EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

Tecnológico Nacional de México

Centro Nacional de Investigación y Desarrollo Tecnológico

Tesis de Maestría

Mejoramiento de imágenes para sistemas de videovigilancia de baja calidad

presentada por

Ing. Rebecca Yuziel Zumaya Lanuza

como requisito para la obtención del grado de

Maestra en Ciencias de la Computación

Directora de tesis

Dra. Andrea Magadán Salazar

Codirectora de tesis

Dra. Daniela Alejandra Moctezuma Ochoa

Cuernavaca, Morelos, México. Noviembre de 2023.

cenidet
Centro Nacional de Investigación y Desarrollo Tecnológico

Cuernavaca, Mor., 27/noviembre/2023

OFICIO No. DCC/202/2023

Asunto: Aceptación de documento de tesis
CENIDET-AC-004-M14-OFICIO

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial de REBECCA YUZIEL ZUMAYA LANUZA con número de control M21CE072, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado **"SISTEMA PARA EL MEJORAMIENTO DE IMÁGENES PARA SISTEMAS DE VIDEO VIGILANCIA DE BAJA CALIDAD"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

ANDREA MAGADÁN SALAZAR
Directora de tesis

DANIELA ALEJANDRA MOCTEZUMA OCHOA
Codirectora

RAÚL PINTO ELÍAS
Revisor 1

SAID POLANCO MARTAGÓN
Revisor 2

C.c.p. Depto. Servicios Escolares.
Expediente / Estudiante

EDUCACIÓN | TECNOLÓGICO NACIONAL DE MÉXICO
04 DIC 2023
CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO
SERVICIOS ESCOLARES
RECIBIDO

Cuernavaca, Mor., 15/diciembre/2023
No. De Oficio: SAC/208/2023
Asunto: Autorización de impresión de tesis

REBECCA YUZIEL ZUMAYA LANUZA
CANDIDATA AL GRADO DE MAESTRA EN CIENCIAS DE LA COMPUTACIÓN
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"SISTEMA PARA EL MEJORAMIENTO DE IMÁGENES PARA SISTEMAS DE VIDEO VIGILANCIA DE BAJA CALIDAD"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
"Conocimiento y tecnología al servicio de México"

EDUCACIÓN | TECNOLÓGICO NACIONAL DE MÉXICO
CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO
DIRECCIÓN

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO

C. c. p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz

Dedicatoria

A mi ma, por ser la luz de mi camino.
A mi mamá por su apoyo, amor y cariño.

Agradecimientos

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por el apoyo económico otorgado para realizar mis estudios de maestría.

Al Centro Nacional de Investigación y Desarrollo Tecnológico (TecNM/CENIDET), por brindarme un espacio de trabajo y los recursos necesarios que me permitieron realizar mis estudios de maestría.

A mi directora de tesis la Dra. Andrea Magadán Salazar y a mi codirectora la Dra. Daniela Alejandra Moctezuma Ochoa por su asesoramiento durante el desarrollo de este trabajo de tesis, por brindarme sus consejos, su apoyo, su tiempo y su paciencia.

A mis revisores el Dr. Raúl Pinto Elías y el Dr. Said Polanco Martagón, por sus comentarios y revisiones que fueron fundamentales para la realización de esta tesis.

A mis compañeros Omar Trujillo y Fernando Luna por su disposición y paciencia para ayudarme y explicarme siempre que lo necesite.

A mis compañeros Paul, Sagrario, Bryan y Fernando Sánchez por todas las pláticas y momentos de diversión que hicieron estos dos años más amenos.

A mi ma, por siempre desvelarse conmigo y acompañarme en esta experiencia como lo ha hecho siempre.

A mi mamá y Miguel por escucharme, hacerme reír y apoyarme siempre.

Finalmente agradezco a cada una de las personas que conocí a través de este trabajo de tesis donde directa o indirectamente contribuyeron en esta tesis.

Resumen

La seguridad es una condición indispensable para el bienestar de los seres humanos, es por ello que la sociedad busca medidas que sean capaces de prevenir e identificar a las personas que cometan atentados en contra de su seguridad. Una de estas medidas es el uso de los sistemas de videovigilancia, ya que, permiten ver imágenes (de espacios específicos) de forma directa y cómoda, a través de internet, utilizando una computadora o cualquier otro dispositivo móvil. Sin embargo, algunas limitantes que pueden presentar las imágenes son la baja resolución, los cambios en la intensidad luminosa como la claridad o la oscuridad de la imagen, perspectiva de la imagen y escala, lo que se dificulta el reconocimiento de las personas presentes en el escenario.

Por ello, en el presente trabajo de tesis se presenta un sistema de visión artificial para el mejoramiento de rostros en imágenes de baja resolución adquiridas por medio de sistemas de videovigilancia de baja calidad.

El trabajo desarrollado considera todas las etapas de un sistema de visión artificial tradicional, permitiendo el ingreso de imágenes desde un archivo de la computadora, realiza el mejoramiento de la imagen por medio del algoritmo de súper resolución (llamado EDSR por sus siglas en inglés de Enhanced Deep Residual Networks), posteriormente localiza el rostro de la persona mediante Mediapipe y realiza la reconstrucción sólo del área del rostro.

La experimentación realizada muestra que el sistema mejora notablemente de manera visual la imagen; sin embargo, los resultados cuantitativos no reflejan dicha mejora al ser diferente a la imagen utilizada como *ground truth.* Durante el proceso de prueba también se aplicó un algoritmo de reconocimiento facial que logra una precisión del 94% al utilizar imágenes con mejoras y reconstrucciones en el área del rostro. Es crucial resaltar cómo esta mejora concreta ha permitido al sistema lograr resultados más precisos en su función de reconocimiento facial, lo que a su vez tiene el potencial de tener un impacto positivo en diversas aplicaciones.

Índice

Índice de figuras

Índice de tablas

Lista de acrónimos

BI: Interpolación Bicúbica.
BIQI: Blind Image Quality Index, por sus siglas en inglés.
CNN: Convolutional Neural Networks, por sus siglas en inglés.
CT: Computed Tomography, por sus siglas en inglés.
DNN: Red Neuronal Profunda, por sus siglas en inglés.
DWT: Discrete Wavelet Transform, por sus siglas en inglés.
EDSR: Enhanced Deep Residual Networks, por sus siglas en inglés.
ESPCN: Efficient Sub-Pixel CNN, por sus siglas en inglés.
FDDB: Face Detection Data Set and Benchmark, por sus siglas en inglés.
FSRCNN: Fast Super Resolution Convolutional Neural Network, por sus siglas en inglés.
FR: Full Reference, por sus siglas en inglés.
GFP-GAN: Generative Facial Prior-Generative Adversarial, por sus siglas en inglés.
GPU: Graphics Processing Unit, por sus siglas en inglés.
HR: High Resolution, por sus siglas en inglés.
LapSRN: Laplacian Pyramid Super-Resolution Network, por sus siglas en inglés.
LR: Low Resolution, por sus siglas en inglés.
MTCNN: Multi-Task Cascaded Convolutional Neural Networks, por sus siglas en inglés.
NCC: Normalized Cross-Correlation, por sus siglas en inglés.
NIQE: Natural Image Quality Evaluator, por sus siglas en inglés.
NR: Non Reference, por sus siglas en inglés.
PSF: Point Spread Function, por sus siglas en inglés.
PSNR: Peak Signal-to-Noise Ratio.
SSIM: Structural Similarity Index, por sus siglas en inglés.
SR: Súper resolución.
SRCNN: Super-Resolution Convolutional Neural Network, por sus siglas en inglés.
SSD: Single-Shot Object Detector, por sus siglas en inglés.
SVA: Sistema de Visión Artificial.
VDSR: Very-Deep Super-Resolution, por sus siglas en inglés

Capítulo 1 Introducción

1.1 Introducción

La seguridad es una condición indispensable para el bienestar de los seres humanos y el tema de inseguridad en México es conocido a través de noticieros, periódicos, redes sociales, etc. A nivel nacional, en junio de 2023, 62.3 % de la población de 18 años y más consideró inseguro vivir en su ciudad (INEGI, 2023).

La sociedad busca medidas para poder prevenir atentados en contra de su seguridad e identificar a las personas que cometan estos delitos y una de estas medidas es el uso de sistemas de videovigilancia. El rendimiento de dichos sistemas depende de varios aspectos, por ejemplo, la resolución que algunos poseen, la intensidad luminosa presente en la imagen para identificar a la persona en la imagen, la posición y altura de las cámaras, etc. Además, pueden presentar posiciones que oculten parcialmente el rostro (por ejemplo, mirar hacia el piso). Otra de sus dificultades es que las personas que cometen estos atentados buscan cubrirse el rostro mediante el uso de pasamontañas, gorras, cubrebocas, lentes, entre otros accesorios, por lo que su correcta identificación se vuelve aún más compleja.

Para dar solución a algunas de estas problemáticas, se desarrolló un sistema de visión artificial que mejora la calidad de las imágenes con problemas de iluminación y contraste. El sistema realiza diversas tareas como el aumentado de su tamaño y reconstrucción del área del rostro de la persona, con la intención de identificar cómo luce en una imagen que tenga mejor calidad.

Para ello, el problema principal del sistema es detectar el rostro en una imagen de baja calidad y mejorar esa zona.

La complejidad del problema es alta debido a los siguientes aspectos:

- La imagen puede presentar variabilidad en la iluminación y escala al considerar que el sistema debe funcionar en un entorno real.
- El tamaño de la imagen y la proporción del rostro con respecto a la imagen puede provocar la perdida de detalles o que se dificulte apreciar con claridad las facciones del rostro.
- La localización del rostro en la imagen es difícil ya que la cara puede tener un tamaño pequeño (por las diferentes escalas) lo que dificulta la visibilidad de los principales componentes faciales. Además de que la posición de la misma y/o el uso de algunos

aditamentos en el rostro (como lentes graduados o de sol, gorras, etc.) también pueden ocultar u ocluir los componentes del rostro.

1.2 Problema

El problema es detectar el rostro de una persona, que puede estar parcialmente ocluido, en la imagen analizada bajo problemáticas de iluminación, contraste y tamaño del mismo.

Este problema es originado debido a la poca o nula iluminación de la escena por elementos naturales como puede ser el sol o por elementos artificiales, que pueden provocar la aparición de sombras y la pérdida o ganancia de contraste en determinadas zonas de una imagen, así como la lejanía entre la localización de la cámara con la persona, dando así otra problemática de escala de la persona en relación con las dimensiones de la imagen.

1.3 Objetivos

Objetivo general

Desarrollar un sistema de visión artificial que realice el mejoramiento de imágenes de rostros de baja calidad.

Objetivos específicos

- Revisar en el estado del arte técnicas de súper resolución, detección de rostros y técnicas de mejoramiento de imágenes.
- Analizar herramientas para detectar un rostro en ambientes reales e implementar uno.
- Estudiar y seleccionar técnicas de mejoramiento de imágenes.
- Diseñar e implementar un sistema que realice el mejoramiento de las imágenes.
- Evaluar el sistema de mejoramiento de imágenes con métricas seleccionadas con base al estado del arte.

1.4 Alcances y limitaciones

Alcances

- ✓ El sistema es capaz de trabajar con rostros de baja calidad.
- ✓ El sistema trabaja con fondos complejos.
- ✓ El sistema considera que las imágenes pueden presentar diferentes condiciones de iluminación, baja resolución, perspectiva del rostro, escala, y desenfoque.

- ✓ Se trabaja con conjuntos de imágenes públicos.

Limitaciones

- ❖ No se garantiza la detección del rostro si este se encuentra cubierto u ocluido.
- ❖ Se busca el mejoramiento de la imagen para llevar a cabo la identificación de la persona, pero no se pretende implementar un sistema de reconocimiento facial.

1.5 Metodología

En esta sección se describe el enfoque para dar solución a este trabajo de tesis, para ello se revisaron varias técnicas de súper resolución, herramientas de localización de rostros, algoritmos de mejora de iluminación y un algoritmo de reconocimiento de personas.

Las técnicas que fueron implementadas y evaluadas, que permitieron desarrollar el método de solución, se describen a continuación. El enfoque final empleado en este trabajo fue estructurado como se muestra en la Figura 1.1.

