con datos previamente entrenados y un algoritmo desarrollado en PCA.

CAPÍTULO 2

2. ESTADO DEL ARTE

2.1. Emoción

Una emoción es una reacción psicofisiológica causada por la evaluación de eventos que son importantes como un objeto, persona, lugar; y pueden incluir cambios corporales y mentales(Oatley, 2008).

Las emociones que se experimentan más comúnmente son las llamadas emociones reactivas debido a que ocurren repentinamente y duran unos pocos segundos; están basadas en las cosas que sabemos, incluyendo pensamientos de cosas que ya sucedieron y de predicciones. Las emociones también nos hacen actuar impulsivamente en relación a alguien más, creando estados de inmediata disposición para actuar(Frijda, 1986).

Cuando las emociones tienen un tiempo de mayor duración se crean procesos denominados estado de ánimo, los cuales pueden durar horas o hasta varios días, y se diferencia de una emoción reactiva principalmente porque puede ser provocado por un evento desconocido o que son de mayor complejidad para nosotros por lo que no solemos entenderlas rápidamente.

Existe un estado emocional de mayor duración que el estado de ánimo, el cual se conoce como sentimiento, puede ser un tipo de amor o desconfianza que permite formar las bases de una relación de muchos años(Oatley, 2008).

Es así que las emociones, estados de ánimo y sentimientos están presentes en todo momento influyendo directamente en la cotidianidad de las personas, creando o deshaciendo relaciones humanas. Desde la antigüedad las emociones en estudios de las primeras civilizaciones

fueron descartadas por ser consideradas una parte no racional del ser humano. Tomándolas como parte inferiores de los seres humanos, pero al ser estados afectivos, indican estados internos de la persona, deseos, necesidades e incluso objetivos. Como cualquier tema de complejidad mental es imposible definir la conducta de una persona a través de sus emociones, pero estas ayudan a intuir cuál será su reacción hacia un problema en específico.

Fisiológicamente se pueden distinguir tres tipos de respuestas emocionales: respuestas automáticas producida de manera rápida muy difícil de detectar para receptores, respuestas controladas por el sistema nervioso automático que incluyen el movimiento órganos internos y músculos y cambios en secreción de hormonas y respuestas neuronales que se pueden apreciar a través de electroencefalogramas. Es decir, los tres cambios internos en el sistema endocrino que se dan en cualquier ser humano de manera automática pueden ser descritos fácilmente como respuestas de cambios químicas de nuestro cuerpo y reflejados a través de gestos del rostro.

2.2. Lenguaje corporal

Es una comunicación que se la realiza a través de movimientos gestos y posturas tanto del cuerpo como del rostro con el propósito de transmitir información generada por el emisor. Por lo general se lo realiza inconscientemente, ya que es un indicador del estado emocional de las personas, va de la mano junto con la entonación y comunicación no verbal. Sin embargo el lenguaje corporal es un excelente espejo de nuestras emociones reales (Navarro, 2014).

Tanto la entonación y gestos de nuestro rostro a la hora de interactuar en nuestra sociedad activa son indicios de comunicación no verbal muy importante para comprender al prójimo, así como los ambientes, las emociones también son alteradas por muchos factores ambientales que

pueden influir con el idioma que es expresado a través del cuerpo, de ahí existen diferentes claves para distinguir el lenguaje corporal de las personas.

2.2.1. Gestos del Rostro Significado

El rostro de las personas son la lupa de sus emociones, razón por la cual se dice que el rostro es el reflejo del alma, para ser interpretado en el lenguaje no verbal se debe ir con cuidado, y no se debe evaluar los gestos del rostro por separado ya que la unión de los mismos son los que nos indican las emociones en sí (Navarro, 2014).

Gracias a los continuos estudios de los gestos presentes en el rostro de varias culturas, en diferentes países y décadas, se ha confirmado que la mayoría de las mismas tiene expresiones similares, incluso la carencia de tecnología en la ambigüedad y la poca observación de personas en detalles ínfimos solo nos habría paso a seis gestos destacados, con el avance tecnológico podremos ya clasificar este tipo de características que conforman una emoción, sin embargo a simple vista podemos clasificarlas en forma general, sobre todo si una persona se lleva las manos a la cara se ha cerca más a las características de inseguridad, desconfianza o de algún pensamiento negativo a continuación se presentan algunos ejemplos:

- **Taparse la boca.** - todo depende del momento en que se realiza esta acción si se la realiza cuando está escuchando es signo de que esa persona cree que le está ocultando algo, de manera contraria si lo realiza una persona que está hablando es signo de intenta ocultar algo.
- **Tocarse la oreja.** - la interpretación de esta acción es la representación inconsciente de bloquear las palabras que se está escuchando.

- **Tocarse la nariz.** - Cuando una persona miente se liberan catecolaminas, sustancia que suele inflamar el tejido interno de la nariz y provoca picazón(Navarro, 2014).

A través del rostro se expresan las emociones mismas que forman parte de uno de los factores más importantes del lenguaje no lingüístico, es por eso, que al tener más rasgos se tiene un enfoque prioritario, cabe recalcar que las actitudes y gestos de una persona adulta no es para nada parecida a los gestos y actitudes de un niño como por ejemplo, cuando un niño interpreta una emoción de susto el inmediatamente se tapa los ojos o se tapa las orejas cerrando los ojos, cosa que no se puede visualizar en las personas adultas, se puede apreciar algo parecido pero no tan expresivo como lo hace el niño, es decir en grado de menor expresividad (Pale., 2016) .

El reconocimiento de expresiones faciales emocionales (REFE) es un proceso que implica percibir la expresión , hacer una formación y categorización del concepto previo de cada una de las emociones, mientras que la percepción de una expresión facial involucra las expresión de como se muestra dicha expresión en el rostro (Adolphs, 2002).

Se ha identificado la participación de la corteza prefrontal media, el surco temporal superior en la parte posterior, la corteza somatosensorial primaria y la secundaria, la insulina, la amigadala, la unión temporo-parietal, y la corteza temporal anterior en el reconocimiento de expresiones faciales emocionales (Ebner, 2013).

Es muy importante mencionar que, entre la niñez, la adolescencia y la vida adulta los niveles de activación o reclutamiento de las mismas áreas ya mencionadas anteriormente se muestran diferentes patrones, los cuales difieren tanto en intensidad de activación como en la cantidad del área involucrada de acuerdo a la edad (Mills, 2014).

2.3. Tipos de expresiones

Según *Ekman and Friesen* existen 6 emociones que se pueden identificar como son: miedo, feliz, triste, enojo, asco, sorpresa. Tal como se muestra en la Figura 1. Existen otras evidencias que afirman que las expresiones faciales de las emociones tienen origen biológico y genético. Por ejemplo, cuando se activan emociones de manera espontánea en personas con ceguera congénita se producen las mismas expresiones faciales que se generan en los rostros de individuos videntes (Cole, 1989).



Figura 1. Las seis emociones básicas y sus expresiones universales.

Nota: Imágenes con copyright © David Matsumoto 2008. Reproducido con permiso

2.3.1. Expresiones faciales emocionales

Las expresiones faciales emocionales son cambios transitorios morfológicos en las características de la cara, producidos por la movilidad de músculos específicos, especialmente en el área de los ojos y la boca. Los cambios y movimientos producidos transmiten información acerca de los sentimientos y emociones que lleva día a día cada persona en su interior como por

ejemplo intenciones y tendencias de acción, (Ekman y Cordaro, 2011; Izard, 2011; Levenson, 2011; Panksepp y Watt, 2011) en consecuencia las expresiones faciales cumplen una gran función comunicativa por parte de quien los emite, a su vez sirven para una importante función adaptativa por parte de quien las recibe. Para una interacción social eficaz y satisfactoria, quien emita debe ser capaz de expresar con precisión sus deseos y necesidades, en correspondencia, quien reciba debe ser capaz de reconocer el estado motivacional subyacente a la expresión. En la vida diaria los cambios expresivos en el rostro pueden ser en ocasiones muy intensos, pero en la mayor parte de casos las normas sociales imponen la dimensión de las expresiones emocionales y en consecuencia los cambios manifestados son relativamente sutiles. Es importante, pues, que el observador detecte e interprete adecuadamente también las señales expresivas de baja intensidad.

La mayoría de investigaciones realizadas de neurociencia, cognitiva y social previa sobre el reconocimiento de expresiones faciales se las ha realizado fundamentándose en una concepción categorial de las mismas. En una de estas concepciones, hay las seis expresiones faciales diferenciadas, básicas y universales como son: feliz, triste, miedo, enojo, asco y sorpresa (Ekman, 2003) siendo cada una de estas, representadas por combinaciones de movimientos musculares faciales propios; por ejemplo, la emoción feliz se caracteriza principalmente por la sonrisa en la boca lo que conlleva el estiramiento de la comisura de los labios hacia atrás y hacia arriba, también una elevación de los caninos y estrechamiento de la abertura de los ojos acompañado de arrugas en los bordes interiores y exteriores de los ojos; se tiene también la expresión de enojo que se caracteriza principalmente por “fruncir el ceño” con las cejas bajadas y contraídas, cierre y parpados tensionados, los labios entreabiertos pero tensos o apretados (Ekman. P., 2002).

