



INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA

MAESTRÍA EN SISTEMAS COMPUTACIONALES

**“Arquitectura para identificación de emociones
a partir de secuencias de vídeo”.**

**TESIS
PARA OBTENER EL GRADO DE
MAESTRO EN SISTEMAS
COMPUTACIONALES**

**PRESENTA
GUSTAVO MARTINEZ ALCAZAR**

**ASESORES:
DR. LUIS ALBERTO MORALES ROSALES
DR. IGNACIO ALGREDO CASTILLO**

MISANTLA, VERACRUZ

Agosto, 2015

Agradecimientos

A mi familia por apoyarme en todo el proceso de la tesis, darme la fuerza y aliento necesario para concluir mis estudios.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) de México, que me apoyo económicamente para poder concluir mis estudios de maestría.

Al Instituto Tecnológico Superior de Misantla, por permitirme realizar mis estudios de maestría.

A mis maestros, los cuales me brindaron sus conocimientos, su tiempo y su experiencia.

A mis revisores de tesis Dr. Ignacio Algreto Castillo y al MIA Roberto Ángel Meléndez Armenta por el tiempo, los consejos y el apoyo para presentar un buen trabajo de tesis.

Al Dr. Luis Alberto Morales Rosales por ser un excelente mentor y asesor de tesis, que me oriento en cada una de sus clases y revisiones de tesis.

Resumen

Este trabajo presenta una arquitectura para el reconocimiento de emociones de una persona en una secuencia de vídeo. Se establece la identificación de las seis emociones básicas especificadas por Ekman, a partir de secuencias de vídeos con rostros frontales y sin oclusión, para personas caucásicas de entre 18 y 40 años de edad.

La arquitectura se basa en la identificación y extracción de características, reducción de los valores obtenidos y finalmente la clasificación de la emoción.

La identificación del rostro y las secciones de interés se realiza utilizando el algoritmo de Viola Jones, permitiendo reducir a más de la mitad el área de trabajo. Utilizando solo las secciones del contorno de los ojos, cejas y boca.

Posteriormente los valores de las regiones de interés se representan como valores matriciales por medio de las transformadas wavelets.

Prosiguiendo con la reducción de los valores matriciales a través del Análisis discriminante (LDA), el cual representa el nuevo rostro como una función discriminante.

Por último, la clasificación de la emoción por medio de la máquina de vectores de soporte (SVM) que clasifica la función discriminante en una de las 6 posibles emociones.

El propósito de la arquitectura es proporcionar robustez al clasificar las emociones, ya que el análisis discriminante presenta desventajas que pueden ser cubiertas con el uso de las máquinas de vectores de soporte (SVM). Los SVM permiten mejorar los resultados al manejar de mejor manera un número grande de variables y al volver el problema linealmente separable para una mejor clasificación.

El trabajo utiliza la base de datos cohn-kanade para crear la base del conocimiento. La principal ventaja es presentar esta base de datos es que las imágenes presentadas son una secuencia de la emoción y no solo una imagen representativa.

Para ello utilizo 160 imágenes, de las cuales 98 son mujeres y 62 hombres. 120 imágenes fueron utilizadas en el entrenamiento y 40 en las pruebas.

Con lo que se obtuvo resultados diferentes para cada emoción: para felicidad 75%, tristeza 50%, sorpresa 60%, disgusto 50%, enojo 83% y miedo 66%

Índice

CAPITULO 1 Introducción.....	8
1.1. Problemática.....	9
1.2 Objetivo.....	11
1.3 Justificación.....	12
1.4 Hipótesis.....	13
1.5 Alcances y limitaciones.....	13
1.6 Metodología propuesta.....	14
1.7 Organización de la tesis.....	16
CAPITULO 2 Estado del arte.....	17
2.1 Introducción.....	17
2.2 Clasificación de trabajos realizados en el reconocimiento de emociones.....	18
2.2.1 Sistemas de codificación de acciones faciales (FACs).....	18
2.2.2 Parámetros de animación facial (FAPs).....	22
2.2.3 Otras técnicas de reconocimiento de emociones.....	24
2.2 Sistemas de aprendizaje automático supervisado en vídeo.....	27
2.3 Conclusiones del estado del arte.....	28
CAPITULO3 Reconocimiento de emociones en vídeo.....	30
3.1 Trabajo relacionado.....	31
3.2 Descripción de las bases de datos usados en el trabajo.....	31
3.3 Definiciones de la arquitectura.....	32
3.3.1 Identificación del rostro de áreas de interés.....	33
3.3.2 Representación de valores de características en forma matricial.....	35
3.3.3 Obtención de función discriminante.....	36
3.3.4 Clasificación de la emoción.....	38
3.4 Conclusiones del capítulo.....	38
CAPITULO 4 Aplicación del algoritmo.....	39
4.1 Características de la arquitectura.....	39
4.2 Diseño de la arquitectura.....	40
4.2.1 Identificación de rostros y áreas de interés.....	40
4.2.2 Representación de valores de características en forma matricial.....	41
4.2.3 Obtención de la función discriminante.....	42

4.2.4 Clasificación de la emoción.....	42
4.3 Resultados.....	42
4.4 Conclusiones del capítulo.....	43
CAPITULO 5 Resultados y conclusiones.....	44
5.1 Conclusiones.....	44
5.3 Trabajo futuro.....	45
CAPITULO 6 Referencias.....	47
6. Referencias.....	48

Índice de imágenes

Figura 2.1. Ejemplo de acciones faciales de cejas, ojos y contorno de los ojos.	19
Figura 2.2. Análisis de rasgos faciales.	20
Fig. 2.3 Ilustración de un modelo Candide.	26
Figura 2.4. Rostro dividido en una subregión de 6 x 7.	27
Figura 3.1 Imágenes de muestra de base de datos Cohn-Kanade	32
Figura 3.2. Creación de la base del conocimiento.	32
Figura 3.3 Ejemplo Haar like features	33
Figura 3.4 Haar like features en imagen de rostro.	34
Figura 3.5 Ecuación del haar like features.	34
Figura 3.6 Figura Cálculo de la imagen integral.	35
Figura 3.7. Obtención de matriz representativa.	36
Figura 3.8. Obtención de medias y formación de nueva matriz	37
Figura 3.9. Obtención de función discriminante	37
Figura 3.10. Matriz final.	38
Figura 4.1 Reducción de la imagen.	41

Índice de tablas

Tabla 5.1 Tabla de resultados.	42
Tabla 5.2 Tabla de resultados.	43
Tabla 5.3 Aplicaciones del reconocimiento de emociones	45

CAPITULO

1 Introducción

La identificación de emociones es un área de investigación con gran utilidad para el desarrollo de nuevas tecnologías, como lo son los Sistemas de Tutor Inteligentes, los videojuegos, los monitores de personas enfermas a través de vídeo. Donde el conocer la emoción presentada por una persona proporciona mejor conocimiento del usuario, y mejora la interacción con este.

Gran parte del uso de los sistemas de reconocimiento de emociones por parte de los investigadores se debe a lo accesible que son los instrumentos para la obtención de datos, siendo los micrófonos y cámaras los más utilizados. Sin embargo, no son los únicos medios que existen, ya que también se encuentran aquellos que obtienen datos a través de monitores de pulsaciones cardiovasculares o voz [8].

La identificación de emociones a través de imágenes y vídeo es de las más utilizadas, hoy en día, hay dos importantes vertientes para su identificación:

- FACs (Facial Action Coding Systems). Enfoque basado en el rostro, que involucra la identificación de músculos faciales individualmente o en grupo.

[11]

- FAPs (Facial Animation parameters). Comprende de un conjunto de parámetros para representar las variaciones del rostro en secuencias de vídeos animadas, a partir de un rostro en estado neutro al mismo con una expresión, con el fin de identificar las variaciones [40].

El proceso en estas dos técnicas se diferencia en que los FACs no necesitan un conocimiento previo del rostro de la persona a identificar, ya que toman las características discriminantes de la persona, para poder clasificar la emoción. Por su parte, las FAPs necesitan de un conocimiento previo del individuo, conocer como es su rostro en estado neutro y así identificar las variaciones en el rostro.

Además de estas técnicas, existen otras para la identificación de emociones como lo son los de aprendizaje automático. Estas consisten en sistemas capaces generalizar comportamientos a partir de la información proporcionada, que en este caso en específico sería el conjunto de imágenes que representen cada emoción.

La relevancia de los sistemas de aprendizaje automático se debe a los buenos resultados que han obtenido al reconocer emociones, como lo visto en [39] y los trabajos citados en el estado del arte.

Siendo la arquitectura propuesta catalogada en este sector. Donde se utilizan conjuntamente las transformadas wavelets, Análisis discriminante (LDA) y Maquinas de Vectores de Soporte (SVM) para la identificaciones de emociones.

1.1. Problemática.

El reconocimiento de emociones, sigue siendo un área en desarrollo, tanto para las áreas de las ciencias sociales como para el área computacional. Su estudio se

puede realizar de distintas maneras, a través de vídeo, imágenes, audio, pulsaciones cardiovasculares o una combinación de estas.

