



INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA

ARQUITECTURA PARA EL RECONOCIMIENTO DE EMOCIONES BASADO EN CARACTERÍSTICAS FACIALES

TESIS

QUE PARA OBTENER EL GRADO DE
MAESTRA EN SISTEMAS COMPUTACIONALES
P R E S E N T A

ELSA IRENE HERRERA SANTIAGO

ASESOR:

DR. LUIS ALBERTO MORALES ROSALES
DR. IGNACIO ALGREDO BADILLO

Agradecimientos

Este documento es el resultado de una investigación realizada con el apoyo de una beca del Consejo Nacional de Ciencia y Tecnología CONACYT, así como el apoyo económico de becas SUBES, por ello quiero agradecer a las personas que han contribuido a la realización del mismo.

Agradezco en primer lugar al Dr. Luis Alberto Morales Rosales quien fuera mi asesor de tesis, por su apoyo total y paciencia desde los inicios de mi maestría, así como el haber compartido conmigo su experiencia y dedicación.

A dios ser maravilloso, por haberme otorgado una familia hermosa, quienes han creído en mí siempre, dándome ejemplo de superación, humildad y sacrificio; enseñándome a valorar todo lo que tengo. A todos ellos dedico el siguiente trabajo, porque han fomentado en mí, el deseo de superación y de triunfo en la vida. Lo que a contribuido a la consecución de este logro. Espero contar siempre con su valioso e incondicional apoyo.

A mi hija por su afecto y cariño que son los detonantes de mi felicidad, de mi esfuerzo, de mis ganas de buscar lo mejor para ella. Aun a su corta edad, me ha enseñado y me sigue enseñando muchas cosas de esta vida. Le agradezco por ayudarme a encontrar el lado dulce y no amargo de la vida. Fuiste mi motivación más grande para concluir con éxito este proyecto de tesis.

Índice

1. INTRODUCCIÓN

- 1.1. Problemática
- 1.2. Objetivos
- 1.3. Justificación
- 1.4. Propuesta de solución
- 1.5. Hipótesis
- 1.6. Metodología
- 1.7. Alcances y Limitaciones

2. ESTADO DEL ARTE

- 2.1. Reconocimiento de emociones
 - 2.1.1. Detección de rostros
 - 2.1.1.1. Algoritmo Viola-Jones
 - 2.1.2. Detección de expresiones faciales emocionales
 - 2.1.2.1. Enfoque basado en rasgos faciales o características locales
 - 2.1.2.2. Enfoque híbrido
 - 2.1.2.3. Enfoque holístico basado en la imagen
 - 2.1.2.3.1. Análisis de Componentes Principales
- 2.2. Aprendizaje supervisado
 - 2.2.1. Aprendizaje por corrección de errores
 - 2.2.2. Con refuerzo
 - 2.2.3. Aprendizaje estocástico
 - 2.2.3.1. Máquina de Vector Soporte

3. METODOLOGÍA

- 3.1. Selección de Variables.
- 3.2. Extracción de características con Algoritmo Viola-Jones.
- 3.3. Reducción de dimensionalidad de la imagen con Transformada de Wavelet.
- 3.4. Obtención de componentes principales con PCA.
- 3.5. Reducción de componentes principales con EigenFace
- 3.6. Clasificación de emociones con SVM.

4. RESULTADOS

5. CONCLUSIONES Y TRABAJO A FUTURO

- 5.1. Conclusiones
- 5.2. Trabajo a Futuro

REFERENCIAS BIBLIOGRÁFICAS

Capítulo 1

1. INTRODUCCION

Actualmente, el reconocer la emoción en una persona es una actividad cotidiana de los seres humanos, resulta hasta simple mirar a una persona y predecir su comportamiento de acuerdo a la emoción que reconocemos en ella. Esta acción puede ser riesgosa y hasta engañosa, ya que una emoción depende de diferentes factores propios de los humanos como son el envejecimiento, expresiones faciales, cambios de iluminación, puntos de vista inducidos por el movimiento del cuerpo y oclusiones. Sin embargo, lo que para el cerebro resulta ser una actividad rápida y sencilla, para la visión artificial resulta ser una operación mucho más compleja basada en estímulos visuales, detalles y matices de los rostros. Desde el siglo antepasado, ha sido interesante construir un sistema automático para la identificación de individuos a través de su cara, de tal manera que, en 1889 Sir Galton buscaba fórmulas matemáticas para solucionar el problema.

Es por ello que, en los últimos 20 años de estudio, muchos investigadores establecen el reconocimiento de caras y emociones como un tema lejos de resolverse, ya que el replicar esta conducta humana tan cotidiana parece ser un problema computacional muy complicado. Dicho tema ha dado la pauta para un gran número de artículos publicados en revistas y conferencias enfocados al área de visión artificial.

En esta tesis se propone una técnica para el reconocimiento de emociones creando una arquitectura capaz de contemplar métodos específicos para la solución del problema, es por ello que se basará en el Algoritmo de Viola-Jones para la obtención de componentes de una región especificada (en este caso el rostro humano) tales como cejas, ojos, nariz, boca, etc.

Con lo anterior, se aplica una transformada Wavelet, con la cual se pueden crear a base de Wavelet lineales una o varias Wavelets ortonormales fundamentadas en la resolución y las escalas de la imagen para la obtención de un plano preciso. Esto lleva a la elección de un método de análisis de componentes principales o PCA (Sirovich y Kirby, 1987; Kirby & Sirovich, 1990) el cual está basado típicamente en dos fases: formación y clasificación. Pero aunado a esto, se puede constituir un clasificador de emociones como es SVM, que ayuda a clasificar las características específicas conforme a similitudes o diferencias y crear así un vector a comparar con las plantillas creadas de emociones.

Dichos métodos serán los constituyentes de la arquitectura propia a desarrollar para el reconocimiento de emociones, como base de múltiples investigaciones. Sin embargo, no se pretende encasillar su uso a una sola área, sino dar la flexibilidad de implementación en diferentes ramas del área de visión artificial.

1.1. Problemática

La interacción diaria con diferentes seres humanos es parte del desarrollo integral del ser humano. No sólo los gestos faciales codifican el estado de ánimo de un ser humano, sino que las características físicas de la cara nos identifican como individuos únicos en un mundo superpoblado. Pero qué pasaría si se diseña una arquitectura capaz de realizar una de las actividades mentales más comunes en nosotros como es el reconocimiento de emociones.

Lo anterior lleva a la importancia de las emociones dentro de un papel como sociedad, ya que forman parte de una interacción entre un grupo de individuos, aprendiendo unos de otros de acuerdo a gestos faciales -características físicas de las caras- haciéndonos únicos del resto de las personas, dando la oportunidad de conocernos más a fondo y detectar los estados de ánimo que influirán en la toma de decisiones diarias.

En este último punto muestra que es imprescindible tener un conocimiento previo de las emociones humanas ya que de ello el ser humano es predecible en sus decisiones pudiendo, de hecho, mejorarlas. Antes de reconocer sus emociones se deben plantear los tipos de emociones que el ser humano demuestra a lo largo de su vida. Con base en la teoría de Ekman y Friesen de 1978 que sustenta los estados emocionales del ser humano constituyéndolos como cambios de estados de ánimos marcados facialmente y corporalmente, se logran agrupar las emociones en 6 categorías: alegría, sorpresa, miedo, disgusto, enojo y tristeza.

De acuerdo a la clasificación de emociones antes descrita, se puede agregar el reconocimiento de dichos estados emocionales como una de las tareas más difíciles para el ser humano de simular. Es decir, este es un proceso computacional complejo que requiere de gran cantidad de recursos temporales y espaciales. Es posible dar un comienzo a dicha tarea empezando con el análisis de una imagen facial donde Gabor en 1980 fue más allá y logró crear filtros capaces de obtener información espacial y la orientación de la textura de dicha imagen, dando lugar a analizar factores relevantes como la iluminación, postura, distancia, simetría y laterabilidad, teniendo influencia en la obtención de una imagen del estudiante como parte del estudio del reconocimiento de emociones.

Por lo anterior, se debe tener en cuenta las grandes diferencias de los más de 7000 millones de habitantes en el mundo, por ejemplo, su nacionalidad, raza, género y hasta su gesticulación. Sin embargo, aun cuando se tienen detectados los tipos de

emociones y sus factores de interpretación, es preciso ubicar en ¿qué se va a reconocer? Se sabe que de primera instancia sería una emoción, pero ¿cómo se sabe que es o no una emoción?, es en este punto es donde se debe considerar la detección de singularidades, es decir, reconocer una cara de una no cara y cada elemento que se encuentre dentro de ésta tomando en cuenta características específicas como son ojos, cejas, nariz y boca como lo mencionaron Lien en 1998, y Kulas y Kanade en 1981, para caracterizar las partes que serán útiles en el reconocimiento de la emoción. Precisamente en esta situación es en donde se presenta un problema más, ya que en base al reconocimiento de la emoción en un rostro donde ya se han detectados cada uno de sus componentes, la forma de clasificación representa un papel primordial, ya que la información obtenida no es suficiente para distinguir una cara de otra, por ello es preciso definir si dicha clasificación debe ser en base a sus similitudes o sus diferencias.

1.2. Objetivos

En esta sección se describen el objetivo general y los objetivos específicos para realizar este proyecto de tesis.

1.2.1. Objetivo general

Diseñar y desarrollar una arquitectura para el reconocimiento de emociones en personas basada en las características faciales presentadas en una foto o video, usando el algoritmo de Viola-Jones, la transformada Wavelet, PCA y SVM.