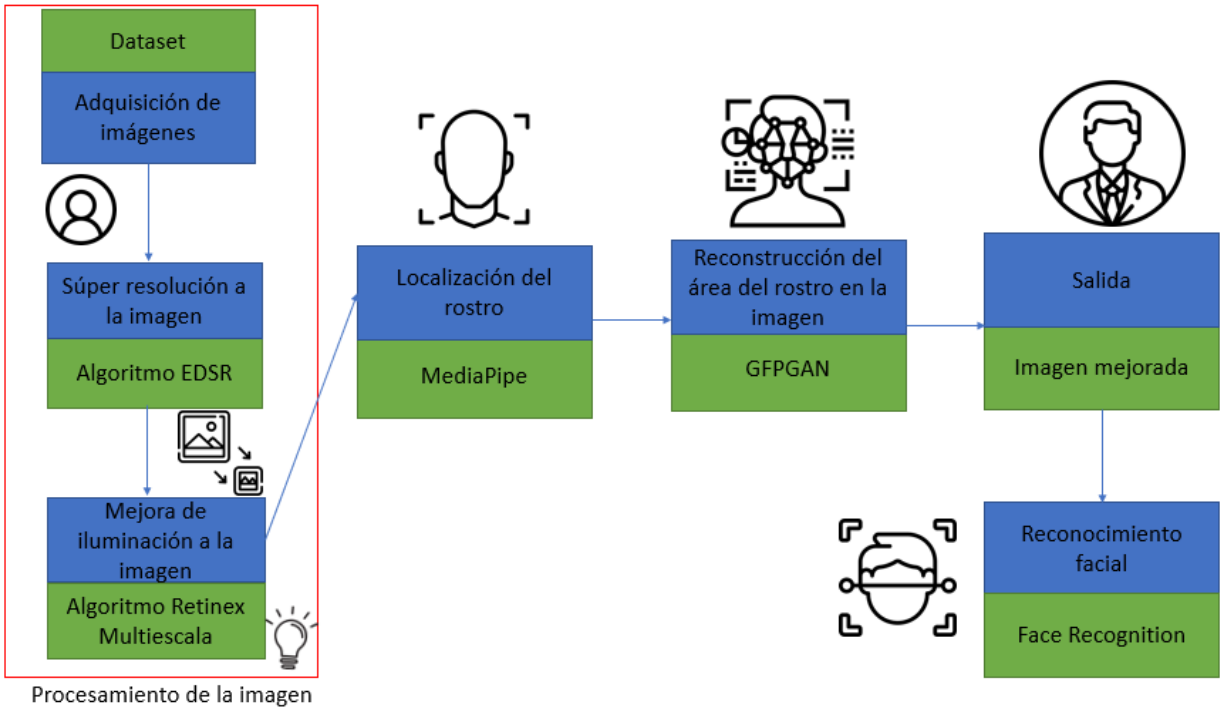

Figura 1.1 Esquema de metodología de solución empleada

1. **Adquisición de la imagen:** Se utilizaron tres conjuntos de imágenes públicas, dos de entornos reales de videovigilancia con la finalidad de experimentar y probar el sistema y el

tercer conjunto fue utilizado para probar la fase de reconocimiento facial y medir el nivel de mejora que obtuvo la imagen.

2. **Súper resolución a la imagen:** Para agregar más información a la imagen, se agregan más píxeles, con la finalidad de tener una imagen de mayor tamaño con una cantidad mayor de píxeles. En esta etapa, se implementaron cuatro algoritmos: **EDSR (**Enhanced Deep Residual Networs, por sus siglas en inglés) (Lim et al., 2017), **ESPCN (**Efficient Sub-Pixel CNN, por sus siglas en inglés) (Shi et al., 2016), **LapSRN** (Laplacian Pyramid Super Resolution Network por sus siglas en inglés) (Lai et al., 2018) y **FSRCNN** (Fast Super Resolution Convolutional Neural Network) (Dong et al., 2016) que no requieren un entrenamiento previo debido a que cuentan con modelos pre entrenados en la mejora de imágenes. El algoritmo que mejor desempeño tuvo fue **EDSR** y es el que el sistema usa para aplicar súper resolución en la imagen analizada.
3. **Mejora de la iluminación de la imagen:** Esta etapa tiene como objetivo mejorar la iluminación de la imagen, en el ámbito de la videovigilancia se cuenta con imágenes muy oscuras, muy brillantes o una combinación de ambas iluminaciones. Se estudió e implementó un algoritmo que, de acuerdo con el estado del arte, es el más utilizado para mejorar la iluminación de las imágenes, este es **Retinex multiescala** (Jobson et al., 1997). Con el uso de este algoritmo el brillo y contraste de las imágenes se equilibró.
4. **Localización del rostro:** Esta etapa tiene como objetivo localizar el rostro de la persona para poder proceder a la etapa de reconstrucción. Se implementaron cuatro herramientas: MediaPipe (MediaPipe, 2023), Dlib (Johnson et al, 2018), Detector Facial DNN (Agarwal, 2021) y MTCNN (Zhang et al., 2018), estas no requieren un entrenamiento previo ya que cuentan con modelos pre entrenados en la detección del rostro.
5. **Reconstrucción del área del rostro en la imagen:** Esta etapa tiene como objetivo reconstruir el rostro que ha sido localizado previamente. Se implementó la herramienta GFPGAN que no requiere un entrenamiento previo puesto que cuenta con dos modelos pre entrenados en la reconstrucción y detección de rostros.
6. **Salida:** El resultado final del sistema es la imagen que ha pasado por cada uno de los procesos mencionados anteriormente. La imagen del rostro de la persona con una mejor calidad es el resultado.
7. **Reconocimiento facial:** Esta etapa tiene como objetivo el verificar si después de que la imagen ha sido mejorada, la identidad de la persona se reconoce y si pese a la modificación de las facciones principales de la persona, se conserva su identidad.

1.6 Organización de la tesis

En el Capítulo 1 se aborda la problemática que llevó a desarrollar este proyecto, los objetivos y la metodología propuesta para la solución del problema.

El Capítulo 2 presenta el marco teórico de las técnicas y herramientas utilizadas en este trabajo.

El Capítulo 3 muestra el análisis del estado del arte y los trabajos relacionados realizados en el TecNM/CENIDET.

El Capítulo 4 contiene la experimentación realizada en este trabajo para cada una de las etapas de la metodología final propuesta; también se realiza la discusión de los resultados obtenidos en este mismo capítulo.

En el Capítulo 5 se proporcionan las conclusiones generales, los logros de los objetivos, las aportaciones, trabajos futuros y las actividades académicas realizadas en el presente trabajo.

En la sección de los anexos se puede encontrar el diseño del sistema y las herramientas utilizadas.

Capítulo 2 Marco teórico

A continuación, se presenta el marco teórico de las técnicas que fueron utilizadas en la metodología explicada en la sección 1.5. Además de las técnicas que fueron revisadas, implementadas y evaluadas.

2.1 Etapas de un sistema visión artificial

De acuerdo con (González & Woods, 1996) un Sistema de Visión Artificial (SVA) consta de cinco etapas, ver Figura 2.1.

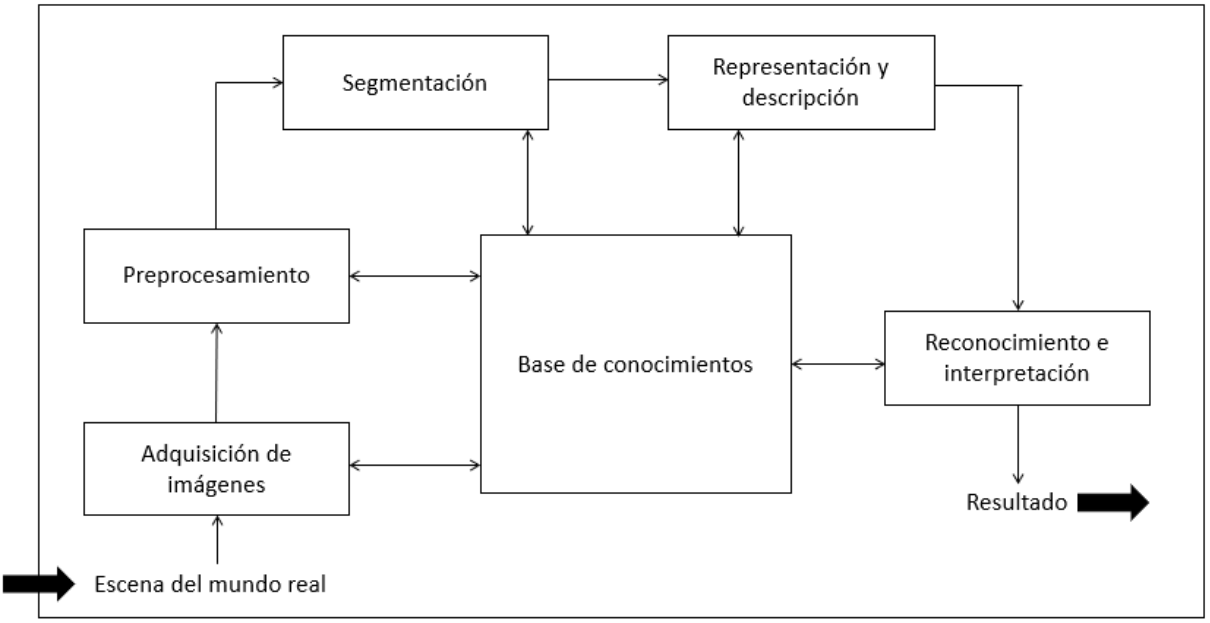

Figura 2.1 Módulos que constituyen un sistema de visión artificial (González & Woods, 1996)

Estas etapas se desarrollan en el sistema implementado en este trabajo de tesis y se estudiarán a detalle las técnicas utilizadas en cada fase.

2.2 Súper resolución

De acuerdo con (Alkanhal et al., 2020) la súper resolución es una técnica computacional utilizada para recuperar imágenes HR a partir de imágenes de LR. Es decir, a través de una imagen de baja calidad se busca aumentar su cantidad de píxeles con la finalidad de obtener más información de la imagen, dando como resultado una imagen en alta resolución. Los métodos se dividen en dos categorías: usar una sola imagen o varias imágenes. En el primer enfoque sólo se obtiene una

única imagen capturada y se mejora a través del aprendizaje de la relación entre baja y alta resolución. El segundo enfoque considera varias imágenes de baja resolución del mismo objeto y genera una imagen de alta resolución a través de la combinación de baja y alta resolución.

Este trabajo sigue el primer enfoque; es decir, se centra en implementar una técnica de súper resolución utilizando una sola imagen de baja calidad adquirida en un entorno de videovigilancia. Para la implementación de esta técnica, se evaluaron cuatro algoritmos de súper resolución de los más utilizados de acuerdo con el estado del arte presentado en el capítulo 3 en la sección 3.2, estos algoritmos fueron implementados con los modelos pre entrenados de cada uno, es decir ningún algoritmo fue entrenado con imágenes propias.

- **EDSR** (Lim et al., 2017): Es una red de súper resolución profunda, realizada con métodos de súper resolución actuales en el estado del arte. Trabaja con una sola imagen de entrada. Esta red obtiene una mejora significativa debido a la optimización mediante la eliminación de módulos innecesarios en las redes residuales. En la Figura 2.2 se presenta un ejemplo del resultado obtenido al utilizar este algoritmo.

Figura 2.2 Implementación de EDSR. a) Imagen con baja resolución b) Imagen con la implementación de EDSR.

- **ESPCN** (Shi et al., 2016)**:** Es la primera red neuronal convolucional capaz de realizar súper-resolución en tiempo real de vídeos de 1080p. Una eficiente capa de convolución sub-píxel fue introducida para que aprenda una serie de filtros de escalado para los mapas de características en la salida de las imágenes mejoradas. Ver en la Figura 2.3 el ejemplo de la aplicación del algoritmo.

Figura 2.3 Implementación de ESPCN. a) Imagen con baja resolución b) Imagen con la implementación de ESPCN.

- **FSRCNN** (Dong et al., 2016): Esta red fue creada con el objetivo de reducir el coste computacional y optimizar la red existente SRCNN. Esta red consiste en una capa de convolución, posteriormente un mapeo que se aprende directamente a partir de una imagen original de baja resolución. Esto genera imágenes con súper-resolución de una forma más rápida. Ver en la Figura 2.4 el ejemplo de la aplicación del algoritmo.

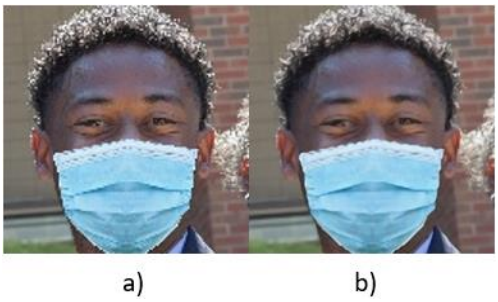

Figura 2.4 Implementación de FSRCNN. a) Imagen con baja resolución b) Imagen con la implementación de FSRCNN.

- **LapSRN** (Lai et al., 2018)**:** Esta es una red convolucional de súper-resolución de pirámide Laplaciana profunda, rápida y precisa de imágenes. Esta red reconstruye progresivamente los residuos de sub-bandas de imágenes de alta resolución en múltiples niveles de la pirámide. Fue entrenada con supervisión profunda utilizando las funciones de pérdida robustas de Charbonnier. Ver en la Figura 2.5 el ejemplo de la aplicación del algoritmo.

