



Instituto Tecnológico
de Buenos Aires

Instituto Tecnológico de Buenos Aires - ITBA
Escuela de Ingeniería y Gestión

Detección de Emociones Faciales

Autores

Pérez Sammartino, Francisco. 57256
Di Giovanni Martínez, Andrés. 57274

Tutora

Dra. Gambini, María Juliana.

TRABAJO FINAL PRESENTADO PARA LA OBTENCIÓN DEL TÍTULO DE
INGENIERO EN INFORMÁTICA

Buenos Aires
7 de Diciembre de 2020

Índice

Resumen	3
Introducción	3
Estado del Arte	4
Métodos existentes	4
SVM sobre características faciales	6
<i>Extreme Sparse Learning</i> (secuencia de imágenes)	7
Modelo de Apariencia Activa con Redes Dinámicas Bayesianas	7
<i>Facial Emotion Recognition in Videos using HOG and LBP</i>	9
<i>Algoritmos en tiempo real para el reconocimiento de emociones faciales</i> .	10
<i>Emotion Recognition using Dynamic Grid-based HOG Features</i>	11
Descripción de los Métodos Utilizados	12
Preprocesamiento de imágenes	13
Detección del Área de Interés - HOG	14
Modificación a la Implementación de Dlib	14
Alineación Planar	15
Redimensionamiento y Filtro Sobel	17
Descripción de Características - HOG	18
Clasificación de la imagen - SVM	19
Búsqueda de Hiperparámetros	21
Análisis de Videos	22
Múltiples Caras	23
Mejoras Propuestas	25
Bases de datos Utilizadas	26
KDEF	26
BAUM1	26
Resultados	27
Pruebas preliminares	27
Resultados entre pares de emociones	27
Resultados en conjuntos de 3 emociones	28
Resultados en conjuntos de 4 emociones	29
Resultados en conjuntos de 5 emociones	30
Resultados en conjuntos de 6 emociones	31
Resultados en conjunto de 7 emociones	32
Clasificación de videos	33

Análisis de ruido	33
Conclusiones y Trabajos Futuros	36

Detección de Emociones Faciales

Andrés Di Giovanni Martínez, Francisco Pérez Sammartino

Tutor: Dra. María Juliana Gambini

Instituto Tecnológico Buenos Aires (ITBA)

Resumen

Existe una gran variedad de aplicaciones para la detección de emociones faciales. Entre ellas se encuentran las utilizadas para obtener la satisfacción del cliente en un banco, o el nivel de fatiga de un conductor de un vehículo. En algunas es necesario que la detección se realice con eficiencia en tiempo mientras que en otras la precisión en la detección es más importante. En el caso de que sea necesario un procesamiento en tiempo real, la rapidez de los resultados será prioritaria, mientras que en el análisis 'off-line', la precisión es lo más importante. En este trabajo, se realiza una exposición de los métodos existentes para detección de emociones, utilizados actualmente con diferentes objetivos y enfoques. A partir de ello, se presentan las mejoras propuestas, y el desarrollo de un método que permite el reconocimiento de las emociones de las personas en videos, o en tiempo real. La aplicación utiliza máquinas de vectores de soporte (SVM) para realizar la clasificación de las emociones, técnicas de rotación planar para corregir la pose, y sistema de seguimiento de caras para poder aplicarlo en presencia de múltiples personas. Dado que el medio por el cual se transmiten los videos o imágenes, como el WiFi, puede contaminarlos con ruido, es importante que el sistema sea resiliente a dichas alteraciones. Para evaluar esta capacidad se realiza un estudio sobre el impacto del grado de contaminación de la información recibida, sobre el resultado. Se alcanza una precisión muy adecuada al distinguir, en imágenes, las emociones de miedo, alegría, disgusto, neutralidad, tristeza, sorpresa, y enojo.

Introducción

El conocimiento del estado emocional de una persona posee múltiples aplicaciones, como por ejemplo:

- **En el área de atención al cliente, sirve para derivar al cliente con un empleado capacitado para atenderlo en determinado estado de emoción.**

- En el área de la educación a distancia [5]: un caso de uso sería, utilizando las cámaras en las computadoras de los alumnos, la monitorización de la atención de los alumnos, o evaluar el nivel de cansancio.
- Detección del dolor en telemedicina.
- En el área de robótica social: utilizar un robot para el área de atención al cliente, quien tiene la primera interacción con el cliente, y según el objetivo de éste y las emociones detectadas por el robot, tomar determinada acción; por ejemplo, derivar al cliente a algún área en particular.
- Detección de fatiga en un operador (Ej. conductor automovilístico).
- Estudio de Marketing para detectar la reacción de una persona al probar un nuevo producto
- Detección de Mentiras

El objetivo del proyecto es la detección de emociones en personas a través del análisis facial. Incluyendo la detección de múltiples caras de forma simultánea. Para la captura de imágenes faciales se dispone de una cámara frontal de una computadora Macbook Pro.

Estado del Arte

Métodos existentes

La gran mayoría de estudios y publicaciones sobre reconocimiento de emociones faciales se basan en el estudio del Dr. Ekman's, donde concluye que existen 6 emociones universales que son: alegría, tristeza, miedo, disgusto, enojo, sorpresa [8]. Todas las demás se forman como combinación de esas principales.

El Dr. Ekman, junto con el Dr. Friesen desarrollan F.A.C.S. (Facial Action Coding System)[10, 9], uno de los estudios más destacados y conocidos del área. En este estudio se explicita un sistema para detectar sutiles cambios de la apariencia facial causada por las diferentes contracciones de los músculos.

Los métodos de detección automática de emociones constan de 3 pasos principales. Realizamos una breve descripción de los mismos, para más detalles ver [16, 3]. Estos son:

1. Preprocesamiento de las imágenes:
 - a) Corrección Gamma: Se aplica una función de corrección a cada canal. La corrección gamma es una operación realizada sobre los píxeles de una imagen, permite corregir las diferencias entre la forma en que la cámara captura el contenido y la forma en que éste se muestra. Este método tiene un parámetro denominado gamma; valores menores a uno están asociados a oscurecimiento

de la imagen, mientras que valores superiores a 1 aumentan el brillo de la imagen. En la figura 1 se puede observar el efecto de la corrección.

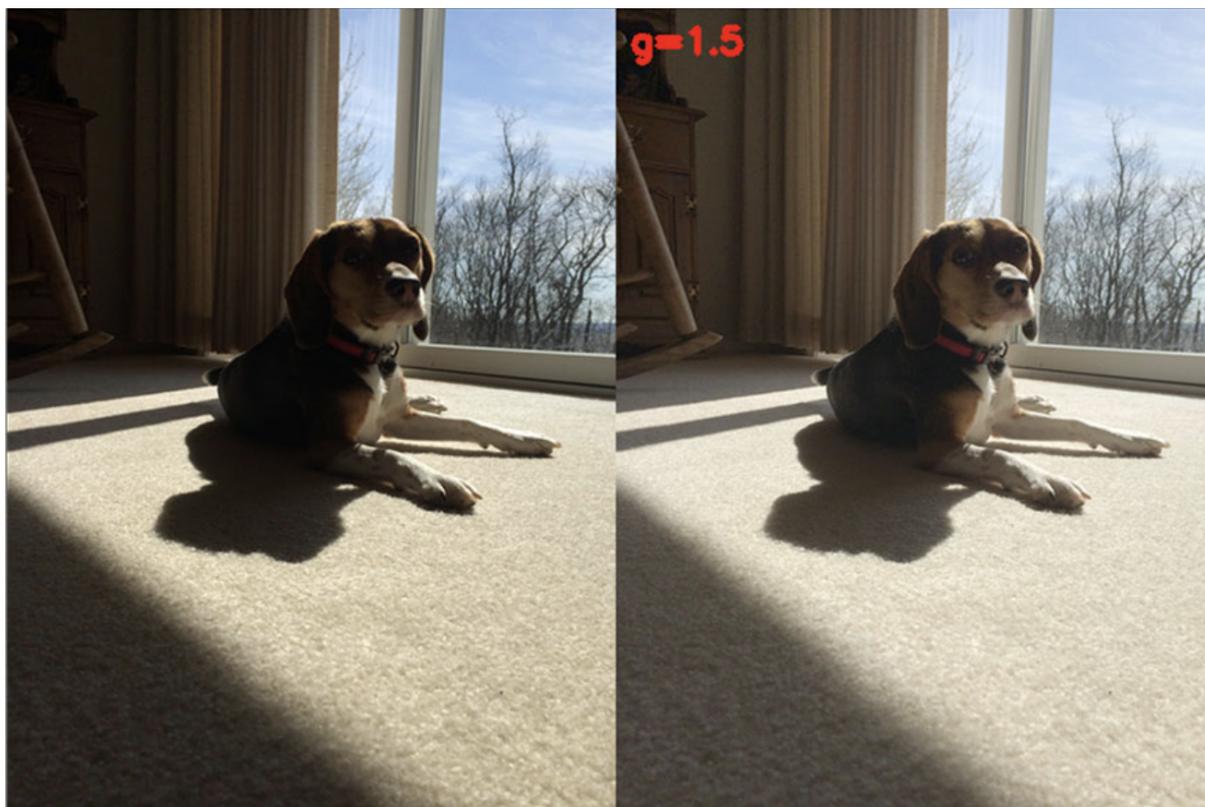


Figura 1: Corrección gamma aplicada a una imagen con valor de gamma igual a 1.5. Izquierda: imagen original, derecha: imagen con la corrección.

2. Detección facial en la escena, es decir, reconocer dónde se encuentra la cara que se desea analizar en la imagen. Existen 2 tipos de enfoques para esta tarea:
 - a) Enfoque Holístico: tiene en cuenta la cara como un todo. Un ejemplo es el uso de las Eigenfaces [21].
 - b) Enfoque basado en Características: intenta extraer características específicas de la cara.

El enfoque holístico tiene ventajas asociadas con la identificación de características prominentes en la imagen, como también con la habilidad de automáticamente encontrar características. Sin embargo, su desempeño se encuentra afectado significativamente por variaciones en iluminación, orientación o escala; además, puede encontrar características que no formen parte de la expresión facial, sino del fondo. El enfoque basado en características incluye entre sus ventajas la selección de las características faciales a evaluar, de esta forma logra mayor independencia de la iluminación y orientación. Las desventajas se asocian con la definición de las características evaluadas. Para más detalles ver [20].

3. Extraer información relacionada con la expresión facial y las emociones. En este paso está incluido la detección de la denominada ROI (*region of interest*).
4. Categorización de emociones.

Uno de los desafíos más grandes del área es poder, limitándose únicamente al análisis facial, reconocer emociones que en circunstancias normales se reconocen en conjunto con gestos en el resto del cuerpo.

A continuación se mencionan algunos de los métodos existentes en la resolución del problema mencionado.

SVM sobre características faciales

Un método para el reconocimiento de emociones faciales se logra utilizando el algoritmo de Support Vector Machine (SVM). Este algoritmo se aplica sobre un subconjunto de las características faciales que se consideren importantes o clasificadoras. Un proyecto que utiliza este tipo de algoritmos se presenta en [11].

La tasa de reconocimiento del proyecto mencionado anteriormente, donde se entrena el sistema con 265 imágenes de 42 personas, fue de 84.55% clasificando en 7 categorías (felicidad, sorpresa, enojo, tristeza, disgusto, miedo y neutral).

En el trabajo de V. M. Alvarez, et al. [6] se realiza detección de la emoción, comparando los puntos de interés de la imagen en cuestión, con los de la imagen de la misma persona en estado neutral.