1.2.2. Objetivos específicos

Para alcanzar el objetivo general, el presente proyecto de tesis ha de llegar a otros objetivos específicos que se describen a continuación:

- Determinar el medio o contexto a considerar del que se obtendrá la cara a reconocer, para evaluar factores como la iluminación, postura, distancia, simetría y laterabilidad y obtener la emoción que esté presente en el individuo.
- Seleccionar las características faciales de análisis de acuerdo a la nacionalidad, raza y género de los individuos para la definición de tipo de cara a reconocer y las emociones que serían discriminadas de la misma.
- Extraer datos y características de un espacio dimensionado llamado cara, para detectar elementos como ojos, nariz, boca, cejas y contorno como base de estudio y clasificación posterior.

- Crear un vector de características principales de acuerdo a cada rostro, para identificar sólo los rasgos específicos de cada emoción.
- Seleccionar un clasificador para establecer los tipos de emociones como son: felicidad, sorpresa, miedo, disgusto, cólera y tristeza.

1.3. Justificación

La interacción diaria con diferentes seres humanos es parte del desarrollo integral del ser humano. No sólo los gestos faciales codifican un estado de ánimo, sino que las características físicas de la cara permiten la identificación como individuos únicos en un mundo superpoblado. Lo anterior lleva a la formalización del análisis de los rasgos biométricos, los cuales pueden ser divididos en: 1) rasgos fisiológicos como pueden ser las huellas dactilares, el iris, el ADN, etcétera, o 2) rasgos del comportamiento como la firma, el modo de andar, la voz, etc. En específico, los primeros tienen una menor variabilidad a lo largo del tiempo, así como también son mucho más invasivos y requieren de la cooperación del individuo para su obtención. Los rasgos faciales suelen ser más difíciles de duplicar o falsificar frente a otros métodos. Esto se debe a que estos rasgos utilizan características propias de cada individuo y únicas de modo que pueden diferenciarlo de otras personas que intenten falsificar su identidad.

Dentro de los rasgos biométricos el reconocimiento facial se ha convertido en un área de investigación bastante popular. Debido a la naturaleza del problema, no sólo los investigadores del ámbito de la ingeniería están interesados en él, sino además otros del ámbito de la neurociencia y la psicología.

La detección de caras lleva desarrollándose desde hace algún tiempo. En la década de los 70 surgieron los primeros algoritmos, pero las investigaciones se abandonaron debido a la falta de utilidad debida a la falta de desarrollo de la técnica. Sin embargo, en la actualidad esta línea de investigación tiene numerosas aplicaciones tanto en el ámbito comercial con las cámaras de fotos, como en el de la seguridad y la investigación criminal, el control a zonas restringidas, el control de fronteras, entre otras. Por otro lado, la mayoría de los algoritmos de detección facial pueden extenderse para el procesamiento de otros objetos como coches, peatones, etcétera, o para la localización de otras áreas faciales.

Sin embargo, la cara no es lo único que identifica a los seres humanos, sino también las emociones ya que permiten relacionarnos con el resto de individuos, reconociéndolas en los otros y comportándonos acorde a ellas. Asimismo, los estados afectivos regulan las interacciones de acuerdo a ciertas reglas sociales que regulan el comportamiento y pueden inhibir o desarrollar algunas emociones más que otras.

El cuerpo humano es una evidencia permanente de las emociones, por lo que estas reacciones fisiológicas pueden influir poderosamente en el estado psicológico de las

personas, pudiendo llevarlas desde un estado de estabilidad a un estado menos ajustado.

En cuanto al comportamiento emocional, éste se desencadena frente a situaciones emocionales. Sin embargo, éstas pueden ser incongruentes con lo que las personas verdaderamente están sintiendo, por lo tanto, no basta con la observación de la acción llevada a cabo, sino incluirla dentro de un determinado contexto. Es por ello que desde el Siglo XX, surge la inteligencia emocional como resultado al éxito académico y laboral. Es precisamente basándose en este dato donde nace la inquietud de reconocer las emociones de cada persona debido a la toma de decisiones, ya que de acuerdo a la situación emocional que puede estar una persona, esta última puede tomar decisiones y afectar a segundas personas o incluso a sí mismo. El reconocimiento de emociones es influenciado por el estado emocional en el que se encuentre inmersa la persona al momento de la toma de foto o video una persona.

Con lo anterior, la creación de una arquitectura que sea capaz de reconocer emociones es de importancia para el sustento de aplicaciones basadas en la toma de decisiones de una persona. Ya que tomando como base dicha arquitectura su implementación se hace adaptable en diferentes áreas de visión artificial, donde la base de aplicación es el reconocimiento de emociones de acuerdo a los estados de ánimo y emocionales que presente cada persona, dando lugar a un rendimiento académico incrementable en medida de su utilización.

1.4. Propuesta de solución

Se desarrolla una arquitectura capaz de reconocer los estados emocionales de una persona. Para la obtención de dicha arquitectura se fundamenta en el modo de extracción de una emoción y por medio de qué técnica se puede lograr. Es decir, se considera que la extracción de las emociones debe ser un proceso transparente para la persona y que a través de una imagen o video se pueden obtener las características necesarias para la determinación de rasgos específicos de las emociones.

Partiendo de la toma de una imagen o video, se utilizan métodos de extracción de datos y características punto a punto, como la transformada de Wavelet que ayuda a localizar pequeñas variaciones en imágenes que el ojo humano no detecta a simple vista. Por lo anterior se segmentó la imagen en detalles a diferentes niveles de resolución y orientaciones, sin repeticiones, comenzando a realizar la similitud cuadro por cuadro, localizando las que se asemejen y las que formen la función a analizar.

Obteniendo factores determinados por la misma transformada de Wavelet de paquetes introducida por Coifman & Meyer en 1990, ésta constituye paquetes basados en una resolución múltiple ortogonal compuesta por Wavelet lineales, que proporcionan la información espacial, orientación y posición de la imagen contemplando la textura de la misma a diferentes escalas y resoluciones. Estos son

resultado del contexto que se le dio a la imagen de manera física para la experimentación, dicho contexto será determinado por el diseñador, implementando variables como iluminación, postura, distancia, simetría y laterabilidad.

Se plantea un método de clasificación por medio de los rasgos obtenidos y clasificados, determinando las características de cada rostro por medio de un Análisis de Componentes Principales (PCA), propuesto por Sirovich y Kirby en 1987. Adicionalmente, se utiliza 1DPCA propuesto en el trabajo de Kirby & Sirovich (1990), y 2DPCA de Turk y Pentland (1991) los cuales proporcionan datos cuantitativos que serán comparados para la obtención de similitudes o diferencias, donde dichos rasgos serán discriminados de la imagen como ojos, nariz, boca, cejas y contorno, dimensionando las características para su especificación posterior.

En este sentido, se da paso a separar la imagen en dos planos con características específicas cada uno, conformando un vector a base de puntos de similitud y diferencias especificados como muestra y clasificador. Esto es logrado por una Máquina de Vector Soporte (SVM) basada en Min-Max (M3-SVM) y descomposición de tareas según Lu, BL, Ma, Q., Ichikawa, M., Isahara, H., en 2003, que proporciona una ventaja muy importante sobre la SVM convencional, ya que implementa un “divide y vencerás”, permitiendo que el riesgo a error entre la muestra y el clasificador sea menor.

Por último, dicha arquitectura puede ser implementada en cualquier área de visión artificial, tomando como referencia la toma de decisiones a partir de una emoción.

1.5. Hipótesis

Es posible desarrollar una arquitectura para reconocer los estados de ánimo de una persona basada en sus características faciales aplicando los algoritmos de Viola-Jones, Transformada de Wavelet, PCA y SVM.

1.6. Metodología

1. Seleccionar los rostros de análisis de acuerdo a la nacionalidad, raza y género de los individuos.
2. Extraer los componentes de la región cara, como son: ojos, nariz, boca, cejas y contorno, usando el algoritmo Viola & Jones.
3. Evaluar factores como la iluminación, postura, distancia, simetría y laterabilidad.
4. Detectar la dimensionalidad del rostro del individuo, con la finalidad de distinguir entre una cara y una no cara.

5. Establecer un plano de la imagen, dimensionando las características del rostro cambiando texturas y resolución para la obtención completa de elementos.
6. Definir las singularidades dentro de un espacio dimensionado y detectado, dichos elementos a identificar serían los extraídos por el algoritmo Viola-Jones.
7. Detectar los rostros de acuerdo a sus características tomando en cuenta sus similitudes o diferencias.
8. Establecer características y no diferencias, con lo cual conlleva una reducción de la dimensionalidad de la imagen, por medio de análisis de componentes principales.
9. Realizar la clasificación de emociones por medio de una máquina de vector soporte alimentado por un vector que contiene características específicas de acuerdo a la descomposición de la imagen.

1.7. Alcances y limitaciones

El proyecto en desarrollo tiene como alcance la creación de una arquitectura capaz de reconocer emociones de acuerdo a las características faciales de una persona.