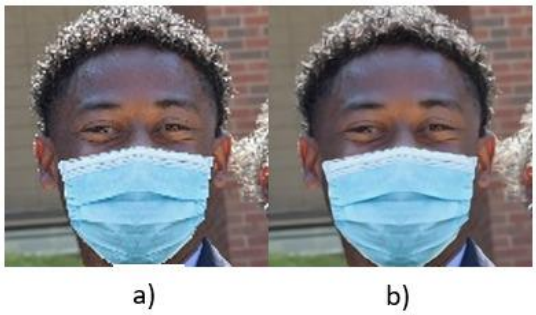

Figura 2.5 Implementación de LapSRN. a) Imagen con baja resolución b) Imagen con la implementación de LapSRN.

2.3 Mejora de iluminación de la imagen

Retinex Multiescala (Jobson et al, 1997):

Este algoritmo está basado en la teoría del color de Edwin H. Land (Land & McCann, 1971), donde se demuestra que el color se construye por medio de la comparación de reflectancias sobre superficies contiguas y, propuso un algoritmo que incluía recorridos de imágenes para calcular la luminosidad relativa. Surgieron trabajos más complejos como retinex multiescala (Jobson et al., 1997), donde se amplía el diseño anterior a una versión multiescalar, que logra la compresión simultánea de la gama dinámica y que logra simultáneamente la compresión del rango dinámico, la consistencia del color y la interpretación de la luminosidad. Esta extensión no produce una buena reproducción del color para un tipo de imágenes que contienen alteraciones de la suposición de mundo gris implícita en el fundamento teórico de la retina. Por lo tanto, se define un método de restauración del color que corrige esta deficiencia a costa de una modesta dilución en la consistencia del color. La escala más pequeña es fuerte en detalles y compresión de la gama dinámica y es débil en cuanto a la reproducción de los tonos y el color. Lo contrario ocurre con la escala más grande. El retinex multiescala combina los puntos fuertes de cada escala y mitiga los puntos débiles de cada una. En la Figura 2.6 se muestra un ejemplo de los componentes de retinex multiescala.

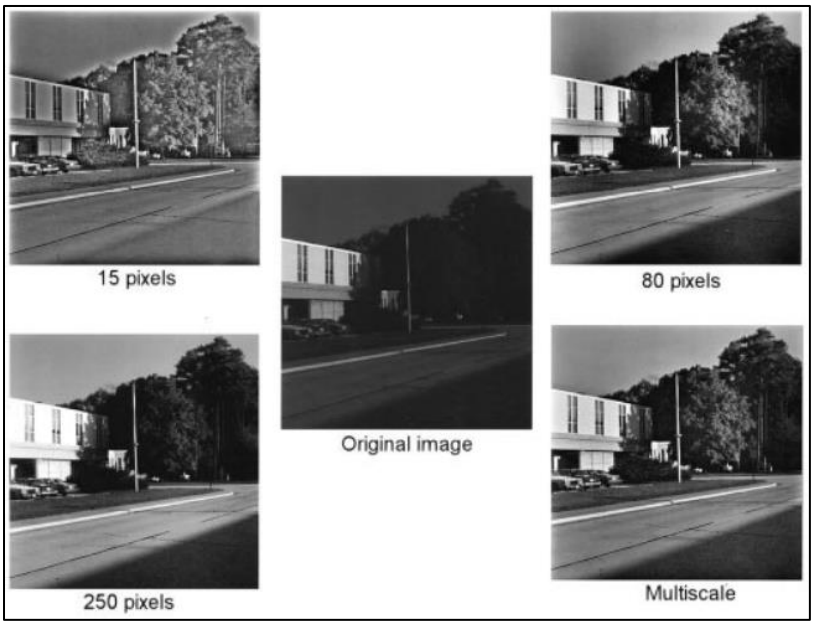

Figura 2.6 Componentes de retinex mutltiescala que muestran su contenido informativo complementario (Jobson et al, 1997)

2.4 Herramientas de detección de rostros

A continuación, se presenta la información de las herramientas de detección de rostros implementadas.

- **Dlib** (Johnson et al, 2018): Es una librería multipropósito, entre sus propósitos está la detección de rostros, tiene 68 puntos clave que detecta en el rostro, tomando el contorno de la cara, cejas, nariz y boca. Aunque está escrito en C++ tiene traductores de Python. En la Figura 2.7, se muestran los puntos clave que necesita para su detección, donde funciona utilizando características extraídas por el Histograma de Gradientes Orientados (HOG por sus siglas en inglés).

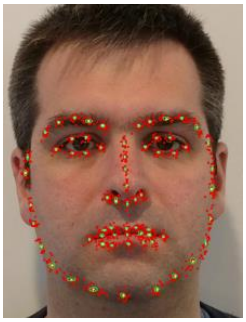

Figura 2.7 Puntos clave de detección de rostros para Dlib (Johnston et al, 2018)

- **MTCNN** (Zhang et al., 2018): Utiliza una estructura en cascada con tres etapas de CNN. En primer lugar, utilizan una red convolucional completa para obtener las ventanas candidatas y sus vectores de regresión de caja delimitadora, y los candidatos muy superpuestos se superponen utilizando la supresión *on-máxima* (NMS). A continuación, estos candidatos se pasan a otra CNN que rechaza un gran número de falsos positivos y efectúa la calibración de las cajas delimitadoras. En la etapa final, se ejecuta una detección de puntos de referencia faciales. En la Figura 2.8 se muestran los 5 puntos clave que toma MTCNN para su detección.

Figura 2.8 Puntos clave de detección de rostros para MTCNN (Zhang et al, 2016)

- **Detector facial DNN en OpenCV** (Agarwal, 2021): Es un modelo de Caffe que se basa en el Detector de Tiro Único-Múltiple (SSD) y utiliza la arquitectura ResNet-10. Fue introducido después de OpenCV 3.3 en su módulo de redes neuronales profundas. También hay una versión cuantificada de Tensorflow que se puede utilizar. En la Figura 2.9, se muestra un ejemplo del resultado de este modelo implementado.

Figura 2.9 Ejemplo de detección de rostro con el detector facial DNN en OpenCv (Rosebrock. 2018)

- **MediaPipe** (MediaPipe, 2023)**:** Es una solución de detección facial ultrarrápida que soporta múltiples rostros. Se basa en BlazeFace (Bazarevsky et al., 2019)**,** un detector de rostros liviano y de buen rendimiento. Cuenta con 6 puntos clave para la detección de rostros. A causa de estos 6 puntos, se deriva una malla facial de 480 puntos que toma en cuenta todo el contorno de la cara, cejas, nariz, boca y ojos. En la Figura 2.10 se muestran los puntos clave que necesita MediaPipe.

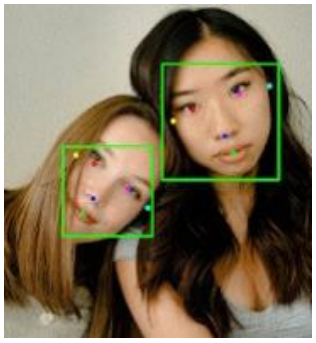

Figura 2.10 Puntos clave de detección de rostros para MediaPipe (MediaPipe, 2020)

2.5 Reconstrucción de imágenes

GFP-GAN (Wang et al., 2021)**:** Son las siglas de Generative Facial Prior-Generative Adversarial Network, es una red neuronal capaz de restaurar retratos dañados y de baja resolución. Combina información de dos modelos pre entrenados, para completar los detalles realistas faltantes, mientras mantiene una alta precisión y calidad. Este enfoque utiliza una versión previamente entrenada de un modelo existente (StyleGAN-2 de NVIDIA) para informar el propio modelo creado por la red neuronal. La técnica tiene como objetivo preservar la "identidad" de las personas en una foto, con un enfoque particular en los rasgos faciales como los ojos y la boca. En la Figura 2.11, se muestra un ejemplo del resultado de aplicar GFP-GAN a una imagen.

Figura 2.11 Ejemplo de la implementación de GFP-GAN (Wang et al, 2021)

2.6 Métricas de evaluación

Esta sección se divide en dos, las métricas investigadas para medir el mejoramiento de las imágenes y las métricas para medir el rendimiento de los algoritmos de detección de rostros.

A continuación, en la Tabla 2.1 se presentan 26 métricas que evalúan la calidad de la imagen. Donde se detalla la métrica, lo que evalúa y si es de referencia completa (FR) o no tiene referencia (NR). Que tenga referencia completa significa que necesita de una imagen de referencia para evaluar el nivel de mejora y, que no tenga referencia significa que no toma ninguna imagen como referencia para medir el nivel de mejora, también se muestra el número de citas con el que cuenta cada métrica y los autores que la utilizan, en algunos casos no se encontró la referencia original, por lo que se seleccionó a los autores con mayor número de citas evaluando la métrica.

Tabla 2.1 Métricas revisadas en el estado del arte que miden la calidad de la imagen

Métrica	¿Qué evalúa?	FR O NR	No. Citas	Autores
Squared Error (MSE)	Error cuadrado medio entre la imagen original y la imagen procesada (no es válido como criterio de calidad para imágenes procesadas para el sistema de la visión humana).	FR	4,791	(Sara U. et al, 2019), etc.
Peak Signal-to-Noise Ratio (PSNR)	Índice de evaluación objetivo más común y ampliamente utilizado para las imágenes, sin embargo, se basa en el error entre los píxeles correspondientes, es decir, se basa en una evaluación de calidad de imagen sensible al error (cuanto mayor sea el valor de PSNR, menor será la distorsión).	FR	5,068	(Huynh-Thu, et al, 2008). (Hore, A., & Ziou, D, 2010), (August et al, 2008), etc.
Signal-to-Noise Ratio (SNR)	Resultado de dividir la media por la desviación típica de los fotones registrados. Cuanto mayor sea la SNR, mejor es la calidad de la imagen.	FR	1,384	(Yuan et al, 2019).
Structural Content (SC)	Se define como la relación entre el cuadrado de la suma de la imagen original y la imagen de referencia.	FR	105	(Zhang et al, 2010), (Jain et al, 2011), etc.
Maximum Difference (MD)	Proporciona el máximo de la señal de error (es decir, la diferencia entre la imagen procesada y la imagen de referencia). Cuanto mayor sea el valor de la diferencia máxima, peor será la calidad de la imagen.	FR	381	(Penny, D., & Hendy, M. D. 1985), (Rao, T. V. N., & Govardhan, A. 2013), etc.
Average Difference (AD)	Proporciona la media del cambio respecto a la imagen procesada y la de referencia **(Lo ideal es que sea cero).**	FR	266	(Memon, F., Unar, M. A., & Memon, S. 2015).
Normalized Absolute Error (NAE)	Se define como la relación entre la suma del valor absoluto de la imagen de diferencia y el valor absoluto de la imagen original. Un valor de NAE más alto indica que la imagen es de mala calidad.	FR	1,035	(Mendo, L. 2009), (Avcibas, et al, 2002), etc.
R-Averaged Maximum Difference (RAMD)	La diferencia máxima promedio es calculado a la suma del valor máximo del número R y dividido por R. Depende del valor que se quiera asignar.	FR	181	(Galbally, J., & Marcel, S. 2014), Saranya et al, 2016), etc.

Tabla 2.1 Métricas revisadas en el estado del arte que miden la calidad de la imagen

Métrica	¿Qué evalúa?	FR O NR	No. Citas	Autores
Laplacian Mean-Squared Error (LMSE)	La relación entre el cuadrado de las diferencias de dos valores a la suma del valor real de la imagen.	FR	254	(Laparra et al, 2016), (Galbally, et al, 2014), etc.
Normalized Cross-Correlation (NCC)	Las imágenes se normalizan para variar el brillo de la imagen y la plantilla debido a la exposición y las condiciones de iluminación. Se estima en cada paso restando la media y dividiendo la desviación estándar.	FR	570	Galbally et al, 2014), (Tsai, et al, 2003), (Pei, L., Xie, Z., & Dai, J. 2010), etc.
Mean Angle Similarity (MAS)	Es la medida de similitud de ángulo medio entre la imagen real y la imagen de referencia.	FR	544	(Qian et al, 2004), etc.
Mean Angle Magnitude Similarity (MAMS)	Es la medida de similitud de la magnitud del ángulo medio entre la imagen real y la imagen de referencia.	FR	1,015	Galbally et al. (2014), (Avcibas et al, 2002), etc.
Total Edge Difference (TED)	Se denota como la relación entre el número total de diferencias de bordes de las dos imágenes con el número total de píxeles.	FR	525	(Miškuff et al, 2007), (Ahmad et al, 2012), etc.
Total Corner Difference (TCD)	Se define como la relación entre el número total de diferencias de esquina entre las dos imágenes con respecto al número total de píxeles.	FR	177	(Galbally, J., & Marcel, S. 2014), etc.
Spectral Magnitude Error (SME)	La varianza entre la transformada de Fourier de la imagen real a la transformada de Fourier de imagen de referencia se promedia utilizando el número total de píxeles.	FR	638	Pravallika et al, 2016), (Raheem et al, 2019), etc.
Spectral Phase Error (SPE)	La varianza entre las imágenes reales transformadas por el ángulo de Fourier a la imagen de referencia transformada en ángulo de Fourier se promedia utilizando el número total de píxeles.	FR	289	(Galbally et al, 2014), (O'leary et al, 1991), etc.