El vídeo es una opción apropiada para sistemas donde una cámara se encuentra siempre frente a la persona que se va a analizar, pero al mismo tiempo implica analizar gran cantidad de información comparado con la identificación de emociones en una sola imagen. Ya que el vídeo representa una secuencia de fotogramas (frames) que han de ser analizadas, siendo en un vídeo convencional el uso de 24 fotogramas por segundos, lo cual representa muchas imágenes en un periodo corto de tiempo.

Las principales desventajas de la identificación de emociones a través de vídeo están relacionadas con la gran cantidad de variables presentes en el proceso de reconocimiento, ejemplo de ello son: la calidad de la imagen, la iluminación, la posición de la cara y hasta las facciones propias de determinados grupos étnicos.

Por tal motivo, diversas investigaciones se realizan sobre ambientes controlados. Estos ambientes corresponden a habitaciones con una buena iluminación, para la toma de imágenes o vídeos con la presencia de una sola persona que se encuentre de frente a la cámara.

El uso de vídeos con ambientes controlados limita su uso a aplicaciones que controlen la iluminación y ángulo del rostro de la persona. Pero ofrece mejores resultados que el involucrar ambientes dinámicos, donde la presencia de más variables hace que sea menos exacto.

Es por ello que se manejan ambientes controlados, donde la persona se encuentra en una habitación con poca o nula variación del ambiente, la iluminación difícilmente varía y la persona se encuentra la mayor parte del tiempo en una misma posición.

Pero, ¿Cómo se puede reconocer las emociones faciales gesticuladas por una persona presente en una secuencia de vídeo?

Las arquitecturas se enfocan a resolver un problema en específico, más tratándose en el caso de una secuencia de vídeo, ya que el proceso de

identificación se realiza durante todo el lapso que dure el vídeo, generando múltiples fotogramas que deben ser analizados.

Siendo las técnicas para la identificación de rostros, extracción de facciones y clasificación de emociones una parte importante para definir una arquitectura.

El estudio de emociones es muy variado, ya que es necesario detallar el entorno en el que se realiza el estudio, ser capaz de identificar la presencia o ausencia de rostros, definir las secciones del rostro que son determinantes para la identificación de emociones y la elección de los algoritmos de extracción y clasificación.

Adicionalmente, la arquitectura tendrá que afrontar las dificultades de la carga computacional que genera el uso del vídeo como método de entrada y la reducción en lo posible de la dimensionalidad del problema.

1.2 Objetivo

Generales

Desarrollar una arquitectura para el reconocimiento de emociones faciales basada en las transformadas wavelets, Análisis discriminante (LDA) y Support Vector Machine.

Específicos

- Determinar la existencia o ausencia de un rostro en una secuencia de vídeo, para la extracción de imágenes del rostro de la persona.
- Extraer las características determinantes de los ojos, contorno ojos, nariz y boca, para su representación matricial.
- Generar una matriz con la proyección del rostro, que permita identificar las diferencias entre las secciones del rostro.
- Clasificar las emociones faciales gesticuladas por una persona, para identificar de felicidad, enojo, miedo, sorpresa, tristeza y disgusto.

1.3 Justificación

El estado de ánimo de una persona afecta su memoria e influye en su juicio para la toma de decisiones. Es por ello que el reconocimiento de emociones ha sido objeto de diversas investigaciones, ya que su aplicación se encuentra en diversas áreas:

- En los sistemas de tutor inteligente permite conocer el estado de ánimo del estudiante, con el fin de realizar intervenciones cuando la situación lo amerite y así mejorar el aprendizaje [5].
- Los vídeo juegos se apoyan en esta tecnología para conseguir un mayor realismo, ya sea adaptando la modalidad del juego a la experiencia del jugador o simplemente reaccionando a las emociones captadas en pantalla.
- Los sistemas de ayuda vial con ayuda de una cámara frontal obtienen información del conductor para conocer su estado emocional y si es necesario aplicar alguna estrategia para la seguridad de los pasajeros en el medio de transporte.
- El monitoreo de enfermos también es un área de oportunidad, ya que el identificar y notificar molestias o complicaciones de un paciente, hace posible la atención simultanea de una gran número de pacientes.

El desarrollo de esta tesis puede representar a largo plazo, el modelo a seguir para el desarrollo de una de las aplicaciones anteriormente mencionadas. Mientras que a corto plazo, representa la definición de una arquitectura para el campo de investigación en el reconocimiento de emociones, que sea punto de referencia para futuros trabajos en el área.

El aporte es la definición y desarrollo de una arquitectura que defina los rasgos faciales necesarios para clasificar emociones en secuencias de vídeos, utilizando las transformadas wavelet, el análisis discriminante y las máquinas de soporte de vectores.

1.4 Hipótesis

Es posible desarrollar una arquitectura para el reconocimiento de emociones en secuencias de vídeo, utilizando el algoritmo Viola Jones para la identificación de los ojos, cejas y boca de una persona y, a partir de estos rasgos aplicar las transformadas wavelets, para ser reducidos por el análisis discriminante y clasificadas por medio las máquinas de vectores soporte de vectores.

1.5 Alcances y limitaciones

Alcances

Identificar las emociones que son extraídas a partir de imágenes que contengan un solo rostro y sin un tipo de oclusión.

Clasificar una emoción presentada en una secuencia de vídeo por una persona de la cual no se tenga registro previo.

Limitaciones

El grupo de personas presentes en los vídeos para la identificación de emociones de la tesis, corresponden a personas caucásicas de entre 18 y 40 años de edad, ya que a partir de los 40 hay una gradual disminución en la detección del miedo, y a partir de los 50 el enojo [20].

El vídeo para la extracción de rostros debe permitir la extracción de imágenes de por lo menos 320 x 240 pixeles [7], donde las personas presentes.

1.6 Metodología propuesta.

Se propone una nueva arquitectura para la identificación de emociones, el cual consta de varios algoritmos de acuerdo a cada etapa, se encuentra dividida en 4 etapas: Identificación de rostros, la extracción las áreas de interés en forma de valores, la creación del vector de proyección y finalmente la clasificación de la emoción.

Identificación de rostros y áreas de interés. Consiste en el pre-procesamiento de la imagen. Esta etapa se encarga de eliminar las secciones ajenas al rostro, con lo que se reduce a menos de la mitad el tamaño de la imagen. Esto se realiza a través del algoritmo Viola Jones, identificando las áreas de interés, que son los ojos, contorno de los ojos y boca, ya que estas secciones proporcionan la mayor información de las emociones [9].

Algunos autores sugieren la utilización de la nariz. Sin embargo, esa sección solo proporciona información para emociones de ira y disgusto.

Adicionalmente, el algoritmo de Viola-Jones es capaz de identificar la ausencia o falta de un rostro, lo representa una ventaja al conocer si es necesario o no aplicar el reconocimiento de emociones. El proceso consiste en obtener una imagen de color en formato RGB, la cual es cambiada a escala de grises, con el fin de ser más fácil localizar las comisuras de los ojos y boca. Estas secciones de la cara son las seleccionadas, ya que estas han demostrado ser las que proporcionan la mayor cantidad de información de una emoción para personas caucásicas.

El funcionamiento de la arquitectura presentada se centra en imágenes con iluminación controlada, ya que el algoritmo Viola Jones es sensible a los cambios de la intensidad de luz, lo que ocasionaría que no fuera posible la identificación del rostro de una persona y, por consecuente no se aplicaría el reconocimiento de emociones.

Es por ello que la problemática de depender del reconocimiento del rostro, hace que una mala detección de las secciones del rostro, resultará en la clasificación errónea de la emoción. Por este motivo el problema está delimitado, para

ambientes controlados de iluminación y sin oclusiones.

La elección del algoritmo Viola Jones sobre otros que realizan las mismas funciones es debido a la rapidez y porcentaje de acierto que presenta en condiciones de buena iluminación [7].

Representación de valores de características en forma matricial. Las wavelets permiten extraer frecuencias y texturas de las imágenes a través de filtros. Este proceso elimina las impurezas de la imagen y obtiene los valores matriciales de cada una de las zonas de interés identificadas por el algoritmo Viola Jones.

Dentro de nuestra arquitectura se define el uso de transformadas wavelets del tipo analítico. Es decir, que la aplicación de las wavelets se realiza solo sobre las regiones identificadas en la fase anterior y no sobre toda la imagen, realizando un procesamiento de una menor cantidad de datos.

Aunque las wavelets son robustas a cambios de luminosidad [1], la propuesta establece la elección de imágenes de pruebas con iluminación controlada, ya que las secciones de interés que se obtienen con la primera fase, si presenta dificultades a los cambios de iluminación.

Obtención de la función discriminante.- La matriz de rasgos discriminantes obtenida con transformadas wavelets se procesará con Análisis Discriminante (LDA), con el fin de reducir el número de características obtenidas con la rejilla. El procesamiento de las características permite la obtención de una matriz de proyección con los rasgos más sobresalientes.