Se han revisado estudios publicados entre los años 1992 y 2010 por Nelson y Russell en donde se han encontrado muestras de datos de 20 sociedades occidentales y 18 no occidentales en total 38 muestras de las cuales la emoción feliz fue la mejor identificada con un 89% de exactitud, seguidas por la de sorpresa con un 83% luego la emoción triste con un 71%, para luego continuar con un 68% de coincidencia con enojo, y por último asco, miedo con un 65 y 59% respectivamente, las emociones representadas a través del rostro se las logra identificar mediante las microexpresiones que tiene cada individuo, aclarando que cada una de ellas es diferente dependiendo de la región y la ciudad.

Como complemento a todo lo ya mencionado anteriormente existe posibles diferencias en las expresiones faciales de mujeres y hombres en los umbrales de reconocimiento facial. Estudios previos realizados acerca de este tema de las diferencias sexuales en el procesamiento de las expresiones faciales emocionales (sacado de revisión y meta-análisis reciente de Thompson y Voyer, 2014; meta-análisis previos de Hall, 1978). Los resultados obtenidos de Thompson y Voyer demostraron un efecto del sexo, obteniendo como resultado del mismo una ventaja a favor de las mujeres en comparación a los hombres. Primero, esta ventaja femenina mencionada es mayor para las expresiones negativas (enojo, miedo, y triste, aunque no para el asco) que las positivas (alegría) y ambiguas (sorpresa). Segundo, dicha ventaja es mayor cuando los estímulos conllevan una integración de información audiovisual que cuando la información es únicamente visual. Tercero, la ventaja de las mujeres es mayor cuando los modelos son masculinos que cuando son femeninos. Cuarto, la ventaja se produce principalmente entre los 13 y los 30 años de edad de los observadores(Voyer, 2014).

Los sistemas automáticos de reconociendo facial de emociones tiene como misión detectar la emoción de una persona en función de sus expresiones faciales y en general, suelen estar divididos en varias etapas.

2.3.2. Microexpresiones

Uno de los descubrimientos importantes también es la existencia de las microexpresiones. Al reflejarse una emoción en el rostro no existe una razón específica que lleve al individuo a cambiar o disimular su reflejo en su rostro, el tiempo de una expresión característica dura entre 0,5 y 4 segundos aproximadamente mostrándose en todo el rostro (Ekman, 2003).

Las microexpresiones posiblemente son señales de emociones que se quiere ocultar, se producen en un instante de tiempo que es muy complicado identificarlas y reconocerlas con claridad. La existencia de las microexpresiones data desde 1872 en la raíz de la hipótesis de inhibición de Darwin, el cual sugiere que las expresiones faciales se producen de manera espontánea e incluso si la persona está tratando de ocultar sus emociones reales. El tracto piramidal impulsa acciones voluntarias faciales y tiene su origen en la franja motora cortical, mientras que el tracto extrapiramidal impulsa expresiones emocionales involuntarias y se origina en las áreas subcorticales del cerebro. Cuando las personas se encuentran en situaciones intensamente emocionales, pero necesitan controlar sus expresiones, se activan ambos sistemas, que se enzarzan en un “tira y afloja” neutral por el control del rostro, permitiendo la filtración rápida y fugaz de microexpresiones (Rinn, 1984).

Ekman y Friesen (1974) demostraron la existencia de las microexpresiones mediante un estudio en el cual analizaron fotograma por fotograma diferentes grabaciones de personas deprimidas.

2.4. Técnicas de detección de rostro

En este apartado se va a ser una breve revisión a las diferentes técnicas de detección de rostro que se puede encontrar, enlistadas a continuación:

2.4.1. Análisis de componentes principales (PCA)

Análisis de componentes principales es un procedimiento que se encarga de transformar un numero de (posiblemente) variables que se encuentran relacionadas a un número (más pequeño) de variables no relacionadas denominadas componentes principales. El componente que representa la mayor cantidad de datos es el primer componente.

El análisis de componentes principales es similar a otro procedimiento multivariado denominado factor análisis con el cual se confunden y muchas veces no lo entienden.

2.4.2. LDA (Linear Discriminant Analysis)

La función principal que tiene esta técnica es la de convertir un problema de alta dimensión en uno de baja, además busca maximizar la separación de la proyección conformado por un grupo de imágenes no semejantes (Martinez, 2011)

2.4.3. Viola-Jones

En un método de detección de objetos que destaca por su bajo coste computacional, lo que permite que sea empleado en tiempo real. Su desarrollo fue motivado por el problema de la detección de rostros donde sigue siendo ampliamente utilizado (Klette, 2014).

Este método puede aplicarse a otra clase de objetos para este proyecto se lo utilizó específicamente para la detección de rostros en la sección 3.3 se explica ampliamente como se utiliza este método.

2.5. Unidades de Acción (AU)

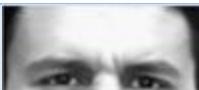
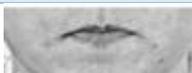
En el 1978, *Paul Ekman y Wallace V. Friesen* desarrollaron el Sistema de Codificación de Acciones Faciales (FACS) (Pantic2006). Este consiste en un profundo estudio de los músculos faciales y una representación de las contracciones de estos en un conjunto de valores llamado Unidades de Acción (AU), definidas como un sistema de codificación de acciones faciales que sirve como taximizador de los gestos en el semblante humano por su expresión superficial. Una unidad de acción por ejemplo, es el levantamiento de la parte interior de la ceja (AU1), y otra el levantamiento de la mejilla (AU6). Las unidades no necesariamente tienen una equivalencia con un músculo, para una unidad de acción pueden intervenir varios músculos faciales, o un mismo músculo determinar por sí solo más de una unidad.

El sistema incluye valores como intensidad, duración y asimetría para cada movimiento. Este estudio se continuó y en el 2002 los propios autores hicieron el lanzamiento de la última revisión. El sistema original consiste de 44 unidades, y requiere mediciones en 3D, lo cual complejiza un procesamiento en tiempo real. Otros autores (Kawakami, Yamada, 1995) presentaron una simplificación del uso de solo 20 unidades, las cuales se presentan en la Tabla 1, pero continuaba el requerimiento de análisis 3D. Áreas comunes en la selección de estas regiones han sido los ojos y la boca, intuitivamente se entienden como estructuras muy expresivas dentro del rostro, además del hecho de que en torno a ellas giran la mayoría de las unidades de acción. Estudios han arrojado como dato que las estructuras faciales superiores (ojos, cejas) parecen tener más peso en la clasificación que las estructuras inferiores (boca, mejillas). (Pantic, 2000)

En el 1978, *Paul Ekman y Wallace V. Friesen* desarrollaron el Sistema de Codificación de Acciones Faciales (FACS) (Pantic 2006). Este consiste en un profundo estudio de los músculos

faciales y una representación de las contracciones de los mismos en un conjunto llamados AUs (actions units o Unidades de Acción). Es decir, las unidades de acción nos muestran y clasifican los movimientos particulares de los músculos de algunas partes del rostro que tienden a cambiar de manera muy diferenciada a comparación de otros músculos del semblante, así pues, el agrupamiento de estos construyen una expresión emocional pudiendo nombrar solamente seis principales ya que fueron las expresiones más básicas expresadas por todos los seres humanos de prueba alrededor del mundo y son: enojo, miedo, asco, feliz, triste y sorpresa, que se producen por el agrupamiento de unidades de acción principalmente en ojos, cejas, boca, barbilla y nariz, Como se puede visualizar en la Tabla 1 el levantamiento de la parte interior de la ceja se expresa mediante (AU1), en el caso de levantamiento de la mejilla es conocida como (AU6). Las unidades no necesariamente tienen una equivalencia con un músculo, para una unidad de acción pueden intervenir varios músculos faciales, o un mismo músculo determinar por sí solo más de una unidad. El sistema incluye valores como intensidad, duración y asimetría para cada movimiento siendo así como se clasifican las emociones.

Tabla 1
Unidades de acción según los FACS

| AUs | Imagen representativa | Definición | AUs | Imagen representativa | Definición |
|------|---|--|------|--|--|
| AU1 |  | Interior de la cejas elevado | AU15 |  | Comisura de los labios hacia abajo |
| AU2 |  | Exterior de las cejas elevado | AU16 |  | Labio inferior hacia abajo |
| AU4 |  | Cejas bajadas | AU17 |  | Barbilla elevada |
| AU5 |  | Párpado superior elevado | AU20 |  | Labios estrechados y estirados en horizontal |
| AU6 |  | Mejillas elevadas | AU22 |  | Labios crateriformes |
| AU7 |  | Párpados estrechados | AU23 |  | Labios tirantes |
| AU9 |  | Nariz arrugada | AU24 |  | Labios presionados |
| AU10 |  | Labio superior elevado | AU25 |  | Labios separados |
| AU11 |  | Nasolabial pronunciado | AU26 |  | Boca entreabierta |
| AU12 |  | Comisura de la boca estirada y elevada | AU27 |  | Boca abierta |

Fuente:(Sanchez, D., 2016)

A través de las AU se ha realizado un listado de patrones faciales basado en las 6 emociones típicas básicas universales, tal como muestra la Tabla 2.