Los aspectos puntuales que comprende la investigación son los siguientes:

- Se establece los tipos de emociones a clasificar: felicidad, sorpresa, miedo, disgusto, cólera y tristeza. De acuerdo a FACS por los movimientos faciales, se clasifican de manera subjetiva.
- Obtención de características de una región rostro, como son: ojos, nariz, boca, cejas, etcétera, por medio del algoritmo Viola-Jones.
- Establecimiento de un plano con resolución ortonormal basado en Wavelet lineales para la obtención de un paquete establecido por diferentes características de la imagen.
- Obtención de características específicas basadas en similitudes o diferencias, seleccionadas por PCA basada en 1DPCA y 2DPCA.
- Clasificación de emoción de acuerdo a un SVM, basado en M3-SVM dividiendo el problema en subclases de análisis y comparación.

Las presentes limitaciones restringirán la investigación:

- Se selecciona la nacionalidad mexicana, la raza Amerindios y el género Hombre y Mujer de estudiantes con edades correspondientes entre 18-20.
- Descartar obstrucciones como: envejecimiento, gesticulación, cicatrices, objetos.

Capítulo 2

2. ESTADO DEL ARTE

Una emoción es la alteración del estado de ánimo de una persona producida por un sentimiento. La representación de la emoción se denota con la modificación de los rasgos faciales, ademanes de rasgos característicos físicos, cambio de la frecuencia cardíaca y tono de la voz.

El área del conocimiento encargado de estudiar el comportamiento de los seres humanos al mostrar las emociones es la psicología. Donde el reconocimiento de emociones a través de los rasgos faciales muy definidos, por ser una técnica no invasiva al usuario y que no hace uso de herramientas de alto costo.

Estas técnicas tomaron auge a partir de la clasificación inicial que hizo Ekman a las emociones, definiendo 6 emociones básicas: felicidad, enojo, miedo, sorpresa, tristeza y disgusto. Y en parte también por el establecimiento del sistema de codificación de acciones faciales (FACs).

Con las FACs fue posible determinar la emoción presentada de una persona con solo observar su rostro. Esto sirvió de base para que áreas como la inteligencia artificial se interesaran en el reconocimiento de emociones y se crearán sistemas automáticos para su identificación.

Muchos de los trabajos empezaron a centrarse en las FACs para definir sus sistemas de reconocimiento. Sin embargo, no son las únicas formas de identificar emociones. También se encuentran:

2.1 Reconocimiento de emociones

2.1.1 Detección de rostros

- Guevara J. Aplica la Transformada Wavelet de Morlet, en la detección de rostros, además describe algunas técnicas comunes utilizadas para la detección de rostros en una escena, también describe un algoritmo para el proceso de detección de rostros utilizando las Wavelet de Morlet. Sin embargo, la detección de rostros en una escena no permite minimizar la imagen, ya que comprende la resolución completa de la imagen, incorporando características fuera de la región cara.

- Guevara M. L., Echeverry J. D. y Ardila U. W. implementan y aplican un método de detección de rostros y características faciales sobre imágenes digitales, basado en la extracción de características con filtros de base Haar y clasificadores en cascada que permitió un porcentaje de detección del rostro y de los ojos del 100% y 92% respectivamente en el mejor de los casos, con un tiempo de procesamiento bajo. Sin embargo, cuando se prueba con características específicas como boca o cejas, el porcentaje de detección disminuye ya que maximiza la imagen y se vuelve poco detectable.

- Serrano A, Conde C., De I. M., Cabello E., Bai L., Shen L. y demás proponen mejorar la tasa de error de igualdad para el experimento de la oclusión en comparación con un método de disminución en la resolución Gabor PCA y obtienen resultados similares en los otros experimentos con menores coeficientes después de la etapa de reducción de dimensión PCA. Sin embargo, cuando se vectorizan las características principales la tasa de error es confusa ya que hay falsos positivos y falsos negativos que disminuyen el funcionamiento, por ello se debe especificar la dimensionalidad del vector de características.

- Shen L., Bai L. y demás presentan un estudio detallado de los algoritmos de reconocimiento facial 2D utilizando Wavelets de Gabor para la extracción de características, tomando como su base de estudio a los rasgos específicos como las distorsiones locales causadas por la varianza de iluminación, expresión y postura, que han sido exitosamente aplicadas para el reconocimiento facial. Sin embargo, aún no sale del pre-procesamiento de la imagen, por ello no existe un porcentaje de detección aproximado.

- Arandjelovi O. y Cipolla R. realizan reconocimiento con precisión de los rostros humanos en presencia de cambios grandes e impredecibles de iluminación. Su objetivo es hacer esto en una configuración realista para aplicaciones más prácticas, es decir, sin restringir excesivamente las condiciones en que los datos de la imagen se adquieren. En concreto, esto significa que el movimiento y la cabeza de la gente no son controlados, por ello, la cantidad de datos de entrenamiento disponibles se limita a una sola secuencia por persona, proporcionando una calidad de imagen baja. En condiciones como éstas, el cambio de iluminación es quizás el desafío más significativo para los algoritmos de reconocimiento facial. Sin embargo, con este método se obtendrían mejores resultados si la base de datos de reconocimiento se limita a determinadas variabilidades de iluminación las cuales podrían ser controladas con éxito, y dependerá de que la detección sea o no en tiempo real.

- Campadelli P., Lanzarotti P. y Lipori G. presentan un método automático para la extracción de características faciales que se utiliza para la inicialización de técnicas de reconocimiento de rostros. Para extraer los componentes faciales iguales se determinan puntos característicos, por ejemplo, el centro y la esquina de los ojos, la punta de la nariz, etc. Se hace especial hincapié a la localización de las características más representativas de la cara. Un aspecto importante de cualquier algoritmo de localización es su precisión, por ello aseguran el detalle de tales características, pues

piensan que maximizando los detalles lograrán mayor precisión en el reconocimiento facial. Sin embargo, al vectorizar las características se vuelve tedioso, ya que la imagen se ha maximizado por el nivel de detalle de la obtención de puntos de localización de rasgos faciales.

- Dao Dai Q. y Yan H. presentan un resumen de la transformada Wavelet en paquetes para su uso en el reconocimiento facial. Ellos extraen características a través de algunas transformaciones lineal o no lineal de los datos con la selección de características para reducir la dimensionalidad de las imágenes faciales, de modo que la característica extraída sea lo más representativo posible. Las transformadas Wavelets se han utilizado con éxito en el procesamiento de imágenes, su capacidad de localizar y dar información en tiempo real de una imagen motiva su uso para la extracción de características. La descomposición de los datos en diferentes rangos de frecuencia permite aislar la frecuencia de los componentes introducidos por deformaciones intrínsecas debido a factores como la iluminación en ciertas sub bandas, que contienen la información más relevante para mejorar y representar los datos.

2.1.1.1 Algoritmo viola & jones

- Bogdan A. *et al.* (2010) implementan un algoritmo capaz de detectar objetos (caras) sobre imágenes además de poder ser evaluado en diferentes condiciones y entornos de trabajo. Implementan y evalúan un detector facial y un detector de ojos en imágenes, basados en el modelo de Viola-Jones. Como punto de partida realizaron un estudio y evaluación de los sistemas biométricos y un estudio del estado del arte de los sistemas de detección facial desarrollados hasta el momento, presentando finalmente el mecanismo seleccionado. Sin embargo, aunque realizan la clasificación de rostros, les falta el pre-procesamiento de la imagen con lo cual podrían disminuir el tiempo de ejecución y elevar el porcentaje de detección.

2.1.2 Detección de expresiones faciales emocionales

- Du S., Martínez A. M., *et al.* (2012) diseñaron un modelo que justifica los resultados reportados en la literatura de la ciencia cognitiva y la neurociencia. Según este modelo, la tarea principal para la clasificación de las expresiones faciales de la emoción es la detección precisa y detallada de los puntos de referencia faciales en lugar de reconocimiento. Sin embargo, no hace a un lado oclusiones ni cambios de iluminación, lo cual entorpece los resultados.

- Du S., Martínez A. M., *et al.* (2011) demuestran un estudio sistemático de la dimensionalidad de las imágenes de las expresiones faciales con emoción. En particular, se investiga cómo el reconocimiento se degrada cuando se reduce la resolución de la imagen, es decir, cuando se reduce el número de píxeles cuando se analiza cuadro a cuadro. Sin embargo, la comprensión de cómo cada ser humano

concibe una emoción es variable de acuerdo a su contexto sociocultural, por lo tanto, encasillar las emociones deja fuera la existencia de otras detectables.

- Cereceda S., Pizarro I., Valdivia V., Hurtado F. E., Ibáñez A., *et al.* (2010) estudian la posible asociación implícita entre el procesamiento de estímulos emocionales de rostros y un contexto semántico específico. Se investigó utilizando el paradigma IAT (Investigación Acción Transformadora). Los hallazgos señalan que los tiempos de reacción de los sujetos son menores en los bloques congruentes que en los incongruentes, por lo tanto, favorecen implícitamente la asociación congruente por sobre la incongruente. En resumen, el IAT de emociones dio como resultado un fuerte y consistente efecto IAT a nivel comportamental, sugiriendo que existe una asociación a nivel intuitivo. Tomando la Teoría de Ekman como base en la interpretación y clasificación de emociones, esto no puede ser posible, ya que una emoción está constituida por rasgos faciales.

- Pantic M. y Stewart M. analizan a las expresiones faciales, es decir las emociones mostradas diariamente, en donde se presenta que el rostro humano es un medio por excelencia de comunicación y la comprensión, así como las intenciones sobre la base de la expresión facial de muestra afectiva de alguien (Keltner y Ekman, 2000). Por lo tanto, el rostro humano es una entrada-salida multi-señal conformada por un sistema de comunicación con gran flexibilidad y especificidad (Ekman y Friesen, 1975). Sin embargo, el establecer una plantilla preestablecida para cada emoción deja a un lado características contextuales de los individuos.