Clasificación la emoción. Una vez que la función discriminante es obtenida, es posible clasificarla a través de las máquinas de soporte (SVM), ya que han demostrado ser un clasificador efectivo en la clasificación de emociones [10].

El entrenamiento de la máquina de vectores es también una etapa a considerar al momento de la implementación. El tipo de SVM propuesto para reconocer la emoción es un clasificador uno frente a todos, donde el número de clasificadores es igual al número de emociones disponibles, es decir, seis clasificadores en nuestra arquitectura. Para determinar cuál es la emoción que presenta la persona, hace una ponderación entre todas las clasificaciones, para finalmente dar un resultado de la emoción presente.

1.7 Organización de la tesis

Este trabajo se encuentra dividido en 5 capítulos. El capítulo 1 contiene la introducción de la problemática y variables utilizadas. El capítulo 2 incluye el estado del arte de trabajo de reconocimiento de emociones a través de imágenes y vídeo. El capítulo 3 contiene las definiciones que conforman esta tesis. Continuando con el capítulo 4 que corresponde con la descripción de la arquitectura propuesta y los resultados obtenidos. Por último, el capítulo 5 presenta las conclusiones y el trabajo futuro.

CAPITULO

2 Estado del arte

En este capítulo realizamos una revisión al estado del arte. En la sección 2.1 mencionamos una introducción al reconocimiento de emociones, la sección 2.2 muestra una clasificación de trabajos que utilizan sistemas para el reconocimiento de emociones a través de sistemas de aprendizaje automático. Finalmente definimos las conclusiones del capítulo en la sección 2.3 .

2.1 Introducción.

Una emoción es la alteración del ánimo de una persona producida por un sentimiento. La representación de la emoción se denota con la modificación de los rasgos faciales, ademanes con pies y brazos, cambio de la frecuencia cardiaca y tono de la voz [41].

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 popular, 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 crearan 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:

- Sistemas basados en los parámetros de animación facial (FAPs)
- Sistemas de aprendizaje automático.

2.2 Clasificación de trabajos realizados en el reconocimiento de emociones.

2.2.1 Sistemas de codificación de acciones faciales (FACs).

Los FACs son técnicas utilizadas por expertos en el campo de la identificación de emociones. Centran su estudio principalmente en la modificación de los músculos de la cara, que corresponde a ojos, boca, nariz y comisuras del rostro.

Las modificaciones de uno o varios músculos del rostro dan origen a las acciones unitarias. Se han definido 46 acciones unitarias, con las cuales es posible identificar la emoción presentada por un individuo en el momento que se observa en [11].

<i>NEUTRAL</i>	AU 1	AU 2	AU 4	AU 5
				
Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5
				
Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7
				
Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids cheeks are raised.	Brows, eyelids, and cheeks are raised.

Figura 2.1. Ejemplo de acciones faciales de cejas, ojos y contorno de los ojos [43]

Hoy en día este proceso está automatizado. Y la ventaja que presenta es la versatilidad al reconocer emociones de personas que no se han observado anteriormente.

Sin embargo, presentan problemas al identificar las emociones de personas con algún tipo de parálisis facial, ya que las acciones unitarias no están diseñadas para estos casos o cualquier otro con algún tipo de traumatismo facial que altere la posición, forma o dimensiones de la boca u ojos.

Y aunque las FACs fueron diseñadas para que el análisis lo realice un individuo, los conceptos sirvieron para la creación de sistemas computacionales. A continuación se presentan trabajos basados en FACs subdivididos en:

- Análisis de flujo óptico.
- Análisis holísticos.

2.2.1.1 Análisis de emociones faciales por flujo óptico

Este método se basa en representar una nube de puntos en el espacio. Los puntos pueden estar constituidos en una malla, donde todos se encuentran unidos por aristas generando polígonos, generalmente triángulos por ser una forma

simple para representar la superficie presente en la cara.

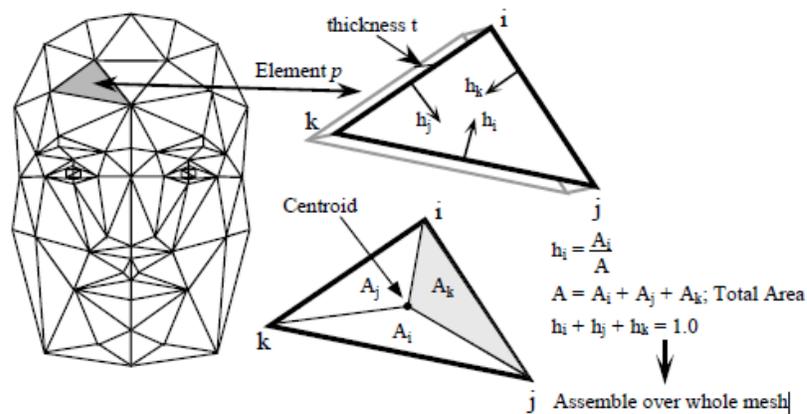


Figura 2.2. Análisis de rasgos faciales [44]

En [13] crean una versión extendida de las FACs, su trabajo se basa en estimaciones ópticas que junto con la geometría, física y dinámica de los movimientos de los modelos generados a partir de las estructuras faciales, pueden determinar la emoción expresada.

Para ello hacen uso de un procesamiento de flujo óptico como base para la percepción y medición del movimiento facial presentado. Detectando las regiones de ojos, nariz y boca por medio del autoespacio modular definido por Pentland and Moghaddam [41].

El modelando la cara utiliza el método Simoncelli 's para el cálculo de flujo óptico, este utiliza un filtro Kalman de multi-escala, que proporciona una buena estimación de movimiento e información de la covarianza del error.

Análisis holísticos.

Los análisis son aquellos modelos que su funcionamiento está dado por el análisis del rostro entero y no por el análisis de partes individuales.

Para lograrlo es necesario clasificar a través de las características relevantes a una imagen. Este proceso se puede realizar por análisis de componentes principales como en [14], por transformadas wavelets tal como se muestra en [15], eigencafe o por Fisherface [17], por mencionar algunos.

[19] Realizo estudios de las diferentes técnicas para el reconocimiento de acciones faciales utilizando análisis de componentes principales (PCA), Análisis de componentes independientes (ICA), Análisis de características locales (LFA), análisis discriminante (LDA) y esquemas locales como las transformadas wavelet, obteniendo los mejores resultados con representación por medio de wavelets y el uso de ICA.

Sin embargo, la desventaja de ICA se encuentra en que las expresiones presentadas en Unidades de Acción pueden ser lineales o no lineales. ICA puede presentar dificultades con datos no lineales como lo son la combinación de Unidades de Acción.

También hay que mencionar que las wavelets hacen uso intenso de la memoria y del tiempo de computo, ejemplo de ello es el trabajo de [16], donde la representación de cada cara en una resolución de 48x48, obtuvo una alta dimensionalidad de $O(10^{25})$.

Este problema es tratado en nuestra arquitectura es solucionado por medio de la reducción de la dimensionalidad, reduciendo el área de trabajo, al enfocarnos solo en las secciones de cejas, ojos y boca se reduce más de la mitad de la imagen del rostro.

2.2.1.2 Otras técnicas de reconocimiento de emociones utilizando FACs

REDES BAYESIANAS

Las redes bayesianas modelan un fenómeno mediante un conjunto de variables y las relaciones de dependencia entre ellas. Con lo que se estima la probabilidad para nuevas variables no conocidas.

Es muy utilizado en áreas de reconocimiento de patrones. Sin embargo, es poco

utilizada en la identificación de emociones, debido a que este algoritmo necesita una gran cantidad de información para poder realizar un buen entrenamiento.

A continuación se presentan trabajos utilizando redes bayesianas:

Tomando trabajos que utilizan el vídeo como su medio de entrada se encuentra [28] que adiciona la detección de las pupilas utilizando infrarrojos y filtros kalman, los cuales le dan robustes y mayor insensibilidad a variaciones de luz, movimientos de cabeza. La representación de las variables integra las redes bayesianas con las ACs. Y finalmente el reconocimiento se realiza por medio de un framework realiza una fusión de la información activa y dinámica

Seguidor de Características

[32] Por medio de un “feature tracker” localizan los puntos de extracción en expresiones espontaneas en tiempo real. Extrayendo la posición de 22 características para clasificarlo por medio de un SVM previamente entrenado con ejemplos de entrenamiento.

2.2.2 Parámetros de animación facial (FAPs)

Esta forma de identificación de emociones se realiza mediante la animación de rostros en el estándar MPEG-4. Se utilizan parámetros para estandarizar los modelos de las medias del rostro a medidas antropométricas, y así definir la deformación que sufre el rostro.

Los parámetros de animación facial, Facial Animation Parameters (FAP), son un conjunto de parámetros que permiten la animación de modelos de cara en forma sintética. Con lo que es posible obtener la deformación que sufre el rostro en base al estado neutro.