Tabla 2.*Correspondencia de AU para cada expresión.*

| Expresión | Descripción | AUs | |
|-----------|--|-------------------------------|---|
| Miedo | Interior y exterior de cejas elevado, párpados superior elevado, labios estrechados o boca entreabierta | AU1, AU2, AU5, AU20, AU26 |  |
| Enojo | Cejas bajadas, párpados estrechados, barbilla elevada y labios presionados | AU4, AU7, AU17, AU24 |  |
| Triste | Interior de cejas elevado, párpados estrechados, comisura de los labios hacia abajo, barbilla elevada o labios presionados | AU1, AU6, AU15, AU17, AU24 |  |
| Asco | Cejas bajadas, nariz arrugada y labio superior elevado | AU4, AU8, AU10 |  |
| Feliz | Mejillas elevadas, comisura de la boca estirada y elevada y con labios separados | AU6, AU12, AU25 |  |
| Sorpresa | Interior y exterior de cejas elevado, párpado superior elevado y boca entreabierta o abierta | AU1, AU2, AU5, AU26, AU27 |  |

Fuente:(Sanchez, D., 2016)

2.6. Machine Learning

Es una disciplina científica que forma parte de la inteligencia artificial creada con el exaltante propósito de emplear sistemas de aprendizaje automáticamente siendo la palabra aprendizaje como una personificación de lo que conocemos comúnmente como el reconocimiento de patrones formados en este caso reconocimiento fácil a través de la información

proporcionada (base de datos), siendo esta ejecutada por la aplicación *ClassificationLearner* del programa Matlab, misma que entrena modelos para clasificar datos, en la sección 3.4 se explica a detalle la utilidad de la misma. Con esta aplicación, se puede explorar el aprendizaje automático supervisado utilizando varios clasificadores. Se consigue explorar sus datos, seleccionar características, especificar esquemas de validación, entrenar modelos y evaluar resultados. Puede realizar un entrenamiento automatizado para buscar el mejor tipo de modelo de clasificación, incluidos los árboles de decisión, el análisis discriminante, las máquinas de soporte vectorial, la regresión logística, los vecinos más cercanos y la clasificación por conjuntos.

Puede realizar un aprendizaje automático supervisado proporcionando un conjunto conocido de datos de entrada (observaciones o ejemplos) y respuestas conocidas a los datos (por ejemplo, etiquetas o clases). Utiliza los datos para entrenar un modelo que genera predicciones para la respuesta a nuevos datos. Para usar el modelo con nuevos datos, o para aprender sobre la clasificación programática, puede exportar el modelo al área de trabajo o generar un código MATLAB® para recrear el modelo entrenado.

2.7. Clasificadores

Los clasificadores son programas empleados para reconocer cada una de las unidades de acción en diferentes formas y ordenes por lo que encontraremos dentro de los mismo dos variedades utilizadas para el programa, uno de los más utilizados para el sistema de reconocimiento facial es SVM(Support Vector Machine o Máquinas de Soporte Vectorial).

“Se trata de un método de clasificación basado en un entrenamiento supervisado (se conoce cuál es el resultado de la muestra). A partir de un conjunto de muestras de entrenamiento se puede etiquetar clases y entrenar” (Olmedillo, Diego Zapatero, 2016)

Donde gráficamente el modelo es representado como un conjunto de puntos en el espacio, se puede visualizar en la Figura 2. donde se separan los puntos que pertenecen a clases diferentes en conjuntos más pequeños, constituyendo así varios planos unidimensionales.

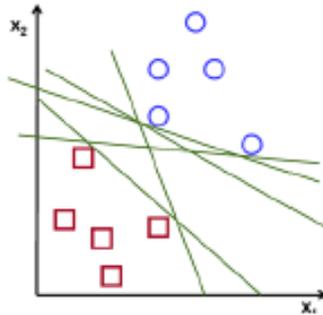


Figura 2. Ejemplo SVM

Fuente (Zapatero, O., 2016)

Es así que cuando el modelo sintetiza una nueva muestra, sitúa esta muestra en el espacio y lo asocia entre una función específica o funciones con cercanía a ellas. Esta cercanía es uno de los puntos principales de este modelo.

Otro de los programas que podemos mencionar es el KNN (k nearest neighbors o k vecinos más cercanos) que consiste en una herramienta basada en la probabilidad de que las características sean más cercanas. El valor del parámetro k es sumamente importante, ya que si es un valor muy bajo se posee poca información, pero si es muy alto se puede incluir muestras irrelevantes y que generen confusión. Este parámetro por lo tanto debe elegirse en función a los datos de entrada y los resultados de las primeras pruebas realizadas con distintos valores de " k ". En la Figura 3. se puede observar un ejemplo muy práctico: si se elige un $k=3$ la nueva muestra es clasificada como un triángulo ya que dentro de la muestra de tiene más objetos triángulo, un

resultado completamente diferente se puede visualizar al elegir un $k=5$ la nueva muestra es clasificada como cuadrado, ya que dentro de la muestra de tiene más objetos cuadrado.

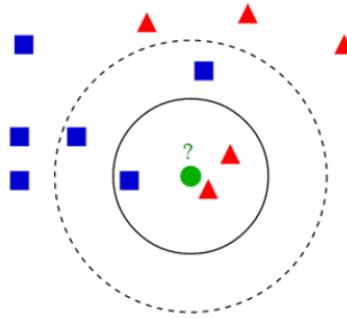


Figura 3. Ejemplo de clasificador KNN
Fuente (Zapatero, O., 2016)

En el Clasificador KNN a diferencia del SVM la información se encuentra mezclada.

Como todo programa al no tener razonamiento lógico puede llegar a confundir muchas de las acciones realizadas por lo que es necesario que el programa cuente con más aplicaciones o tipos de clasificadores para que puedan cumplir con su función de manera más óptima.

CAPÍTULO 3

3. MÉTODOS Y MATERIALES

3.1. Descripción del modelo de investigación

Con el fin de identificar las 6 categorías básicas de expresiones descritas por Ekman y Friesen tales como: ira, asco, feliz, miedo, triste y sorpresa, se utilizó la base de datos de expresión codificada por la AU de Cohn-Kanad, es de acceso libre, y se encuentra publicada en la red. Como primer paso se procede a realizar la detección del rostro de la base de datos, rostros que serán entrenados mediante una aplicación de Matlab llamada *Classificationlearner* en donde mediante aprendizaje automático se obtiene el modelo de entrenamiento total como se explica en el diagrama de bloques de la Figura 4, mismo que será de mucha ayuda para realizar la predicción de la emoción mediante el diagrama mostrado en la Figura 5, y la probabilidad de poder ser otra emoción.

Entrenamiento:

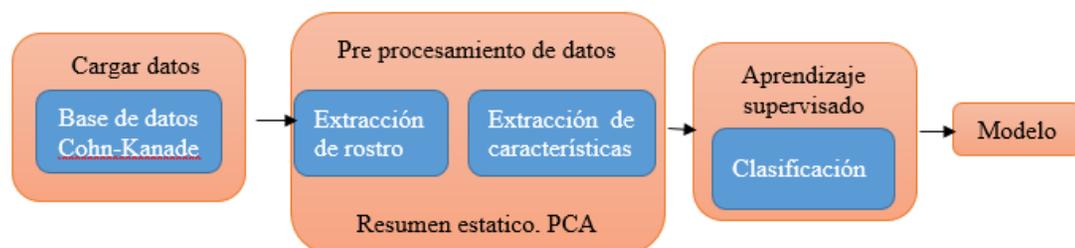


Figura 4. Diagrama de bloques del entrenamiento

Predicción:

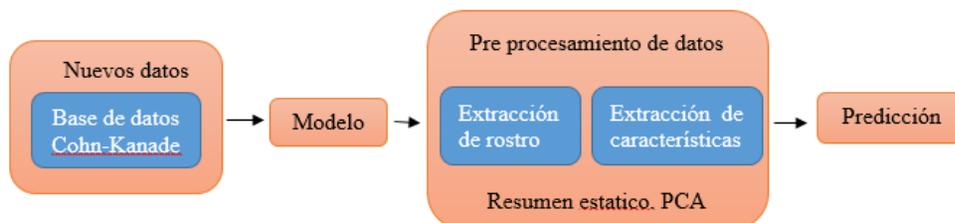


Figura 5. Diagrama de bloques de la predicción

Para la realización del presente proyecto se utiliza el *classificationlearner* de Matlab que permite realizar aprendizaje automático, el aprendizaje automático es una técnica que utiliza datos para crear un programa y así realizar una tarea.

3.2. Base de datos

Una investigación previa del proyecto se optó por utilizar una base de datos de ojos y bocas en binario para realizar el entrenamiento, opción que fue descartada ya que esta base de datos contaba simplemente con características de 3 emociones en comparación a las 6 que fueron procesadas en este proyecto, entonces se procedió a realizar un programa para incrementar la base de datos con usuarios superiores a los de la base de datos tomada inicialmente, luego de realizar una investigación de las bases de datos libres en la red, encontrando como mejor opción la base de datos Cohn-Kanade.