2.1.2.1 Enfoque basado en rasgos faciales o características locales

- Blázquez L., *et al.* (2013) estudian, implementan y evalúan un sistema automático de detección y corrección de puntos característicos faciales, obtenidos mediante un sistema comercial automático. Debido a la existencia de ruido en las imágenes, desarrollaron un método de pre-procesado de imágenes, el cual lleva a cabo el alineamiento, normalización y minimización a escala de las imágenes. Sus estudios se han centrado en observar el potencial del detector y del corrector de puntos faciales desarrollado en el proyecto, tanto en entornos controlados, como en entornos no controlados. Sin embargo, el aplicar dicha técnica conlleva a la corrección de puntos característicos faciales y esto al derroche de recursos de cómputo.

- Cruz P. C., *et al.* (2008) proponen un sistema de reconocimiento de rostros en ambientes con condiciones interiores realistas y un tiempo de respuesta adecuado para un robot móvil. El sistema es capaz de aprender en línea un nuevo rostro basado en una sola imagen, la cual es usada posteriormente para reconocer a la persona bajo diferentes condiciones en el ambiente. Una etapa de pre-procesamiento es usada para reducir el efecto de las diferentes condiciones de iluminación, y entonces identificar tres regiones en el rostro: ojo izquierdo, ojo derecho y nariz-boca. Las características son extraídas de cada región y son almacenadas en un vector de características, el cual es

usado para su reconocimiento. La estrategia de correspondencia es capaz de descartar rostros desconocidos y un enfoque Bayesiano mejora la precisión sobre múltiples imágenes. Sin embargo, no es ideal en el tiempo de pre-procesamiento puesto que la búsqueda de características es lenta.

- Aguerrebere C., Capdehourat G., *et al.* (2006) proponen una técnica donde forman agrupamientos debido a rasgos similares o condiciones comunes. Destaca una importante característica del descriptor y guarda la información de la persona. Por lo tanto, ratifica como algoritmo el reconocimiento de caras. Las cotas de desempeño (*performance*) obtenidas son muy prometedoras, donde se destacan las propiedades del descriptor vistas con el agrupamiento. Sin embargo, dicha técnica mejora el reconocimiento si la imagen es maximizada, y tiene casos con éxito en posición frontal, por lo tanto, para los casos de mayor dimensión es una mejor opción, dejando a un lado casos de dimensiones pequeñas.

2.1.2.2 Enfoque híbrido

- Mazanec J., Melisek M., *et al.* (2008) presentan resultados de los diferentes algoritmos estadísticos utilizados para el reconocimiento de caras, PCA (Análisis de Componentes Principales), LDA (Análisis discriminante lineal) y SVM (*Support Vector Machines*), así como para el pre-procesado (normalización del tamaño, la posición y la rotación, la optimización del contraste y el rostro enmascarado unificado). También se propone una combinación de métodos de PCA y LDA con SVM que produce resultados interesantes desde el punto de vista del éxito reconocimiento, tasa, y la robustez del algoritmo de reconocimiento de la cara. Utilizando diferentes clasificadores para que coincida la imagen de una persona a una clase (un sujeto) obtenido a partir de los datos de entrenamiento. Sin embargo, si estas técnicas se unieran se lograría una mayor eficiencia en el uso de los algoritmos.

- Eleyan A. y Demirel H. proponen dos sistemas de reconocimiento facial, una basada en la PCA seguido de una red neuronal (FFNN) llamado PCA-NN, y el otro sobre la base de LDA seguido por un FFNN llamada LDA-NN. Los dos sistemas consisten en dos fases que son la PCA o LDA función de la fase de extracción, y la fase de clasificación de la red neuronal. Los sistemas introducidos proporcionan una mejora en las prestaciones de reconocimiento sobre el LDA convencional y PCA se enfrentan los sistemas de reconocimiento. Después de generar los vectores propios utilizando métodos PCA y LDA, se calculan los vectores de proyección de imágenes de la cara en el conjunto de entrenamiento y luego se usa para entrenar la red neuronal. Estas arquitecturas se llaman PCA-NN y LDA-NN. Sin embargo, el utilizar una red neuronal conlleva el realizar un vector de entrenamiento y el reconocimiento de caras no podría ser en tiempo real.

2.1.2.3 Enfoque holístico basados en la imagen

2.1.2.3.1 Análisis de componentes principales

- Rodríguez S., *et al.* (2010) proponen una solución al problema del reconocimiento facial mediante la técnica de extracción de características PCA, estudiando el efecto de diferentes preprocesadores aplicados sobre las imágenes. Por lo tanto, los esfuerzos de investigación se centran, por una parte, en la obtención de un algoritmo robusto para la identificación de características, y por otra, en el desarrollo de diversas técnicas de pre-procesado de imágenes que mejoren el rendimiento del algoritmo propuesto. Sin embargo, de acuerdo al planteamiento de la técnica específica la robustez de la misma, lo cual llega a requerir mucho procesamiento de cómputo, llevándolo a ser inaccesible.

- Serrano A., Conde C., De Diego I. M., Cabello E., Shen L., Bai L., *et al.* (2007) proponen una fusión de la función Gabor basado en la Máquina de Vector Soporte (SVM) como clasificador para la verificación de la cara. 40 wavelets se utilizan en paralelo para extraer características para la representación de la cara. Estos vectores extraídos se implementan primero en el correspondiente Análisis de Componentes Principales (PCA), y luego se introducen en SVM para la clasificación y la fusión. En comparación con los tres métodos de referencia desarrollada en la literatura, es decir, PCA, Gabor PCA y Gabor PCA basado en funciones, el algoritmo propuesto logra los mejores resultados en la expresión neutra y experimentos de oclusión. En comparación con un método Gabor-PCA, dicho algoritmo también obtuvo tasas de error similares con una dimensión inferior característica. Sin embargo, no estandarizan el tamaño ni el contenido del vector con características, ya que el vector entrenador es el que determina la eficacia del clasificador.

- Conde C., Ruiz A., Cabello E., *et al.* (2003) experimentaron con tres clasificadores considerados para comparar los resultados obtenidos mediante la PCA versus los resultados obtenidos con las imágenes de baja resolución. Un primer conjunto de caras ubicadas se ha utilizado para el cálculo de la matriz de PCA y para la formación de todos los clasificadores. Los clasificadores considerados son vecinos k-más cercano (KNN), redes neuronales artificiales: función de base radial (RBF) y de Apoyo Vector Machine (SVM). Los resultados muestran que SVM siempre logra mejores resultados que los otros clasificadores. Con diferencia la verificación correcta SVM entre PCA y el procesamiento de baja resolución es sólo un 0,13% (99,52% frente a 99,39%).

2.2 Aprendizaje supervisado

- Serrano A., De Diego I. M., Conde C., Cabello E., Shen L., Bai L. *et al.* (2007) proponen un sistema de verificación de la cara utilizando Parallel Gabor Análisis de Componentes Principales (PGPCA) y fusión de máquinas de soporte vectorial (SVM). El algoritmo ha sido probado en dos bases de datos: XM2VTS (imágenes frontales con frontal o iluminación lateral) y FRAV2D (imágenes frontales con difusa o iluminación cenital, variando poses y oclusiones). Este método supera a los demás cuando se mantienen menos coeficientes PCA. También cuenta con la tasa de error equivalente más bajo en experimentos utilizando imágenes frontales con oclusiones. Sin embargo,

se puede elevar la efectividad si se contemplan oclusiones que discriminen las imágenes.

2.2.1 Aprendizaje por corrección de errores

- Graña M., Beristain A., *et al.* (2007) describen un sistema formado por la adquisición automática de la imagen facial de los sujetos de cualquier edad, la etnia y la apariencia y proponen la utilización de un clasificador SVM. Es un sistema robusto a la variación en los rasgos y de caras parcialmente con oclusiones, donde no hay marcadores especiales o maquillaje. Las posiciones son estables con movimientos de cabeza rígidos logrando con esto detección de rostros automático. La extracción de datos de la expresión facial ha logrado la clasificación de la expresión facial automático dando la oportunidad de distinguir todas las expresiones posibles con cambios faciales unilaterales, obediendo las reglas anatómicas. Sin embargo, se obtiene un 71% de porcentaje de detección en expresiones faciales, clasificando sólo 4 emociones.

2.2.2 Con refuerzo

- Nguyen T., Li M., Bass I., Sethi I. K., *et al.* (2005) realizan una combinación de apoyo de SVM y el aprendizaje de árbol de decisión para reconocer cuatro emociones, que son: neutral, enojado, feliz y triste. Se propusieron tres métodos de combinar aprendizaje máquinas de vectores soporte y clasificadores de árboles de decisión, es decir, la clasificación errónea mínima, máxima precisión, y la clase dominante. 67.4% en detección de emociones, que se obtiene mediante el uso de vectores de soporte de aprendizaje de la máquina y el árbol de decisión solo. Por lo tanto, al clasificar sólo cuatro emociones se tiene una limitante, anexando el excesivo consumo de recursos de cómputo al realizar el árbol de decisiones.

2.2.3 Aprendizaje estocástico

2.2.3.1 Máquina de vector soporte

- Aswathy. R., *et al.* (2013) proponen un nuevo sistema de reconocimiento de la expresión facial basado en el espacio de color perceptual. Se utiliza un lugar del espacio de color RGB, ya que no puede trabajar bien con la iluminación y plantean variaciones, realizando una clasificación con SVM. Los resultados experimentales se compararon mediante la precisión y el método propuesto, y muestran una mejora significativa en términos de estos factores.