Tabla 2.1 Métricas revisadas en el estado del arte que miden la calidad de la imagen

Métrica	¿Qué evalúa?	FR O NR	No. Citas	Autores
Gradient Magnitude Error (GME)	La varianza entre el gradiente de la imagen real al gradiente de la imagen de referencia es promediada utilizando el número total de píxeles.	FR	3,839	Zhang, L., Mou, X., & Zhang, D. (2011).
Gradient Phase Error (GPE)	La varianza entre el ángulo de gradiente de la imagen real al ángulo de gradiente de la imagen de referencia se promedia utilizando el número total de píxeles.	FR	1,540	(Moussavi et al, 2014), (Wahl et al, 1994), etc.
Structural Similarity Index Measure (SSIM)	Índice de evaluación de calidad de imagen de referencia completa, que mide la similitud de la imagen en términos de brillo, contraste y estructura.	FR	4,930	(Wang et al, 2003)
Visual Information Fidelity (VIF)	La medida de VIF está en la base de la hipótesis de que las imágenes visuales humanas son un escenario natural y, por lo tanto, tienen propiedades estadísticas del mismo tipo.	FR	3,702	(Sheikh, H. R., & Bovik, A. C. 2006).
Reduced Reference Entropic Difference (RRED)	Mide la diferencia de contenido de información local entre la imagen de referencia y el pronóstico de la imagen poco clara. Estima la varianza promedio entre entropías locales calculadas de coeficientes de ondícula de referencia e imágenes poco claras pronosticadas de forma dispersa.	FR	349	(Soundararajan, R., & Bovik, A. C. 2011).
JPEG Quality Index (JQI)	Estima la calidad en imágenes exageradas por el bloque habitual artificial que se encuentra en muchas series de algoritmos de compresión a bajas tasas de bits, como el JPEG.	NR	113	(Babu, R. V., Suresh, S., & Perkis, A. 2007).
High-Low Frequency Index (HLFI)	Es sensible a la nitidez de la imagen calculando la diferencia entre la potencia en las frecuencias inferiores y superiores del espectro de Fourier.	NR	177	(Galbally, J., & Marcel, S. 2014).

Tabla 2.1 Métricas revisadas en el estado del arte que miden la calidad de la imagen

Métrica	¿Qué evalúa?	FR O NR	No. Citas	Autores
Blind Image Quality Index (BIQI)	Utiliza un conocimiento previo tomado de la escena natural de la imagen libre de alteración. La razón de ser detrás de esta tendencia cuenta con la hipótesis de que las imágenes claras del mundo presentan naturalmente ciertas propiedades regulares que están dentro de un subespacio asegurado de todas las imágenes posibles. Si se calcula correctamente, la desviación de la regularidad puede ayudar a estimar la calidad perceptiva de la imagen.	NR	1,151	(Moorthy, A. K., & Bovik, A. C. 2010).
Natural Image Quality Evaluator (NIQE)	Se utiliza para analizar la calidad de la imagen ciega sobre la base de crear conciencia de calidad al recopilar características de estadísticas aliadas a muchas variaciones.	NR	830	Galbally, J., & Marcel, S. (2014), (Zhang, L., & Bovik, A. C. 2015).
Gradient Similarity Metric (GSM)	Calcula el mapa de calidad local (LQM). Después de calcularlo, se calcula una única puntuación de calidad utilizando la desviación estándar como estrategia de agrupación. El GSM es mucho más rápido que la mayoría de los demás parámetros de calidad de imagen.	NR	713	(Liu, A., Lin, W., & Narwaria, M. 2011).

Se seleccionaron cuatro métricas de las 26 investigadas, estas cuatro métricas fueron seleccionadas de acuerdo con el análisis del estado del arte, dónde se encontró que fueron ampliamente utilizadas por diversos trabajos relacionados, además de los criterios que evalúan. Se consideraron dos métricas con referencia (FR) y, dos métricas sin referencia (NR).

A continuación, se detalla a profundidad cómo funciona cada una de las métricas seleccionadas, así como la fórmula para calcular cada una de ellas.

- **SSIM** (Wang et al., 2004): Mide la similitud de la imagen en términos de brillo, contraste y estructura. Es una métrica que necesita de una imagen de referencia para evaluar la calidad de la imagen que se proporcione. Sus valores están en el rango de 0 y 1, siendo 1 el más alto y 0 el más bajo. La forma de calcular esta métrica se muestra en la ecuación 2.1:

$$SSIM(x,y) = \frac{(2\mu_x\,\mu_y + C_1)\,(2\,\sigma_{xy} + C_2)}{(\frac{\mu_x}{2} + \frac{\mu_y}{2} + C_1)(\sigma_y^2 + C_2)} \qquad (Ec\ 2.1)$$

μ_x es la media de la muestra de píxeles de x, μ_y es la media de píxeles de y y σ_y^2 es la varianza de y, σ_y es la covarianza de y y, C_1 y C_2 son dos variables para estabilizar la división con denominador débil, mientras que x representa las columnas de píxel de la imagen y y las filas de píxeles.

- **NCC** (Yoo et al., 2009)**:** Las imágenes se normalizan en términos de brillo, exposición y condiciones de iluminación. Se calcula restando la media de la imagen de entrada con la imagen de referencia y dividiendo la desviación estándar de la imagen de referencia, como se muestra en la ecuación 2.2. Es una métrica que necesita de una imagen de referencia para evaluar la calidad de la imagen que se proporciona. Sus valores están en el rango de 0 y 1, siendo 1 el mejor valor y 0 el peor valor.

$$NCC = \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{(A_{ij} - B_{ij})}{A^2 ij} \qquad (Ec\ 2.2)$$

Donde *A* y *B* son dos matrices de la imagen original y la imagen mejorada, ij son el valor de las posiciones de la columna y fila en píxeles de la imagen.

- **NIQE** (Wu et al., 2021)**:** Se utiliza para analizar la calidad de la imagen ciega sobre la calidad de la imagen al recopilar características de estadísticas aliadas a muchas variaciones. Es una métrica de no referencia. Sus valores están en el rango de 0 y 1, siendo 1 el mejor valor y 0 el peor valor. Compara con un modelo por defecto calculado a partir de imágenes de escenas naturales.
- **BIQI** (Moorthy et al., 2010)**:** Utiliza un conocimiento previo tomado de la escena natural de la imagen libre de alteración. Considera que las imágenes claras del mundo presentan naturalmente ciertas propiedades regulares que están dentro de un subespacio de todas las imágenes posibles. Si se calcula correctamente, la desviación de la regularidad puede ayudar a estimar la calidad perceptiva de la imagen. Es una métrica que no necesita referencia, es decir, no necesita ninguna imagen de comparación para evaluar la calidad de la imagen proporcionada. Sus valores están en el rango de 0 y 1, siendo 1 el mejor valor y 0 el peor valor. La forma de calcular esta métrica se muestra en la ecuación 2.3:

$$BIQI = \sum_{i=1}^{5} \rho i\ x\ qi \qquad (Ec\ 2.3)$$

$\boldsymbol{\rho}$ es el valor de todas las columnas de píxel, mientras que $\boldsymbol{q}$ representa el valor de las filas, lo que significa que realiza una sumatoria de todas las columnas del píxel.

La segunda parte de esta sección trata de las métricas de clasificación consideradas para medir el rendimiento de las herramientas de localización y reconocimiento de rostros:

- **Matriz de confusión** (Barrios, 2019): Es una herramienta que permite visualizar cómo se ha desempeñado algún algoritmo donde, en cada columna de la matriz se representa el número de predicciones que obtuvo cada clase y, en cada fila se representan las instancias en la clase real. Con motivo de esta representación se pueden saber los aciertos y errores que un modelo ha tenido. En la Figura 2.12 se muestra la representación de esta matriz.

Figura 2.12 Matriz de confusión (Barrios, 2019)

La matriz de confusión está conformada por 4 valores las cuales son:

  - **Verdadero positivo**: El valor real es positivo y la prueba predijo también que era positivo.
  - **Verdadero negativo:** El valor real es negativo y la prueba predijo también que el resultado era negativo.
  - **Falso negativo:** El valor real es positivo, y la prueba predijo que el resultado es negativo.
  - **Falso positivo:** El valor real es negativo, y la prueba predijo que el resultado es positivo.

A causa de la matriz de confusión se pueden calcular las siguientes métricas:

- **Accuracy** (Barrios et al., 2019)**:** Esta métrica hace referencia a lo cerca que está el resultado de la medición del valor verdadero, es decir, está relacionada con el sesgo de

una estimación y, es representada como la proporción de resultados verdaderos (positivos y negativos) entre el número total de casos examinados (positivos y negativos). La forma de calcular esta métrica se muestra en la ecuación 2.4:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (Ec\ 2.4)$$

Las variables de la fórmula son las detalladas en la matriz de confusión. Su valor final está en un rango de 0 a 1 o de 0 a 100, dependiendo de la normalización aplicada.

- **Recall** (Barrios, 2019): Esta métrica hace referencia a la sensibilidad y la especificidad que son dos valores que indican la capacidad del modelo para discriminar los casos positivos, de los negativos. La sensibilidad es representada como la fracción de verdaderos positivos, mientras que la especificidad, es la fracción de verdaderos negativos y, es representada como la proporción de resultados verdaderos positivos entre la misma proporción de resultados verdaderos positivos más el valor de los falsos negativos obtenidos. La forma de calcular esta métrica se muestra en la ecuación 2.5:

$$Recall = \frac{TP}{TP + FN} \qquad (Ec\ 2.5)$$

Las variables de la fórmula son las detalladas en la matriz de confusión. Su valor final está en un rango de 0 a 1.

- **Precision** (Barrios, 2019)**:** Hace referencia a la dispersión de un conjunto de valores obtenidos a partir de mediciones repetidas en una magnitud. Cuanto menor es la dispersión mayor es su precisión y, es representada como la proporción de resultados verdaderos positivos entre la misma proporción de resultados verdaderos positivos más el valor de los falsos positivos obtenidos. La forma de calcular esta métrica se muestra en la ecuación 2.6:

$$Precision = \frac{TP}{TP + FP} \qquad (Ec\ 2.6)$$

Las variables de la fórmula son las detalladas en la matriz de confusión. Su valor final está en un rango de 0 a 1.

- **F1-Score** (Barrios, 2019): Es una métrica muy utilizada, ya que resume la precisión y el recall del modelo en una sola métrica. En el caso de tener clases con distribuciones desiguales resulta de gran utilidad. Básicamente es una media ponderada de la precisión y recall. La forma de calcular esta métrica se muestra en la ecuación 2.7:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recal} \qquad (Ec\ 2.7)$$

Capítulo 3 Estado del Arte

El presente capítulo contiene la revisión de la literatura referente a técnicas de súper resolución, herramientas de localización de rostros y algoritmos de mejora de iluminación. Así como trabajos previos realizados en el TecNM/CENIDET referentes a algoritmos de súper resolución, mejoramiento de imágenes y detección de rostros.

3.1 Trabajos relacionados

En esta sección se presentan los trabajos realizados en el TecNM/CENIDET referentes a algoritmos de súper resolución, mejoramiento de imágenes y detección de rostros.

Cejudo (2020) en su trabajo “Desarrollo de un FrameWork para la experimentación con algoritmos de súper resolución”, desarrolló un framework que permite a los usuarios obtener imágenes de alta resolución a partir de imágenes de baja resolución. En esta tesis se ejecutaron cinco algoritmos clásicos de súper resolución: Interpolación del vecino más cercano, bilineal, bicúbica y Lanczos, así como wavelet haar, también se permite realizar funciones de lectura y escritura de formatos gráficos como: BMP, JPG, PNG, TIIF y de archivos .txt, para que en ellos se puedan almacenar comandos para posteriormente ser ejecutados por el intérprete. Así como la evaluación de las imágenes mediante cuatro métricas MSE, RMSE, PSNR y SSIM.

En el trabajo de Valderrama (2019) titulado “Reconocimiento automático del rostro para verificación de identidad para evaluación en línea” se desarrolló un sistema de visión artificial que realiza reconocimientos faciales de manera automática, para la verificación de la identidad de aspirantes que presentan un examen de admisión en línea. El sistema se integra de cinco algoritmos de preprocesamiento que son evaluados para concluir cuál es el que obtiene mejores resultados. Para la verificación de la identidad, se compararon cuatro clasificadores: máquinas de vector soporte (SMV), perceptrón multicapa, redes bayesianas y random forest, concluyendo que la combinación del algoritmo propuesto en el procesamiento (foto receptores con retinex multi-escala) y el clasificador SVM alcanzan el mejor rendimiento.