Para ello es necesaria la obtención previa del rostro en estado neutro, lo que implica que la imagen se encuentre:

- Con los músculos de la cara relajados
- Los párpados tangentes al iris
- Pupila a un tercio del iris
- Labios en contacto. La línea de contacto es horizontal de la misma longitud que la distancia entre comisuras.
- Boca cerrada y dientes de arriba y abajo se tocan.

El rostro en estado neutro está definido por un conjunto de puntos característicos estandarizados, denominados parámetros de definición facial (FDP) que se miden en FAPU (Parámetro unitario de animación facial).

A partir de conocer estos valores, es posible comparar la cara con la emoción a identificar con el rostro de la misma persona en estado neutro, y así, hacer posible identificar una sonrisa pequeña de una grande [9].

Diversas técnicas son utilizadas en el reconocimiento de emociones utilizando FAPs.

Ejemplo de ello, es un sistema de fuzzificación [22], que utiliza los valores de las FAP, el conocimiento previo del rostro y los perfiles de las expresiones para deliberar la expresión mostrada.

Otra técnica muy utilizada por diversos autores es el uso de HMC (hidden Markov model)[21]-[24]. En el trabajo [21] se usó solo la ceja y contorno de los labios para realizar las clasificaciones. Como resultado obtuvieron un 100% de certeza en detectar la sorpresa y 70.6% (emoción con menor porcentaje de acierto) en detectar el miedo.

También se encuentra el uso de redes bayesianas [26] utilizadas para unificar el análisis de las expresiones faciales y sintetizarlo en una estructura correcta. Al finalizar el análisis, se toman las FAP y FAC en una red bayesiana dinámica, para considerar las incertidumbres en la extracción de FAPs y modelar la evolución de

las expresiones faciales.

2.2.3 Otras técnicas de reconocimiento de emociones

Sin embargo, las técnicas basadas en FACs y FAPs no son las únicas utilizadas, también se encuentra con aquellos que se basan a partir del aprendizaje automático que generan los algoritmos, como es el caso de las redes neuronales, las redes bayesianas, los métodos holísticos anteriormente mencionados que no hacen uso de las FACs.

Este tipo de trabajos son principalmente los basados en apariencias, presentes en los cambios de la cara, que son extraídos aplicando filtros al rostro entero o a parte de esta.

Los filtros más utilizados en el reconocimiento de emociones son las wavelets, principalmente por ser robustos a cambios de iluminación y a cambio de facciones [30]. Esta técnica se combina otras que permitan el reconocimiento de patrones o de clasificación.

Esto se puede observar en [31] donde probaron con diferentes técnicas posteriores a los filtros. Y los mejores resultados los encontraron utilizando filtros Gabor usando AdaBoost y SVM alcanzando un 93% para el reconocimiento de 7 categorías básicas.

Sistemas de aprendizaje automático.

PCA

El análisis de componentes principales es una técnica estándar para el reconocimiento de patrones estadísticos y la reducción de datos en el procesamiento de señales.

Por tal motivo, es también utilizado en el reconocimiento de emociones. Una forma

de utilizarlo [27] es por medio de imágenes que representen una emoción de repositorio como JAFEE (El cual esta constituido por 273 imagenes de 10 modelos japonesas mostrando las 6 emociones basicas y neutral), extrayendo las características del rostro de interés, para posteriormente aplicar un clasificador euclidiano, que calcule las distancias entre la imagen proyectada y encuentre la imagen entrenada que más se le parezca.

Las imágenes de la base de datos JAFEE, se extraen las características importantes y se clasifican por medio de un clasificador euclidiano.

Utilizando PCA en el reconocimiento a través de imágenes se presenta mejor precisión al detectar estado neutro, triste y enojado.

ICA

Análisis de componentes individuales (ICA) es una generalización del análisis de componentes principales (PCA) que separa las dependencias de alto orden en comparación con las dependencias de segundo orden.

Ejemplos de estos trabajos se puede observar en [29] donde se utiliza un FastICA para permitir un manejo simple en covarianzas cubicas.

LDA

El análisis discriminante (LDA) es una técnica que describe las diferencias entre un conjunto de grupos con respecto a las variables observadas. Con estos datos, trata de encontrar una buena función discriminante que sea una combinación lineal de las variables originales. Al aplicar la función a un dato nuevo nos dice a qué grupo pertenece.

Debido a ello es posible utilizarlo en trabajos de reconocimiento de patrones y en el área de clasificación de emociones. A continuación se presentan trabajos relacionados que hacen uso de esta técnica:

[27] Utiliza un modelo candide (ver imagen 2.3), el cual consiste en un marco

triangular 3D del rostro humano con 113 puntos de referencia y 184 triángulos representando un rostro humano. Adicionado un conjunto de 14 facial units (FU), para evaluar los pesos de los coeficientes desconocidos de los 11 action units definidos, y construir las características dimensionales que serán clasificadas a través del análisis discriminante.

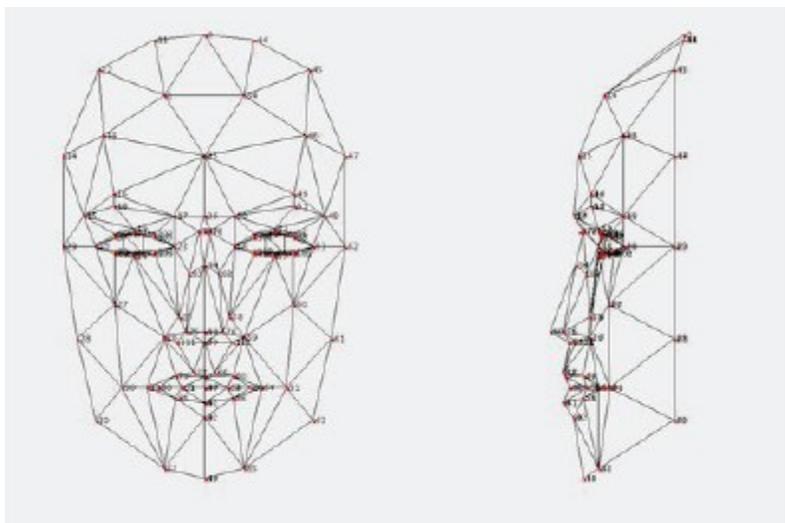


Figura 2.3 Ilustración de un modelo Candide [27]

Representación local

La representación local (Local Binary Pattern) fue utilizado para el reconocimiento de textura, pero con el paso del tiempo fue aplicado para la representación de imágenes faciales. La parte importante de este algoritmo es su tolerancia a los cambios de iluminación y computo simplificado.

El algoritmo funciona creando un histograma, el cual contiene la información acerca de la distribución de los micro-patronos locales, como bordes, manchas y áreas planas, más de toda la imagen, por lo que se puede utilizar para describir estadísticamente características de la imagen.

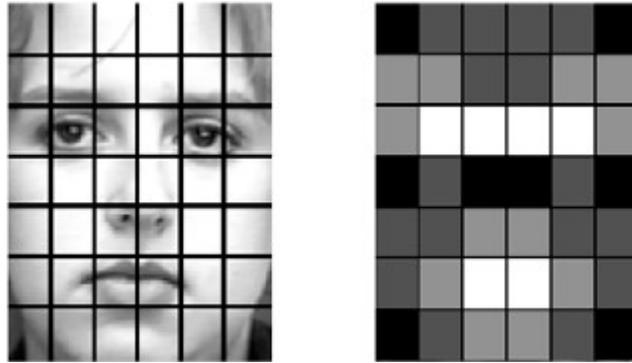


Figura 2.4. Rostro dividido en una subregión de 6 x 7 [18]

Trabajos relacionados respecto a esto se puede observar en [18] donde se utilizaron diversas máquinas de aprendizaje y diversas bases de datos. Para ello se realizaba extracción de las características más discriminantes por medio de un boosted Local Binary Pattern y, la clasificación por medio de máquinas de soporte, las cuales presentaron mejores resultados, en comparación con el análisis discriminante, redes neuronales y máquinas de soporte.

La combinación de LBP y SVM les presento robustez y estabilidad con imágenes con baja resolución (110x150, 55x75 36x48 27x37) y buen desempeño en vídeos en vídeos con baja resolución.

2.2 Sistemas de aprendizaje automático supervisado en vídeo.

Hay diversos trabajos para el reconocimiento de rostros utilizando las transformadas wavelets [1], [2] que se apoyan del uso de LDA, PCA, PCA+ LDA, PCA+ SVM, Adaboost+ LDA, Adaboost + SVM. Sin embargo, no se han utilizado wavelets+LDA+SVM conjuntamente en el reconocimiento de emociones como lo planteado.

Entre los trabajos más relacionados se encuentran

[33] Que hacen uso de técnicas similares, ya que realizan el análisis de

características por medio de LDA y SVM para el reconocimiento de rostros.

[34] A partir imágenes de Cohn and Kanade realizan la localización de la cara y a partir de este se le aplico un filtro Gabor de 40 filtros para posteriormente ser procesado por una máquina de vectores. En dicho trabajo se realizaron pruebas con diferentes tipos de SVM. Obteniendo mejores resultados con una combinación de SVM lineales en altos y bajos Gabor sobre toda la cara, con un rendimiento del 87.5%

2.3 Conclusiones del estado del arte.

En la presente unidad se mostró 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.