El proceso de clasificación se puede ver en la figura 25. El proceso se divide en 3 partes:

1. Landmark Detector: es el proceso mediante el cual se distinguen los puntos claves de la imagen y se obtienen los puntos de interés.
2. Comparison Algorithm: se compara la ubicación de los puntos de interés expresando emoción con los de la misma persona en estado neutral
3. Classifier: en base a las diferencias encontradas en el paso anterior, se utiliza una máquina de soporte de vectores (SVM), para realizar la clasificación. Para más detalles sobre SVM ir a la sección *Descripción de características*.

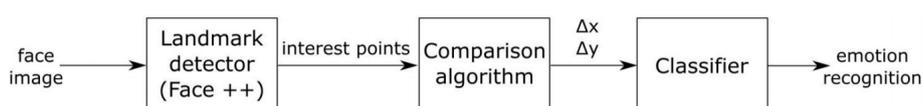


Figura 2: Diagrama de reconocimiento facial de emociones.

Extreme Sparse Learning (secuencia de imágenes)

El artículo [19] está enfocado en el reconocimiento de las emociones en situaciones más realistas, es decir con ruido en la imagen, datos e imágenes imperfectas, etc. El verdadero reto del reconocimiento de emociones en el mundo real es la correcta extracción de características faciales, y que éstos no varíen por la pose, iluminación, obstrucción o cualquier otro factor que no tenga relación con la clasificación. En otros términos, que no exista variación dentro de cada una de las clases de emociones, pero sí entre las diferentes clases para los fines de clasificación.

En este proyecto se reconocen las emociones de personas en movimiento, que presenta un reto a la hora de extracción de las características y el seguimiento y movimiento de las mismas, pero una vez superado ese reto proporciona más información para la clasificación. Además el reconocimiento de emociones en movimiento se acerca más a la realidad que las fotos estáticas.

Algunas técnicas mencionadas para tener en cuenta el factor temporal en el análisis son: Modelos de Markov Ocultos, y Redes Bayesianas Dinámicas, entre otras.

En este artículo los autores comparan el método propuesto con otros 3, obteniendo resultados mejores: del 92% para ECK+ Database [14], 61.8% para AVEC DB [17] y y 21.8% para EmotiW DB [7]; estos dos últimos resultados son notablemente inferiores al primero y se deben a las características de sendas bases, que almacenan imágenes realistas y no de laboratorio. Se detectan algunas fuentes de errores que son:

- Fallas en la detección de la cara.
- Si los movimientos de cabeza son muy amplios se dificulta el seguimiento de los puntos de interés.

Modelo de Apariencia Activa con Redes Dinámicas Bayesianas

El artículo [13] está orientado a la detección de emociones en una secuencia de imágenes. Está basado en el modelo de apariencia activa y en F.A.C.S. (*Facial Action Coding System*).

Para detectar la posición de la cara, se usan algunos puntos claves como puntos de referencia. Algunos de estos puntos pueden ser la ceja o la esquina de la boca.

En la figura 3 se muestra un diagrama de la secuencia de tareas que involucra el procesamiento para el método propuesto. El procesamiento de reconocimiento consta de 3 etapas principales que son la de extracción de características faciales utilizando el modelo de apariencia activa, análisis de las características aplicando el modelo del Dr. Ekman F.A.C.S. y finalmente la clasificación utilizando redes bayesianas. Para más detalles ver [13].

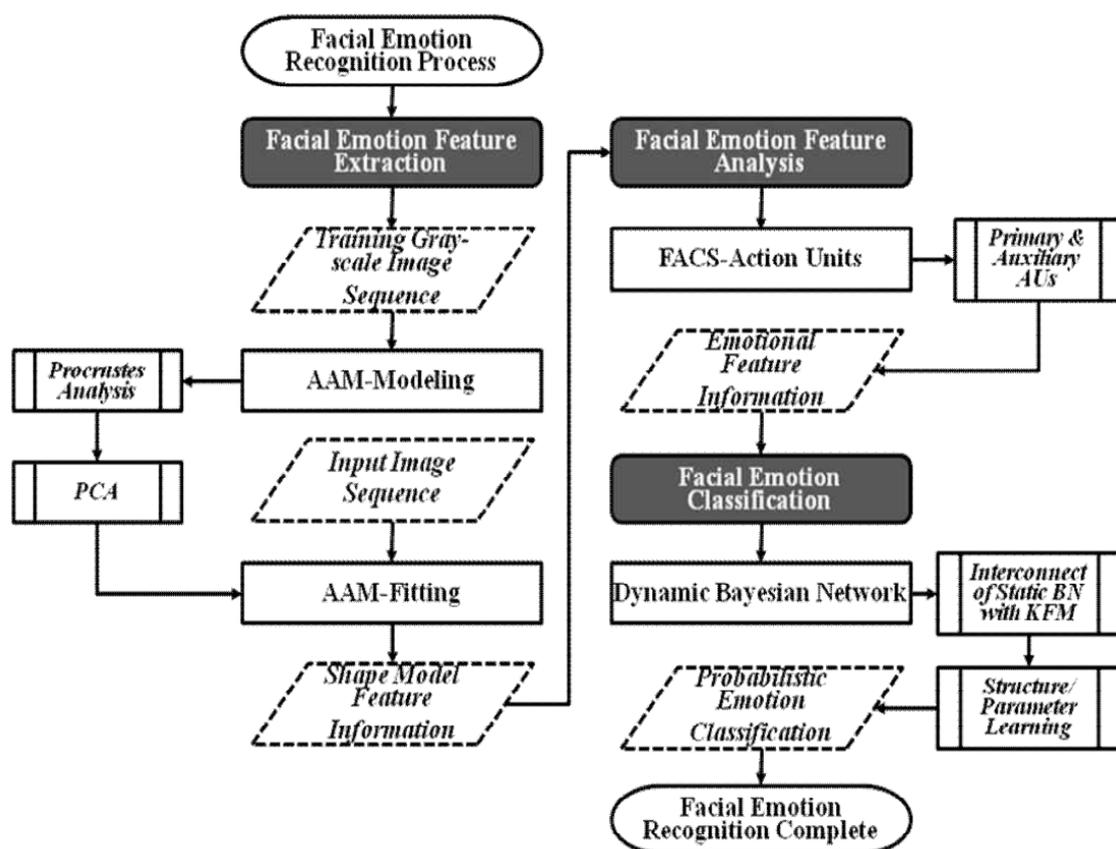


Figura 3: Proceso de clasificación de emociones faciales utilizando un modelo de apariencia activa para la extracción de características, y redes Bayesianas dinámicas para la clasificación.

En este trabajo se utiliza la secuencia de imágenes que se muestra en la figura 4.

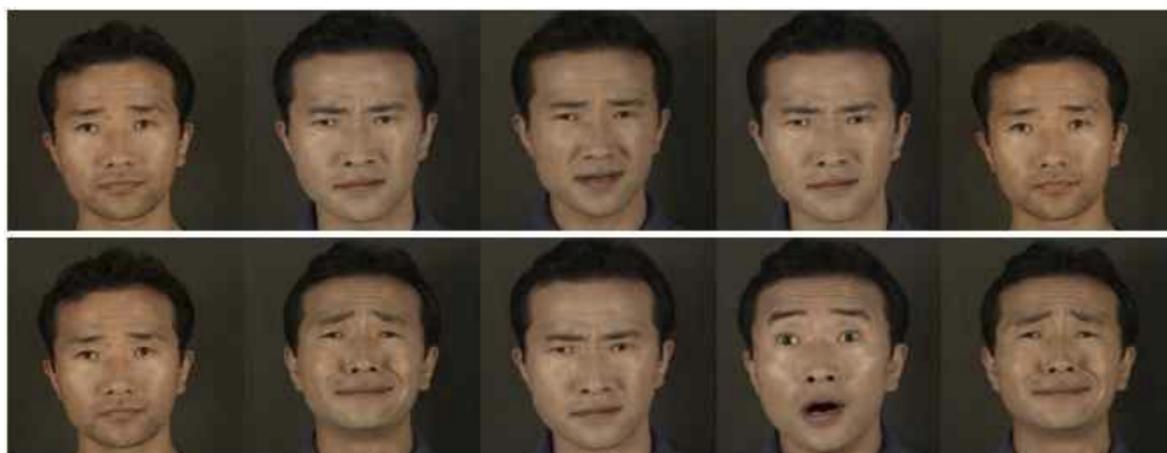


Figura 4: Secuencia de 10 imágenes de una persona mostrando diferentes emociones.

Los resultados se muestran en la tabla 1. En cada fila se puede ver la probabilidad de que la emoción de la columna se esté expresando. Es decir para el frame 1 hay un 12 % de probabilidades de que exprese felicidad, y uno 85 % de que exprese estado normal.

Frame	Happiness	Sadness	Surprise	Anger	Normal
1	0.12	0.62	0.12	0.12	0.85
2	0.24	0.51	0.21	0.21	0.78
3	0.26	0.33	0.34	0.52	0.32
4	0.24	0.51	0.51	0.96	0.71
5	0.12	0.62	0.52	0.75	0.96
6	0.12	0.61	0.43	0.74	0.95
7	0.21	0.91	0.95	0.65	0.23
8	0.23	0.61	0.94	0.52	0.72
9	0.42	0.21	0.72	0.42	0.12
10	0.11	0.92	0.27	0.12	0.32

Tabla 1: Matriz de confusión con probabilidades para cada emoción.

Facial Emotion Recognition in Videos using HOG and LBP

El proyecto [3] busca clasificar en tiempo real las emociones en 3 categorías: enojo, felicidad, tristeza.

En la figura 5 se muestra la arquitectura del sistema.

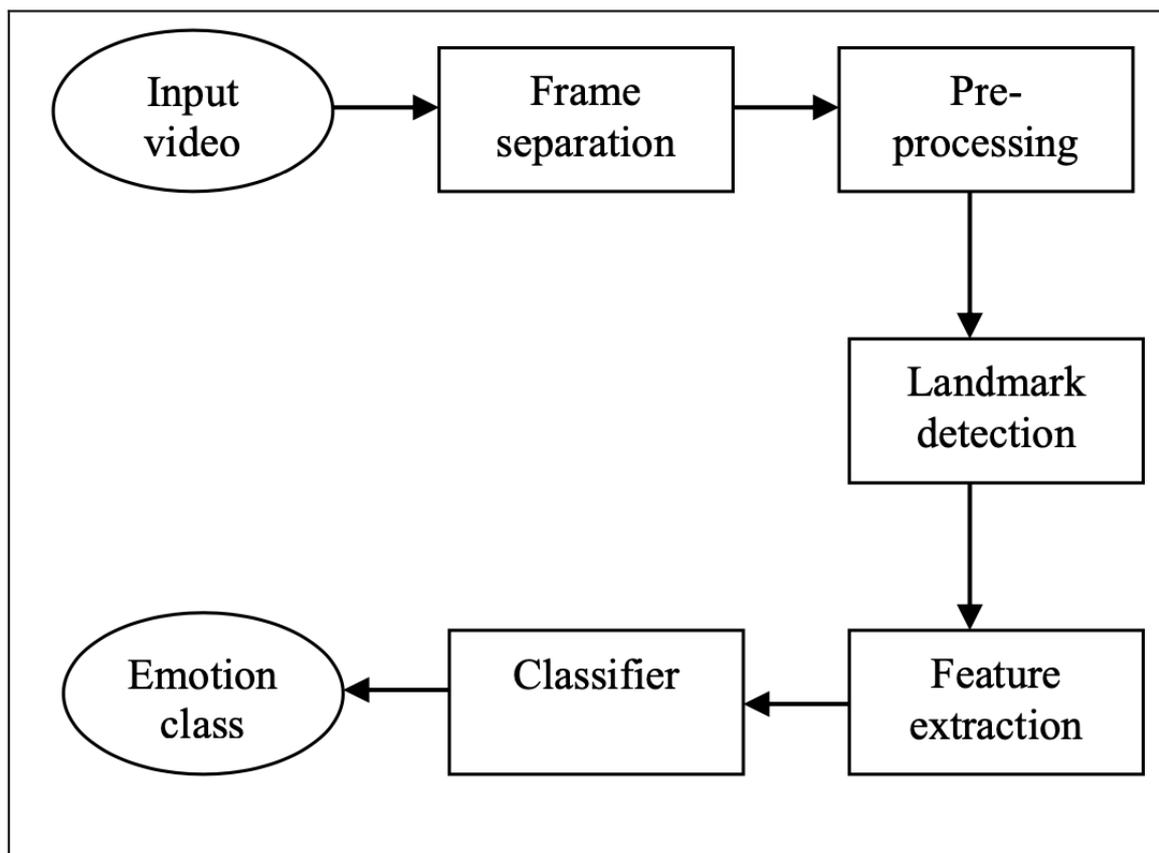


Figura 5: Se puede ver la arquitectura para la clasificación de emociones sobre un video en tiempo real.