- Weimin X., *et al.* (2008) proponen un algoritmo basado en el filtro de Gabor y de SVM para el reconocimiento de la expresión facial. Las características de las expresiones faciales pasaron por un filtro de Gabor, luego se utilizaron las características para entrenar el clasificador SVM. Por último, la expresión facial se clasificó por las máquinas SVM. Utilizando el algoritmo, una alta tasa de reconocimiento puede ser alcanzado. Los resultados experimentales mostraron que el algoritmo es factible y aplicable.

- Kharat G. U., Dudul S. V., *et al.* (2008) desarrollaron un "Humanoid Robots" que puede llevar a cabo una conversación intelectual con los seres humanos. Se llevó a cabo un reconocimiento de las emociones humanas por un ordenador utilizando una red neural. En este trabajo las seis emociones básicas universalmente reconocidos a saber son enojado, asco, miedo, alegría, tristeza y sorpresa, junto con un neutro son reconocidos. Se basa en técnicas de extracción de características tales como Transformada Discreta del Coseno (DCT), transformada rápida de Fourier (FFT), descomposición en valores singulares (SVD) se utilizan para extraer las características útiles para el reconocimiento de las emociones de las expresiones faciales y por último se utiliza SVM para el reconocimiento de la emoción usando los rasgos faciales extraídos y se compara el rendimiento de varias técnicas de extracción de características. Lograron alcanzar 100% de precisión en el reconocimiento conjunto de datos de entrenamiento y 94,29% en el conjunto de datos de validación cruzada.

- Conde C., Ruiz A., Cabello E., *et al.* (2004) analizan la influencia de las condiciones de iluminación en los rostros, utilizando máquinas de vectores soporte (SVM) y vecinos k-más cercano, se analizó a base de un experimento en el que las imágenes se adquieren en condiciones de iluminación controladas o no controladas y se consideró el análisis de componentes principales (PCA) para llevar a cabo la reducción dimensional. Sin embargo, las técnicas SVM ofrecen mejores resultados, incluso si se consideran núcleos lineales.

- Chittora A., Mishra O., *et al.* (2003) proponen las SVM como una estrategia de reconocimiento de emociones, y utilizan un árbol binario para abordar el problema de reconocimiento de rostros. Se compara el rendimiento de núcleo basado SVM en otras funciones del kernel para el reconocimiento de rostros. Sin embargo, al comparar los resultados de las diferentes técnicas disponibles, con las de un clasificador con SVM, no se equilibran los porcentajes de precisión.

- Michel P., Kaliouby R. E. *et al.* (2003) proponen un enfoque en tiempo real para el reconocimiento de emociones a través de la expresión facial capturada en un vídeo. Utilizaron un programa de seguimiento automático de rasgos faciales para extracción de características. Los desplazamientos de características faciales en el flujo de vídeo se utilizan como entrada a una máquina clasificadora de vectores de soporte. A su vez evaluaron un método propuesto por ellos en términos de precisión de reconocimiento para una variedad de escenarios de interacción y clasificación. Los resultados de sus experimentos con personas dependientes e independientes demuestran la eficacia de una máquina de soporte vectorial. Sin embargo, una de las mejoras que ellos citan es optimizar el vector de alimentación al clasificador SVM.

- Churchillcollege P. M., *et al.* (2003) aplican una Máquina de Vector Soporte para la clasificación de las emociones a través de la expresión facial mediante el diseño, implementación y evaluación de una solicitud para reconocer las emociones expresadas en imágenes de la cara. La aplicación realiza la extracción de los datos en un conjunto de entrenamiento definido por el usuario de imágenes de rostros

agrupados por la emoción expresada, utiliza los datos para entrenar un clasificador SVM y así permitir la posterior clasificación de las imágenes. Sin embargo, el problema es en base de datos no controladas, por el tamaño y tipo de las imágenes.

Se mostraron las técnicas para el reconocimiento de emociones más utilizadas basándose en las publicaciones más sobresalientes en esta área, los resultados obtenidos y las bases de datos utilizadas.

Las técnicas varían de la cantidad de pasos realizados para alcanzar el objetivo, el cómputo para lograrlo y las características necesarias para realizarlo. Estos factores definen el tipo de problemática al que este enfocado, es decir, las características del problema definirán los algoritmos utilizados.

Cada técnica tiene sus ventajas y desventajas. Ya que algunos presentan buenos resultados con el coste de un procesamiento mayor, o aquellos que realizan el análisis con un análisis con menos procesamiento, pero es necesario conocer el rostro neutro de la persona.

Es importante mencionar que las técnicas mencionadas no son utilizadas solo para el reconocimiento de emociones, sino también del reconocimiento facial, ya que todo comienza ahí.

Capítulo 3

3. METODOLOGÍA

La metodología se basa en la propuesta de una arquitectura para el reconocimiento de emociones. Ésta se divide en seis etapas como se muestra en la figura 1.

- La primera etapa se realiza la elección de la base de datos de conocimiento, en donde se tengan diversas muestras de caras o rostros a realizar la detección y clasificación de emociones, se pretende que las imágenes capturen las seis emociones de estudio de una persona, es decir tener seis imágenes por persona.
- La segunda etapa es la detección de caras dentro de imágenes. A partir de la detección con el algoritmo Viola-Jones es cuando se puede crear cualquier sistema que analice la información contenida en las caras: ojos, nariz, boca, cejas, y pómulos. La detección facial se encarga de determinar si hay o no alguna cara en una imagen dada y, en caso de que exista, de extraer la localización y el contenido de dicha cara y sus elementos contenidos.
- En la tercera etapa se evalúa factores como la iluminación, postura, distancia, simetría y laterabilidad. Detectando la dimensionalidad del rostro de la persona, con la finalidad de establecer a través de una transformada Wavelet un plano de la imagen, dimensionando las características del rostro cambiando texturas y resolución para la obtención completa de elementos y la determinación de una cara y una no cara.
- En la cuarta etapa se definen las singularidades dentro de un espacio dimensionado, dichos elementos a identificar serían extraídos por el algoritmo PCA, obteniendo características de acuerdo a sus similitudes o diferencias. Se establecen características y no diferencias, lo cual conlleva una reducción de la dimensionalidad de la imagen. Por último, se realiza la comparación de plantillas de las emociones con las de un vector basado en características específicas de acuerdo a la descomposición de la imagen.
- En la quinta etapa se incorporó a la arquitectura de reconocimiento de emociones un sistema basado en el algoritmo básico de Eigenfaces. Dicho algoritmo será incorporado en la fase de entrenamiento dentro de la arquitectura con el fin de reducir y obtener los componentes que serán los necesarios para la reconstrucción de la imagen y la clasificación de una emoción.

- En la sexta etapa se realiza la comparación de plantillas de emociones de acuerdo a un vector creado con características específicas conforme a la descomposición de la imagen. Por último, se crean clases de comparación con características específicas.