Y es esta amplia gama de posibilidad que es posible detectar emociones a partir de rostros se puede realizar sobre imágenes o vídeo, con aplicaciones de aprendizaje supervisado o no supervisado, con bases de datos que contienen personas representando emociones o con el uso de emociones auténticas.

Siendo la base de datos importante en el proceso. Como se observó en el presente capítulo, las más utilizadas son JAFFE, FERRET y Cohn-Kanade. Sin embargo, recientes estudios han empezado a realizar trabajos con expresiones auténticas, por medio de su propia base de datos o utilizando las bases de datos conocidas pero tomando la mitad de la duración de la expresión [3].

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. Ejemplo de ello es su uso en identificación de rostros [1], [2].

CAPITULO

3 Reconocimiento de emociones en vídeo

Los trabajos realizados para el reconocimiento de emociones son muy variados como se observó en el capítulo 2. Pero el trabajo aquí presente se enfoca al reconocimiento de emociones a través de vídeo por medio de un aprendizaje supervisado.

El capítulo aquí presente muestra las definiciones involucradas en la creación de la arquitectura para el reconocimiento de emociones. El capítulo está dividido en 4 secciones. La sección 1, presenta el trabajo más relacionado para el reconocimiento de emociones. La sección 2, describe la base de datos utilizada, En la sección 3, se presentan las definiciones que se utilizaron para el desarrollo de la arquitectura. Y finalmente la sección 4, contiene las conclusiones del capítulo.

3.1 Trabajo relacionado.

En esta sección, se presentan los trabajos que han propuesto soluciones para el reconocimiento de emociones utilizando métodos de aprendizaje supervisado en vídeo.

Trabajos en este ámbito se busca reducir los datos de entrada, ya que el vídeo o las secuencias de imágenes son una representación tridimensional. Para ello se toma el conocimiento de los objetos de interés (áreas de interés) [36].

Trabajos que buscan a través del LDA y un SVM clasificar las emociones [38]. Mientras que aplicaciones que hacen uso de las técnicas similares a las planteadas, se encuentra [37] que realiza el análisis de características por medio de las wavelets, obtiene las características con el PCA y utiliza un clasificador SVM

3.2 Descripción de las bases de datos usados en el trabajo.

Las bases de datos que son más utilizadas para el reconocimiento de emociones son la de JAFFE (Japanese Female Facial Expression), FERRET (The Facial Recognition Technology) y Cohn-Kanade.

Para propósito de esta tesis, la base de datos de JAFEE no fue utilizada por contener rostros de personas con rasgos asiáticos, lo cual representa una diferencia en los rasgos presentados al desarrollar una emoción en comparación con una persona caucásica.

Y la decisión de utilizar Cohn-Kanade sobre FERRET se debe a que las imágenes con las que cuenta son secuencias de imágenes desde un estado neutro a la emoción en cuestión, siendo similar a extraer imágenes desde una secuencia de vídeo. Además de su uso en trabajos similares de detección de emociones a partir de un vídeo [35][36].

Todas las imágenes son frontales con iluminación controlada, fondo liso y blanco.

Esto con el fin de evitar la mala detección del rostro y el entrenamiento del sistema sea de forma correcta.

Esta base de datos contiene 97 modelos que presentan de una a seis emociones básicas (Figura 3.1). Está constituido por hombres y mujeres de diversas ascendencias étnicas, por lo que fue necesario eliminar un conjunto de 20 modelos que presentaban rasgos afroamericanos o el uso de una oclusión como lo es el pelo sobre los ojos o gorra.



(Figura 3.1 Imágenes de muestra de base de datos Cohn-Kanade) [42]

Con la selección de imágenes fue posible crear el repositorio y posteriormente la base del conocimiento, conformado con 160 imágenes de hombres y mujeres, para representar cada una de las emociones básicas definidas por Ekman. 120 utilizadas para el entrenamiento y 40 para pruebas.

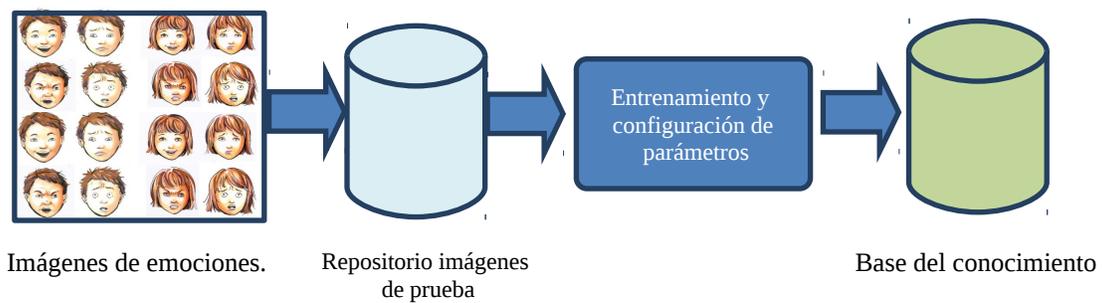


Figura 3.2. Creación de la base del conocimiento. Aportación propia

3.3 Definiciones de la arquitectura.

La arquitectura está compuesta por 4 secciones:

- Identificación del rostro de áreas de interés
- Representación de valores de características en forma matricial.
- Obtención de función discriminante
- Clasificación de la emoción.

Cada una presenta definiciones que se presentan a continuación:

3.3.1 Identificación del rostro de áreas de interés.

Basando en el uso del algoritmo Viola Jones es posible detectar el rostro y área de interés para trabajar.

Esto lo logramos a través de la identificación de las características o *features* f de una imagen I , codificada para llevar a cabo una tarea computacional.

El algoritmo busca entre varios *features* aquellos que le den mayor información acerca de la cara. Ya que la evaluación de estos pueden determinar si hay o no un rostro.

Se hace uso de los Haar like features que a través de dos, tres o cuatro rectángulos encuentran en la imagen integral las zonas blancas y las zonas negras. Se denomina F al conjunto de tipos de features posibles.

$$\mathbb{F} = \left\{ F_1 : \begin{array}{|c|} \hline \blacksquare \\ \hline \end{array}, F_2 : \begin{array}{|c|} \hline \square \\ \hline \end{array}, F_3 : \begin{array}{|c|c|} \hline \blacksquare & \square \\ \hline \end{array}, F_4 : \begin{array}{|c|c|} \hline \square & \blacksquare \\ \hline \end{array}, F_5 : \begin{array}{|c|} \hline \square \\ \hline \blacksquare \\ \hline \end{array} \right\}$$

Figura 3.3 Ejemplo Haar like features [7]

Los features de Haar que capturan muchas de las características presentes en

una cara, ejemplo de ello son las cuencas de los ojos que suelen ser más oscuras que la parte de las mejillas y la zona de la nariz ser más claras.

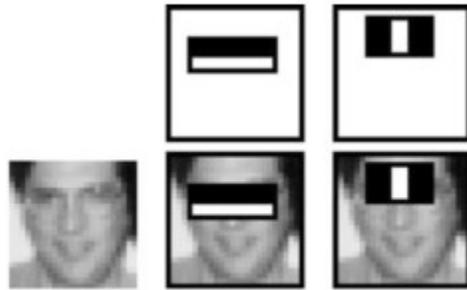


Figura 3.4 Haar like features en imagen de rostro. [7]

Estas características son sencillas de calcular y aportan poca información individualmente, pero su cálculo es muy veloz y se puede combinar con otras características.

Todas las imágenes utilizadas son en escala de grises, proporcionando imágenes de dos dimensiones de tamaño $m \times n$, donde el valor de la intensidad en la posición (x,y) es $1 \leq x \leq n$, $1 \leq y \leq m$.

Y la función de un feature de haar está definida por la ecuación:

$$f_F(I) = \sum_{r_i \in F} c_i \cdot \text{RecSum}(I, r_i) \quad , \quad \text{donde } F \in \mathbb{F}$$

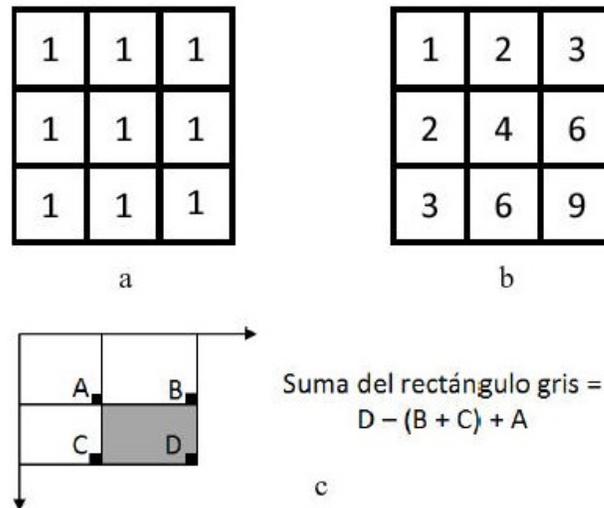
$$c_i = \begin{cases} -1 & \text{si } r_i \text{ es blanco} \\ 1 & \text{si } r_i \text{ es negro} \end{cases}$$

Figura 3.5 Ecuación del haar like features.