Para la detección facial se utiliza el algoritmo de Viola-Jones [2].

Para el reconocimiento se utilizaron 2 técnicas de detección de características de imagen:

- HOG[5]: histograma de gradientes orientados.
- LBP[1]: Patrón Local Binario.

Se obtuvieron mejores resultados utilizando HOG.

Se propone la combinación con cámaras térmicas infrarrojas para obtener mejores resultados, ya que estas no están influenciadas por las condiciones lumínicas.

Algoritmos en tiempo real para el reconocimiento de emociones faciales

El objetivo del proyecto [12] es comparar diferentes métodos para reconocer, en tiempo real y a partir de la cara, 4 emociones: felicidad, tristeza, enojo y miedo.

Se comparan los resultados de 5 métodos alternativos:

- Basados en CNN:

- AlexNet CNN.
- Affdex CNN.
- FER-CNN.
- Basados en HOG:
 - SVM de características HOG.
 - MLP de características HOG.

Los resultados se muestran en la tabla 2 .

<i>Facial emotion recognition algorithm</i>	<i>Total accuracy [%]</i>
Affdex CNN	85.05
Fine-tuned AlexNet CNN	76.64
SVM classification of HOG features	63.55
MLP classification of HOG features	56.07
FER-CNN	55.14

Tabla 2: Resultados de precisión para los diferentes métodos de clasificación.

Emotion Recognition using Dynamic Grid-based HOG Features

En el artículo [4] se comparan los resultados de detección de emociones faciales basándose en LBP y en HOG. Para realizar la comparación se ven los resultados de ambos para un conjunto de videos.

Para la detección facial y de ojos se utiliza Open CV, tanto para el caso de LBP como para HOG.

En el caso de HOG para la extracción de características, se reduce la imagen a una porción de interés, ROI. En las figuras 6 y 7, se puede ver la secuencia de imágenes mostrando el proceso de selección de la región de interés y la definición de las celdas de la matriz y luego el cálculo de gradientes en cada celda. Se pueden ver 4 etapas importantes que forman parte del proceso de reconocimiento y clasificación cuando se utiliza HOG, que son:

1. Reconocimiento de la cara en la imagen y detección de puntos “ancla” , como pueden ser los ojos.
2. A partir de las relaciones entre las distancias de los diferentes puntos “ancla” de la cara, se establece la región de interés (*ROI*) y se subdivide en diferentes celdas.

3. Cálculo los gradientes para cada celda.

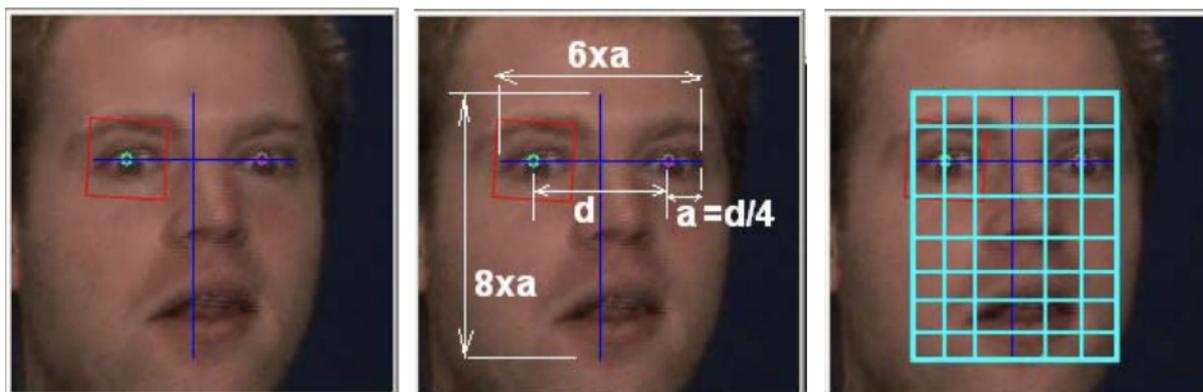


Figura 6: Proceso de detección facial y definición de región de interés.

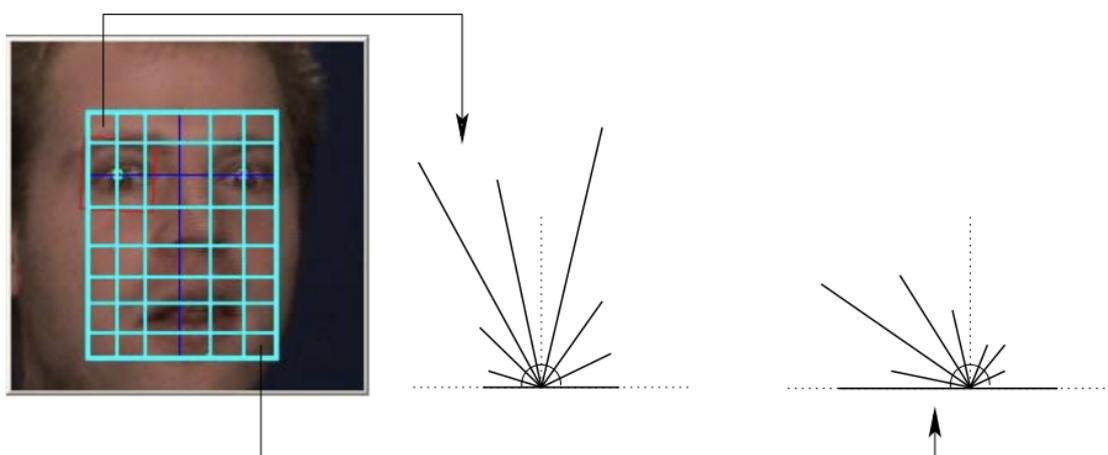


Figura 7: Cálculo de gradientes para cada celda.

Descripción de los Métodos Utilizados

Como fue mencionado en los métodos existentes, el proceso de reconocimiento de emociones consta de una serie de pasos, a continuación se enumeran los métodos utilizados en cada uno de ellos:

1. Preprocesamiento de imágenes: en primer lugar las imágenes pasan por una serie de procesos para mejorar su posterior clasificación, por ejemplo mejora del contraste, corrección de condiciones de la cámara o la eliminación del ruido.
2. Detección del área de interés: para encontrar el ROI (área de interés), se utiliza el método de histograma de gradientes orientados (HOG).
3. Descripción de características: para describir las características relevantes de la imagen, se utiliza el método HOG de histogramas de gradientes orientados.

4. Clasificación de la imagen: en ésta etapa será utilizado el algoritmo de aprendizaje automático SVM.

Para evaluar la efectividad de los diferentes algoritmos aplicados en las diferentes etapas del procesamiento y clasificación, se utiliza el parámetro *score* o precisión. Este valor es calculado como el porcentaje de muestras correctamente clasificadas, sobre el total de las mismas.

Preprocesamiento de imágenes

El preprocesamiento consta de los siguientes pasos:

1. Cambio de Escala de Colores: la imagen es convertida a la escala de grises. Esto aumenta la independencia con la iluminación y al mismo tiempo reduce los datos para el entrenamiento.
2. Normalización: la imagen es normalizada. Esta operación reduce el rango de valores de cada píxel a un valor entre 0 y 1. Esto acelera la convergencia en la etapa de entrenamiento del modelo.

En la figura 8(b) se puede ver el resultado al aplicar las operaciones mencionadas sobre la figura 8(a).



(a) Imagen original.



(b) Imagen preprocesada.

Figura 8: Preprocesamiento de imagen.

Detección del Área de Interés - HOG

El algoritmo de detección facial utilizado en el proyecto está basado en el Histograma de Gradientes Orientados (HOG) [6], combinado con un clasificador lineal. Este método es general y capaz de detectar muchos tipos de objetos semi-rígidos además de las caras humanas.

Modificación a la Implementación de Dlib

La función Dlib es la encargada de la detección del área de interés utilizando gradientes orientados. Retorna como resultado el conjunto de rectángulos que contienen a cada una de las caras presentes en la imagen.

Como el objetivo del proyecto es la detección de emociones, es necesario obtener aquellas secciones de la cara asociadas con la expresión, como lo es la sección de las cejas. Para ello, es modificada la relación entre el alto y el ancho de los rectángulos resultantes. La proporción fue modificada en un 5% respecto a la original.

En la figura 9 se encuentra dibujado el rectángulo resultante del área de interés sobre la imagen. El número ubicado sobre el rectángulo hace referencia a la numeración de las caras en caso de que se encuentre más de una.



Figura 9: Área de Interés sobre imagen facial.

Alineación Planar

Una vez detectados los cuadrantes contenedores de caras, en caso de que exista una rotación planar en las mismas, se corrige para que quede alineada con los ejes cartesianos. Para definir el ángulo de rotación necesario, son tomados como puntos de referencia las líneas que forman los ojos y los labios.

En la figura 10 se muestra el proceso de cálculo del ángulo de rotación en el plano en base a la línea de los ojos. En el método de Dlib, dicha medición es corregida utilizando la posición de los labios.

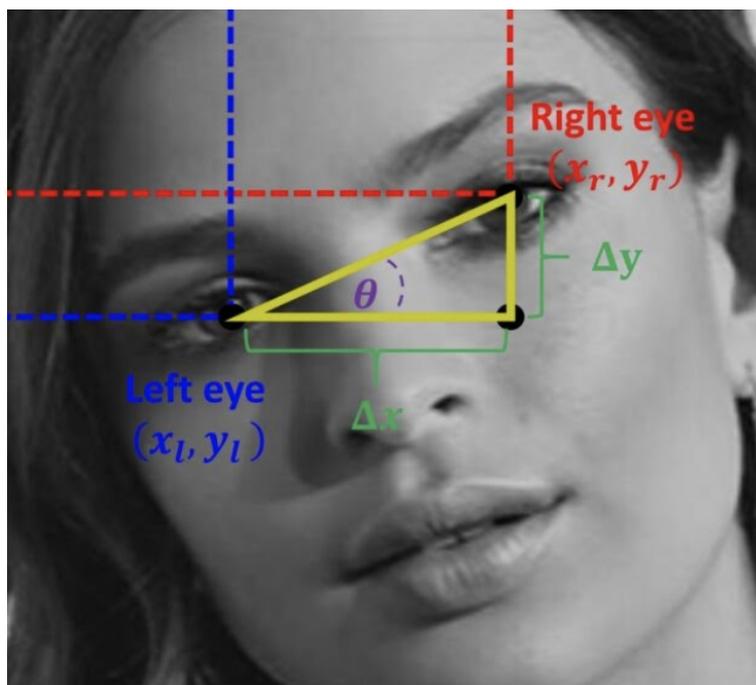


Figura 10: Cálculo del ángulo de la rotación planar en base a los ojos.

En la figura 11(b) se puede ver el resultado al aplicar las operaciones mencionadas sobre la imagen preprocesada, figura 11(a).



(a) Imagen preprocesada.