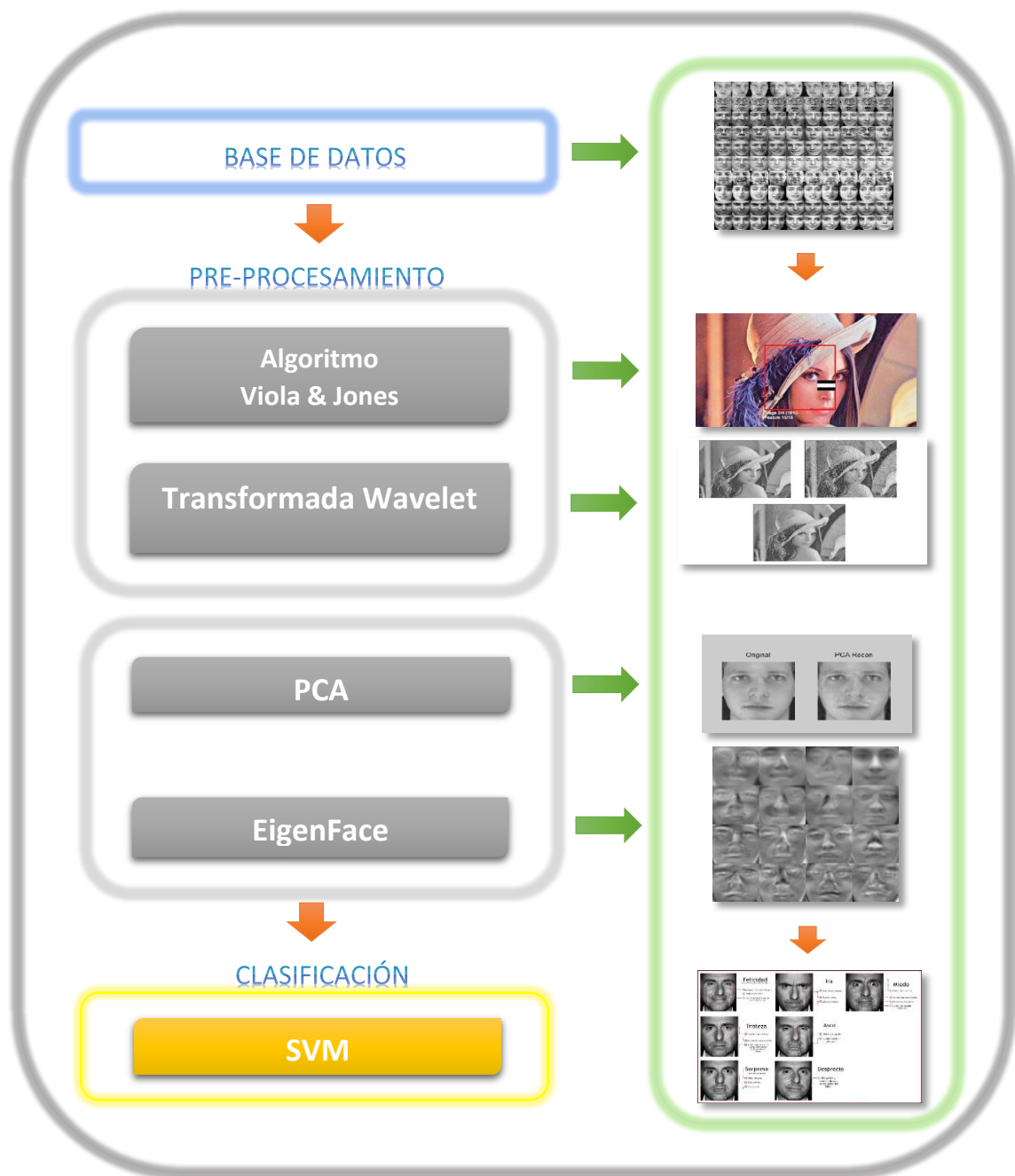


Figura 1. Diagrama de bloques

2.1. Selección de Variables

La base de datos de estudio pertenece a la Universidad Técnica de Munich (Alemania), y está conformada por imágenes de rostros de 18 personas que muestran las 6 emociones básicas definidas por Eckman y Friesen (alegría, sorpresa, enojo, tristeza, desagrado y miedo). Esta base de datos fue generada como parte del proyecto FG-NET (*Face and Gesture Recognition Research Network*), cada imagen es de 320x240 píxeles a 8 bits y está en formato JPEG.

Esta base de datos sobrelleva el paradigma que presentan algunas bases de datos en donde muestran emociones distintas a las naturales, debido a que en esta base de datos se pide a las personas reaccionar lo más natural o comportarse lo más espontáneo posible mientras se les estimula con videos o imágenes.

2.2. Extracción de características faciales con Viola-Jones

La segunda fase del sistema de detección de emociones consiste en aplicar el algoritmo de Viola-Jones para desarrollar un detector facial/rasgos en tiempo real, el cual proporciona un alto porcentaje de acierto la imagen y posición de uno de los ojos, boca, nariz y pómulo, de los cuales se extraerán las posiciones de los rasgos en ellos.

Primero se establece el obtener la cara de la persona, y para ello se utiliza el algoritmo de Viola-Jones, ver Figura 2, una explicación básica de lo que hace este algoritmo es ésta:

- Se transforma la imagen a escala de grises
- Recorre la imagen a procesar mediante ventanas de 24x24 píxeles a diferentes escalas.
- Para cada una de estas imágenes obtiene una serie de características, que son los resultados de la diferencia de los valores de sus píxeles entre áreas.

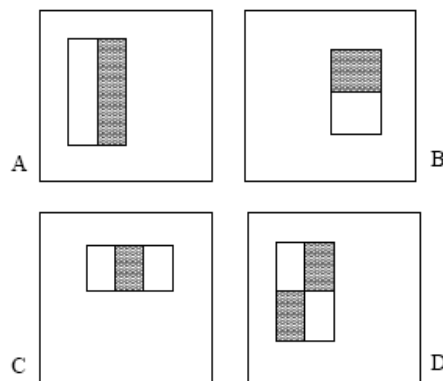


Figura 2. Características utilizadas en el algoritmo Viola-Jones para la detección de caras

Una vez obtenida la cara del sujeto, se procesa la imagen de la cara para obtener la posición de los ojos, nariz, boca y pómulos. Para ello, se utiliza nuevamente el algoritmo de Viola-Jones anteriormente mencionado pero entrenado en este caso para detectar los rasgos humanos, ver Figura 3.

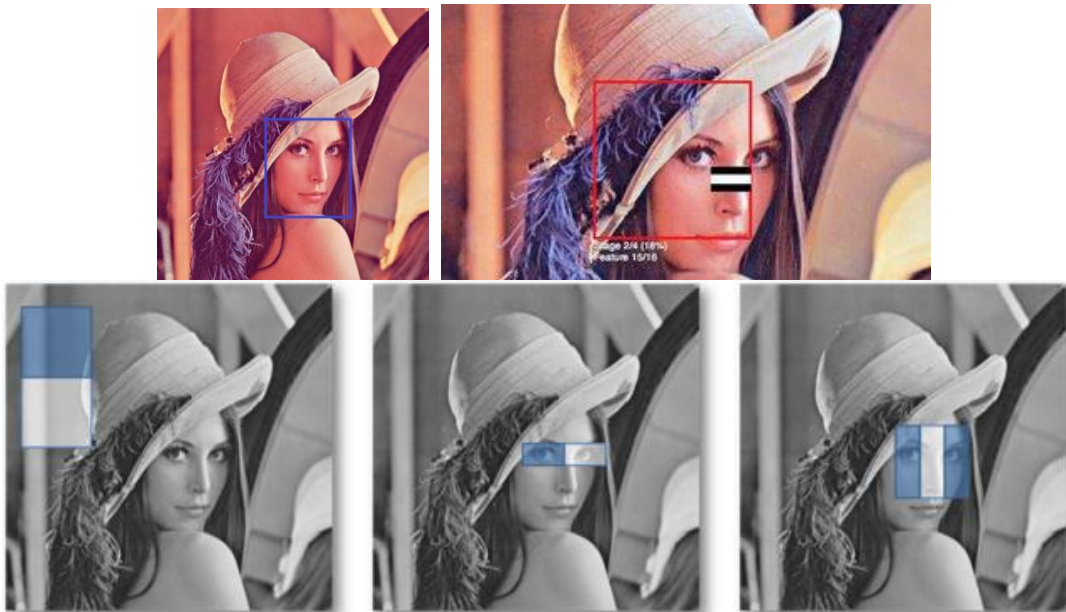


Figura 3. Cara detectada mediante algoritmo Viola&Jones

3.3. Reducción de dimensionalidad de la imagen con Transformada de Wavelet

Una vez localizados los elementos característicos de la imagen, la tercera fase comprende una técnica usando Transformada Wavelet, la cual ayuda a localizar pequeñas variaciones en imágenes que el ojo humano no detecta a simple vista. Por ello se segmenta la imagen en detalles a diferentes niveles de resolución y orientaciones, sin repeticiones, comenzando a realizar la similitud cuadro por cuadro, localizando las que se asemejen.

Pero esto se obtiene factores determinados por la misma transformada de Wavelet de paquetes, dichos paquetes se basan en una resolución múltiple ortogonal compuesta por Wavelet lineales, que proporcionan la información espacial, orientación y posición de la imagen contemplando la textura de la misma a diferentes escalas y resolución. Dichos factores son resultados del contexto que se le dará a la imagen de manera física para la experimentación, dicho contexto será determinado por el diseñador, implementando variables como iluminación, postura, distancia, simetría y laterabilidad.

Con lo anterior, el concepto de ruido como ese efecto “indeseable” es el elemento que se debe eliminar de nuestras imágenes, éste consiste en la aparición aleatoria de señales ajenas a la imagen original, especialmente apreciable en las zonas de sombra de la imagen, donde se puede hacer visible la separación del contexto que rodea, el elemento caro, ver Figura 4.



Figura 4. Discriminación de datos, como entorno del elemento cara, por medio del cambio de texturas de la imagen.

Tomando lo anterior, se puede dar respuesta al por qué se utilizan los Wavelets de Morlet para la detección de rostros en este trabajo, y es que está en el hecho de que son buenos extractores de características de las imágenes, forman una especie de firma del rostro, ya que se basan en la respuesta de cómo las células de la corteza visual tienen campos receptivos, las cuales son pequeñas regiones altamente estructuradas; investigaciones por parte de los científicos Hubel y Wiesel describieron a esas células como detectores de bordes, algunas investigaciones más recientes como la hecha por Jhones y Palmer mostraron que el comportamiento de las respuestas de estas células en los gatos, correspondían a medidas locales en la frecuencia. También se notó que la respuesta dependía de la frecuencia y orientación del campo visual. En los experimentos hechos por Jhones y Palmers, la respuesta de estas células primarias fueron medidas con electrodos, mientras que los campos receptivos de estas células primarias fueron medidos por proyectar estímulos parecidos a un punto en una ventana homogénea.

En otro trabajo, dos científicos Pollen y Ronner (1974) examinan la relación de fase entre células adyacentes de la corteza visual de los gatos, ellos concluyen que las células de un par de células adyacentes tienen ciertas simetrías definidas por frecuencias similares, magnitudes similares, y especificaciones similares de las direcciones: Una de ambas tiene simetría par y otra simetría impar. Esto permite modelar ambos campos receptivos de los pares de células por funciones de valor complejo.

Sin embargo, el algoritmo es muy limitado, pues sólo trabaja con imágenes de tamaño pequeñas, y busca rostros con una medida estándar, por lo tanto, tendrá mal desempeño en imágenes grandes, también tendrá mal desempeño si se buscan

localizar imágenes con rostros muy grandes, o imágenes con rostros muy pequeños. Se puede seguir investigando para proponer un algoritmo más robusto que busque las áreas candidatas a ser rostro, pero en imágenes a diferentes niveles de resolución para poder reconocer rostros pequeños y rostros muy grandes, ver Figura 5.