Donde F es uno de los tipos de features, I es la imagen donde se evalúa el feature, r_i son los rectángulos que componen al tipo F y $\text{RecSum}(I, r_i)$ es la función que calcula la suma de los píxeles de r_i en la imagen I .

En el algoritmo se hace uso de la imagen integral, que permite calcular rápidamente los features de la imagen. La cual es una imagen donde cada punto contiene el resultado de la suma de los valores de todos los puntos situados por

encima y a su izquierda en la imagen original.



3.6 Figura Cálculo de la imagen integral [7]

(a)Imagen original, (b) imagen integral, (c) Cálculo de la suma en una imagen integral

La imagen integral sirve además para normalizar la imagen y reducir los efectos causados por la iluminación.

3.3.2 Representación de valores de características en forma matricial.

El proceso de detección de áreas de interés nos proporciona una imagen recortada, sin embargo es necesario extraer los valores que representan esas emociones.

Para ello se aplica una transformada wavelet de una dimensión, con el fin de obtener un vector transformado que contiene la mitad con alta frecuencia y la otra con baja frecuencia.

Aquí la parte de alta frecuencia presenta información pequeña y puede descartarse, permitiendo una compresión de los datos.

Para extraer la información a partir de los datos, se le aplican filtros digitales. La información de baja frecuencia se obtiene aplicando un filtro G , denominado pasa-

bajo, determinado por los coeficientes $\{g_i\}_{i=1}^L$

Una vez obtenido las wavelet de cada imagen, se agrupan por emoción. Cada emoción tendrá un numero N de imágenes, de tamaño $w \times h$. Cada imagen de la emoción X tomara una fila de cada imagen de N y formará una nueva matriz de tamaño $N \times hw$.

De tal forma que se obtendrán 6 matrices una por cada emoción.

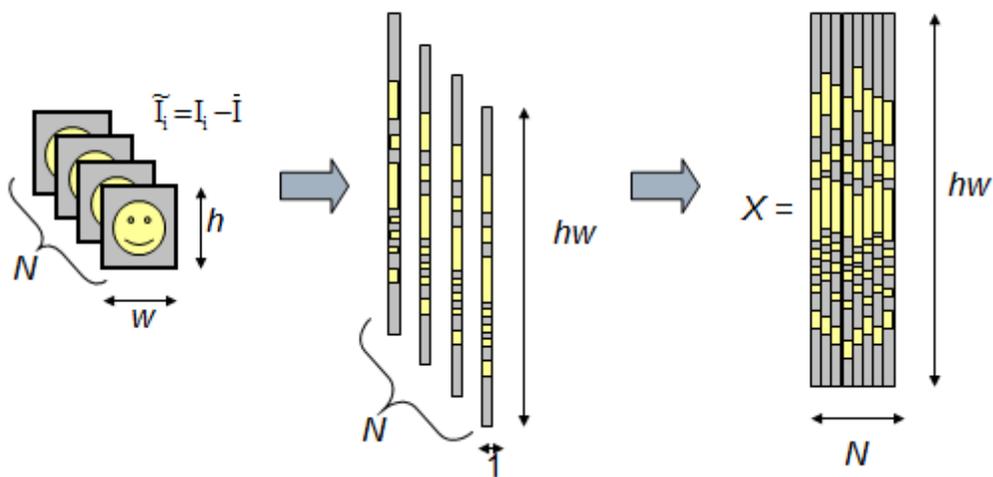


Figura 3.7 Obtención de matriz representativa de una emoción. [43]

3.3.3 Obtención de función discriminante.

La aplicación de la transformada wavelets a cada imagen y compresión en una matriz por cada emoción hace posible aplicar el análisis discriminante, siguiendo los pasos del algoritmo:

Se obtiene la cara media

Sea I_1, \dots, I_N el conjunto de matrices de emociones, tamaño $h \times w$.

Sea la imagen media (“cara media”).

Restamos la media a todas los valores de la matriz: $\tilde{I}_1, \dots, \tilde{I}_N$

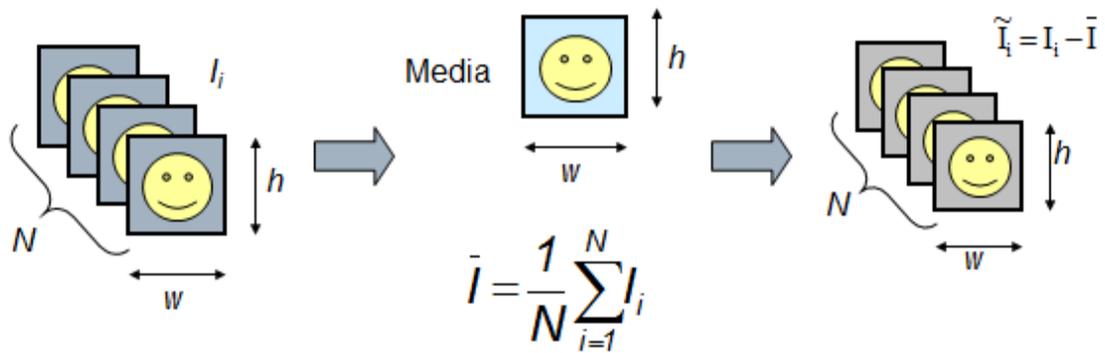


Figura 3.8 Obtención de medias y formación de nueva matriz [43]

Para posteriormente calcular las matrices de covarianzas

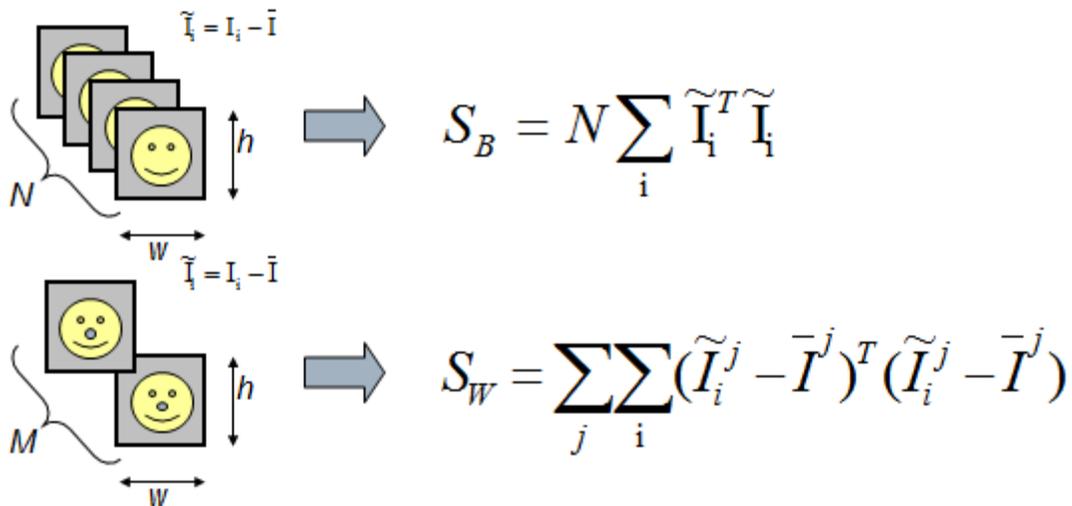


Figura 3.9 Obtención de función discriminante. [43]

Donde se maximiza con la expresión

$$tr\left(\frac{S_B}{S_W}\right)$$

Calculando los autovectores

$$S_W^{-1} S_B = P \Delta P^T$$

Reducción de la dimensionalidad al obtener la función discriminante.

Se proyecta la imagen con la función discriminante para obtener otra matriz

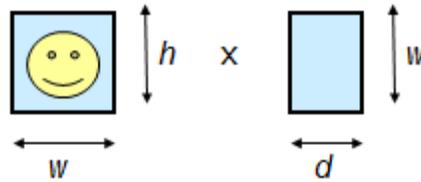


Figura 3.10 Matriz final. [43]

3.3.4 Clasificación de la emoción.

Una vez que la función discriminante es obtenida, es posible clasificarla a través de las máquinas de soporte (SVM), ya que han demostrado ser un clasificador efectivo en la clasificación de emociones [10].

El entrenamiento de la máquina de vectores es también una etapa a considerar al momento de la implementación. El tipo de SVM propuesto para reconocer la emoción es un clasificador uno frente a todos, donde el número de clasificadores es igual al número de emociones disponibles, es decir, seis clasificadores en nuestra arquitectura.

3.4 Conclusiones del capítulo.

Presentamos las características de la base de datos utilizada, así como la forma en que se construyó nuestra base del conocimiento.

Las definiciones utilizadas en todo el proceso de la detección de emociones, comenzando por la identificación de áreas de interés con el algoritmo Viola Jones, la presentación matricial con las transformadas wavelets, el proceso para adopción a matrices y obtención de la función discriminante y finalmente la clasificación por medio de las máquinas de soporte de vectores.