(b) Imagen alineada y redimensionada.

Figura 11: Rotación planar y redimensionamiento.

Redimensionamiento y Filtro Sobel

Al encontrar la región de interés, la imagen se recorta a esa sección. La imagen, a continuación, es redimensionada y se le aplica el filtro Sobel del paquete opencv que resalta los bordes de la imagen. Esto permite lograr una mayor independencia con respecto a la iluminación, y a la vez resaltar las características relevantes asociadas a la expresión de las emociones faciales. La detección de bordes reduce significativamente la cantidad de información, filtrando la información menos relevante, mientras que preserva las propiedades estructurales [18].

En la figura 12(b) se puede ver el resultado al aplicar el filtro sobel sobre la imagen alineada y preprocesada en la figura 12(a)



(a) Imagen alineada y redimensionada.

(b) Filtro sobel sobre figura 12(a)

Figura 12: Rotación planar y redimensionamiento.

Para encontrar el tamaño ideal de la imagen se realizan pruebas con las diferentes dimensiones que se pueden ver en la tabla 3.

Para cada dimensión es calculado el score sobre el conjunto de evaluación que se puede ver en la columna Score. El conjunto de evaluación representa el 30% del conjunto total. Para cada dimensión son realizadas 5 pruebas y reportado el valor medio de las mismas.

Dimensión de Imagen	Score
50	0.7929
60	0.8119
70	0.7833
80	0.8190

Dimensión de Imagen	Score
90	0.85
100	0.85
110	0.8595
120	0.8405
130	0.8714
140	0.8476
150	0.8619
160	0.8476
170	0.8405
180	0.8405
190	0.8357
200	0.8238

Tabla 3: Resultados de clasificación variando la dimensión de la imagen cuadrada. Fueron utilizadas 80 imágenes como conjunto de entrenamiento y 60 para la evaluación

Descripción de Características - HOG

Una vez obtenida la región de interés en la imagen, es necesario reducir el conjunto total de píxeles que conforma la imagen a un conjunto menor que siga siendo representativo para la posterior clasificación. Es necesario que la función a aplicar mantenga las características relevantes para el objetivo de la detección. El reducir la información permite no solo conservar los rasgos asociados al fin clasificatorio sino descartar aquellos que dificulten la tarea.

HOG es un método que se basa en el cálculo de histogramas a partir de gradientes orientados. Los gradientes dejan en evidencia los bordes encontrados en la imagen, que a su vez son representativos para las emociones faciales.

En la figura 7 se puede ver el cálculo de gradientes para un recuadro determinado en la imagen.

El histograma de gradientes orientados calcula en primer lugar los gradientes horizontal y vertical para cada píxel de la imagen, al compararlos con los píxeles vecinos.

A partir de estos gradientes verticales y horizontales se calculan el módulo y dirección resultante por cada píxel utilizando trigonometría.

Obtenidos los gradientes con sus módulos y respectivas orientaciones, se divide la imagen en celdas y para cada celda se suman los gradientes según la orientación formando así el histograma para cada celda. En la figura 13 se puede ver el cálculo del histograma para una celda en particular a partir del módulo y orientación de los gradientes en cada píxel

contenido. Para cada píxel de la celda se suma su valor en módulo en el compartimiento según la orientación. En caso de que la orientación no coincida con ninguno de los compartimientos se distribuye de forma ponderada según la cercanía con cada uno de ellos. Por ejemplo, en caso de que sea equidistante se asigna el 50 % de su módulo a ambos compartimientos.

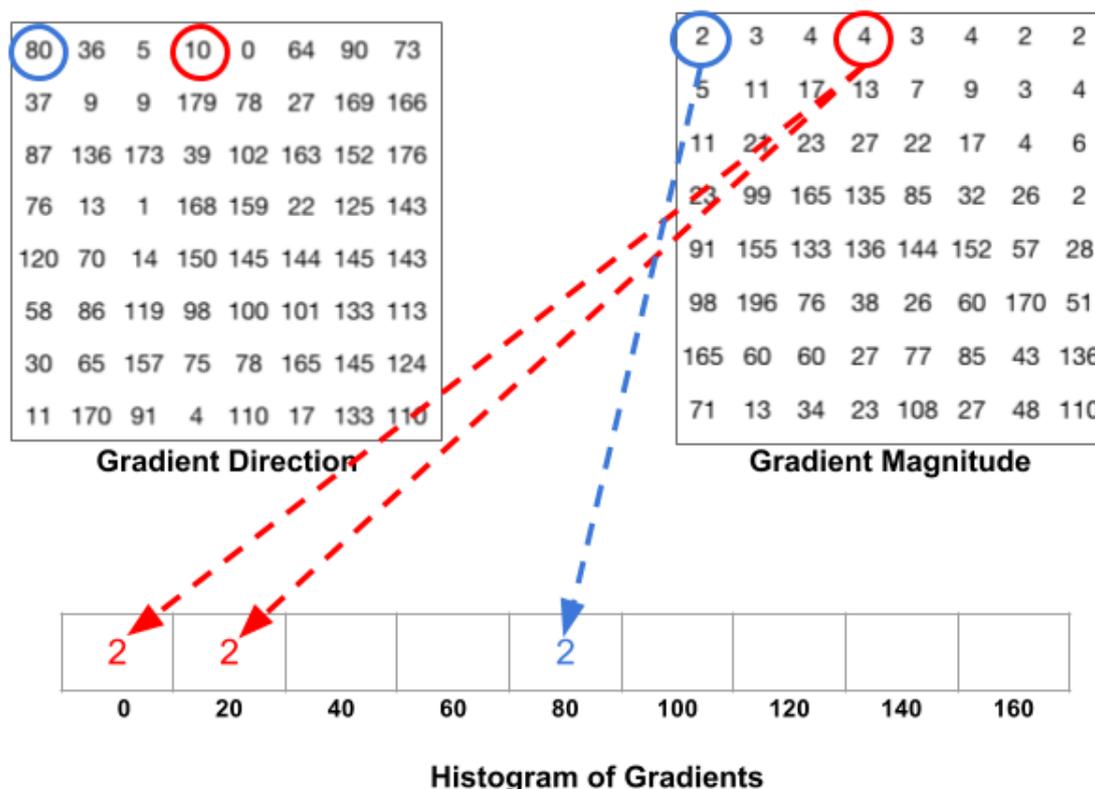


Figura 13: Cálculo del histograma de gradientes orientados para una celda.

El vector de características para la totalidad de la imagen se obtiene concatenando los histogramas de cada celda.

Fue aplicado utilizando el paquete skimage.

Clasificación de la imagen - SVM

El método utilizado para la clasificación de las emociones es Máquinas de Vectores de Soporte (SVM). Es un conjunto de algoritmos de aprendizaje supervisado que fue desarrollado por Vladimir Vapnik. Sirve tanto para clasificación como para regresión; en este proyecto es utilizado para la clasificación.

Un SVM construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad alta para luego poder clasificar muestras desconocidas. Dado un conjunto de muestras de N diferentes clases, la máquina busca el hiperplano óptimo para separar las

muestras con el mayor margen posible; define como margen la distancia mínima entre las muestras y el hiperplano. En la figura 14 se puede ver el funcionamiento de la máquina.

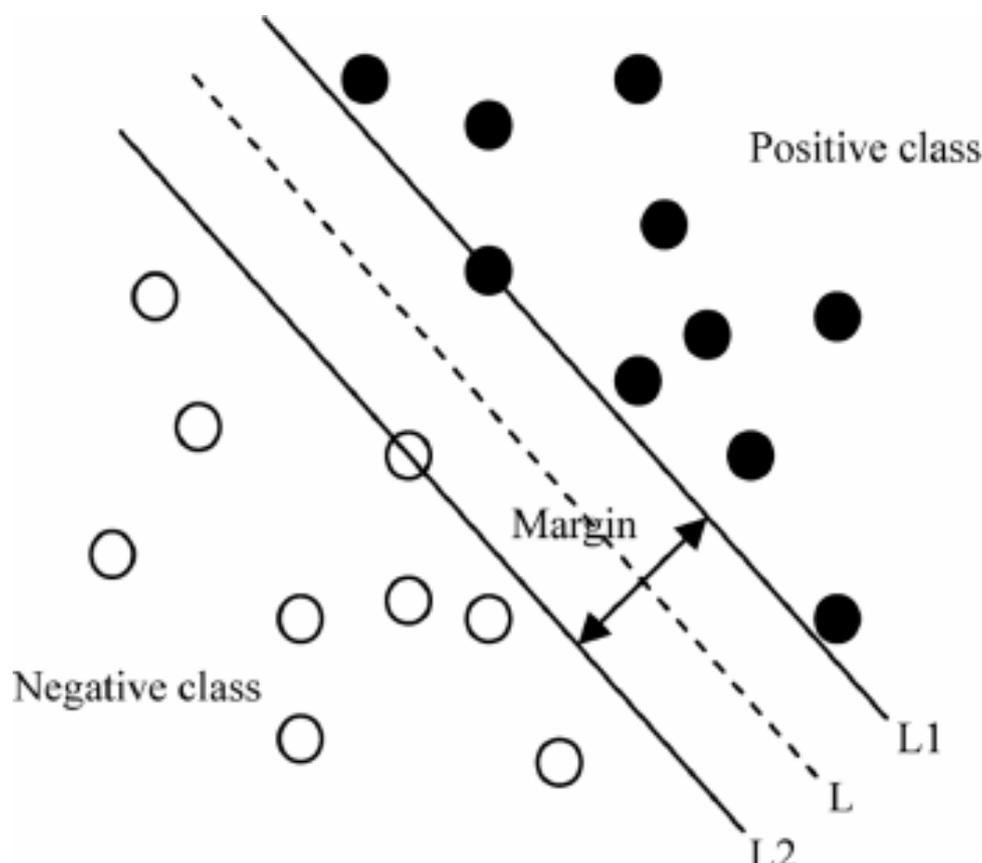


Figura 14: Clasificador SVM. Búsqueda del hiperplano que maximice el margen.

Para aquellos casos donde no existe un plano que permite la separación entre clases, es añadida una dimensión nueva a las muestras con el objetivo de poder encontrar un hiperplano que separe las clases en el nuevo espacio que es de mayor dimensión.

El algoritmo depende de 3 hiperparámetros principales, que son:

- Función de Kernel: la función relacionada con las dimensiones que se agregan a las muestras. Existen del tipo polinomial, gaussiano, rbf, etc.
 - Kernel Polinomial: $K(x, y) = (x^T y + c)^d$; donde x e y son vectores en el espacio de entrada, $c \geq 0$ es un parámetro libre, d es el grado polinomial.
 - Kernel RBF: $K(x, y) = \exp(-\gamma \|x - y\|^2)$; donde γ es el parámetro libre.
- C: la clasificación perfecta del hiperplano no siempre es posible, y en los casos en que existe, puede resultar en un modelo que no puede ser generalizado para otros casos, lo que se conoce como sobreajuste. Para esto existe el parámetro C, que le otorga flexibilidad al modelo. Este parámetro compensa entre errores en el entrenamiento

y los márgenes rígidos, creando los márgenes blandos, permitiendo errores en la clasificación, penalizándolos.

- Gamma: define qué tan lejos alcanza la influencia de una muestra sola. Valores bajos significan una radio de influencia mayor, mientras que valores más altos, una influencia más cercana.

En la figura 15, se puede ver la influencia de los hiperparámetros C y Gamma en la construcción del hiperplano utilizando el kernel RBF.