Cohn-Kanade es una base de datos con codificación AU para la investigación en el análisis de imágenes faciales automáticas, esta base de datos cuenta con dos versiones disponibles la utilizada en este proyecto es la versión 2 denominada CK + incluye expresiones

planteadas y no planteadas (espontáneas). Esta base de datos proporciona protocolos y resultados para el seguimiento de características faciales y la unidad de acción y el reconocimiento de emociones. (Pittsburgh, 2010).

3.3. Viola-Jones

3.3.1. Detección y extracción de la cara

En esta etapa como primer punto la finalidad es separar la cara del fondo de las imágenes de la base de datos con la cual se va a realizar este proyecto, para extraer los componentes de rostro se ha tomado en como herramienta computacional a Matlab ya que es un software matemático que ofrece en uno de sus entornos, el procesamiento de imágenes, por medio de un lenguaje de programación.

Posteriormente también se necesitará realizar la separación de las características de la cara como son los ojos y la boca. Para sustentar la emoción que fue predicha, para esto utilizaremos el método de viola-jones, ya que es rápido y nos presenta grandes resultados en la detección de rostros, también cuenta con ventajas muy significativas como es su robustez, tiene una tasa de detección muy alta y tiene bajas tasas de falsos positivos.

Viola-Jones se basa en la clasificación de diferentes ordenadores uno detrás de otro en forma de cascada, ya que juntos permiten buscar y encontrar las áreas con mayor cantidad de información, más precisa y poder excluir las regiones que no son de interés, en la Figura 6. se puede visualizar cómo funciona el método de viola-jones en (a) se muestra la imagen original tal cual la obtenemos de la base de datos, en (b) se muestra el rostro sin el fondo de la imagen.



Figura 6. Método de Viola Jones de rostro.

3.3.2. Detección y extracción de la boca

Para poder reconocer las diferentes emociones que reflejan el estado de ánimo de las personas es necesario contar con una base de datos que refleje dichas emociones y que sea de varias personas, para luego hacer la comparación entre la predicción realizada a través de aprendizaje automático ya sacar la emoción mediante la comparación entre la imagen predicha y los AUs de dicha imagen.

Para extraer la sección de la boca se lo realiza con el mismo método Viola-Jones ya que este método realiza la detección y extracción de la zona de interés, como es en este caso la boca. Para realizar esta extracción hay que definir los valores para el margen de la imagen puede variar desde 10 hasta 200, por lo general cuando se ingresa un margen mayor al indicado no se logra obtener la segmentación adecuada de la boca. Tal como se muestra en la Figura 7 se puede apreciar la detección de la boca con diferentes márgenes como sucede en el caso de (a) se utilizó un margen de 70. Mientras que en el caso de (b) se utilizó un margen de 50 y en el caso de (c) se utilizó un margen de 60.

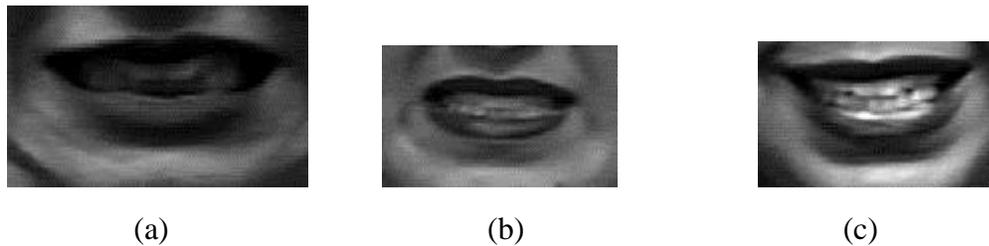


Figura 7. Método de Viola Jones de boca

Es muy importante mencionar que los márgenes que se asignan para la selección de las diferentes características como son ojos y boca deben ir de acuerdo a lo que se ha especificado ya que en la detección de la boca de la emoción sorpresa se ha presentado algunos inconvenientes.

3.3.3. Detección y extracción de los ojos

De la misma manera que se lo realiza para la boca se utiliza la base de datos que refleje las emociones y que sea de varias personas, para luego hacer la comparación entre la predicción realizada a través de aprendizaje automático ya sacar la emoción a través de la comparación entre la imagen predicha y los AUs de dicha imagen.

Para extraer la sección de los ojos se utiliza un algoritmo diferente al de la boca, ya que en este caso son dos objetos los que van a ser diferenciado de todo el rostro, de igual manera se utiliza el mismo método Viola-Jones ya que este método realiza la detección y la extracción de la zona de interés. Para realizar esta extracción hay que definir los valores para el margen de la imagen puede variar desde 10 hasta 200, por lo general cuando se ingresa un margen mayor al indicado anteriormente no se logra obtener la segmentación adecuada.

Tal como se muestra en la Figura 8 como es en el caso de (a) se utilizó un margen 90 en el caso de (b) uno de 50 y en el caso de (c) un margen de 120.

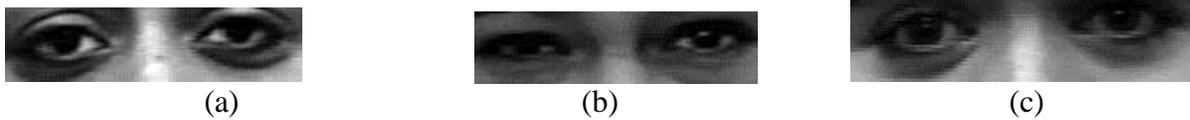


Figura 8. Método de Viola Jones de boca

3.4. Machine Learning (aprendizaje automático)

Como se a especificado anteriormente este proyecto se utilizó el algoritmo de aprendizaje automático (Machine Learning) en dos pasos, el primero es entrenar el modelo para lo cual se utiliza *classificationlearner*, que es una aplicación disponible en Matlab mismo que permite realizar la extracción de las características de la imagen y el segundopaso es el entrenamiento de las características extraídas para después obtener un modelo de entrenamiento con los datos y a través de ellos realizar la predicción de las emociones.

3.4.1. Classificationlearner

Es una herramienta interactiva que permite realizar tareas comunes del aprendizaje automático, como, por ejemplo: explorar los datos de manera interactiva, seleccionar características, entrenar los modelos y evaluar los resultados. Como primer punto se exportan los datos que ya fueron clasificados a través de viola jones, específicamente de la cara, el siguiente paso es realizar un pre procesado de nuestros datos utilizando extracción de características este paso es muy importante y que se trasforman en los datos de entrada mediante un algoritmo de clasificación que el sistema pueda entender, después de esto el sistema ya está preparado para aprender de estos datos, se procede a entrenar nuestro sistema mediante aprendizaje supervisado y con esto ya se crea el modelo.

Esto es un proceso interactivo lo cual requiere regresar a la parte de pre procesado de los datos probar diferentes algoritmos de aprendizaje automático y diferentes parámetros.

El siguiente paso del flujo de trabajo de este proyecto es utilizar el modelo con nuevos datos a la entrada seguido del pre procesado de los datos utilizado en el entrenamiento, seguido del modelo que se obtuvo en el entrenamiento, este modelo va a ser primordial en la predicción sobre los datos nuevo acerca de la emoción.

3.4.2. Extracción de características básicas del rostro-Bag of words (banco de palabras)

Para realizar la extracción de características se usa bag of words (banco de palabras) de Matlab, mismo que utiliza las funciones de Computer Vision System Toolbox para realizar la clasificación de las categorías de imágenes generando una banco de palabras visuales. Este proceso de generación de banco de palabras origina un histograma con similitudes de ocurrencias de palabras visuales que representan una imagen, mismo que es utilizado para entrenar un clasificador de categorías de imágenes. conforme los pasos que se enlistan a continuación.

➤ Paso 1: Configurar conjuntos de categorías de imágenes

Ordena y distribuye las imágenes en subcategorías de entrenamiento y de prueba. Las mismas que luego son almacenadas en un clasificador para posteriormente ser utilizadas en el entrenamiento. En la Figura 9 se visualiza a estas organizadas en categorías para hacer un manejo de grandes conjuntos más sencillo.

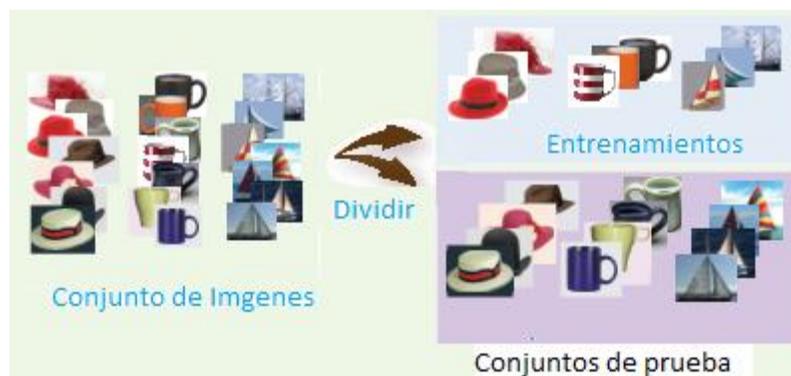


Figura 9. Imágenes organizadas por categorías.
Fuente MathWorks,2019.

➤ **Paso 2: Crear banco de características**

Genera un vocabulario visual, o un banco de características, extrayendo descriptores de características de imágenes más representativas de cada categoría.