CAPITULO

4 Aplicación del algoritmo

En este capítulo presentamos el algoritmo propuesto para el reconocimiento de emociones en secuencias de vídeo. Este se centra en la detección de un momento específico y se utilizan las descripciones mencionadas anteriormente. El capítulo se divide en 4 secciones, en la primera sección mencionamos las características con las que trabaja la arquitectura, el segundo se describe la arquitectura, el tercero contiene los resultados y finalizamos con las conclusiones del capítulo.

4.1 Características de la arquitectura.

La arquitectura está enfocada a reconocer emociones con las siguientes condiciones o características.

La arquitectura permite identificar una de las seis emociones básicas definidas por Ekman a partir de una secuencia de vídeo, extrayendo momentos específicos como imágenes y realizando la identificación.

Las imágenes extraídas son de una toma frontal de personas caucásicas de entre 18 y 40 años de edad, en un ambiente controlado y sin ningún tipo de oclusión.

La imagen debe contener un solo rostro.

Está diseñada para funcionar en resoluciones no tan demandantes, como lo son 320 x 240, de la cual se recorta las áreas de interés y queda un área de trabajo aún más pequeño.

Se utilizan imágenes .png en escala de grises, en caso de no contener ese formato son transformadas con el fin de volver la dimensión de las imágenes de n x m.

Es necesario seleccionar un conjunto de imágenes para crear la base del conocimiento.

Se utiliza un entrenamiento supervisado, que permite identificar emociones de personas de las que no se tiene un registro, lo que evita el entrenamiento cada que aparezca una cara nueva.

4.2 Diseño de la arquitectura.

La arquitectura está diseñada en cuatro módulos independientes, que en conjunto realizan la identificación de emociones. Estos módulos corresponden a:

- Identificación de rostros y áreas de interés.
- Representación de valores de características en forma matricial.
- Obtención de la función discriminante.
- Clasificación de la emoción.

4.2.1 Identificación de rostros y áreas de interés

La base de datos proporciona imágenes de 640 x 490, sin embargo se probó con imágenes de 320 x 240 sin presentar cambios en los resultados.

A partir de esta imagen proporcionada se obtiene el área de la cara, esto para buscar los ojos y boca en el área correcta y evitar la detección errónea en una sombra.

Los features para detectar el par de ojos son más exactos que los de la boca, por este motivo se aplica en primera instancia la detección de los ojos y

posteriormente la detección de la boca.

La detección de la boca puede presentar complicaciones y confundir las cuencas de los ojos como boca (ver imagen 4.2). Para evitar eso se aplica el cálculo:

$$Y' + A' \neq X$$

Dónde:

X es el la posición en el plano x de la boca detectada

Y' la posición y de los ojos

A' es la altura de la sección de los ojos.

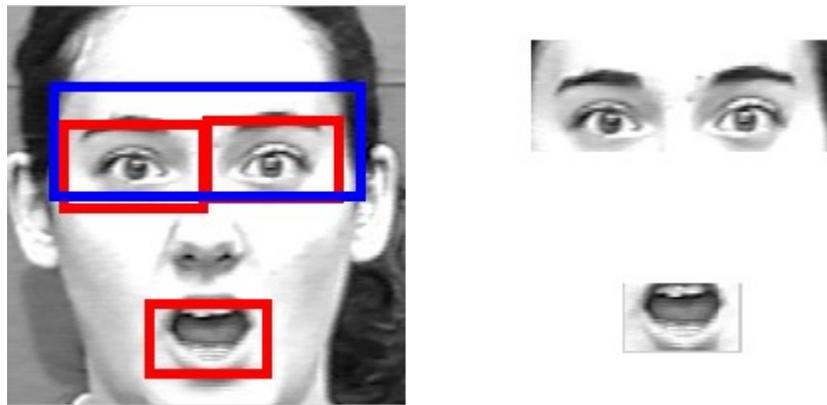


Figura 4.1 Reducción de la imagen. Aportación propia utilizando imagen de base de datos cohn-kanade como base

Una vez identificados son redimensionados para presentarse en un arreglo y no en una matriz

4.2.2 Representación de valores de características en forma matricial

La representación de valores consiste en tomar el vector y aplicarle una transformada wavelet para que puede resaltar las los rasgos y sea mas representativo los valores.

Cada imagen genera una matriz, se toma una fila de cada imagen para formar una nueva matriz que represente a todo la emoción.

4.2.3 Obtención de la función discriminante

La función discriminante se realiza por medio del proceso de sacar la función discriminante de cada matriz.

4.2.4 Clasificación de la emoción.

Se obtiene una función discriminante por cada emoción del paso anterior, pero con el fin de hacer el problema linealmente separable y mejorar la clasificación con emociones que presenten características similares se aplica un SVM.

El SVM que mejores resultados obtuvo fue el 1 contra todos, donde se crea un SVM por cada emoción, siendo el caso particular 6 de ellas.

Para discernir cuando dos o más emociones que sean positivas para una misma imagen, se hace uso de reglas de asociación que le dan prioridad a ciertas emociones que presentan mayor probabilidad de acierto.

4.3 Resultados

Se presentan los resultados obtenidos con 3 distintos tipos de SVM, uno contra todos, uno a uno y el multi SVM. Los resultados se muestran a continuación:

Clasificador	% de Acierto
MultiSVM	50%
1 vs 1	62%
1 vs all	54%

Tabla 5.1 Tabla de resultados

Los mejores resultados se encontraron con el SVM 1 vs 1 , donde el mejor resultado fue el identificar el enojo y el peor la tristeza.

Emoción	SVM 1vs todos
Felicidad	75%
Enojo	83%
Tristeza	50%
Miedo	66%
Disgusto	50%
Sorpresa	50%

Tabla 5.2 Tabla de resultados

4.4 Conclusiones del capítulo.

En esta capítulo se definió el funcionamiento de la arquitectura para el reconocimiento de emociones en sus cuatro fases. Se delimita la función de cada fase y labor específica que realizará.

La arquitectura está diseñada para funcionar bajo las condiciones antes mencionadas, por lo que la modificación en uno de sus variables podría resultar en un desempeño menor al reportado.

Gran parte del buen funcionamiento se debe a la buena elección de imágenes de entrenamiento.

CAPITULO

5 Resultados y conclusiones

5.1 Conclusiones

Como se observó en los resultados es una arquitectura que es muy bueno para identificar unas emociones pero no tanto para otras.

La tesis ha planteado una arquitectura para el reconocimiento de emociones, y el proceso que involucra realizarlo. Se abarca desde la identificación de emociones a través de la arquitectura propuesta. Abarcando las fases de identificación del rostro, extracción de características, análisis de los rasgos y finalmente la clasificación de la emoción.

La arquitectura propuesta plantea el uso de un aprendizaje automático supervisado, que sea capaz de identificar emociones a partir de las características de los ojos, cejas, boca y contorno de la boca. Centrando el estudio en imágenes de base de datos para el reconocimiento de emociones, donde se presenta una iluminación controlada, rostros frontales y sin oclusiones.

Recomendaciones Ya que la arquitectura es del tipo de aprendizaje supervisado,

la base del conocimiento debe estar bien constituida, para ello es fundamental la elección de las imágenes que formarán el repositorio, ya que el proceso del enrolamiento hace uso de esto.

Siendo la entropía entre las imágenes lo que proporciona la diferencia entre ellos, y mientras mayor sea la entropía entre las emociones, mejor será la función discriminante de cada emoción obtenida por el LDA.

5.3 Trabajo futuro

El trabajo presentado tiene múltiples direcciones a seguir, muchas de ellas en la modificación de las variables con las que se definió, y otras en posibles aplicaciones (Ver tabla 5.2).

Área	Aplicación
Entretenimiento	vídeo juegos de realidad virtual. Interacción de robots con humanos.
Salud	Vigilancia de personas enfermas en hospitales. Detección de cansancio en conductores.
Educación	Incorporación con sistemas de tutor inteligentes.
Comercio	Monitoreo de satisfacción del cliente en establecimiento. Monitoreo de comportamiento de empleados. Obtención de reacciones de personas en películas

Tabla 5.3 Aplicaciones del reconocimiento de emociones.

Las variables que pueden ser modificadas son la iluminación, manejo de oclusiones, imágenes de resoluciones menores, identificación a otro grupo étnico, uso de base de datos de expresiones espontaneas, funcionamiento en tiempo real, entre muchas otras.

Además tiene la flexibilidad de poder combinarse a otros sistemas de

reconocimiento como lo son los basados en el tono de la voz, en el ritmo cardiaco y en funciones neuronales.

CAPITULO

6 Referencias

6. Referencias

[1] Chengjun Liu, Harry Wechsler (2002) Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. IEEE, trans, image processing, Vol. 11(4). pages 467-476. ISSN :1057-7149

[2] Ángel Serrano Sánchez de León (2009). Parametrización óptima de un banco de filtros de Gabor para su aplicación a un problema de reconocimiento facial. Tesis magisterial. Universidad Rey Juan Carlos.