En el trabajo de Arismendi (2021) titulado “PCNN en la mejora de imágenes a color” se implementó una red neuronal de pulso acoplado (PCNN), para el mejoramiento de imágenes médicas digitales a color. Se implementaron varios modelos de redes neuronales para su evaluación y así elegir con qué modelo debería desarrollarse este trabajo. Posteriormente, se realizaron experimentos con cada

modelo utilizando imágenes generadas de forma artificial, también se implementaron diversas técnicas de ajuste automático de parámetros. Por último, se evaluaron y compararon los resultados obtenidos para comprobar el mejoramiento de las imágenes.

3.2 Estado del arte

La sección se integra del análisis de 31 artículos relacionados con técnicas de algoritmos de súper resolución, técnicas de mejoramiento de imágenes y detección de rostros en ambientes de videovigilancia por medio de videos e imágenes.

3.2.1 Técnicas de súper resolución

Súper resolución de imágenes: Las técnicas, las aplicaciones y el futuro (Yue et al., 2016)

En este trabajo de investigación se habla de cómo la súper resolución (SR) se ha desarrollado durante más de tres décadas, tanto la SR multi-frame como la single-frame. El objetivo de este artículo es ofrecer una revisión de la SR desde el punto de vista de las técnicas, las aplicaciones, y en especial de las principales contribuciones de los últimos años. Para obtener una imagen de alta resolución, una de las soluciones es desarrollar dispositivos ópticos más avanzados. Los principales avances en las técnicas de SR pueden dividirse en tres etapas. En la primera década, los investigadores cambiaron los métodos del dominio de la frecuencia a los algoritmos del dominio espacial. En la segunda etapa, el marco de la SR regularizada multi-frame fueron el principal foco de atención. Sin embargo, en los últimos años el desarrollo de la SR multi-frame se ha ralentizado, y los buscadores se han centrado principalmente en la reconstrucción de la SR en los distintos campos de aplicación. Algunas de sus aplicaciones son: Mejora periódica de la información de vídeo vigilancia, diagnósticos médicos, teledetección para la observación de la Tierra, observación astronómica e identificación de información biométrica. Los autores concluyen en que, en el ámbito de la videovigilancia se obtienen resultados de reconstrucción eficaces en términos de verosimilitud visual, y obtienen bordes más nítidos, sin embargo, se cuentan con capacidades limitadas para modelar texturas visualmente complejas.

Súper resolución de imágenes: un estudio (Sha et al., 2012)

En este estudio se menciona cómo la mejora de la calidad de la fotografía siempre ha sido una cuestión de tecnología de la imagen. Para algunas aplicaciones es esencial la calidad, por ejemplo, en el departamento forense, donde la imagen debe mejorarse para extraer información diminuta incrustada en la misma, como el rostro de un delincuente, la matrícula de una licencia, etc. El aumentar el tamaño de la imagen no es la solución ya que sólo da lugar a una imagen borrosa. La

causa más posible de este problema es la limitación del hardware que incluye principalmente la velocidad de muestreo del dispositivo de carga acoplada y esto se vuelve crítico cuando se captura un objeto en movimiento a gran velocidad. Es por ello que en este artículo se distinguen algunos de los enfoques de la súper resolución, estas técnicas se clasifican para grandes rasgos en dominio de la frecuencia o basadas en el dominio espacial. Los métodos del dominio de la frecuencia se basan en tres principios fundamentales, la propiedad de desplazamiento de la transformada de Fourier, la relación de *aliasing* entre la transformada continua de Fourier y la transformada discreta de Fourier. Mientras que técnicas basadas en el dominio espacial se basan en Máximo a posterior, enfoques basados en el aprendizaje, enfoque basado en wavelets, enfoque de retroproyección iterativa, enfoque basado en la dispersión y proyección sobre conjunto convexo. La mayoría de los enfoques utilizan una secuencia de imágenes de baja resolución para extraer una imagen de súper resolución.

Algoritmo de súper resolución de una sola imagen basado en la autosimilitud estructural y las características de los bloques de deformación (Chen et al., 2019)

En este artículo se ha propuesto un algoritmo de súper resolución de una sola imagen basado en el modelo de escala y características de deformación (SISR-SMDF). En primer lugar, el algoritmo SISR-SMDF propuesto puede construir el modelo de escala, ampliar el tamaño del espacio de búsqueda tanto como sea posible, y superar el defecto causado por la limitación de los recursos de entrenamiento de súper resolución de una sola imagen. En segundo lugar, el tamaño del diccionario interno se incrementa con las características de deformación de la muestra. Por último, con el fin de la restauración de imágenes, se ha utilizado el método de aprendizaje disperso de grupo para reconstruir la imagen. Los resultados experimentales demuestran que con su propuesta se puede obtener un rendimiento superior en comparación con algoritmos conocidos como el de Brainfield, interpolación bicúbica (BI), la codificación dispersa (SC), la red convolucional recursiva profunda Convolucional Recursiva Profunda (DRCN), la Red de súper resolución Profunda Multiescala (MDSR), la Red Convolucional de Super-Resolución (SRCNN) y Variación Direccional Total Generalizada de segundo orden (DTGV). Las imágenes de súper resolución con efectos visuales subjetivos, pero con una mayor evaluación objetiva se pueden obtener mediante el método propuesto. En comparación con los algoritmos existentes, la red estructural converge más rápidamente, los efectos de reconstrucción de bordes y texturas de la imagen mejoran de forma evidente, y la evaluación de la calidad de la imagen, como el pico de señal y como la relación señal-ruido máxima (PSNR), el error cuadrático medio (RMSE) y la similitud estructural (SSIM), son también superiores y populares en la evaluación de imágenes.

Human face super-resolution on poor quality surveillance video footage (Farooq et al., 2021)
Este artículo narra sobre la problemática que se tiene al no utilizar imágenes de baja calidad naturales en súper resolución ya que, la mayoría de los artículos reportan una baja calidad (LR) sintética generada a partir de imágenes de alta resolución (HR). Enfatizando en cómo este tipo de trabajos no son útiles cuando se trabaja en sistemas de videovigilancia.

En el artículo se propone un método de aprendizaje de transferencia de estilo para enfrentar súper resolución utilizando conjuntos de datos de baja y alta resolución que comparten propiedades importantes. A comparación de otros enfoques, no se requirió un par de imágenes HR-LR alineadas de forma sintética, ni la reconstrucción de súper-resolución a partir de imágenes de baja resolución. El método propuesto requiere una imagen única, con ello se demostró que la idea de generar conjuntos datos HR-LR no alineados y apareados conduce a capacidades de reconstrucción de SR sorprendentemente fuertes cuando se combina con un medio efectivo de aprendizaje no supervisado. Se encontró que los conjuntos de datos de imágenes HR-LR del mundo real producen mejores modelos súper resolución que los pares degradados HR-HR cuando la resolución y la tasa de bits del flujo LR son bajas.

Estudio comparativo de algoritmos de súper resolución de una sola imagen basados en aprendizaje profundo (Ochoa Domínguez et al., 2020)
Este trabajo presenta un estudio comparativo de cuatro métodos recientes del estado del arte de súper resolución utilizando aprendizaje profundo y un algoritmo como punto de referencia para la comparación. Los cinco métodos evaluados son: Red Neuronal Convolucional para Súper Resolución (SRCNN), Red Neuronal Convolucional Rápida para súper resolución (FSRCNN), Red muy profunda para súper resolución (VDSR) y Red Profunda y Recursiva para súper resolución (DRCN) y el algoritmo de interpolación bicúbica.

Los algoritmos comparados fueron implementados en Matlab y Pytorch. Los conjuntos utilizados para la fase de pruebas fueron Set5, Set14, Urban100 y BSD100, los cuales son ampliamente usados para las pruebas en trabajos de SR. Para comparar el rendimiento de los algoritmos de SR se manejaron las métricas Proporción Máxima de Señal a Ruido (PSNR) e Índice de Similitud Estructural (SSIM).

Cualitativamente considerando la similitud estructural y cuantitativamente con la relación señal-ruido máxima, el método de interpolación bicúbica proporciona los resultados más modestos. Los métodos muy profundos obtienen mejores resultados en comparación con las arquitecturas poco

profundas. Los métodos muy profundos logran, en promedio, mejorar el PSNR al compararlos contra la interpolación bicúbica en 2.39 dB (decibelios), mientras que los métodos poco profundos aventajan a la interpolación bicúbica en 2.17dB. Los métodos VDSR y DRCN tienen rendimientos similares; pero superiores a los métodos poco profundos, la arquitectura DRCN es más compacta que VDSR. Los menos profundos tienen una menor cantidad de capas, esto implica una menor complejidad computacional.

Image super-resolution by extreme learning machine (Bhanu et al., 2012)

En esta investigación se trabajó con súper resolución y sus diferentes enfoques. Uno de ellos es el algoritmo tradicional donde se requieren múltiples imágenes en baja resolución para generar una imagen de alta resolución integrando la información de todas las imágenes anteriormente requeridas. Otro tipo de algoritmo es donde solo se requiere una sola imagen de baja resolución como entrada. La desventaja de estos tipos de algoritmos es que se degradan fácilmente cuando el factor de aumento es grande, es por eso por lo que se propuso un algoritmo basado en una máquina de aprendizaje extrema que consta de dos pasos: entrenamiento y prueba. En la fase de entrenamiento las características se extraen de imágenes de alta resolución inicialmente interpoladas y se aprende un modelo que asigna las imágenes interpoladas a los componentes de alta frecuencia. Para la fase de prueba se estiman estos componentes de alta frecuencia utilizando el modelo aprendido durante el entrenamiento. Mediante la imagen interpolada y los componentes de alta frecuencia se genera fielmente una imagen con suficientes detalles.

Súper resolución de imágenes en color con poca luz mediante un sensor RGB/NIR (Honda et al., 2018)

En esta investigación, se propuso un método para la súper resolución de imágenes en color de baja resolución tomadas en escenas con poca luz. Este método se basa en la técnica de súper resolución multiframe, que fusiona múltiples imágenes de baja resolución tomadas en diferentes posiciones de la cámara para sintetizar una imagen en color de alta resolución. La información NIR (información infrarroja cercana) se pudo capturar con claridad, esto permitió reducir eficazmente los artefactos de imagen causados por el ruido y el desenfoque de movimiento al realizar la súper resolución de imágenes con poca luz. En la Figura 3.1 se muestra un ejemplo de los problemas comunes que enfrentan las imágenes de baja calidad que además cuentan con deficiencia de luz. Por ello, se buscó realizar un método que fuera capaz de resolver esta problemática y así observar partes específicas de una imagen con claridad.

Figura 3.1 Problemas comunes en imágenes con poca luz y baja resolución (Honda et al., 2018)

Se realizó una eliminación de ruido y una imagen en color de alta resolución. Se realizaron experimentos con imágenes reales demostrando cómo el método propuesto fue más eficaz con respecto a otros métodos reportados en el estado del arte. Los autores resaltan que este trabajo es el primer intento de realizar súper resolución en imágenes en color con poca luz.

Research for Face Image Super-Resolution Reconstruction Based on Wavelet Transform and SRGAN (Cao et al., 2021)

En este trabajo, se propone un método de reconstrucción de imágenes faciales basado en la transformada wavelet y la red generativa adversarial de súper-resolución (SRGAN), para reducir el impacto de la imagen de baja resolución causada por el hardware de imagen, el ancho de banda de la red y el entorno de muestreo en la precisión del reconocimiento facial. A continuación, se utiliza GAN para aprender sobre los coeficientes wavelet, se aplica la restricción de preservación de identidad a la imagen de salida y, se realiza la función de pérdida perceptual de los coeficientes wavelet de fusión. Por último, se utiliza el modelo de aprendizaje profundo basado en SRGAN para obtener imágenes faciales de alta resolución. Para la fase de experimentación se utilizó el dataset The Muct Face, con 750 imágenes de 50 personas con diferente género, edad, iluminación y expresión.

Una vez implementada la reconstrucción de imágenes faciales, se necesita evaluar el rendimiento de la calidad de la imagen reconstruida, para ello se utilizaron las métricas PSNR, SSIM y FSIM, dando como resultado una mejora en 2.05 dB (decibelios), 0.03 y 0.04 respectivamente.
Los resultados experimentales muestran que el método propuesto puede lograr la restauración (con súper resolución) de imágenes faciales de baja resolución y cumplir con los requisitos de precisión de reconocimiento facial. Ver en la Figura 3.2 una comparación de los resultados de reconstrucción de los dos algoritmos.