Figura 5. Compresión de imágenes con Transformada Wavelet

3.4. Obtención de Componentes Principales con PCA

El éxito de esta arquitectura para el reconocimiento de emociones depende fuertemente de los elementos utilizados para representar las imágenes para su posterior clasificación. El patrón que represente a una imagen debe estar compuesto por los elementos más sobresalientes de ella, lo cual es obtenido en esta cuarta etapa, permitiendo reducir la cantidad de datos usados en el proceso de clasificación, y aumentar la diferencia entre ellas para que actúe como un poderoso discriminante o clasificador. Una de las técnicas más utilizadas para seleccionar un subconjunto de elementos que cumpla con esas condiciones es el Análisis de Componente Principal (PCA), la cual genera un conjunto de vectores ortonormales que maximizan la dispersión entre todas las muestras proyectadas, reduciendo al mismo tiempo su dimensión, ver Figura 6.

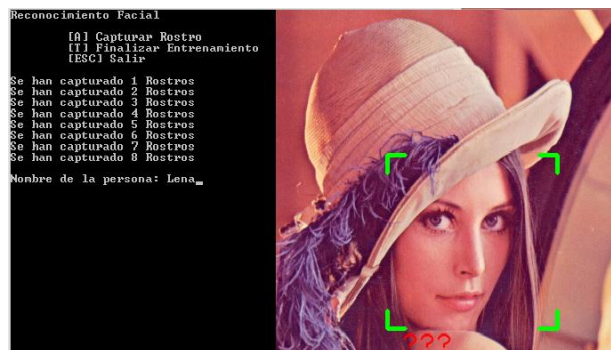


Figura 6. Obtención de Componentes Principales

3.5. Reducción de componentes principales con EigenFace

Tanto la etapa de entrenamiento como la de reconocimiento de la emoción utilizan una base de rostros compuesta por tres conjuntos de imágenes (20 y 40 personas por grupos de prueba) de dimensión $N \times N$.

Esta quinta fase consiste del proceso de entrenamiento, la cual conlleva los siguientes pasos:

1. Cada imagen es reorganizada como un vector de dimensión N^2 cuyo valor es construido como la concatenación de cada una de las filas de la imagen, formando así una matriz de $N^2 \times M$.

2. Se obtiene el rostro promedio

3. El rostro promedio obtenido es restado a cada una de las imágenes M obteniendo un nuevo conjunto de vectores que conforman la matriz de $N^2 \times M$.

4. En este punto se buscan los autovectores de la matriz de covarianza de dimensión $N^2 \times N^2$. Estos vectores propios son los vectores ortonormales usados para construir la representación de las imágenes. El tamaño de la matriz hace intratable este paso por el espacio y el tiempo requerido). Para resolver este problema se buscan los autovectores de la matriz de covarianza. Debido a su gran dimensión, éstos deben ser aproximados a través de los vectores propios de la matriz de covarianza reducida.

En este trabajo se propone un nuevo método que consiste en formar una imagen de menor tamaño que permita obtenerlos directamente, ver Figura 7.

- a. Cada imagen es dividida en bloques de $P \times P$. Cada uno de estos bloques se promedia y se obtiene una nueva imagen de $D \times D$, que se obtiene de reemplazar cada bloque por su promedio.
- b. Cada imagen es reorganizada como un vector de dimensión 2D cuyo valor es construido como la concatenación de cada una de las filas de la imagen, formando así una matriz de $D^2 \times M$.
- c. Se obtiene el rostro promedio.
- d. El rostro promedio obtenido es restado a cada una de las imágenes obteniendo un nuevo conjunto de vectores que conforman la matriz.
- e. Se obtiene la matriz de covarianza de dimensiones $D^2 \times D^2$.
- f. Se obtienen los autovectores, los que ordenados de mayor a menor según sus correspondientes autovalores, conforman la matriz.

g. Se obtiene un patrón

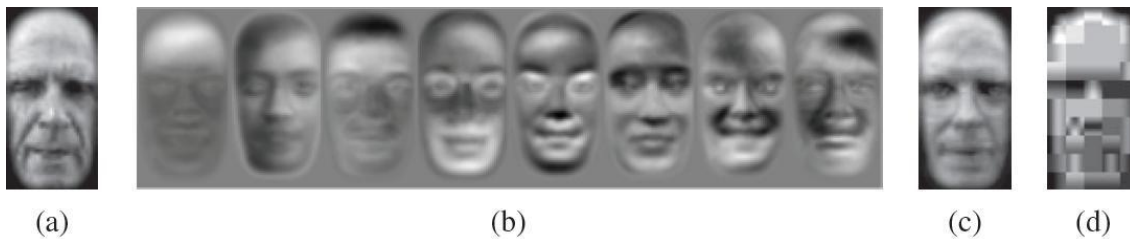


Figura 7. Ejemplo de un eigenface donde se muestran las diferentes clases (correspondientes a las emociones de cada uno de los individuos) y la imagen de entrada a identificar (punto en negro de la gráfica).

3.6. Clasificación de emociones con SVM

La última fase se basa en las dos ideas fundamentales para la construcción de un clasificador SVM, las cuales son la transformación del espacio de entrada en un espacio de alta dimensión y la localización en dicho espacio de un hiperplano separador óptimo. La transformación inicial se realiza mediante la elección de una función kernel adecuada. La ventaja de trabajar en un espacio de alta dimensión radica en que las clases consideradas serán linealmente separables con alta probabilidad, por tanto, encontrar un hiperplano separador óptimo será poco costoso desde el punto de vista computacional. Además, dicho hiperplano vendrá determinado por unas pocas observaciones, denominadas, vectores soporte por ser las únicas de las que depende la forma del hiperplano.

Una de las principales dificultades en la aplicación de este método radica en la elección adecuada de la función kernel. Es decir, construir la función de transformación del espacio original a un espacio de alta dimensión es un punto crucial para el buen funcionamiento del clasificador.

La forma final de la regla de clasificación para un clasificador binario (dos clases, +1 y -1) son parámetros aprendidos por el clasificador durante el proceso de entrenamiento, por lo tanto, nuestro patrón obtenido con el algoritmo eigenface, será el valor de la función kernel. Si la función es mayor que un umbral entonces la emoción estimada para un punto x será una +1 y será -1 en caso contrario.

En el problema del reconocimiento de emociones es típico trabajar con más de dos emociones. Suponiendo que el número de emociones consideradas es n . Es necesario llevar a cabo una generalización del clasificador binario al caso multiclase. En este caso, se dispone de n clasificadores, es decir, n valores de la regla de clasificación para cada objeto. En ambos casos para determinar la emoción que corresponde a cada objeto se realiza una ponderación sobre todas las reglas de clasificación disponibles.

Las SVMs han demostrado ser un método muy efectivo en la clasificación de expresiones faciales espontáneas como describe Bartlett, *et al.* (2001).

Capítulo 4


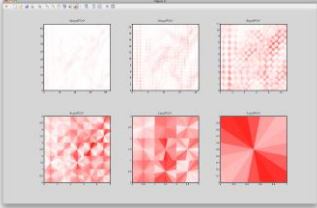
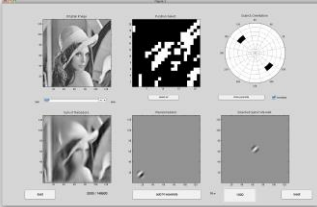
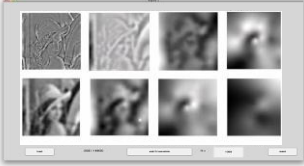
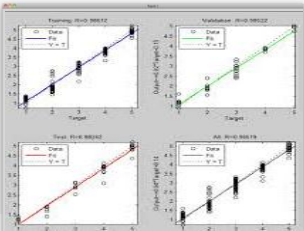
4. RESULTADOS

La siguiente tabla presenta los resultados del reconocimiento de 6 emociones prototipo.

<i>EMOCION</i>	<i>DATOS DE PRUEBA</i>	<i>IMAGEN</i>	<i>POSITIVOS</i>	<i>FALSOS POSITIVOS</i>	<i>EFICIENCIA</i>
Alegría	24		19	2	95%
	48		39	5	97.5%
Sorpresa	24		19	3	95%
	48		39	8	97.5%
Miedo	24		18	3	90%
	48		37	7	92.5%
Enojo	24		19	1	95%
	48		39	3	97.5%
Disgusto	24		19	2	95%
	48		39	5	97.5%
Tristeza	24		19	1	95%
	48		39	3	97.5%
EFICIENCIA EN MUESTRA DE 20					94.16%
EFICIENCIA EN MUESTRA DE 40					96.66%

En la tabla se consignan los resultados de la arquitectura propuesta. Es de observarse que la expresión de miedo es la más difícil de reconocer con un 92.5% debido a que en algunas ocasiones ésta se manifiesta como una expresión ambigua. Esta observación concuerda con lo encontrado en el estudio realizado por Ehrlich en donde se expone que la expresión de miedo puede ser una excepción de las clases básicas de emociones derivadas, ya que esta emoción está restringida al proceso de selección forzada mediante el cual se seleccionan las categorías.