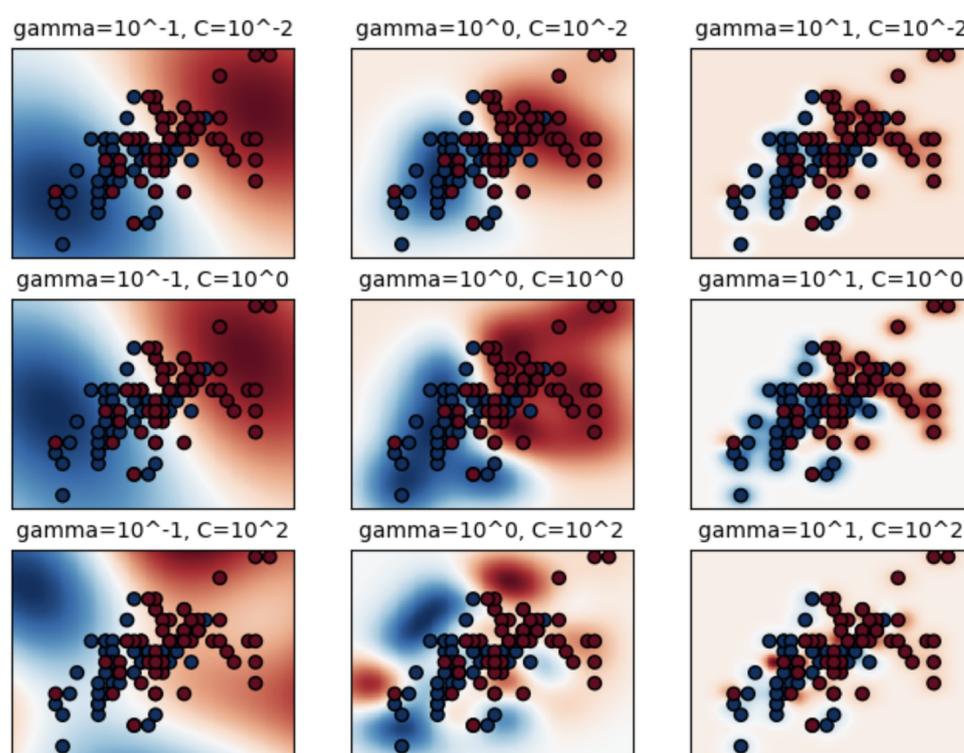


Figura 15: Influencia de hiperparámetros en la construcción del hiperplano, utilizando un kernel RBF

Búsqueda de Hiperparámetros

Para encontrar los hiperparámetros que construyen el modelo óptimo se realizan las pruebas con el conjunto de datos disponible en la base de datos KDEF.

El kernel que se utiliza es el RBF, un tipo de kernel radial. Es el más utilizado en combinación con la modelo SVM.

Se utiliza la función GridSearchCV del paquete sklearn. Ésta función, evalúa las diferentes opciones de parámetros que recibe como entrada en el modelo y, utilizando los datos, aplica validación cruzada para obtener la mejor combinación.

Los resultados se muestran en la figura 16. La combinación óptima de parámetros es la siguiente:

- Hiperparámetro C (param_C) = 10.
- Hiperparámetro Gamma (param_gamma) = 0.01

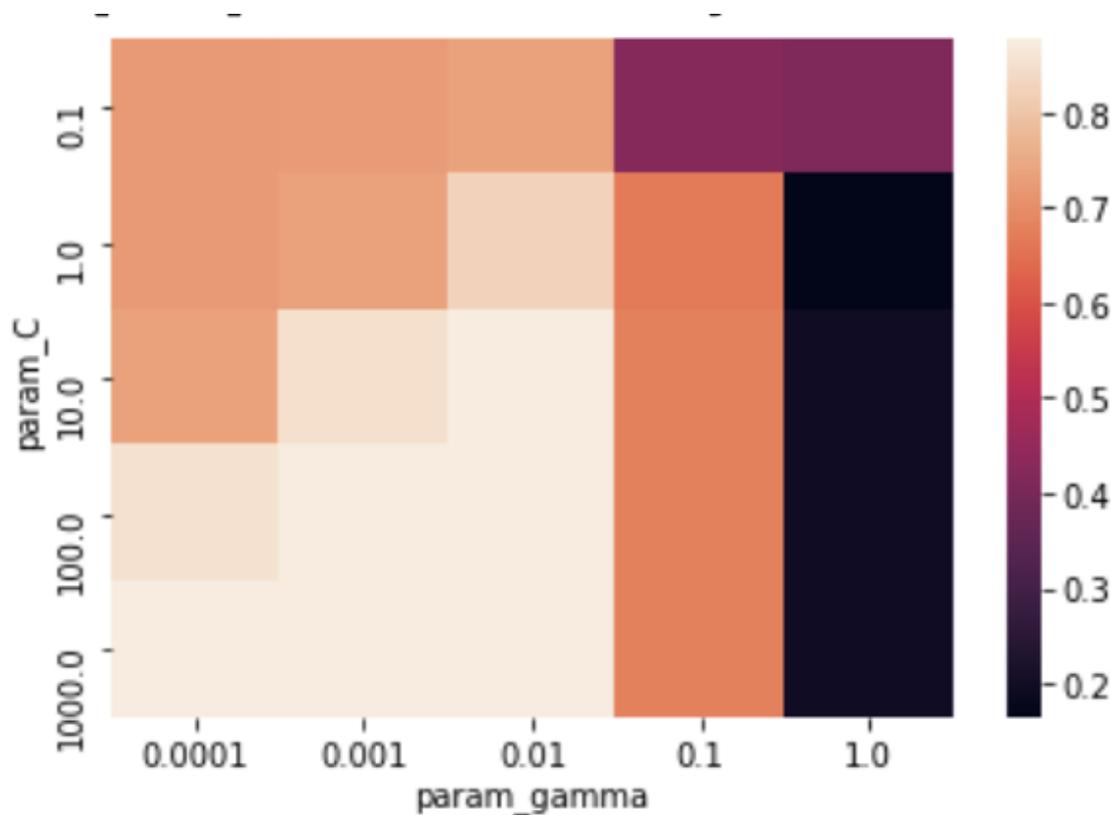


Figura 16: Resultados para las diferentes combinaciones de parámetros. El color de cada celda está asociado al *score* promedio.

Análisis de Videos

El objetivo del proyecto es el análisis y detección de la emoción de una persona, por lo que el análisis instantáneo de las mismas es insuficiente.

Por dicha razón se concluye una emoción global al analizar la evolución de la misma a lo largo del tiempo del video que se disponga.

Para construir la emoción final o global se calcula la moda de las emociones cada 10 frames del video, el objetivo de esto es la eliminación de emociones instantáneas que no perduran en el tiempo. La emoción final se resuelve como aquella que ha sido seleccionada como moda una mayor cantidad de veces.

En la figura 17 se puede ver la determinación de la emoción global a partir de la emoción detectada en cada frame. Se modifica, para el ejemplo, el cálculo de la moda cada 5 frames para simplificar la ilustración. Se parte, en el ejemplo mostrado, de un conjunto de 10

las caras para el cálculo de la moda, para ello se utiliza el seguimiento del centroide facial. Esto corresponde a calcular la posición del centro de cada una de las áreas de interés en cada frame (x, y) , de forma tal que la misma cara asociada a dicha área de interés se encuentra en $(x + \Delta_x, y + \Delta_y)$ en el frame siguiente. Siendo Δ_x, Δ_y , los desplazamientos máximos esperados en ambas direcciones.

En la figura 19 se puede ver el funcionamiento del seguimiento de caras utilizando centroides. Para ello se ejemplifica analizando el seguimiento de 2 objetos: 1, y 2, como se indica en la parte superior de cada recuadro de color en la figura. Existen dos recuadros nombrados como Objeto 1 y dos con Objeto 2, unos con color naranja y otros azul. El color naranja refiere al frame 1 (tiempo t), mientras que el color azul al frame 2 (tiempo $t + \Delta t$). Analizando solo los de color naranja, vemos 2 recuadros, asociados cada uno a un objeto diferente, esta es la situación inicial. Analizando los de color azul, nuevamente tenemos 2 recuadros, Objeto 1 y Objeto 2, y hacen referencia a los mismos objetos de color naranja, pero ahora se encuentran desplazados ya que refieren a una situación posterior a la inicial. El algoritmo permite asociar cada recuadro de la situación inicial, uno naranja, con otro en la situación siguiente, uno azul; para ellos se calculan las distancias entre los centroides naranjas y los azules. Las distancias son nombradas como Δx_i y Δy_i , cuando $i = 1$ refiere al desplazamiento que existe del objeto 1 entre el Frame 1 y el Frame 2, mientras que cuando $i = 2$ refiere al desplazamiento del objeto 2. Serán asociados como iguales aquellos cuyos centroides tengan un desplazamiento menor al máximo esperado. Las distancias entre el objeto 1 al tiempo t y el objeto 2 al tiempo $t + 1$, y la del objeto 2 al tiempo t y el objeto 1 al tiempo $t + 1$ no son dibujadas ya que será mayor que la existente entre los propios objetos, de esta forma es simplificada la figura. Sin embargo en un caso real sí son consideradas ya que no se sabe cuál es cada objeto, ese es el objetivo del método. Aplicado al caso del presente trabajo, el objeto es el de seguir caras, por lo que los objetos representarían las áreas de interés obtenidas de la detección facial. Las distancias máximas esperadas para una cara contenida en un área de interés es del 10% del lado del área.

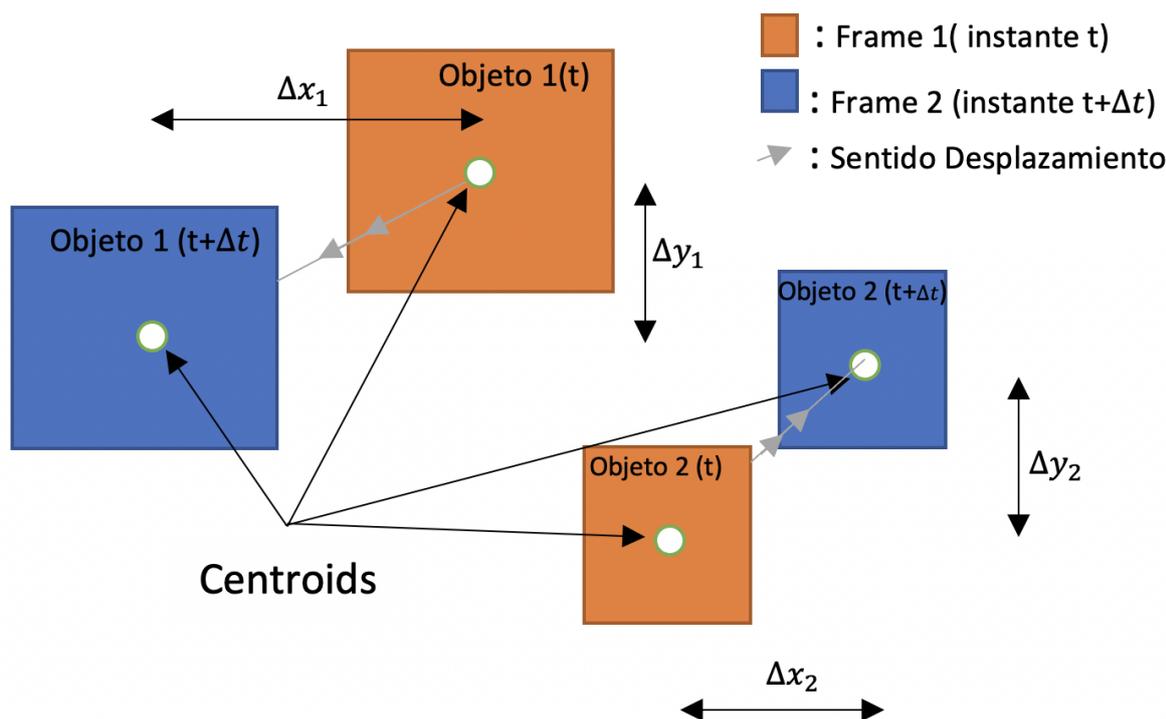


Figura 19: Seguimiento de objetos utilizando distancia entre centroides.