Figura 3.2 Comparación de los resultados de reconstrucción de dos algoritmos (De izquierda a derecha: imagen de baja resolución, imagen original de alta resolución, imagen de reconstrucción basada en SRGAN) (Cao et al., 2021)

Image Super-Resolution Via Wavelet Feature Extraction and Sparse Representation (Álvarez-Ramos et al., 2018)

Este artículo propone una nueva técnica de súper resolución (SR) basada en la extracción de características wavelet y la representación dispersa (propuesta llamada SR-WAFE-SR).

En primer lugar, la imagen de baja resolución (LR) se interpola empleando la operación Lanczos. A continuación, la imagen se descompone en sub-bandas (LL, LH, HL y HH) mediante la Transformación Wavelet Discreta (DWT). Las cuatro sub-bandas tienen la mitad del tamaño debido a la decimación de la DWT. Cada sub-banda de alta frecuencia (HF) se interpoló utilizando interpolación Lanczos para recuperar el mismo tamaño de la imagen inicial. El análisis de componentes principales se aplicó a tres sub-bandas interpoladas de HF (LH, HL y HH) para reducir la dimensionalidad y obtener una banda con los detalles más relevantes.

Para evaluar los resultados y poder realizar una discusión sobre ellos, se evaluó el rendimiento de diferentes técnicas del estado del arte y de la técnica SR-WAFE-SR propuesta empleando los siguientes criterios: Peak Signal-to-Noise Ratio (PSNR) para determinar la supresión de ruido y la mejora en la reconstrucción de la imagen, y el Structural Similarity Index Measure (SSIM), para estimar la calidad visual.

En las simulaciones, la técnica de SR propuesta se probó con imágenes de diferentes tipos (estándar, médicas y aéreas). Los autores concluyen en que, en comparación con las técnicas de mejora de la resolución más avanzadas, la técnica de SR propuesta utiliza la DWT para obtener los detalles de la imagen LR y, a continuación, estos detalles se aplican en la reconstrucción de la imagen HR mediante una representación dispersa con la ayuda de dos diccionarios. Al comparar el nuevo enfoque con las técnicas de mejora de la resolución más avanzadas, el marco de trabajo

SR propuesto parece demostrar un rendimiento superior en términos de criterios objetivos (PSNR y SSIM), así como en la percepción subjetiva a través del sistema visual humano.

Face Super-resolution Guided by Facial Component Heatmaps (Yu et al., 2018)

En este trabajo, se propuso un método que incorpora explícitamente información estructural de rostros en el proceso de super resolución mediante el uso de una red neuronal convolucional (CNN) multitarea. Dicha CNN tiene dos partes: una para la superresolución de imágenes faciales y la otra rama para predecir las regiones salientes de un rostro, denominadas mapas térmicos de componentes faciales. Estos mapas térmicos animan al flujo de muestreo ascendente a generar rostros super resueltos con detalles de mayor calidad. Este método no sólo utiliza información de bajo nivel (es decir, la similitud de intensidad), sino también la información de nivel medio (es decir, la estructura de la cara) para explorar más a fondo las restricciones espaciales de los componentes faciales de las imágenes de entrada de baja resolución (LR).

Cuando la resolución de las imágenes de entrada es demasiado pequeña, los componentes faciales serán aún más pequeños. Por lo tanto, es difícil para los detectores de última generación localizar con precisión los puntos de referencia faciales en imágenes de muy baja resolución. Sin embargo, se propuso predecir los mapas térmicos de los componentes faciales a partir de mapas de características, en lugar de localizar los puntos de referencia en las imágenes de entrada LR; ya que los mapas de características sobre muestreados contienen más detalles y sus resoluciones son lo suficientemente grandes como para estimar los mapas térmicos de los componentes faciales. Aprovechan los 68 puntos de referencia faciales para generar los mapas térmicos. En concreto, cada punto de referencia se representa por un núcleo gaussiano y el centro del núcleo es la ubicación del punto de referencia. Ajustando la varianza estándar de los núcleos gaussianos, de acuerdo con las resoluciones de los mapas de características o las imágenes, se genera un mapa de calor para cada componente.

Con la ayuda de la rama de estimación de componentes faciales, el método resuelve los rostros en diferentes distorsiones causadas por la localización errónea de componentes faciales en las imágenes de entrada LR. Por lo tanto, es capaz de resolver imágenes faciales que no sean de un tamaño de 16×16 píxeles con un factor de ampliación de 8, conservando la estructura del rostro. Experimentos exhaustivos demuestran que esta red logra resultados superiores de mejora de rostros y supera las técnicas del estado del arte de este artículo.

Super-Resolution using Deep Learning to Support Person Identification in Surveillance Video (Alkanhal et al., 2020)

En este trabajo, se propone un sistema de súper resolución basado en aprendizaje profundo que tiene como objetivo mejorar las imágenes de rostros capturadas de videos de vigilancia para apoyar la identificación de sospechosos. El sistema propuesto se basa en una técnica de procesamiento de imágenes llamada súper resolución (SR) que consiste en recuperar imágenes de alta resolución a partir de baja resolución. Se utilizó la red neuronal Very-Deep Super-Resolution (VDSR) para mejorar la calidad de la imagen.

Para entrenar la red VDSR, el primer paso es procesar las imágenes de entrenamiento. Las imágenes son de alta resolución (HR), se transforman en el espacio YCbCr y, a continuación, se extraen los canales de luminancia. A continuación, las imágenes se muestrean en sentido descendente utilizando diferentes factores de escala para obtener imágenes de baja resolución (LR).

Las imágenes LR se redimensionan mediante interpolación bicúbica para que coincidan con los tamaños originales de sus imágenes HR. Por último, se calculan las imágenes residuales y se almacenan junto con las imágenes redimensionadas (imágenes LR).

Cuando se inicia la fase de entrenamiento, los pesos de la red (filtros) se generan aleatoriamente y los sesgos se inicializan en ceros. La red aprende a dar como salida la imagen residual estimada de la entrada, minimizando la función de pérdida, que es el error medio cuadrático.

El modelo propuesto se entrenó con el conjunto de datos de rostros CelebA y se utilizó para mejorar la resolución el conjunto de datos QMUL-SurvFace. Se obtuvo un pico de señal/ruido (PSNR) del 7% y el índice de similitud estructural (SSIM) del 3%. Lo más importante es que aumentó la tasa de reconocimiento facial en un 45.7%; ver Figura 3.3.

Figura 3.3 Resultados de la mejora de la SR (Alkanhal et al., 2020)

Kernel Modeling Super-Resolution on Real Low-Resolution Images (Zhou & Süsstrunk, 2019)

Este trabajo propone una red de súper resolución (SR) de modelado de kernel (KMSR) que incorpora el desenfoque de una imagen de baja resolución (LR) en el entrenamiento. Esta propuesta de KMSR consta de dos etapas: primero se construye un conjunto de núcleos de desenfoque realistas con una red generativa adversarial (GAN) y luego se entrena la red de SR con imágenes de baja y alta calidad (HR) correspondientes con los núcleos generados.

Antes de construir el conjunto de entrenamiento, es necesario estimar núcleos de desenfoque realistas a partir de fotografías reales. Las imágenes LR se obtienen a partir de las imágenes HR, se reduce su tamaño para generar una mala calidad y poder entrenar así la red. Dichos núcleos se utilizan para entrenar mejor el GAN para el modelado. La combinación de los núcleos estimados y de los núcleos generados por el GAN forma el gran conjunto de núcleos utilizado para construir los datos de entrenamiento LR-HR emparejados.

Para validar la capacidad del KMSR propuesto en imágenes con núcleos, se llevaron a cabo experimentos de recopilación de imágenes LR con núcleos de desenfoque realistas no vistos y bajo las métricas PSNR y SSIM, demostrando que el método propuesto es el que obtiene valores más altos.

Para la experimentación se utilizó un objetivo zoom de 24-70 mm para capturar pares de fotos. La foto de 35 mm de distancia focal sirve como imagen LR, y la foto tomada en la misma posición con la distancia focal de 70 mm sirve como imagen HR de referencia (ver Figura 3.4).

Figura 3.4 Comparación cualitativa de SR (Zhou & Süsstrunk, 2019)

Low Resolution Facial Manipulation Detection (Han et al., 2020)

Este artículo propone un módulo de súper resolución de artefactos y enfoque (ASFR) y un extractor de características de dos flujos (TFE). El módulo ASFR está construido para aplicar la súper-resolución (SR) a las entradas de baja resolución (LR) y recuperar la información perdida debido

a la resolución, por ejemplo, las señales faciales y los artefactos de manipulación. El ASFR se implementa basándose en la arquitectura del autocodificador. Para ello se trabajó con la red ResNet, seguido de un decodificador que consta de cinco bloques conectados de forma residual. En cada uno de ellos, el mapa de características de la capa anterior se muestrea mediante una interpolación del vecino más cercano, en la cual se añaden dos capas de convolución 3x3, norma de lote y ReLU. Las conexiones residuales pueden preservar las señales visuales de las capas anteriores y, por tanto, ayudan a mejorar la calidad de las imágenes reconstruidas.

El extractor de características de dos flujos aplica primero dos flujos de capas de convolución para extraer mapas de características para las imágenes reconstruidas y sus correspondientes imágenes de error. A continuación, los dos mapas de características se fusionan a través de capas de convolución posteriores para generar la salida final. Los datasets utilizados para la fase de prueba fueron FaceForensics++ y DeepfakeTIMIT, ambos con videos originales y falsos. Los experimentos han demostrado que los dos módulos pueden mejorar el rendimiento de la detección de manipulación facial en baja resolución. También se muestra la solidez del método en la detección de varios tipos de imágenes manipuladas con diferentes resoluciones (ver Figura 3.5).

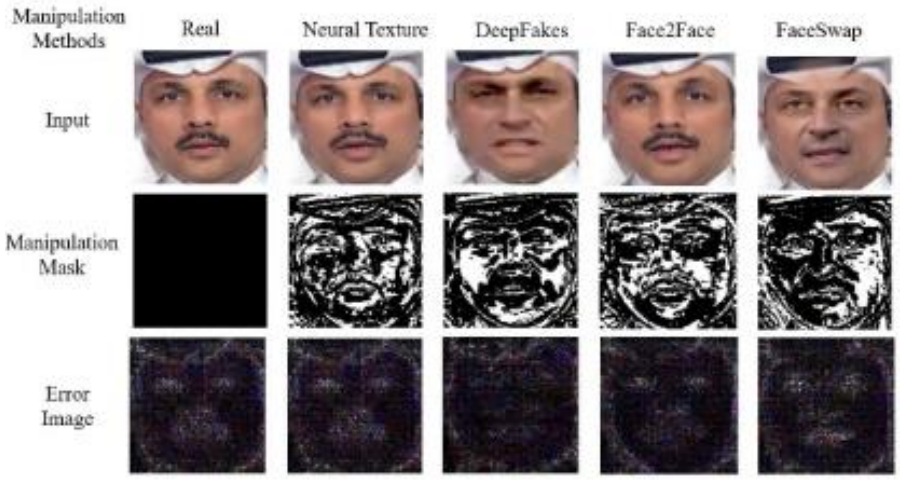

Figura 3.5 Imágenes de error y máscara de manipulación de una imagen real y sus correspondientes cuatro tipos de imágenes manipuladas (Han et al., 2020)

Persistent Memory Residual Network for Single Image Super Resolution (R. Chen et al., 2018)

El artículo se enfoca en analizar la súper-resolución para trabajar problemáticas como la degradación de una imagen, el desenfoque, los ruidos y la deformación geométrica.

Utilizan una red de memoria persistente que demostró ser eficaz en la restauración de imágenes combinada con una red neuronal convolucional residual profunda. Se diseñaron dos tipos de

bloques de memoria para el desafío NTIRE2018 (MemEDSR e IRMem), se incorporaron estos tipos de elementos en el marco de la red de súper resolución mejorada (EDSR). El primer tipo de bloque de memoria es un módulo que contiene cuatro módulos residuales seguidos de una unidad de compuerta que selecciona de forma adaptativa las características necesarias a almacenar. El segundo tipo de memoria es un bloque convolucional dilatado residual, que contiene siete capas de convolución dilatadas unidas a una unidad de compuerta.

Para la fase de experimentación se utilizó el conjunto de datos DIV2K 2018, que contiene 1000 imágenes de alta calidad para tareas de súper-resolución de una sola imagen con cuatro calidades: a) Calidad 1: Bicúbica clásica -×8, en la que la degradación de la imagen se genera mediante una reducción bicúbica con el factor 8. b) Calidad 2: Leve realista -×4, en la que una imagen de baja resolución es una versión reducida con el factor 4 con ruido. c) Calidad 3: Realista difícil -×4 presenta una imagen de baja resolución con el factor 4 junto con ruido y desenfoque. d) Pista 4: Realista salvaje -×4, cuya imagen es de baja resolución con el factor 4 con ruido, desenfoque y desplazamiento.