El objeto *bagOfFeatures* define las características, o palabras visuales, mediante el uso del algoritmo de agrupamiento extrae punto clave de las imágenes para luego agruparlos en descriptores de características extraídos de *trainingSets* (entrenamientos). El algoritmo agrupa iterativamente los descriptores en *kclusters* mutuamente excluyentes. Los grupos resultantes son compactados y separados por características similares. Cada centro de clúster representa una palabra visual.

Puede extraer características en función de un detector de las mismas, o puede definir un método llamado cuadrícula para extraer descriptores. Tal como se muestra en la Figura 10 al realizar extracción de punto clave Este método puede perder información de escala de grano fino (MathWorks, 2019).

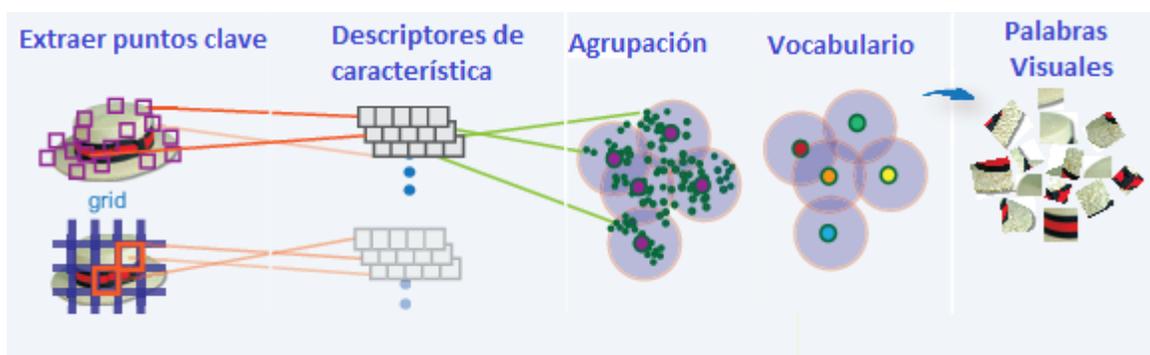


Figura 10. Creación de banco de características
Fuente MathWorks,2019.

➤ **Paso 3: Entrena un clasificador de imágenes con un banco de palabras visuales**

La función *trainImageCategoryClassifier* devuelve un clasificador de imágenes, misma que los entrena en multiclase utilizando el marco de los códigos de salida de corrección de errores (ECOC) con clasificadores de máquinas de soporte vectorial (SVM). *TrainImageCategoryClassifier* usa el banco de palabras visuales devueltas por el objeto *bagOfFeatures* para codificar las imágenes en el conjunto de imágenes en el histograma de palabras visuales. El histograma de palabras visuales se utiliza como muestras positivas y negativas para entrenar al clasificador. (MathWorks, 2019)

Se utilizó el método de codificación *bagOfFeatures* para cifrar cada imagen del conjunto de entrenamiento. Esta función detecta y extrae características de la imagen y luego utiliza el algoritmo del vecino más cercano para construir un histograma de características para cada imagen. Luego, la función incrementa los depósitos de los mismos en función de la proximidad del descriptor a un centro de clúster particular. La longitud de este corresponde al número de palabras visuales que construyó el objeto *bagOfFeatures*, para convertirse después en un vector de características para la imagen. (MathWorks, 2019).

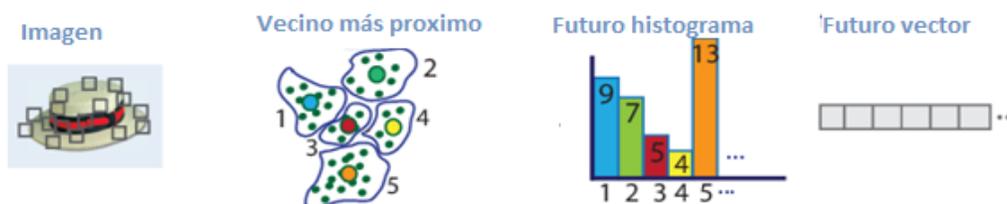


Figura 11. Extracción de características

Fuente MathWorks,2019.

3.4.3. Entrenamiento del sistema

Lo primero que se hizo para entrenar el sistema es importar los datos del workspace de Matlab a *classification learner* el aplicativo para entrenar sistemas, aparece una ventana con tres

pasos a seguir, en el primer paso se debe elegir los datos que son de nuestro interés como en este caso son FaceData, en el paso dos aparece el contenido de las variables ya seleccionada anteriormente Matlab identifica los valores de esta variable dependiendo de si es un predictor o es una respuesta, esto lo hace basándose en la clase. En el paso 3 se eligió el método de validación este paso es muy importante ya que se puede evitar problemas como el sobreajuste, se selecciona *Holdout Validation* ya que la base de datos cuenta con suficientes características en la Figura 12 se puede visualizar cada una de los pasos antes de entrenar un sistema.

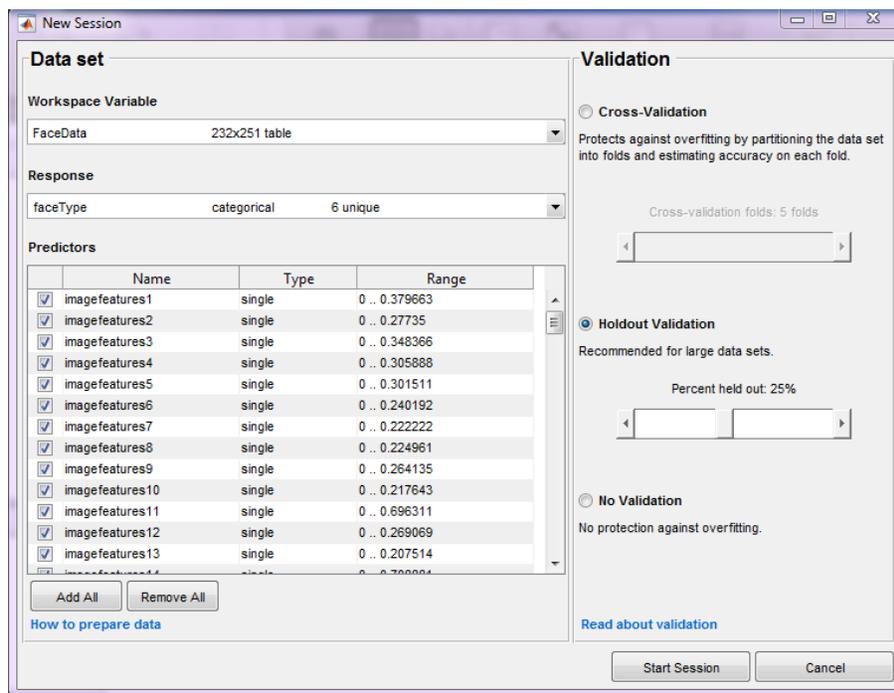


Figura 12. Pasos a seguir antes de entrenar un modelo.

Una vez ya seleccionados los datos procedemos a realizar el entrenamiento utilizando los diferentes clasificadores que nos ofrece esta aplicación de Matlab, contamos con clasificadores como arboles de decisión, regresión lineal, SVM, KNN. Al momento de elegir un clasificador no existe uno adecuado, diferentes clasificadores funcionan mejor dependiendo de los datos de

entrada. Razón por la cual probamos con los diferentes clasificadores con los que contamos, obteniendo como mejor respuesta los clasificadores SVM y KNN, de los cuales fue extruido el modelo entrenado, es decir, se obtuvo un modelo entrenado SVM y otro modelo entrenado KNN. Ya que estos dos modelos son los que nos dan el mejor porcentaje al momento de entrenar los datos.

Una vez que tenemos ya el modelo de entrenamiento como último paso procedemos a realizar la predicción, con una nueva imagen de la base de datos que no fue utilizada en el entrenamiento.

3.4.4. Comparación

Para sustentar la predicción que realiza el sistema se utiliza un programa llamado reconocimiento, en el cual se utilizaron bases de datos tanto de ojos como de boca previamente clasificados manualmente por AUs (unidades de acción), de acuerdo a cada emoción a través del programa *OpenFace*. Cabe mencionar que estas bases de datos de ojos y de boca se la realizo a partir de las imágenes ya entrenadas.