Mejoras Propuestas

Como se expresa en la sección de Descripción de los Métodos Utilizados, se realizan algunas modificaciones en los diferentes métodos y tipos de procesamiento aplicados. Las modificaciones son las siguientes:

- En la detección del área de interés, el cuadrante que contiene a la cara es modificado en su relación alto y ancho un 5% respecto al original en la implementación en Dlib.
- Alineación planar.
- Aplicación de filtro sobel para resaltar los bordes y lograr una menor dependencia con la iluminación.
- Definición del tamaño ideal de redimensionamiento.
- Búsqueda de hiperparámetros para construir el modelo óptimo con el kernel RBF: son realizadas pruebas con diferentes combinación de parámetros gamma y C.
- Cálculo de emoción resultante en intervalo temporal.
- Detección de múltiples caras.

Para más detalles sobre las diferentes mejoras y métodos utilizados ir a la sección Descripción de los Métodos Utilizados.

Bases de datos Utilizadas

Para el correcto funcionamiento del método presentado, dado su carácter de entrenamiento supervisado, es necesaria la elección de un correcto conjunto de datos. El *data set* no solo debe ser completo, abarcando todos los tipos de muestras necesarios, en este caso de las 7 emociones a distinguir, sino, que debe ser correcto, es decir, las muestras deben estar clasificadas correctamente. Errores en el conjunto de entrenamiento lleva a errores en el funcionamiento posterior. Además, el *data set* debe presentar variedad en las características de las personas que expresan las emociones: diferentes edades, sexos, tonos de piel, etc.

KDEF

El *data set* utilizado para el entrenamiento del modelo es el Karolinska Directed Emotional Faces (KDEF) [15]. Consiste en 4900 muestras de expresiones faciales asociadas a las emociones de miedo, disgusto, neutralidad, tristeza, sorpresa y alegría.

El grupo de personas fotografiadas esta compuesto por:

- 70 actores: 35 hombres y 35 mujeres.
- Edades entre 20 y 30 años.

En las imágenes se restringe el uso de barba y bigote, de aros y de anteojos.

BAUM1

La aplicación propuesta, además del análisis de emociones instantáneas o en fotos, consta de un mecanismo para la detección de emociones resultantes expresadas en un intervalo de tiempo. Por ello, es que es necesario un conjunto de datos de videos para poder analizar su efectividad. El *data set* utilizado para dicho propósito es el de BAUM1 [22]. Éste contiene videos de 31 personas de entre 18 y 66 años expresando las emociones de: felicidad, enojo, tristeza, disgusto, miedo, neutralidad, aburrimiento, interés, desconfianza. Fueron seleccionadas solamente las referidas a las 7 emociones estudiadas en el presente trabajo. El conjunto seleccionado esta formado por 240 muestras.

En la figura 20 se pueden ver algunas imágenes obtenidas a partir de las muestras en la base de datos.



Figura 20: Imágenes de las muestras presentes en el conjunto de datos BAUM1.

Resultados

Para poder evaluar la capacidad de clasificación del modelo propuesto, se realizaron las siguientes pruebas en diferentes conjuntos de emociones.

Pruebas preliminares

Utilizando el modelo propuesto anteriormente, se analiza la capacidad de clasificación en diferentes conjuntos de emociones.

Para garantizar fiabilidad en los resultados se utiliza el método de validación cruzada. Esto implica que el algoritmo se ejecuta 5 veces sobre los datos, modificando cada vez los datos para entrenamiento y evaluación.

El porcentaje de muestras utilizado para entrenamiento es de 70 %, y para evaluación el 30 % restante.

Resultados entre pares de emociones

En primera instancia el objetivo es el de distinguir entre pares de emociones. En la tabla 4, se encuentran los resultados sobre el conjunto de evaluación. En cada celda se puede ver el promedio del score utilizando validación cruzada. En esta tabla se puede analizar la efectividad del algoritmo en la distinción entre 2 emociones.

	Miedo	Disgusto	Felicidad	Neutralidad	Tristeza	Sorpresa	Enojo
Miedo							
Disgusto	0.9429						
Felicidad	0.9810	0.9833					
Neutralidad	0.9524	0.9881	0.9929				

	Miedo	Disgusto	Felicidad	Neutralidad	Tristeza	Sorpresa	Enojo
Tristeza	0.8738	0.95	0.9904	0.9166			
Sorpresa	0.8667	0.9928	1.0	0.9833	0.9833		
Enojo	0.9286	0.9143	0.9904	0.9881	0.9476	0.9952	

Tabla 4: Resultados entre pares de emociones. En la celda (i,j) se muestra el porcentaje de aciertos promedio (validación cruzada) sobre las muestras de evaluación entre las emociones i y j.

Resultados en conjuntos de 3 emociones

En estas pruebas se reduce tanto el conjunto de entrenamiento, como el de evaluación para todos los posibles conjuntos de tres emociones; para ello se entrena utilizando el 70 % de las muestras que correspondan al conjunto de las 3 emociones, y luego se evalúa con el 30 % restante. El proceso se realiza 5 veces siguiendo el esquema de la validación cruzada, y cada resultado se presenta en las columnas de la tabla nombradas S_n . De esta manera, se pueden analizar los conjuntos más difíciles de distinguir. En la tabla 5 se encuentran los resultados.

Conjuntos de Emociones	S1	S2	S3	S4	S5	Promedio
Happy,Disgusted,Afraid	0.9206	0.9683	1.0	0.9524	0.9444	0.9571
Neutral,Disgusted,Afraid	0.9286	0.9524	0.9048	0.9286	0.9365	0.9302
Neutral,Happy,Afraid	0.9365	0.9365	0.9683	0.9286	0.9603	0.9460
Neutral,Happy,Disgusted	0.9762	1.0	0.9841	0.9683	0.9603	0.9778
Sad,Disgusted,Afraid	0.9206	0.8016	0.8730	0.8968	0.8889	0.8762
Sad,Happy,Afraid	0.9524	0.9048	0.8413	0.8889	0.9048	0.8984
Sad,Happy,Disgusted	0.9365	0.9365	0.9286	0.9762	0.9524	0.9460
Sad,Neutral,Afraid	0.8651	0.8968	0.8730	0.8333	0.7937	0.8524
Sad,Neutral,Disgusted	0.9127	0.8810	0.9365	0.8810	0.9683	0.9159
Sad,Neutral,Happy	0.9762	0.9603	0.9603	0.9365	0.9286	0.9524
Surprised,Disgusted,Afraid	0.9286	0.8730	0.8968	0.9048	0.9127	0.9032
Surprised,Happy,Afraid	0.9048	0.9365	0.8730	0.9286	0.8968	0.9079
Surprised,Happy,Disgusted	0.9762	0.9841	0.9921	0.9921	0.9841	0.9857
Surprised,Neutral,Afraid	0.9127	0.8968	0.9127	0.9048	0.8810	0.9016
Surprised,Neutral,Disgusted	0.9762	0.9841	0.9841	0.9762	0.9921	0.9825
Surprised,Neutral,Happy	0.9762	0.9762	1.0	0.9921	0.9841	0.9841
Surprised,Sad,Afraid	0.8730	0.8413	0.8571	0.8175	0.8095	0.8397
Surprised,Sad,Disgusted	0.9683	0.9841	0.9603	0.9524	0.9762	0.9683
Surprised,Sad,Happy	0.9762	0.9683	0.9841	0.9841	0.9841	0.9794
Surprised,Sad,Neutral	0.9683	0.9683	0.9603	0.9365	0.9444	0.9556

Conjuntos de Emociones	S1	S2	S3	S4	S5	Promedio
Disgusted,Angry,Afraid	0.9127	0.9286	0.8571	0.8492	0.8413	0.8778
Happy,Angry,Afraid	0.9286	0.9603	0.9683	0.9444	0.9444	0.9492
Neutral,Angry,Afraid	0.9365	0.9444	0.9127	0.9286	0.8651	0.9175
Neutral,Happy,Angry	0.9603	0.9841	1.0	0.9921	0.9762	0.9825
Sad,Angry,Afraid	0.8492	0.8730	0.8810	0.9127	0.8254	0.8683
Sad,Disgusted,Angry	0.8810	0.9206	0.8651	0.8730	0.8492	0.8778
Sad,Happy,Angry	0.9524	0.9365	0.9444	0.9762	0.9683	0.9556
Sad,Neutral,Angry	0.9206	0.9127	0.9127	0.9048	0.9127	0.9127
Surprised,Angry,Afraid	0.9127	0.8730	0.9365	0.8571	0.9286	0.9016
Surprised,Disgusted,Angry	0.9683	0.9524	0.9444	0.9524	0.9286	0.9492
Surprised,Happy,Angry	0.9921	0.9603	0.9762	1.0	1.0	0.9857
Surprised,Neutral,Angry	0.9841	0.9603	0.9683	0.9603	0.9603	0.9667
Surprised,Sad,Angry	0.9524	0.9524	0.9524	0.9603	0.9286	0.9492

Tabla 5: Resultados para conjuntos de 3 emociones. En las columnas representadas con (Sn) se encuentran los resultados de la iteración n en la validación cruzada. En la columna promedio se calcula el promedio de los 5 resultados.

Se observa que la precisión en distinguir tres emociones es muy alta. Sin embargo, el conjunto de emociones con peores resultados, es el de Sorpresa, Tristeza y Miedo.

Resultados en conjuntos de 4 emociones

Al reducir el conjunto de entrenamiento y evaluación a solo 4 emociones, se puede ver aquellos conjuntos en donde el clasificador es más o menos preciso. Los resultados al clasificar en conjuntos de 4 emociones se encuentran en la tabla 6.

Conjuntos de Emociones	S1	S2	S3	S4	S5	Promedio
Disgusted,Happy,Neutral,Afraid	0.9405	0.9286	0.9405	0.9405	0.9166	0.9333
Disgusted,Happy,Sad,Afraid	0.875	0.9524	0.8750	0.8929	0.8452	0.8881
Disgusted,Happy,Surprised,Afraid	0.9345	0.9405	0.9405	0.9345	0.9286	0.9357
Disgusted,Neutral,Sad,Afraid	0.8571	0.8810	0.8393	0.8333	0.8214	0.8464
Disgusted,Neutral,Surprised,Afraid	0.8988	0.9345	0.8690	0.9107	0.8929	0.9012
Disgusted,Sad,Surprised,Afraid	0.8810	0.8631	0.8333	0.8393	0.8631	0.8560
Happy,Neutral,Sad,Afraid	0.9107	0.8929	0.8929	0.8869	0.9167	0.9
Happy,Neutral,Surprised,Afraid	0.8988	0.8810	0.9167	0.8929	0.9107	0.9
Happy,Sad,Surprised,Afraid	0.8393	0.8869	0.8274	0.875	0.8810	0.8619
Neutral,Sad,Surprised,Afraid	0.8214	0.7917	0.8452	0.8571	0.8155	0.8262
Happy,Neutral,Sad,Disgusted	0.9583	0.9107	0.9345	0.9464	0.9405	0.9381