Los resultados experimentales en el conjunto de datos DIV2K muestran que los modelos propuestos logran un mejor rendimiento que el algoritmo EDSR. Logrando una mejora en la superresolución y mitigando la degradación de la imagen debido al ruido y el desenfoque (ver Figura 3.6).

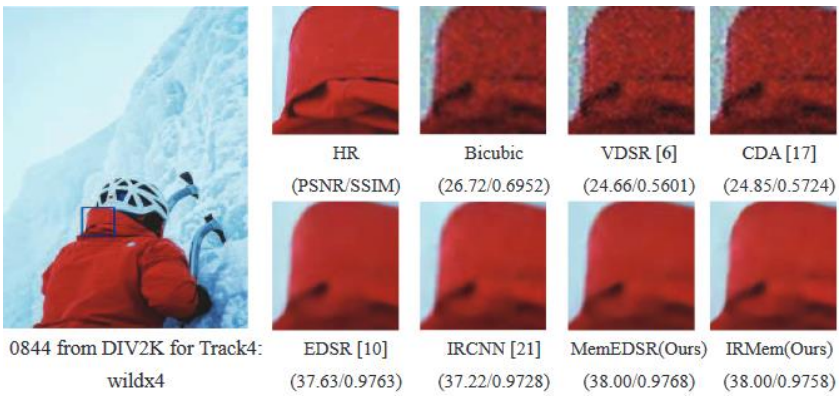

Figura 3.6 Comparación de algoritmos del estado del arte y el trabajo presentado (R. Chen et al., 2018)

3.2.2 Técnicas de mejoramiento de imágenes

Contrast Improvement using Local Gamma Correction (Aditya Acharya & A Venkat Giri, 2020)

En este trabajo se presenta una técnica que intenta complementar las limitaciones del esquema de corrección Gamma convencional. Una imagen con diferentes grados de regiones de intensidad no puede ser mejorada simultáneamente utilizando un único valor gamma. Bajo el método propuesto se busca solucionar este problema, con una técnica de corrección gamma local adaptativa. En este método, el factor de corrección gamma se varía localmente en función de la intensidad media local de una zona.

Como resultado, los valores gamma locales correspondientes a los diferentes grados del rango de intensidad son diferentes, lo que da lugar a una manipulación eficaz del contraste de una imagen. De este modo, se consigue mejorar el contraste de diferentes regiones de la imagen con grados de intensidad distintos. Este algoritmo se centra en realzar las regiones de intensidad oscura y brillante de los diferentes tonos sin perder gran parte de la información original de una imagen.

Los resultados de la simulación revelan que la técnica propuesta proporciona, subjetivamente, un rendimiento superior a muchos de los esquemas de mejora de la imagen existentes. Un ejemplo de sus resultados se puede ver en la Figura 3.7.

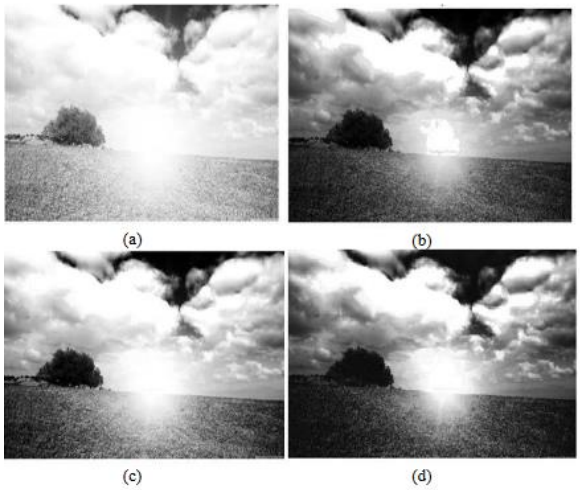

Figura 3.7 Evaluación subjetiva de diferentes algoritmos de mejora para una imagen brillante en exteriores: (a) Original; (b) Ecualización de Histograma (HE); (c) Corrección Gamma- (GC); (d) Propuesta (Aditya Acharya & A Venkat Giri, 2020)

Resonancia Estocástica para el Mejoramiento del Contraste y Calidad en Imágenes Acústicas de Sonar de Barrido Lateral (Sousa et al., 2016)

En este trabajo se presenta una novedosa técnica de resonancia estocástica con un ruido blanco Gaussiano para el mejoramiento de contraste y calidad visual aplicados en imágenes acústicas de sonar de barrido lateral, manteniendo las características representativas de la imagen.

El sonar de barrido lateral (SSS - Side Scan Sonar) es uno de los tipos de sonares de escaneo lateral más difundidos, siendo la principal herramienta de visualización de imágenes acústicas del suelo marino. El método propuesto permite agregar a la imagen de entrada repetidamente ruido estocástico generando imágenes binarias, estas imágenes se combinan conformando una única imagen compuesta promediando sus valores de píxeles, siendo esta la salida. Esta imagen compuesta se evalúa mediante métricas de contraste y calidad, con el objetivo de encontrar la intensidad del ruido que maximice la calidad y contraste de la imagen compuesta.

Para ello se utilizaron dos métricas: contraste por valor cuadrático medio (RMS –root mean square) y coeficiente de correlación entre la imagen de salida y, de entrada. En la fase de experimentación, se adquirieron imágenes en la bahía de Todos los Santos, Brasil. Se utilizó el SSS Starfish 450F de la empresa Tritech. Se propuso una gráfica CvC (contraste versus coeficiente de correlación) para mostrar el efecto de la intensidad de ruido aplicada a la imagen de entrada con respecto al número de representaciones binarias de la imagen compuesta de salida, y por consiguiente determinar cómo varía su contraste y calidad.

Improvement of image quality at CT and MRI using deep learning (Higaki et al., 2019)

En este artículo, se discuten técnicas para mejorar la calidad de las imágenes de diagnóstico por tomografía computarizada y resonancia magnética con la ayuda del aprendizaje profundo.

Dichas técnicas fueron clasificadas útiles para diferentes aspectos sobre el mejoramiento de imagen, al igual que en la reducción de ruido que debe ser eliminado sin degradar el verdadero componente de la señal. Otra técnica es la súper resolución, que consiste en mejorar la resolución espacial de la imagen original, definiendo la nitidez de los bordes que tiende a perderse al ampliar una imagen y que logran conservar con técnicas avanzadas de súper-resolución. La última categoría de clasificación es para la adquisición y reconstrucción de imágenes, la cual está limitada por el tiempo; sin embargo, el aprendizaje profundo se utiliza para complementar la información que falta. Se han publicado estudios sobre la sustitución del proceso de reconstrucción de imágenes, algunos de ellos pueden clasificarse como esfuerzos de "reducción del ruido y los artefactos de la imagen".

Cada una de estas técnicas ayuda en la mejora de imágenes médicas teniendo diversos objetivos como disminuir la exposición a la radiación en los estudios de tomografía computarizada clínica (CT). La súper resolución puede aumentar la capacidad de diagnóstico al mejorar la resolución de las imágenes. Muchas de las técnicas de adquisición y reconstrucción de imágenes reducen el ruido y los artefactos atribuibles a la adquisición incompleta de datos y podrían clasificarse como técnicas de reducción de ruido. Entre las técnicas innovadoras, se discutió un método para reconstruir directamente los datos sin procesar en datos de imagen utilizando una DCNN. Aunque el aprendizaje profundo se espera para varias aplicaciones clínicas, hardware adicional como una unidad de procesamiento gráfico (GPU) es necesario para su aplicación, ya que el coste del cálculo es elevado. En la Figura 3.8 se muestra una comparación de las técnicas de reconstrucción de imágenes.

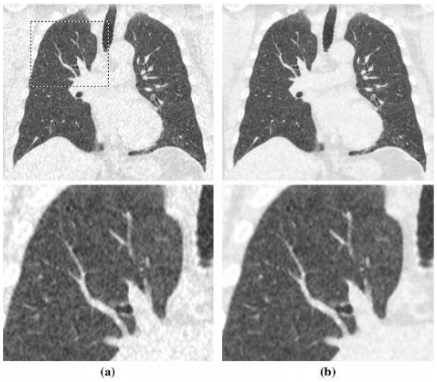

Figura 3.8 Comparación de las técnicas de reconstrucción de imágenes de CT aplicadas a la CT de tórax de baja dosis. a Reconstrucción iterativa híbrida iterativa, b Reconstrucción basada en aprendizaje profundo (Higaki et al., 2019)

Thermal Image Enhancement Algorithm Using Local and Global Logarithmic Transform Histogram Matching with Spatial Equalization (Voronin et al., 2018)

En este artículo, se presenta un nuevo algoritmo de mejora de imagen térmica basado en el procesamiento de la imagen en el dominio de la frecuencia. El enfoque presentado aprovecha el hecho de que la relación entre el estímulo y la percepción es logarítmica. Se aplica logarítmicamente con un enfoque de ecualización espacial en diferentes bloques de imágenes.

Para el proceso de mejora de la imagen se realizó la transformación de la imagen, luego se manipuló el coeficiente de la transformación, y a continuación, se realiza la transformada ortogonal inversa. Se efectúa la transformada discreta de Fourier y la transformada logarítmica ayuda a mapear un estrecho rango de valores de nivel de gris bajo en una gama más amplia del nivel de salida. Este tipo de transformación se emplea para ampliar los valores de los píxeles oscuros de una imagen mientras se comprimen los valores de nivel superior. El mapeo de histogramas es una

versión más generalizada de la ecualización de histograma que permite especificar la forma del histograma que se desea que tenga la imagen procesada.
Los resultados de la mejora de la imagen propuesta se comparan favorablemente con otros enfoques dentro del estado del arte.

Image enhancement with the application of local and global enhancement methods for dark images (Singh et al., 2017)

En este trabajo de investigación, se presenta un nuevo método que utiliza técnicas de mejora local y global en la misma imagen. Esto con la finalidad de abordar y reducir la discrepancia entre las mejoras individuales de la imagen que se tienen actualmente. Este método se simula en MATLAB y los resultados se verifican con los parámetros de calidad de la imagen. El funcionamiento es como sigue: en primer lugar, la imagen local y el resultado se procesa de nuevo con el método de mejora global. La mejora global se utiliza para aumentar el contraste de la imagen. En este proceso, cada píxel de la imagen se ajusta de forma que se obtenga una mejor visualización. En la mejora del contraste espacial, la operación se realiza directamente sobre el píxel. Los píxeles se disponen de manera que se distribuyan en el rango de intensidad deseado.

En el método, primero se toma una imagen y se convierte del espacio de color RGB al espacio de color HSV. Del espacio de color HSV se toma la componente V o la componente de luminancia para aplicar el algoritmo. Con el fin de potenciar los gradientes o los detalles locales, se utilizó un método de realce local existente que utiliza el enmascaramiento como técnica de mejora de los detalles locales. Como su nombre indica, la imagen borrosa se emplea como máscara para intensificar los detalles locales en forma de nitidez de los bordes. La imagen optimizada localmente se emplea como entrada para la mejora global, con el objetivo de aumentar la percepción visual. Este enfoque es eficaz para la mayoría de las imágenes con poca luz.

Hue-Preserving Color Image Enhancement on a Vector Space of Convex Combination Coefficients (Ueda & Suetake., 2019)

Este trabajo propone un nuevo método de mejora de imágenes en color que preserva el tono que se ejecuta en un espacio vectorial de coeficientes de combinación convexa y no genera el problema de la gama de colores. En el método propuesto, en primer lugar, cada píxel de una imagen de entrada se representa mediante una combinación convexa de blanco-negro y su color puro. A continuación, los coeficientes de la combinación se transforman mediante la técnica de especificación del histograma, utilizando los histogramas ponderados basados en la norma de

gradiente de los coeficientes y las distribuciones de los coeficientes se expanden para mejorar adecuadamente los componentes de intensidad y saturación de la imagen.
En la fase de experimentación, se comparó el método con el estado del arte. A través de estas comparaciones, se comprobó la eficacia del método propuesto y se verificó en términos de la intensidad del contraste, el colorido y la calidad de la imagen. Las imágenes resultantes se evalúan utilizando dos métricas objetivas: el valor Q de la métrica de contraste de intensidad y la métrica de colorido C-valor. Los valores Q y C más altos indican que el contraste de intensidad y el colorido de la imagen son más altos, respectivamente. En particular, la imagen resultante del método propuesto es la más colorida.

3.2.3 Detección de rostros

A Novel Technique to Detect and Track Multiple Objects in Dynamic Video Surveillance Systems (Adimoolam et al., 2022)

En este artículo se habla sobre una nueva y novedosa técnica para seguir múltiples objetos. El objetivo principal fue aumentar la precisión y disminuir el tiempo de entrenamiento para contribuir en el área de los sistemas de seguridad humana a través de sistemas de videovigilancia.