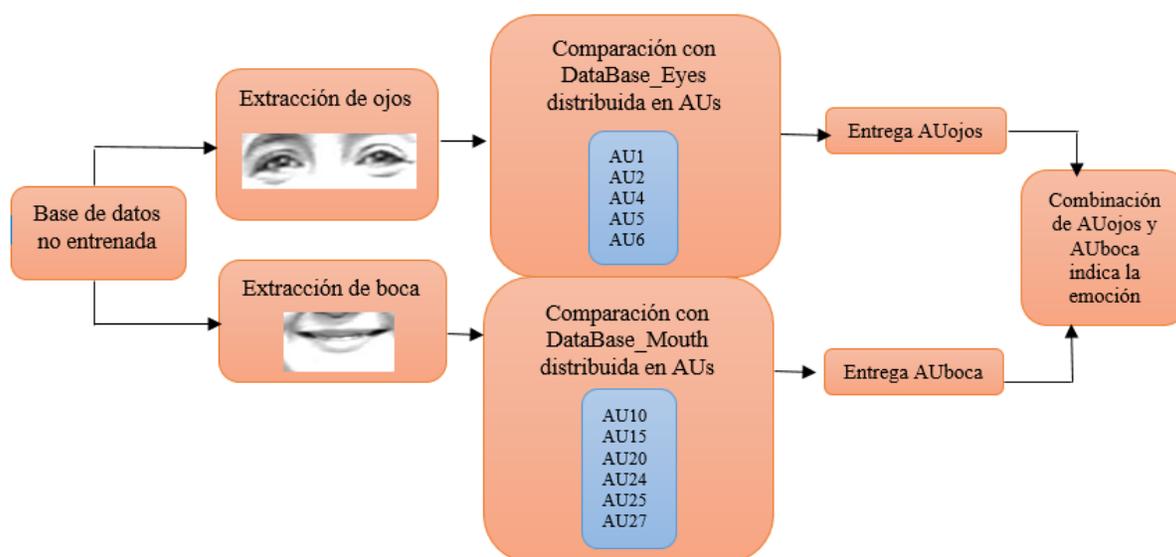


Figura 13. Diagrama de bloques de comparación

OpenFace es una herramienta que a través de localización de puntos de interés en el rostro indica los AUs que se activan en cada emoción al ingresar una imagen. Tal como se muestra en la Figura 14 la imagen ingresada representa a la emoción feliz en la Figura 15 se puede visualizar los AU y la intensidad con la que se activan; los AUs que se activan son: AU6, AU7, AU9, AU10, AU12, AU14, AU20, AU25, AU26. De acuerdo a la Tabla 2 planteada inicialmente por Ekman las AUs a activarse en la emoción feliz son: AU6, AU12 y AU 25, para la ejecución de este proyecto ya que la AU12 (comisura de la boca estiradas y elevadas) y AU25 (labios separados) son específicamente de la boca se ha decidido utilizar el AU25 ya que es la característica más representativa de esta emoción, como se especifica en la Tabla 3. Similar análisis se ha realizado para las 5 emociones restantes mostrando en la Tabla 2, las AUs que fueron utilizadas en el desarrollo de este proyecto.

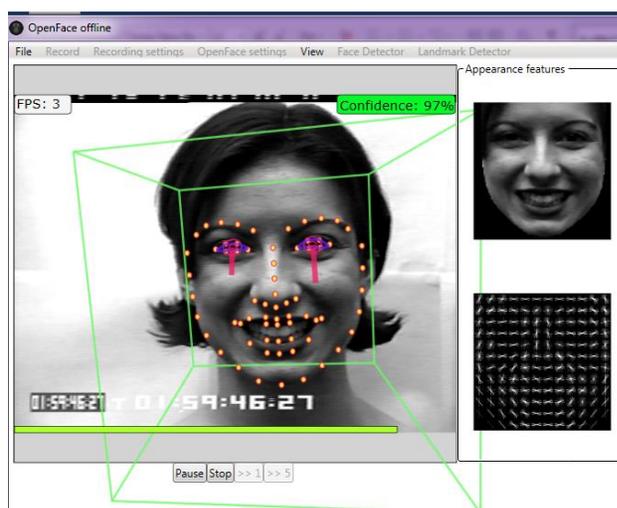


Figura 14. Puntos de interés con *OpenFace*.

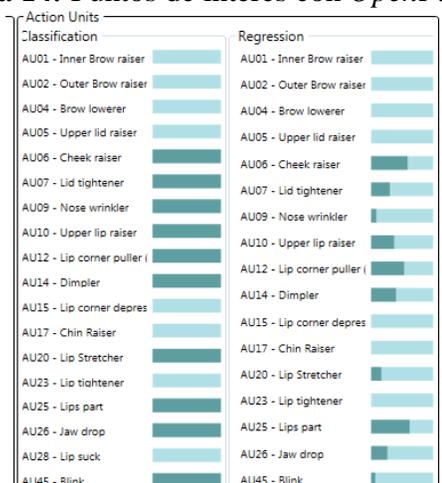


Figura 15. Activación de puntos de interés.

Tabla3

Unidades de acción utilizadas

| Emoción | AUs | Significado |
|---------------|------|------------------------------------|
| Feliz | AU6 | Mejillas elevadas |
| | AU25 | Labios separados |
| Triste | AU1 | Interior de las elevado |
| | AU15 | Comisura de los labios hacia abajo |
| Miedo | AU2 | Exterior de las cejas elevado |
| | AU20 | Labios estrechados y estirados |

Continúa



| | | |
|-----------------|------|--------------------------|
| | AU4 | Cejas bajadas |
| Enojo | AU24 | Labios Presionados |
| | AU5 | Parpado superior elevado |
| Sorpresa | AU27 | Boca abierta |
| | AU4 | Cejas bajadas |
| Asco | AU10 | Labio superior elevado |

La comparación se realiza a través de la extracción de las características del rostro de la imagen a ser predicha, como son ojos y boca. Tal como se muestra en la Figura 16, se visualiza tanto la imagen del rostro utilizada para la extracción y las características extraídas.

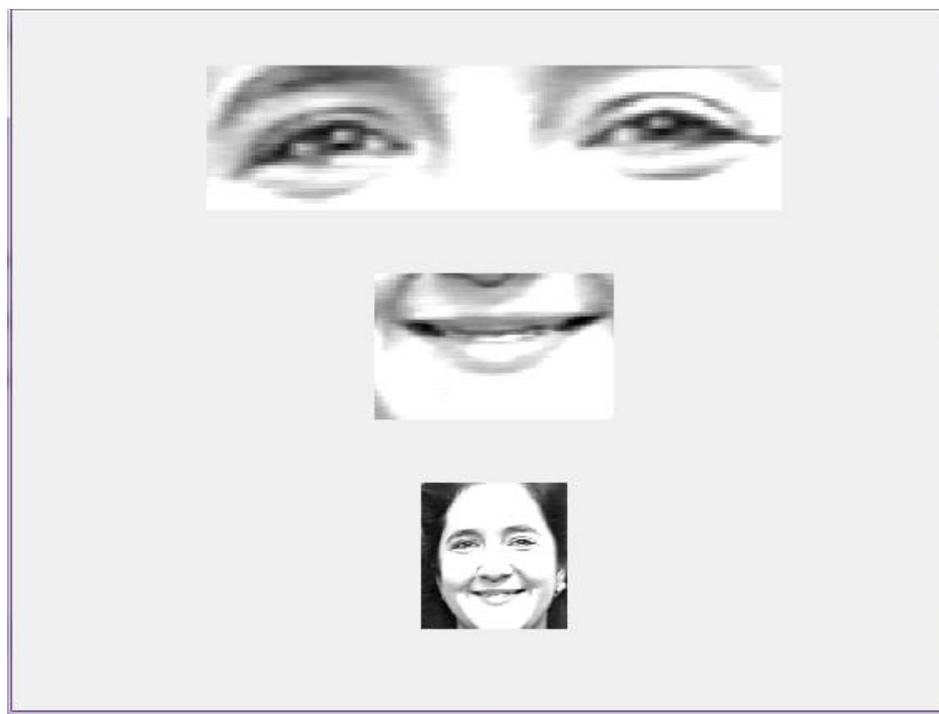


Figura 16. Extracción de ojos y boca.

Para corroborar a la predicción obtenida mediante el modelo previamente entrenado se realizó un programa llamado reconocimiento en el cual se extrae las características de la imagen

que fue predicha, se la dimensiona en una misma medida y la compara con la base de datos de AUs.

Una vez que se extrae las características de la nueva imagen a ser predecida estas imágenes son guardadas como Imagen0021 la de la boca e Imagen0022 la de los ojos y comparada con cada una de las imágenes existentes en la base de datos AUs buscando coincidencias con alguna de esas AUs y mostrando a cual más se asemeja. Cuando Imagen0022 encuentra una similitud con alguna AUs se guarda hasta que se realice la otra comparación con la base de datos de ojos de la misma manera como se realiza con boca, cabe aclarar que esta base de datos no está previamente entrenada simplemente se usa PCA para poder cuando encuentra similitud y con alguna de las AUs tanto de ojos como de boca y se guarda. Al encontrar una combinación en específica como se muestra en la Tabla 2, se muestra como resultado la emoción a la que pertenece.

Cuando no coincide las AUs según se especifica en la Tabla 2 muestra como resultado: “No se ha podido reconocer la emoción”.

CAPÍTULO 4

4. ANALISIS DE RESULTADO OBTENIDOS

Como se pudo observar en el apartado anterior, se explicó detalladamente el proceso que conlleva el algoritmo y la implementación del sistema de entrenamiento, en esta sección se explicara los inconvenientes que se presentaron a lo largo de la ejecución del proyecto de investigación, así como, también las soluciones frente a estos y las pruebas realizadas, a fin de obtener un resultado exitoso.