Conjuntos de Emociones	S1	S2	S3	S4	S5	Promedio
Happy,Neutral,Surprised,Disgusted	0.9703	0.9821	0.9524	0.9821	0.9821	0.9738
Happy,Sad,Surprised,Disgusted	0.95833	0.9405	0.9583	0.9583	0.9761	0.9583
Neutral,Sad,Surprised,Disgusted	0.9464	0.9583	0.9583	0.9048	0.9345	0.9405
Neutral,Sad,Surprised,Happy	0.9286	0.9643	0.9583	0.9286	0.9226	0.9405
Angry,Disgusted,Happy,Afraid	0.9345	0.9107	0.8929	0.9286	0.8988	0.9131
Angry,Disgusted,Neutral,Afraid	0.8571	0.8631	0.8810	0.8929	0.9048	0.8798
Angry,Disgusted,Sad,Afraid	0.8452	0.8214	0.8095	0.8690	0.875	0.8440
Angry,Disgusted,Surprised,Afraid	0.8452	0.8452	0.8690	0.8869	0.9345	0.8762
Angry,Disgusted,Neutral,Afraid	0.9405	0.9524	0.9464	0.9405	0.9345	0.9429
Angry,Happy,Sad,Afraid	0.8988	0.8929	0.8452	0.9048	0.9048	0.8893
Angry,Happy,Surprised,Afraid	0.8929	0.9464	0.8988	0.8810	0.9345	0.9107
Angry,Neutral,Sad,Afraid	0.8988	0.8690	0.8095	0.8631	0.8333	0.8548
Angry,Neutral,Surprised,Afraid	0.8810	0.8810	0.8571	0.9049	0.8988	0.8845
Angry,Sad,Surprised,Afraid	0.8452	0.8690	0.8393	0.8631	0.8155	0.8464
Disgusted,Happy,Neutral,Angry	0.9405	0.9524	0.9405	0.9524	0.9048	0.9381
Disgusted,Happy,Sad,Angry	0.9286	0.8690	0.9226	0.9702	0.91078	0.9202
Disgusted,Happy,Surprised,Angry	0.9524	0.9405	0.9405	0.9524	0.9524	0.9476
Disgusted,Neutral,Sad,Angry	0.9226	0.8869	0.875	0.8869	0.8988	0.8940
Disgusted,Neutral,Surprised,Angry	0.9345	0.9286	0.9286	0.9464	0.9286	0.9333
Disgusted,Sad,Surprised,Angry	0.8810	0.9345	0.9226	0.8869	0.9107	0.9071
Happy,Neutral,Sad,Angry	0.9405	0.9345	0.9643	0.9524	0.9702	0.9524
Happy,Neutral,Surprised,Angry	0.9881	0.9821	0.9881	0.9583	0.9702	0.9774
Happy,Sad,Surprised,Angry	0.9643	0.9464	0.9643	0.9405	0.9643	0.9560
Neutral,Sad,Surprised,Angry	0.9345	0.9643	0.9524	0.9405	0.9167	0.9417
Neutral,Sad,Surprised,Angry	0.9345	0.9643	0.9524	0.9405	0.9167	0.9417

Tabla 6: Resultados en conjuntos de 4 emociones. En las columnas representadas con (Sn) se encuentran los resultados de la iteración n en la validación cruzada. En la columna promedio se calcula el promedio de los 5 resultados.

Se observa que la precisión en distinguir cuatro emociones es muy alta. Sin embargo, el conjunto con peores resultados es el conformado por Enojo, Disgusto, Tristeza y Miedo.

Resultados en conjuntos de 5 emociones

Los resultados al clasificar en grupos de 5 emociones se encuentran en la tabla 7.

Conjuntos de Emociones	S1	S2	S3	S4	S5	Promedio
Happy/Neutral/Sad/Disgusted/Afraid	0.8905	0.8905	0.8619	0.9095	0.8524	0.8810

Conjuntos de Emociones	S1	S2	S3	S4	S5	Promedio
Happy/Neutral/Surprised/Disgusted/Afraid	0.90	0.9143	0.90	0.9286	0.9190	0.9124
Happy/Sad/Surprised/Disgusted/Afraid	0.8762	0.8524	0.9190	0.8571	0.8810	0.8771
Neutral/Sad/Surprised/Disgusted/Afraid	0.8810	.8619	0.8238	0.8429	0.8333	0.8486
Neutral/Sad/Surprised/Happy/Afraid	0.8952	0.8524	0.8714	0.8571	0.8857	0.8724
Neutral/Sad/Surprised/Happy/Disgusted	0.9333	0.9381	0.9238	0.9667	0.9619	0.9448
Disgusted/Happy/Neutral/Angry/Afraid	0.8857	0.8952	0.9286	0.8381	0.8952	0.8886
Disgusted/Happy/Sad/Angry/Afraid	0.90	0.8524	0.8571	0.8667	0.8857	0.8724
Disgusted/Happy/Surprised/Angry/Afraid	0.8429	0.9095	0.8333	0.9048	0.8762	0.8733
Disgusted/Neutral/Sad/Angry/Afraid	0.8381	0.8714	0.8238	0.8286	0.8619	0.8448
Disgusted/Neutral/Surprised/Angry/Afraid	0.8905	0.8524	0.8762	0.8905	0.8619	0.8743
Disgusted/Sad/Surprised/Angry/Afraid	0.8143	0.8429	0.8524	0.8762	0.8571	0.8486
Happy/Neutral/Sad/Angry/Afraid	0.8857	0.8714	0.8571	0.8952	0.8952	0.8810
Happy/Neutral/Surprised/Angry/Afraid	0.9047	0.9238	0.8810	0.8810	0.9429	0.9067
Happy/Sad/Surprised/Angry/Afraid	0.8810	0.8429	0.9048	0.90	0.8810	0.8819
Neutral/Sad/Surprised/Angry/Afraid	0.8333	0.8810	0.8476	0.8667	0.8333	0.8524
Happy/Neutral/Sad/Disgusted/Angry	0.8905	0.9143	0.9095	0.8810	0.8667	0.8924
Happy/Neutral/Surprised/Disgusted/Angry	0.9381	0.9381	0.9429	0.9333	0.9667	0.9438
Happy/Sad/Surprised/Disgusted/Angry	0.9381	0.9619	0.9333	0.9429	0.9190	0.9390
Neutral/Sad/Surprised/Disgusted/Angry	0.9143	0.9238	0.9238	0.8905	0.8857	0.9076
Neutral/Sad/Surprised/Happy/Angry	0.9238	0.9524	0.9429	0.9238	0.9571	0.9400

Tabla 7: Resultados en conjuntos de 5 emociones. En las columnas representadas con (Sn) se encuentran los resultados de la iteración n en la validación cruzada. En la columna promedio se calcula el promedio de los 5 resultados.

Se observa que la precisión en distinguir cinco emociones es muy alta. Sin embargo, el conjunto con peores resultados es el conformado por Disgusto, Neutralidad, Tristeza, Enojo y Miedo.

Resultados en conjuntos de 6 emociones

Los resultados para los conjuntos de 6 emociones se encuentran en la tabla 8.

Emociones	S1	S2	S3	S4	S5	Prom.
Neutral/Sad/Surprised/Happy/Disgusted/Afraid	0.9008	0.8651	0.8770	0.8730	0.8452	0.8722
Happy/Neutral/Sad/Disgusted/Angry/Afraid	0.8889	0.8810	0.8611	0.8730	0.8492	0.8706
Happy/Neutral/Surprised/Disgusted/Angry/Afraid	0.8929	0.8770	0.9048	0.9008	0.8968	0.8944
Happy/Sad/Surprised/Disgusted/Angry/Afraid	0.8770	0.8810	0.8770	0.8810	0.8651	0.8762
Neutral/Sad/Surprised/Disgusted/Angry/Afraid	0.8452	0.8294	0.8651	0.8214	0.8492	0.8421

Emociones	S1	S2	S3	S4	S5	Prom.
Neutral/Sad/Surprised/Happy/Angry/Afraid	0.8968	0.8770	0.8333	0.8690	0.8254	0.8603
Neutral/Sad/Surprised/Happy/Disgusted/Angry	0.8929	0.9087	0.9167	0.9127	0.8968	0.9056

Tabla 8: Resultados para los conjuntos de 6 emociones. En las columnas representadas con (Sn) se encuentran los resultados de la iteración n en la validación cruzada. En la columna promedio se calcula el promedio de los 5 resultados.

Se observa que la precisión en distinguir seis emociones es muy alta. Sin embargo, el conjunto con peores resultados es el conformado por Neutralidad, Tristeza, Sorpresa, Disgusto, Enojo y Miedo.

Resultados en conjunto de 7 emociones

Los resultados para el conjunto de 7 emociones se encuentran en la tabla 9.

Emociones	S1	S2	S3	S4	S5	Prom.
Neutral/Sad/Surprised/Happy/Disgusted/Afraid/Angry	0.8571	0.7993	0.8503	0.8776	0.8231	0.8415

Tabla 9: Resultados para el conjunto de 7 emociones. En las columnas representadas con (Sn) se encuentran los resultados de la iteración n en la validación cruzada. En la columna promedio se calcula el promedio de los 5 resultados.

Sobre el grupo de 7 se realiza un análisis más profundo para ver específicamente qué emociones están siendo confundidas con qué otras.

En la tabla 10, se puede ver la matriz de confusión entre las 7 emociones. En dicha matriz se puede ver cuantas imágenes del total de evaluación fueron categorizadas de forma correcta como también cómo fueron categorizadas aquellas que fueron confundidas.

	Afraid	Disgusted	Happy	Neutral	Sad	Surprised	Angry	Total
Afraid	38	0	0	1	1	2	0	42
Disgusted	1	39	0	0	1	0	1	42
Happy	0	0	41	0	1	0	0	42
Neutral	1	1	0	40	0	0	0	42
Sad	3	0	1	0	38	0	0	42
Surprised	2	0	0	2	0	38	0	42
Angry	0	2	0	0	0	0	40	42

Tabla 10: Matriz de confusión. En cada fila se puede ver la predicción del modelo para las imágenes de cada categoría.

Clasificación de videos

Una vez analizada la precisión al clasificar imágenes se procede a realizar pruebas sobre la clasificación en videos utilizando el mecanismo de la moda detallado en la sección Análisis de Videos. La base de datos utilizada para la evaluación es la BAUM1.

En la figura 21 se presentan los resultados. Se observa que para las emociones de disgusto y felicidad, el método obtiene una precisión superior al 85 %.

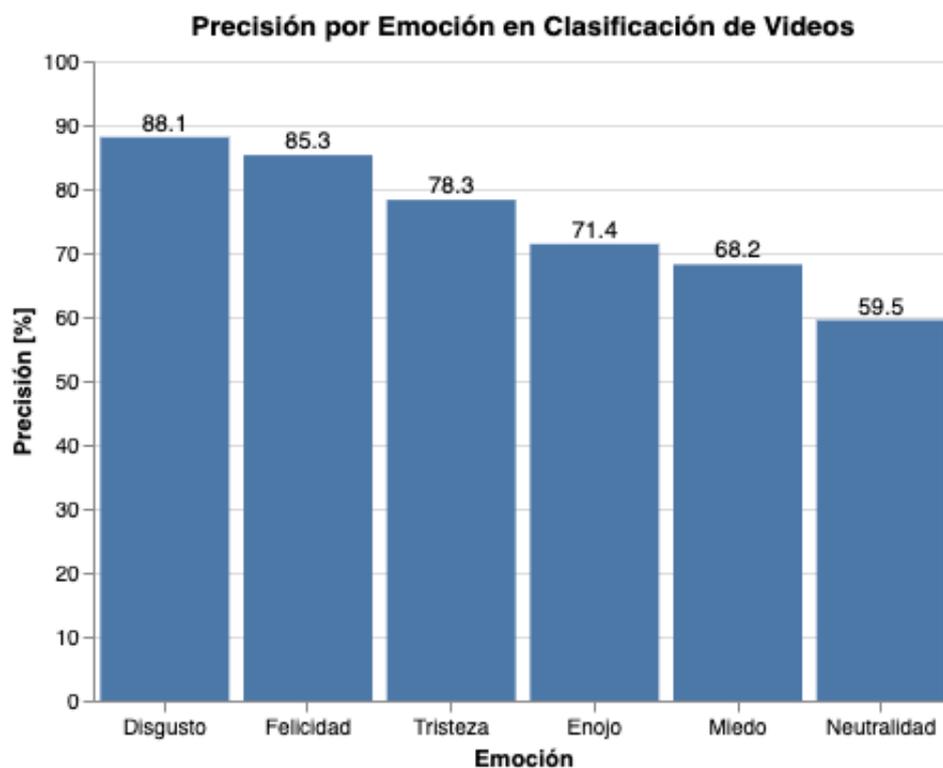


Figura 21: Precisión en la clasificación de emociones en videos.