[3] Bogdan Raducanu (2008) Dynamic vs. Static Recognition of Facial Expressions. Aml '08 Proceedings of the European Conference on Ambient Intelligence. Pages 13-25. ISBN: 978-3-540-89616-6

[4] José M. Buenaposada¹, Enrique Muñoz, Luis Baumela (2007) Recognising facial expressions in video sequences. Pattern Analysis and Applications.. Volume 11, Issue 1 , pages 101-116 DOI 10.1007/s10044-007-0084-8

[5] McQuiggan¹, Bradford W. Mott², and James C. Lester¹ (2008) Modeling Self-Efficacy in Intelligent Tutoring Systems: An Inductive Approach Scott W. User Modeling and User-Adapted Interaction. Volume 18, Issue 1-2 , pages 81-123. DOI 10.1007/s11257-007-9040-y

[6] Michel C. Desmarais, Ryan S. J. d. Baker (2011). A review of recent advances in learner and skill modeling in intelligent learning environments. User Modeling and User-Adapted Interaction. Volume 22, Issue 1-2 , pages 9-38. DOI 10.1007/s11257-011-9106-8.

[7] Paul Viola, Michael J. Jones (2004). Robust Real-Time Face Detection. International Journal of Computer Vision. International Journal of Computer Vision. Volume 57, Issue 2 , pp 137-154. DOI: 10.1023

[8] Shashidhar G. Koolagudi, K. Sreenivasa Rao (2011) Recognition of Emotion from Speech: A Review. International Journal of Speech Technology. Volume 15, Issue 2 , pages 99-117 . 2011 DOI 10.5772/39246.

[9] Vinay Bettadapura (2009) Face Expression Recognition and Analysis: The State of the Art. Emotion, Tech Report, arXiv:1203.6722

[10] P. L. Bartlett et al., Application of support vector machines to the classification of proteinchip system mass spectral data of prostate cancer serum samples (abstract). In Second Annual National Cancer Institute Early Detection Research Network Scientific Workshop.

[11] P. Ekman and W. Friesen (1978), Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press

[12] Interaction between real and virtual humans during walking.

[13] I. A. Essa and A. Pentland (1997) Coding, Analysis, Interpretation, and Recognition of Facial Expressions Trans. Pattern Analysis and Machine Intelligence, IEEE Computer Society, 19(7):757-763, 1997.

[14] S.M. Bartlett, P.A. Viola, T.J. Sejnowski, B.A. Golomb, J. Larsen, J.C. Hager, and P. Ekman (1999). Classifying Facial Action. IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 21 Issue 10, Page 974-989. doi 10.1109/34.799905

[15] Padgett, C. & Cottrell, G. W. (1996), Representing Face Images for Emotion Classification., in Michael Mozer; Michael I. Jordan & Thomas Petsche, ed., 'NIPS' , MIT Press, , pp. 894-900.

[16] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, J. Movellan (2005) Recognizing facial expression: machine learning and application to spontaneous behavior, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2005.297.

[17] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman (1996) Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. Computer Vision — ECCV '96, pages 43-58. DOI 10.1007/BFb0015522.

[18] Caifeng Shana, Shaogang Gong , Peter W. McOwanb (2009) Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing, volume 27, pages 803–816. DOI: 10.1016/j.imavis.2008.08.005

[19] G. Donato, M. Bartlett, J. Hager, P. Ekman, T. Sejnowski (1999) Classifying facial actions. , IEEE Transactions on Pattern Analysis and Machine Intelligence volume DOI 10.1109/34.799905.

[20] Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., et al. (2003). Facial expression recognition across the adult life span. *Neuropsychologia*, 41, 195–202.

[21] Petar S. Aleksic (2006) Automatic Facial Expression Recognition Using Facial Animation Parameters and Multistream HMMs. *Information Forensics and Security IEEE* volume 1, pages 3-11. DOI 10.1109/TIFS.2005.863510

[22] N. Tsapatsoulis, A. Raousaiou, S. Kollias, R. Cowie, E. Douglas-Cowie Venue (2002) MPEG-4 Facial Animation - The standard, implementations, applications. *Emotion Recognition and Synthesis Based on MPEG-4 FAPs* Pages 143-169. DOI: 10.1002/0470854626.ch9

[23] J. L. Landabaso, M. Pardàs, and A. Bonafonte (2003) HMM recognition of expressions in unrestrained video intervals. *Acoustics, Speech, and Signal Processing*, 2003. Volumen 3, pages 197-200. DOI 10.1109/ICASSP.2003.1199141

[24] M. Pardàs, A. Bonafonte, and J. L. Landabaso (2002) Emotion recognition based on MPEG-4 facial animation parameters. In *Proc. ICASSP*, vol. 4, pp. 3624–3627. DOI 10.1109/ICASSP.2002.5745440

[25] M. Pardàs and A. Bonafonte (2002) Facial animation parameters extraction and expression detection using HMM,” in *Signal Process: Image Communication Journal*, 2002, vol. 17, 2002, pp. 675–688.

[26] Yongmian Zhang (2008) Dynamic Facial Expression Analysis and Synthesis with MPEG-4 Facial Animation Parameters. *Circuits and Systems for Video*

Technology, volumen 18, pages 1383-1396. DOI 10.1109/TCSVT.2008.928887

[27] Akshat Garg¹, Vishakha Choudhary (2012) Facial Expression Recognition Using Principal Component Analysis. International Journal of Soft Computing and Engineering (IJSCE), Volume-3, Issue-4. ISSN: 2231-2307.

[28] Y. Zhang, Q. Ji (2005) Active and dynamic information fusion for facial expression understanding from image sequences. Pattern Analysis and Machine Intelligence, IEEE, volume 27, issue 5. DOI 10.1109/TPAMI.2005.93

[29] Beat Fasel, Jürgen Lüttin(1999) Recognition of Asymmetric Facial Action Unit Activities and Intensities. published in ICPR 2000, pages 1100-1103.

[30] Chengjun Liu, Wechsler, H. (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. Image Processing vol.1 no4 pp. 467-476. DOI:10.1109/TIP.2002.999679

[31]. Gwen Littlewort, Marian Stewart Bartlett, Ian Fasel, Joshua Susskind, Javier Movellan (2005) Dynamics of facial expression extracted automatically from video. Image and Vision Computing, pages 615–625. DOI:10.1109/CVPR.2004.53.

[32] Philipp Michel (2003) Real Time Facial Expression Recognition in Video using Support Vector Machines. Proceeding ICMI '03 Proceedings of the 5th international conference on Multimodal interfaces Pages 258-264. DOI:10.1145/958432.958479

[33] Anissa Bouzalmat, Jamal Kharroubi, Arsalane Zarghili (2013)Face

Recognition Using SVM Based on LDA. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 4. ISSN: 1694-0814.

[34] Gwen Littlewort, Ian Fasel, Marian Stewart Bartlett, Javier R. Movellan (2002) Fully Automatic Coding of Basic Expressions from Video.

[35] Hong, K, Chalup, S.K., King, R.A.R. (2010) A Component Based Approach Improves Classification of Discrete Facial Expressions Over a Holistic Approach. On Neural Networks (IJCNN), The 2010 International Joint Conference, pages 1-8. ISSN :1098-7576

[36] Zahid Riaz, Christoph Mayer, Michael Beetz and Bernd Radig (2009) Facial Expressions Recognition from Image Sequences. In 2nd International Conference on Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions. Springer.

[37] Hong-Bo Deng, Lian-Wen Jin, Li-Xin Zhen, Jian-Cheng Huang (2005). A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA, in International Journal of Information Technology Vol. 11 No. 11

[38] Stan Z. Li, Anil K. Jain (2011) Handbook of Face Recognition. ISBN: 978-0857299314

[39] BEN-YACOUB S., ABDELJAOUED, Y., MAYORAZ, E. (1999). Fusion of Face and Speech Data for Person Identity Verification. IEEE Transactions on Neural Networks, 10 (5):1065-1074.

[40] Peter Eisert, Thomas Wiegand(2000) Model-Aided Coding: A New Approach to Incorporate Facial Animation into Motion-Compensated Video Coding. IEEE Transactions on circuits and systems for video technology Vol 10, NO. 3

[41] Mayank Bhargava, Tim Polzehl.(2013) Improving Automatic Emotion Recognition from speech using Rhythm and Temporal feature. CoRR abs/1303.1761

[42] Patrick Lucey , Jeffrey F. Cohn , Takeo Kanade , Jason Saragih , Zara Ambadar , Iain Matthews . The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. DOI:10.1.1.182.3759

[43] Ying-li Tian (2001) Recognizing Action Units for Facial Expression Analysis. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 23, NO. 2, Febrero 2001

[43] Enrique Cabello Pardos (2013). Presentación ciia 2014. Representacion de información.

[44] E.R. Gast (2010). A Framework for Real.Time Face and Facial Feature Tracking using Optical Flow Pre-estimation and Template Tracking. Tesis de maestria. Universidad de Leiden