Resultados por etapas parciales:

<i>ETAPA</i>	<i>DATOS DE PRUEBA</i>	<i>IMAGEN</i>	<i>POSITIVOS</i>	<i>FALSOS POSITIVOS</i>
VIOLA & JONES	24		19	2
	48		39	5
WAVELET	24		19	3
	48		39	8
PCA	24		18	3
	48		37	7
EIGENFACE	24		19	1
	48		39	3
SVM	24		19	2
	48		39	5

Capítulo 5

5.1 CONCLUSIONES

La creación de una arquitectura segmentada por diferentes algoritmos como son Viola-Jones, Transformada de Wavelet, Algoritmo de Componentes Principales, EigenFace y Máquina de Vector Soporte para el manejo de una imagen, ayuda a la obtención de componentes específicos, descriptivos y característicos que permiten la construcción de rasgos faciales, que alimentaran al clasificador obteniendo con ello el reconocimiento de emociones.

En particular, el estudio de patrones de pixeles es aplicado a las regiones detectadas como ojos por el algoritmo de Viola y Jones. Posteriormente se aplica el análisis de componentes principales o PCA (del inglés *Principal Component Analysis*) y se seleccionan las características críticas, las cuales son utilizadas como parámetros de entrada para un clasificador SVM. De esta manera, se logra reducir el número de falsos aciertos/positivos mejorando, por tanto, la tasa de acierto global de sistema.

Gracias a dicha arquitectura, el algoritmo de Viola-Jones, además de ofrecer buenos resultados, ofrece también una tasa de análisis muy elevada, de más de 90% de eficiencia, haciendo del mismo una herramienta útil para sistemas de obtención de características faciales en tiempo real.

En resumen, queda claro que se ha obtenido una importante ventaja en la codificación de la imagen mediante la representación que proporciona la transformada wavelet. No obstante, se ha comprobado que las componentes de alta frecuencia de la transformada wavelet pueden codificarse de forma mucho más burda que los componentes de baja frecuencia sin que aparezcan pérdidas aparentes en la calidad de la imagen. Por ello a diferencia de otros autores, la aplicación de una Transformada Wavelet como parte del pre-procesamiento de la imagen es de ayuda para la minimización de la imagen. Quizás, uno de los grandes problemas de los sistemas de detección de emociones reside precisamente en la capacidad para caracterizar aquello que no corresponde a una emoción. Por ello, se propone usar una máquina de vectores soporte o SVM (del inglés *Support Vector Machine*) de ochos clases para así evitar dicho problema. Adicionalmente, utilizar las auto-caras para caracterizar las imágenes, este modelo obtiene una tasa de acierto superior al 90%.

5.2 TRABAJO A FUTURO

Como trabajo a futuro se pretende enfocarse en la etapa de reducción de dimensionalidad, puesto que, en el tema de reconocimiento de emociones, éste es el punto medular que afecta en general las arquitecturas.

Dentro de la arquitectura expuesta, se puede mejorar la dimensionalidad de la matriz de confusión generada a partir de los componentes principales obtenidos, de dichos componentes. Actualmente, se realiza una reducción directa de dicha matriz ocasionando la pérdida de información ocasional, sin embargo, si la reducción de la matriz se realiza normalizada, esa pérdida no existiría.

Por lo anterior se pretende obtener un algoritmo EigenFace normalizado y no directo, puesto que se optimizarían los porcentajes de resultados de acuerdo a cada emoción.

REFERENCIAS BIBLIOGRAFICAS

Zhao, W.; Chellappa, R.; Rosenfeld, A. y Phillips, P.J. (2003). *Reconocimiento facial: A Literatura Encuesta, ACM Computing Surveys*, Vol. 35, No. 4, diciembre 2003, pp 399-458.

Zhou, S.; Krueger, V. & Chellappa, R. (2003). *Reconocimiento probabilístico de rostros humanos de Video, Visión por Computador y la Interpretación de Imágenes*, Vol. 91, 2003, pp 214-245.

P. Viola and M. Jones, "Robust real-time face detection". In *Proc. Of IEEE Workshop on Statistical and Computational Theories of Vision*, 2001.

Coifman, R. R. & Meyer, Y. (1990). *Bases ortonormales paquetes wavelet, preparativos para la imprenta*.

Liu C. C., P. y D. Dai Yan H. (2007). *Paquetes wavelet discriminante Local coordina para la cara reconocimiento, Journal of Machine Learning Investigación*, Vol. 8 (mayo de 2007) 1165-1195.

Sirovich, L. & Kirby, M. (1987). *Procedimiento de pocas dimensiones para la caracterización de Rostros Humanos, Revista de la Sociedad Americana de Óptica A*, Vol. 4 (3), (marzo de 1987), 519 - 524, 1.084-7.529.

Lu, B.L., Ito, *la descomposición de tareas M. y combinación de módulos en base a las relaciones de clase: una red neural modular para la clasificación de patrones*. IEEE Transactions on Neural Networks, vol.10, (1999) 1244 -1256.

Lu, B.L., Wang, K.A., Utiyama, M., Isahara, H. *Un método parcial contra-parte de forma masiva formación paralela de máquinas de vectores de soporte*. In: Actas de IJCNN '04, Budapest, July 25-29 (2004).

Lu, B.L., Ma, Q., Ichikawa, M., Isahara, H. (2003). *E ± ciente Tagging parcial de voz con un Min-Max Modular Modelo Neural Network, Inteligencia Aplicada*, vol.19, pp.65-81.

Yang, Y., Lu, B. L. (2006). *Predicción de proteínas subcelulares Multi-Lugares con Min-Max Modular Vector Apoyo a máquina*, en Memorias del Tercer Simposio Internacional en redes neuronales (ISNN 2006).

Guevara Díaz J. (2013). *Detección de rostros por medio de las wavelets de morlet*.

Guevara M. L., Echeverry J. D. & Ardila Urueña W. (2008). *Detección de rostros en imágenes digitales usando clasificadores en cascada*.

Serrano A, Conde C., De Diego I. M., Cabello E., Bai L. & Shen L. (2007). *Parallel gabor pca with fusion of svm scores for face verification*.

Shen L. & Bai L. (2004). *En una revisión de wavelets gabor para el reconocimiento facial*.

Sánchez Sánchez A., Cabello Pardos E. & Pérez L. P. (1998). *Reconocimiento de caras humanas mediante una red neuronal con ada95*.

Arandjelovi O. & Cipolla R. *Lograr invariance iluminación por imagen*.

Campadelli P., Lanzarotti P. & Lipori G. *Automático de extracción de características faciales de reconocimiento de caras.*

Dao Dai Q. & Yan H. *Wavelets y reconocimiento facial.*

Bogdan Dragolici A. (2010). *Detección, localización e identificación biométrica de caras en imágenes: métodos y evaluación en el marco nist-mbgg.*

Du, S. & Martínez A. M. (2012). *Un modelo de la percepción de las expresiones faciales de emoción por los seres humanos: resumen investigación y perspectivas.*

Du S. & Martínez A. M. (2011). *La resolución de las expresiones faciales de las emociones.*

Cereceda Danús S., Pizarro I., Valdivia V., Hurtado F. E. & Ibáñez A. (2010). *Reconocimiento de emociones: estudio neurocognitivo.*

Khashman A. (2008). *Intelligent face recognition global.*

Pantic M. & Stewart Bartlett M. *Análisis automático de las expresiones faciales.*

Blázquez Pérez L. (2013). *Reconocimiento facial basado en puntos característicos de la cara en entornos no controlados.*

Cruz Pérez C. (2008). *Reconocimiento de rostros basado en características invariantes.*

Aguerreberre C. & Capdehourat G. (2006). *Reconocimiento de caras con características locales.*

Mazanec J. & Melisek M. (2008). *Support vector machines, pca and lda in face recognition*

Eleyan A. & Demirel H. *Pca y lda basen redes neuronales de reconocimiento de la cara humana.*

Rodríguez Gozábez S. (2010). *Análisis del preprocesado de imágenes en el reconocimiento de caras basado en pca.*

Serrano A., Conde C., De Diego I. M., Cabello E., Shen L. & Bai L. (2007). *Fusion of gabor feature based classifiers for face verification.*

Conde C., Ruiz A. & Cabello E. (2003). *Pca vs low resolution images in face verification.*

Serrano A., De Diego I. M., Conde C., Cabello E., Shen L. & Bai L. (2007). *Influence of wavelet frequency and orientation in an svm-based parallel gabor pca face verification system.*

Revista electrónica teoría de la educación. *Educación y cultura en la sociedad de la información, (2006). Técnicas de reconocimiento automático de emociones.*

Graña M. & Beristain A. (2007). *Emotion from facial expression recognition.*

Albornoz E. M., Milone D. H. & Rufiner H. L. (2013). *Robust emotion recognition using bio-inspired features.*

Nguyen T., Li M., Bass I. & Sethi I. K. (2005). *Investigation of combining svm and decision tree for emotion classification.*

Kobayashi V. (2013). *A hybrid distance-based method and support vector machines for emotional speech detection.*

Aswathy. R. (2013). *Facial expression recognition using svm classification in perceptual color space.*

Weimin X. (2008). *Facial expression recognition based on gabor filter and svm.*

Kharat G. U. & Dudul S. V. (2008). *Human emotion recognition system using optimally designed svm with different facial feature extraction techniques.*

Conde C., Ruiz A. & Cabello E. (2004). *Face verification using svm: influence of illumination.*

Chittora A. & Mishra O. (2003). *Face recognition using rbf kernel based support vector machine.*

Michel P. & Kaliouby R. E. (2003). *Real time facial expression recognition in video using support vector machines.*

Churchillcollege P. M. (2003). *Support vector machines in automated emotion classification.*