Sin hacer distinciones por emoción, la precisión en la clasificación de videos es de 74.03 %.

Análisis de ruido

Para probar la efectividad del método en la práctica real, donde las transmisiones e imágenes pueden ser sometidas a distintos tipos de ruido, se hizo una prueba contaminando las imágenes con ruido sintético. Las imágenes utilizadas para el análisis son las mismas que entrenaron al modelo, pero contaminadas.

El ruido sintético es de dos tipos, un ruido Gaussiano con un valor de desviación estándar $\sigma = 15$ y ruido sal y pimienta con un parámetro de densidad $p = 0,05$. La variable de control fue el porcentaje de ruido aplicado a la imagen, es decir, la proporción de píxeles contaminados. Para el análisis, esta variable se tomó con valores del 0 % al 100 %, con pasos del 10 %.

El ruido sintético fue aplicado durante el preprocesado de la imagen, porque así se considera que llega al sistema de reconocimiento desde, por ejemplo, una cámara. En la Figura 22 se observa un ejemplo de una imagen original y en las Figuras 23 y 24 los resultados después de la contaminación con ruido Gaussiano y Sal y Pimienta, respectivamente.



Figura 22: Ejemplo de imagen original de la base de datos KDEF



Figura 23: Ejemplo de imagen contaminada un 80% con un ruido Gaussiano y un valor de desviación estándar $\sigma = 15$



Figura 24: Ejemplo de imagen contaminada un 50 % con un ruido sal y pimienta y un parámetro de densidad $p = 0,05$

En Figura 25 se muestran los resultados del análisis, comparando como cada tipo de ruido afecta la efectividad del reconocimiento a medida que aumenta su presencia.

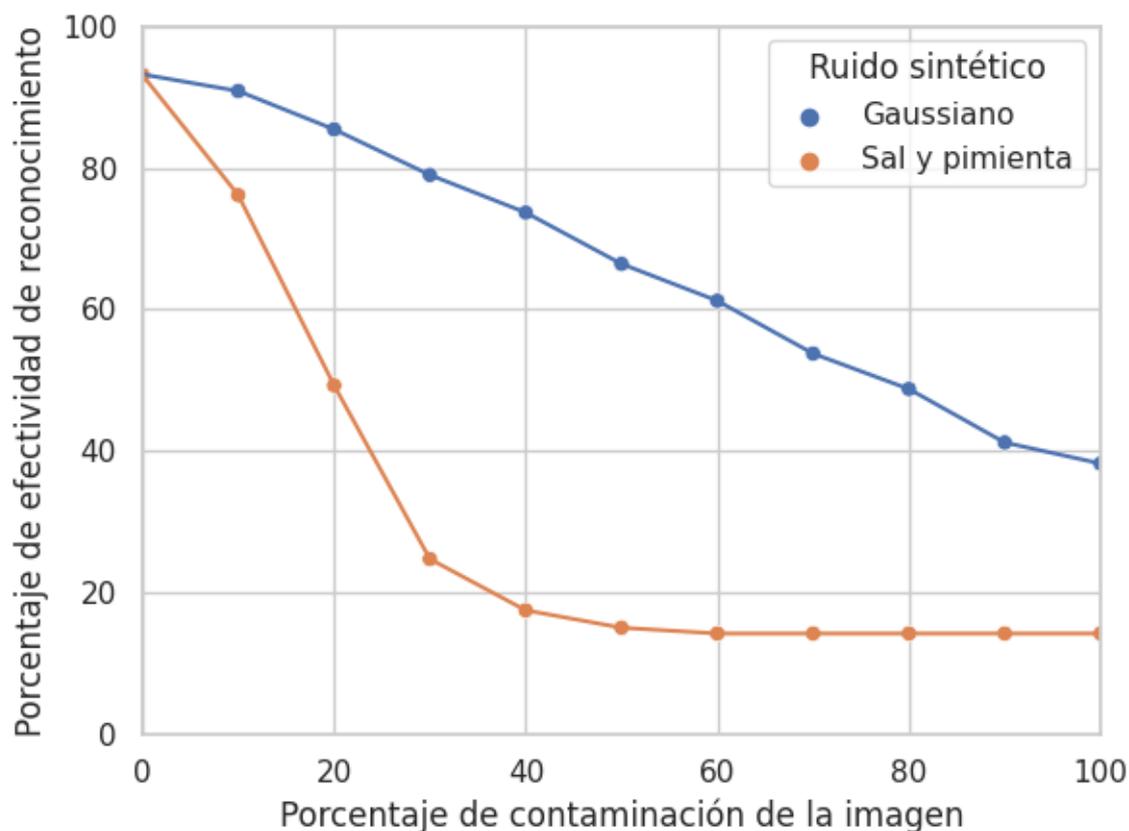


Figura 25: Resultados de reconocimiento facial de emociones en imágenes con ruido sintético agregado.

Se puede ver que cuando la imagen está contaminada con un porcentaje de ruido menor que el 10 % obtenemos más de 70 % de aciertos en la clasificación, tanto para ruido Gaussiano como sal y pimienta. El método soporta un 40 % de ruido Gaussiano con un valor de desviación estándar $\sigma = 15$, si se considera que un 70 % de rendimiento es suficiente.

Conclusiones y Trabajos Futuros

En cuanto a la clasificación de las 7 emociones estudiadas en imágenes, son obtenidas las siguientes mejores combinaciones:

- Para diferenciar entre 2 emociones los mejores resultados se obtienen para distinguir, con más del 99 % de precisión:
 - Tristeza y felicidad.
 - Neutralidad y felicidad.
 - Sorpresa y felicidad.

- Sorpresa y enojo.
- Para diferenciar 3 emociones el mejor resultado se obtiene entre sorpresa, felicidad y disgusto y entre sorpresa, felicidad y enojo, ambas con 98 % .
- Para diferenciar 4 emociones el mejor resultado se obtiene entre felicidad, neutralidad, sorpresa y enojo, con 97 %.
- Para diferenciar 5 emociones el mejor resultado se obtiene entre neutralidad, tristeza, sorpresa, felicidad y disgusto, con 94 %.
- Para diferenciar 6 emociones el mejor resultado se obtiene entre neutralidad, tristeza, sorpresa, felicidad, disgusto y enojo, con 90 %.
- Para diferenciar las 7 emociones, la precisión obtenida es de 84 %.

Dentro del conjunto de las 7 emociones, la que obtiene mayor cantidad de verdaderos positivos es felicidad, mientras que las que obtienen menos cantidad de verdaderos positivos son miedo, tristeza y sorpresa.

Las emociones más confundidas entre sí, son sorpresa y miedo.

En cuanto a la resistencia al ruido, la clasificación de emociones es más vulnerable al ruido de tipo Sal y Pimienta que al ruido Gaussiano. Es importante destacar que las imágenes son contaminadas antes de la detección del área de interés, cuyo funcionamiento depende de los gradientes de la imagen.

En relación a la clasificación en videos, los mejores resultados se obtienen para clasificar disgusto y felicidad entre las 7 emociones, con una precisión superior al 85 %.

En cuanto a trabajos futuros, es importante trabajar sobre las rotaciones en el espacio, además de las planares que son las que se tienen en cuenta en este trabajo.

También es posible lograr una mayor resistencia al ruido Gaussiano al suavizar la imagen con un filtro en la etapa de preprocesado.

Referencias

- [1] Timo Ahonen, Abdenour Hadid y Matti Pietik ain. “Face Recognition with Local Binary Patterns”. En: *IEEE Trans Pattern Anal. Mach. Intell.* 28.12 (2006), págs. 469-481.
- [2] K. Candra Kirana, S. Wibawanto y H. Wahyu Herwanto. “Facial Emotion Recognition Based on Viola-Jones Algorithm in the Learning Environment”. En: *2018 International Seminar on Application for Technology of Information and Communication.* 2018, págs. 406-410.
- [3] Junkai Chen y col. “Facial Expression Recognition in Video with Multiple Feature Fusion”. En: *IEEE Transactions on Affective Computing* PP (jul. de 2016), págs. 1-1.
- [4] M. Dahmane y J. Meunier. “Emotion recognition using dynamic grid-based HoG features”. En: *2011 IEEE International Conference on Automatic Face Gesture Recognition (FG).* 2011, págs. 884-888.
- [5] N. Dalal y B. Triggs. “Histograms of oriented gradients for human detection”. En: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05).* Vol. 1. 2005, 886-893 vol. 1.
- [6] N. Dalal y B. Triggs. “Histograms of oriented gradients for human detection”. En: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)* 1 (2005), 886-893 vol. 1.
- [7] Abhinav Dhall y col. “Emotion Recognition in the Wild Challenge 2013”. En: *Proceedings of the 15th ACM on International Conference on Multimodal Interaction.* Association for Computing Machinery, 2013, págs. 509-516.
- [8] P. Ekman. *Emotion in the Human Face.* Cambridge University Press, 1982.
- [9] P. Ekman y W.V. Friesen. *Facial Action Coding System (FACS): Manual.* Palo Alto, Calif: Consulting Psychologists Press, 1978.
- [10] P. Ekman y W.V. Friesen. *Unmasking the Face; a Guide to Recognizing Emotions from Facial Clues.* Prentice Hall, 1975.
- [11] Jian-Ming Sun, Xue-Sheng Pei y Shi-Sheng Zhou. “Facial emotion recognition in modern distant education system using SVM”. En: *2008 International Conference on Machine Learning and Cybernetics.* Vol. 6. 2008, págs. 3545-3548.
- [12] A. Kartali y col. “Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches”. En: *2018 14th Symposium on Neural Networks and Applications (NEUREL).* 2018, págs. 1-4.
- [13] K. Ko y K. Sim. “Development of a Facial Emotion Recognition Method Based on Combining AAM with DBN”. En: *2010 International Conference on Cyberworlds.* 2010, págs. 87-91.

- [14] P. Lucey y col. “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression”. En: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. 2010, págs. 94-101.
- [15] D. Lundqvist, A. Flykt y A. Öhman. *The Karolinska Directed Emotional Faces - KDEF*. 1998.
- [16] M. Pantic y L. J. M. Rothkrantz. “Automatic analysis of facial expressions: the state of the art”. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.12 (2000), págs. 1424-1445.
- [17] Björn Schuller y col. “AVEC 2011—The First International Audio/Visual Emotion Challenge”. En: *Affective Computing and Intelligent Interaction* (2011). Ed. por Sidney D’Mello y col., págs. 415-424.
- [18] A. Sharma y S. Jaswal. “Analysis of Sobel Edge Detection Technique for Face Recognition”. En: *International Journal of Advanced Research in Computer Engineering Technology (IJARCET)* 4 (mayo de 2015).
- [19] S. Shojaeilangari y col. “Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning”. En: *IEEE Transactions on Image Processing* 24.7 (2015), págs. 2140-2152.
- [20] Stewart Tseng. “Comparison of holistic and feature approaches to face recognition”. En: 2003.
- [21] M. A. Turk y A. P. Pentland. “Face recognition using eigenfaces”. En: *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1991, págs. 586-591.
- [22] S. Zhalehpour y col. “BAUM-1: A Spontaneous Audiovisual Face Database of Affective and Mental States”. En: *IEEE Transactions on Affective Computing* 8.3 (2017), 886-893 vol. 1.