El inicio se trabajó con una base de datos existente de ojos y boca en binario, de solo 4 emociones, no se podía realizar ningún entrenamiento con esos datos ya que primero se necesitaba completar la base de datos con las 2 emociones faltantes, se procedió a completar la base de datos a partir de la cámara un computador portátil Sony Vaio, la base de datos con la que se contaba inicialmente tenía un total de 24 sujetos, a medida que se completaba la base de datos iban presentando inconvenientes tales como la no detección de las características principales como los ojos y la boca, presentaba también problemas de incompatibilidad en cuanto a las dimensiones de la imagen, lo que provocaba tiempos altos de ejecución en algunas ocasiones, provocado por los diferentes tamaños que tenían las imágenes, también hubo complicaciones ya que algunos sujetos que fueron utilizados para completar la base de datos tenían la cara con características similares a una emoción en específica, por ejemplo, tenía la boca hacia abajo lo que daba a entender que siempre estaba triste. Otra complicación que se visualizó en la binarización de las imágenes fue que el programa detectaba cualquier línea como la boca.

Se realizaron varios experimentos antes de obtener el modelo adecuado de entrenamiento:

En el primer experimento, una vez completada la base de datos tanto de ojos como de boca se procedió a entrenar los datos, para realizar un entrenamiento más acertado de la base de datos que debe tomar un clasificador que adecuado a las imágenes que tenemos en este caso el clasificador que fue utilizado es el de árboles de decisión ya que con imágenes en binario no se tenía muchas características que analizar. El clasificador nos entrega un porcentaje mayor en comparación a los otros clasificadores tal como se muestra en la Tabla 4, los porcentajes de respuesta en comparación a los clasificadores con los que contamos para este proyecto.

Tabla 4
Porcentajes de Respuesta

| Clasificador | Porcentaje |
|-----------------------------------|-------------------|
| Árboles de Decisión | 10,1% |
| Análisis de discriminación | 8,5% |
| SVM | 5,6% |
| KNN | 3,21% |
| Regresión logística | Error |

Estos porcentajes son valores más óptimos comparados entre ellos, es decir, entre los 5 clasificadores el que nos entregaba la mejor respuesta era el clasificador de árboles de decisión ya que nos entregaba un valor de aproximadamente el 10,1% de efectividad, mientras que KNN nos entregaba un porcentaje de 3,21%, con los datos ingresados. Con este resultado no se podía sacar un modelo entrenado que funcione de una forma eficiente al momento de hacer las predicciones de las emociones.

Los resultados obtenidos en este primer experimento pese a la cantidad de imágenes con la que se contaba en la primera base de datos, llevó a tomar una decisión o ampliar la base de datos a un mayor número de imágenes o cambiar por una base de datos donde se cuente con imágenes en blanco y negro si dividirla en secciones, es decir, si separar ojos y boca.

Como segundo experimento se realizó una base de datos con imágenes de rostros en blanco y negro sin dividirla, inicialmente se crea una base de datos con imagen capturadas directamente desde la cámara de un computador portátil Sony Vaio, imágenes que son guardadas en carpetas de una emoción específica.

En la tabla 5 se muestran las características de la computadora portátil Sony Vaio

Tabla5

Características de computadora portátil Sony Vaio

| | |
|---------------------------------|---------------------------------------|
| Procesador | Intel Core i5-2450M CPU 2.5GHz |
| Sistema operativo | Windows 7 |
| Versión de Matlab | R2017b |
| Sistema de entrenamiento | Classifier Learner |

Completando esta base de datos con una de acceso libre Cohn-Kanade, base de datos que cuenta con la clasificación de las emociones de 100 sujetos, como se explicó en el apartado anterior la única modificación que se hizo en esta base de datos en extraer el fondo de la imagen y trabajar simplemente con el rostro, para reducir el proceso de entrenamiento.

Se procedió a realizar el entrenamiento con esta base de datos teniendo en cuenta que de los 100 rostros con lo que contamos solo se utiliza 60-40 es decir se entrenó el sistema con el 60% de la base de datos y el otro 40% se lo utiliza para realizar las predicciones, a comparación

del entrenamiento realizado con las imágenes binarizadas en este entrenamiento logramos obtener un mayor porcentaje de eficiencia las imágenes de los rostros completos en blanco y negro. Como se mencionó anteriormente la selección de un clasificador depende de las imágenes que estamos utilizando, para imágenes en blanco y negro son de mas aceptación en la Tabla 6, se puede visualizar que los clasificadores SVM, KNN entregándonos resultados más acertados de las imágenes entrenada.

Tabla 6

Porcentajes de comparación entre clasificadores utilizados.

| Clasificador | Porcentaje |
|----------------------------|-------------------|
| SVM Lineal | 65.5% |
| SVM Cuadrático | 58.6% |
| KNN Coseno | 51,7% |
| KNN Fine | 22.4% |
| Arboles de decisión | 21.3% |

Una vez entrenada la base de datos obtenemos un modelo del clasificador que mayor porcentaje de aceptación nos muestre como en este caso son los clasificadores SVM lineal y KNN Coseno, con estos dos modelos, procedemos a realizar las predicciones con las imágenes que no fueron utilizadas durante el entrenamiento.

Cada uno de estos modelos fueron entrenados en las emociones correspondientes a las indicadas por *Ekman and Friesen*, en este orden: Asco, Enojo, Feliz, Miedo, Sorpresa y Triste.

4.1. Resultado Obtenidos con clasificador KNN

Como se mencionó anteriormente solo se utiliza el 60% de las imágenes para entrenarlas y el 40% de ellas para predecir la emoción, mencionado esto se utilizaron las imágenes restantes para realizar las predicciones, las primeras pruebas se realizaron con el modelo obtenido a través del clasificador KNN.

Como resultado de esta primera predicción podemos decir que el sistema da una predicción correcta en cuanto a 4 emociones como son feliz, triste, sorpresa y enojo, con cierto porcentaje de confusión con otras emociones, en cuanto a las predicciones de las otras 2 emociones como son asco y miedo, muy seguido el sistema predice una confusión errónea confundiendo con las otras emociones, como se muestra en la matriz de confusión representada en la Tabla 7.

Tabla 7
Matriz de confusión de clasificador KNN.

| | Asco | Enojo | Feliz | Miedo | Sorpresa | Triste |
|-----------------|-------------|--------------|--------------|--------------|-----------------|---------------|
| Asco | 26% | 20% | | 39% | 7% | 7% |
| Enojo | 33% | 27% | 5% | 35% | | 3% |
| Feliz | | | 65% | 35% | | |
| Miedo | 40% | 25% | | 25% | 5% | 5% |
| Sorpresa | | | | 13% | 88% | |
| Triste | 18% | | | 24% | 7% | 50% |

En la matriz de confusión podemos evidenciar con más claridad el porcentaje con el cual se confunde las emociones, como mencionamos anteriormente feliz se confunde en un porcentaje

casi imperceptible con sorpresa y con asco ya que en la emoción sorpresa algunas imágenes de la base de datos que fue utilizada, tenía los ojos abiertos mientras se realizaba esta expresión, y se confunde también con asco pues esta expresión por lo general implica la presentación de los dientes, con estos coincidencia el sistema tiende a confundir estas emociones, confusión que no sucede con la expresión sorpresa, el sistema detecta esta emoción sin ningún inconveniente, rara vez se confunde con miedo, pero esto sucede en un porcentaje relativamente bajo.

En la expresión de miedo los sujetos de las imágenes tienden a estirar la comisura de los labios hacia atrás, característica que es muy similar a la de la emoción de felicidad y por ende tiende a confundir también con la emoción asco.

Esta confusión se presenta ya que la base de datos con la que estamos trabajando son expresiones actuadas.

4.2. Resultado Obtenidos con clasificador SVM

Como resultado de esta primera predicción podemos decir que el sistema de predicción es eficiente en cuanto a 4 emociones como son feliz, triste, sorpresa, muy seguido la predicción se confunde entre asco, enojo y miedo, como semuestra en la matriz de confusión representada en la Tabla 8.

Tabla 8

Matriz de confusión de clasificador SVM

| | Asco | Enojo | Feliz | Miedo | Sorpresa | Triste |
|-------|------|-------|-------|-------|----------|--------|
| Asco | 25% | 13% | | 38% | 13% | 13% |
| Enojo | 5% | 70% | 5% | 7% | | 3% |
| Feliz | 6% | | 88% | | 6% | |

Continúa



| | | | | | |
|-----------------|-----|-----|-----|-----|-----|
| Miedo | 50% | 10% | 20% | 20% | |
| Sorpresa | | | 6% | 94% | |
| Triste | 9% | | 7% | 7% | 76% |

En esta matriz de confusión podemos evidenciar en porcentajes que es menos acertada que la matriz de confusión obtenida con el clasificador KNN ya que en la emoción sorpresa en la Tabla 8 muestra un porcentaje del 94% y en la Tabla 7 muestra un porcentaje del 88%, lo mismo sucede con la emoción triste y la emoción feliz existe una disminución en el porcentaje de acertar la emoción a través de la predicación.

Con estos valores presentados podemos evidenciar que el modelo obtenido a través del clasificador SVM es más eficiente al momento de hacer una predicción más acertada a la realidad. A diferencia del modelo obtenido a través del clasificador obtenido KNN ya que este clasificador nos da una predicción menos acertada a la realidad, ya que las emociones que predice de forma correcta son: feliz, sorpresa y triste.

A continuación en la Figura 17, podemos visualizar los resultados del clasificador SVM frente al clasificador KNN.

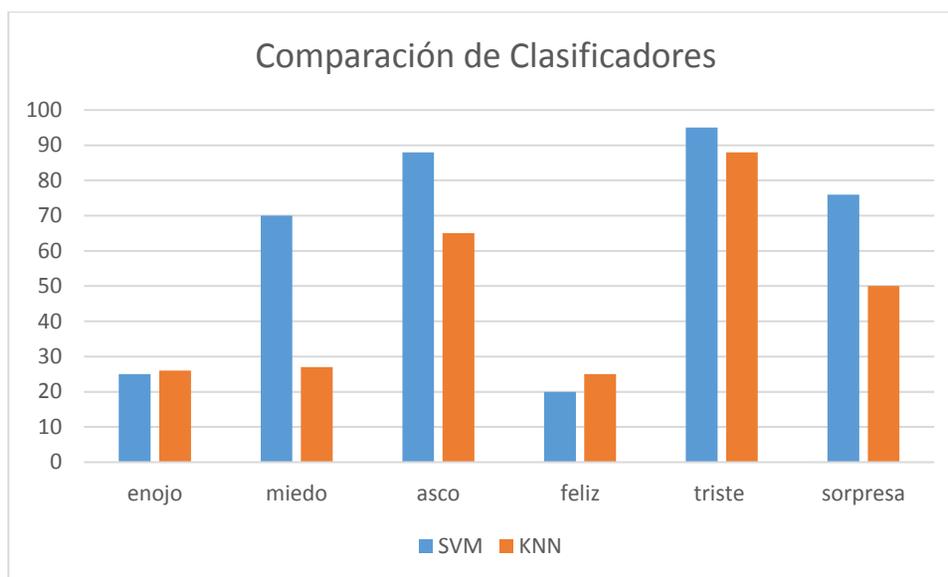


Figura 17. Comparación de clasificadores

4.3. Resultados de Comparación

Al extraer las características de las imágenes de rostro no usadas en el entrenamiento se tuvo unos inconvenientes con la asignación del margen, como ya se mencionó en la sección 3.3.2.

En este proceso específicamente se presentaron ciertos inconvenientes en la detección de la boca, cuando se ingresaba imágenes de la emoción sorpresa específicamente, y que no detectaba con gran facilidad la boca abierta, la solución al presente problema presentado fue jugar con el valor del margen de extracción. Sin embargo, el resultado obtenido en la extracción de la boca de las otras 5 emociones fue satisfactorio, ya que los detectaba sin problemas.

Al realizar la comparación entre los resultados obtenidos de la predicción realizada a través de un sistema previamente entrenado y un programa que lo que hace es comparar con las características extraídas a través de PCA, con una base de datos, como es en este caso “entrenamiento”. En el programa entrenamiento no existe confusión de emociones simplemente

nos entrega un resultado de no se puede reconocer la emoción, lo que no sucede con la predicción en la Figura 17 podemos apreciar el porcentaje de aceptación de la emoción.

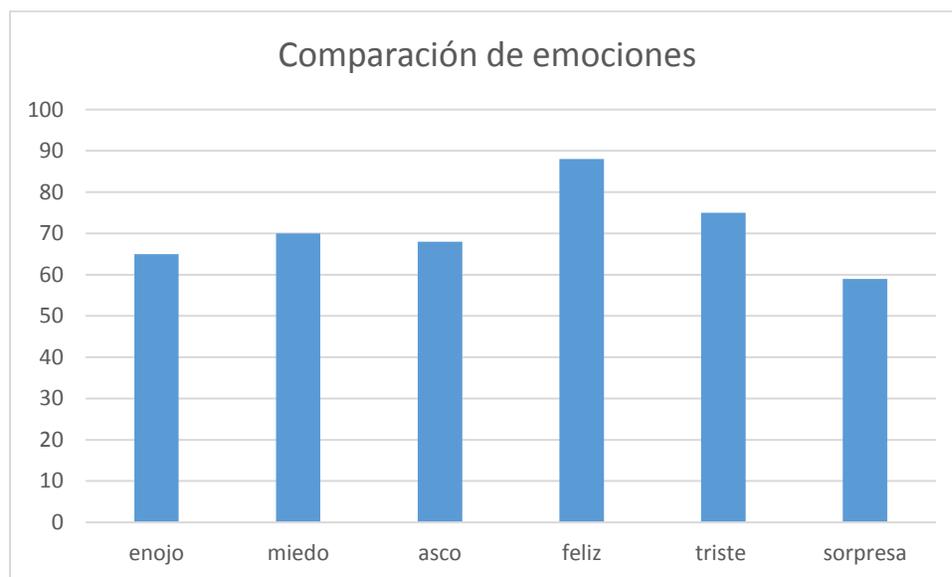


Figura 18. Comparación de Emociones

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Los clasificadores KNN y SVM utilizados para generar modelos de entrenamiento que nos ofrece Matlab, fueron seleccionados porque se ajustaron a los datos con los que contó para realizar el entrenamiento y seguidamente la predicción de los mismos, sin dejar de lado las recomendaciones de usar simplemente el 60% de los datos para entrenar el sistema y el 40% para obtener una predicción del sistema ya entrenado.
- Se obtuvieron dos modelos de entrenamiento el modelo obtenido con el clasificador KNN y el modelo obtenido con el clasificador SVM para realizar la predicción de la emoción, tras evaluar los resultados obtenidos con cada uno de estos modelos el modelo más adecuado para realizar una predicción acertada es el modelo obtenido a través del clasificador SVM ya que este modelo tiene menor porcentaje de error al momento de predecir las emociones, adicional este modelo se equivoca en 2 emociones de manera seguida como son: asco y miedo, en comparación al modelo SVM que se presenta resultados erróneos en 3 emociones como son: miedo, enojo y asco.
- Se logró realizar un programa que a través de la comparación de las características indica la emoción de la imagen que se utiliza para realizar las predicciones.
- Se concluye que sería más eficiente realizar un sistema previamente entrenado con AUs para realizar la predicción de emociones, ya que al trabajar con AUs generaría mayor eficacia a la hora de predecir una emoción.

- Se concluye la mejor técnica para trabajar con imágenes es utilizando el clasificador SVM ya que como se muestran los resultados de este proyecto muestra mayor eficiencia con respecto a otras técnicas probadas.

5.2. Recomendaciones

- Se recomienda pulir este sistema ya obtenido para poder predecir imágenes no solo con bases de datos previamente extraídas sino con imágenes que se tomados directamente desde la webcam desde cualquier ángulo, no simplemente de frente.
- Se recomienda generar una base de datos con las 6 emociones universales básicas como son: Asco, Enojo, Miedo, Feliz, Sorpresa, Triste, de un solo sujeto, para la utilización de este sistema de entrenamiento y posteriormente ampliar esta base de datos para poder realizar predicciones usando este sistema.

Bibliografía

- Pale., M. &. (2016). Análisis del reconocimiento de las expresión emocional en niños y adolescente mexicanos. *Analysis in recognition of facial empresion of emotion in mexican children and adolecents*. Ciudad de Mexico, Mexico.
- Ebner, N. C. (2013). Processing own-age vs. other-age faces. *Processing own-age vs. other-age faces: Neuro-behavioral*.
- Mills, K. L. (2014). Developmental changes in teh estructurree of the social brain in lete chilbood and adolecence. *Developmental changes in teh estructurree of the social brain in lete chilbood and adolecence*.
- Navarro, P. F. (2014, Junio 12). *Habilidad Social*. Retrieved from Habilidad Social: <https://habilidadesocial.com/el-lenguaje-corporal/>
- Adolphs, R. (2002). Recognizing emotion from facial expressions. *Psychological and Neuralogical Mechanisms*.
- Oatley, K. (2008). *Emotions: A Brief History*. Toronto: John Wiley & Sons.
- Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.
- Cole, P. J. (1989). Spontaneous exxpressive control in blind and sighted children. *Child Developed*, 683-688.
- Ekman, P. (2003). Emotios revealed. *New York: Times Books*.
- Rinn, W. E. (1984). The neuropsychology expression. *Psychological Bulletin*, 52-77.
- Ekman. P., F. W. (2002). *Facial action coding system*. Salt Lake City, UT: Investigator's guide.
- Olmedillo, Diego Zapatero. (2016). *herramienta de reconocimiento facial de emociones en Androi*. Madrid.
- MathWorks. (2019). *The MathWorks In*. Retrieved from The MathWorks In: <https://es.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html>
- D.Matsumoto, H. H. (2011). *Science Brief*. Retrieved from Science Brief: <https://www.apa.org/science/about/psa/2011/05/facial-expressions>
- R.W. Picard, E. V. (2014). Towar machne emotional intelligence. *IEEE Transactions on Pattern Analisys an Machine Intelligence*.
- Voyer, T. y. (2014). Meta-Análisis.

Martinez, D. H. (2011). *Una propuesta para incrementar* . Ingenieria y ciencia.

Klette, R. (2014). *Concise computer Vision: An introduction into theory and algorithms*.
Springer.

Pittsburgh, R. L. (2010). *Affect Analysis Group*. Retrieved from Affect Analysis Group:
<http://www.pitt.edu/~emotion/ck-spread.htm>