Durante el proceso para la detección de objetos, primero se extraen los fotogramas de un vídeo y se realiza la sustracción del fondo; las personas pueden ser localizadas a través de varios aspectos como el rostro, el color, la forma y la piel. Posteriormente, se extraen diversas características de los objetos y utilizan métodos bayesianos para compararlos con la información almacenada y reconocer a las personas. Por medio de esta metodología, los autores pueden determinar la ubicación de un objeto dentro de la imagen dada y, en el caso de querer seguir un objeto específico, sólo proporcionan la información de píxeles de la región de interés y se busca la región que tiene mayor similitud.

El objeto detectado se rastrea continuamente mediante el flujo de entrada. Para ello, se trabajó con una red neuronal convolucional de selección de características densas. El resultado de esta metodología obtuvo una precisión del 98% junto con un tiempo de predicción de un minuto y once segundos; dichos resultados fueron comparados con métodos existentes como el filtro Kalman (KF) y una red neuronal profunda (DNN).

Face Detection in Blurred Surveillance Videos for Crime Investigation (Menaka & Yogameena, 2021)

Este artículo aborda cómo la detección de rostros supone un problema importante en los vídeos de vigilancia de baja resolución, principalmente debido al desenfoque. Esto motivó a los autores a presentar el algoritmo de desenfoque para mejorar la tasa de detección de rostros y reducir la tasa de falsos positivos. Por lo tanto, el trabajo se centra en la eliminación del desenfoque mediante la aplicación de un algoritmo de desconvolución ciega, suprimiendo el artefacto de anillamiento en el vídeo de vigilancia mediante la adopción de la Transformada Wavelet Discreta (DWT).

La estructura propuesta comienza transformando la entrada de vídeo en una secuencia de fotogramas. Después, la fase de preprocesamiento elimina el desenfoque del vídeo de vigilancia. La desconvolución ciega se lleva a cabo con la función de dispersión de punto (PSF) de desenfoque gaussiana modelada para identificar los bordes nítidos en la banda de frecuencia más alta. A continuación, el vídeo desenfocado se suaviza con un filtro gaussiano para eliminar el ruido y se suprimen los artefactos de anillamiento con un mapa de bordes adaptativo de DWT en la fase de posprocesamiento. Por último, los rostros se detectan utilizando las redes neuronales Yolo versión 2 y Darknet-19.

El enfoque propuesto para la detección de rostros se ha experimentado con los siguientes conjuntos de datos de vigilancia: OWN, Chokepoint, ATM1, ATM2, ATM3, IISC Bangalore, Non-ATM 1 SC Face, ETH, ILIDS-VID y PRID 2011.

La solución propuesta ha aumentado significativamente la tasa de detección de rostros para su ámbito de aplicación de perfil y rostros frontales. Puede extenderse a la función de reidentificación de una persona para una investigación criminal en la que el crimen pasa desapercibido e identifica el seguimiento del tráfico de la persona sin casco para reducir la presión de la policía de tráfico. Además, el trabajo puede extenderse para reconocer la expresión facial del apuñalador, involucrado en cualquier evento de apuñalamiento o robo, específicamente en sectores basados en las finanzas como cajeros automáticos, bancos y joyerías. En la Figura 3.9 se muestra un ejemplo de los rostros detectados.

Figura 3.9 Rostros detectados con Yolov2 después de quitar el ruido con des convolución ciega (Menaka & Yogameena, 2021)

Face Detection and Extraction from Low Resolution Surveillance Video Using Motion Segmentation (X. Chen et al., 2019)

Este artículo habla sobre algunos problemas a los que se enfrenta la videovigilancia tales como el desenfoque, factores ambientales y el ancho de banda con el que se procesan las imágenes. En busca de contribuir al área de detección de rostros, se creó este trabajo el cual está basado en características Haar, empleando el escalado de la imagen, para así detectar rostros en varias escalas.

Primero se necesita encontrar la diferencia del cuadro de video actual con respecto al anterior para encontrar los píxeles segmentados por movimiento. Además, solo los píxeles segmentados por movimiento están sujetos a la generación de subimágenes para ser analizadas para la clasificación de rostros. Dicho detector, se entrenó usando imágenes de 18x18 píxeles. La detección de rostros multiescala se ha facilitado mediante la aplicación de escalado de imagen. Los valores mínimo y máximo de los factores de escala se calculan en función del tamaño de la imagen de entrada, el tamaño mínimo de la cara que se detectará y el tamaño máximo de la imagen que puede manejar el sistema.

El sistema de ejecutó en una computadora con Windows 8.1 de 64 bits y un procesador Intel Core i3, dentro del entorno de MATLAB versión 8.2.0.701.En la fase de prueba se utilizaron videos de videovigilancia de baja resolución con el conjunto de datos INRIA, con especificaciones de ancho de 384 píxeles y una altura de 288 píxeles junto con una velocidad de 25 fotogramas por segundo.

En resumen, se realiza la diferencia entre fotogramas y los píxeles que tienen la diferencia por encima del nivel de umbral establecido. Se marcan como píxeles segmentados por movimiento, que se someten a la generación de subimágenes para la detección de rostros. La aportación de este trabajo fue lograr una reducción considerable del espacio de búsqueda y un aumento de la eficiencia mediante la técnica de segmentación de movimiento. En la Figura 3.10 se muestra un ejemplo de los rostros detectados con la base de datos CAVIAR.

Figura 3.10 Resultados de detección en algunos fotogramas de vídeo de prueba (Chen et al., 2019)

Small Face Detection Using Deep Learning on Surveillance Videos (Cárdenas et al., 2019)
Este trabajo tiene como objetivo proponer un nuevo modelo para la detección de rostros en videos de baja resolución, basado en la morfología de la parte superior del cuerpo de las personas, haciendo uso de aprendizaje profundo.

La red neuronal entrenada está compuesta por 12 capas, una capa de entrada con un tamaño de 28x28 píxeles, tres capas convolucionales con 16,32 y 64 filtros de un tamaño de 3x3 píxeles. Tres capas de Unidades Lineales Rectificadas (ReLu). Dos capas de agrupación máxima, la primera con un número de píxeles que se desliza sobre la matriz de entrada y, la segunda capa totalmente conectada. A continuación, una capa SoftMax y, finalmente, la capa de clasificación con dos clases (cara y fondo).

Para la fase de entrenamiento se utilizaron la base de datos de videos CAVIAR y la base de datos de videos UCSP. Primero se probó el comportamiento del modelo de aprendizaje profundo con varias imágenes; en segundo lugar, se probaron los mejores modelos en ambas bases de datos y se resumieron los resultados obtenidos.

El modelo de aprendizaje profundo se entrenó en ambas bases de datos por separado, lo que produjo dos modelos diferentes. Las imágenes de rostros se extrajeron manualmente de los conjuntos de datos, para cada uno el 70 % de las imágenes se utilizaron para entrenamiento y el resto de las imágenes para pruebas.

Los resultados mostraron un aumento significativo en la precisión del modelo propuesto con una baja tasa de falsos positivos en videos de baja resolución. Con un promedio de 39% de precisión en el conjunto de datos de CAVIAR y 32% en el conjunto de datos de UCSP. En comparación con otras técnicas, estos resultados son mayores debido a que sólo alcanzan el 1% de precisión. También se obtuvo una menor precisión y tasas de falsos positivos cuando se utilizó el modelo propuesto en la base de datos CAVIAR. En la Figura 3.11 se muestra un ejemplo de los resultados obtenidos con el modelo propuesto.

Figura 3.11 Resultados obtenidos con el modelo propuesto en la red UCSP2 (Cárdenas et al., 2019)

A fast and accurate system for face detection, identification, and verification (Ranjan et al., 2019)

Dentro de este trabajo, se describió una línea de aprendizaje profundo para la identificación y verificación de rostros sin restricciones que alcanza un rendimiento de vanguardia en varios conjuntos de datos de referencia. Se propuso un nuevo detector de rostros, Deep Pyramid Single Shot Face, que es rápido y detecta caras con grandes variaciones de escala (especialmente caras pequeñas). Además, se propuso una nueva función de pérdida, llamada Crystal Loss, para las tareas de verificación e identificación de rostros. La pérdida de cristal restringe los descriptores de características a una hiperesfera de radio fijo, minimizando así la distancia angular entre pares de sujetos positivos y maximizando la distancia angular entre pares de sujetos negativos.

Se evaluó el detector de rostros propuesto en cuatro conjuntos de datos de detección de rostros, se logró el rendimiento más avanzado en el conjunto de datos Pascal Faces y se obtuvieron resultados competitivos en los conjuntos de datos WIDER, UFFDD y FDDB.

Face detection in low-resolution color images. International Conference Image Analysis and Recognition (Zheng et al., 2010)

En este trabajo de investigación, se estudió la relación entre la resolución y la tasa de detección automática de rostros con la transformada de censo modificada, uno de los algoritmos más exitosos para la detección de rostros hasta la fecha. Se propuso una nueva transformada de censo de color que proporciona resultados mejores que la original cuando se aplica a imágenes en color de baja resolución. En sistemas de vigilancia, las regiones de interés suelen estar empobrecidas o borrosas debido a la gran distancia entre la cámara y los objetos. La transformada de censo de

color de 12 bits fue propuesta para aprovechar la información de color y esta proporciona una detección de rostros precisa incluso cuando la resolución es baja.

Se creó un algoritmo para construir una cascada de clasificadores utilizando una variación del algoritmo AdaBoost, donde sólo se utilizaron tres etapas en la detección de rostros de baja resolución. Se observó cómo la transformada de censo de color de 12 bits funciona de mejor forma que la transformada de 9 bits original.

Survey of face detection on low-quality images (Zhou et al., 2018)

En este estudio se revisaron los detectores de rostros en imágenes de baja calidad más avanzados y su rendimiento, estos fueron comparados con los protocolos de diseño de los mismos algoritmos evaluados en este estudio. En segundo lugar, también se investigó la degradación de rendimiento mientras se probaron imágenes de baja calidad con diferentes niveles de desenfoque, ruido y contraste. Los métodos tradicionales de detección de rostros se basan en características elaboradas a mano, y pueden clasificarse en tres clases: métodos en cascada, modelo de partes deformables y las características de canal agregadas. Las redes profundas para tareas de clasificación de imágenes han demostrado ser sensibles a los ejemplos generados, añadiendo intencionalmente pequeñas perturbaciones mediante métodos de gradiente.

La comparación de cada modelo al probarlo en imágenes con diferentes niveles de desenfoque mostró que tanto los métodos tradicionales como los de aprendizaje profundo no son lo suficientemente robustos para las muestras de prueba borrosas, simplemente por las características de desenfoque en los bancos de filtros diseñados o aprendidos.

Face detection techniques: a review (Kumar et al., 2019)

Dentro de este trabajo de investigación, se estudió la necesidad de comprensión por parte de sistemas inteligentes la detección de rostros, ya que estos juegan un papel muy importante para realizar reconocimientos faciales. Se presentaron varias técnicas exploradas para la detección de rostros mediante la estimación de la postura de la cabeza. Los principales desafíos de la detección de rostros son fondos complejos, demasiadas caras en las imágenes, expresiones extrañas, iluminaciones, menor resolución, oclusión de la cara, color de la piel, distancia y orientación. Dentro de las aplicaciones más comunes de este tipo de sistemas son asistencia biométrica, reconocimiento facial, marketing, control de documentos y control de acceso, etc. Para todo ello se necesita de diversas técnicas para poder usarse, la más comunes son:

- Modelo de forma activa: Se centra en características complejas no rígidas.

- Análisis de bajo nivel, donde los factores de colores juegan un papel muy importante.
- Análisis de características, tiene como objetivo encontrar características estructurales.
- Enfoques basados en imágenes, donde involucra métodos con redes neuronales, métodos del subespacio lineal, enfoque estadístico y un estudio comparativo del enfoque basado en características y del enfoque basado en imágenes.

Face detection using deep learning: An improved faster RCNN approach (Sun et al., 2018)

En el trabajo se presentó un nuevo esquema de detección de rostros utilizando el aprendizaje profundo y logrando el rendimiento de detección más avanzado en la conocida evaluación de referencia de detección de rostros FDDB (Face Detection Data Set y Benchmark). El método obtuvo el mejor rendimiento de detección de rostros y se clasificó como uno de los mejores modelos en términos de curvas ROC, publicado en el benchmark FDDB. Esencialmente consta de dos partes, la primera es una red de propuestas regionales para generar una lista de propuestas de regiones que probablemente contengan objetos, o llamadas regiones de interés y la segunda una red Fast RCNN para clasificar la región de la imagen en objetos (y fondo) y refinar los límites de esas regiones. Las dos partes comparten parámetros comunes en las capas de convolución utilizadas para la extracción de características, lo que permite a esta arquitectura realizar tareas de detección de objetos a una velocidad bastante competitiva.

Influence of low resolution of images on reliability of face detection and recognition (Marciniak et al., 2021)

En este artículo, se analizó la fiabilidad del sistema en tiempo real de detección de rostros y reconocimiento facial a partir de imágenes de baja resolución, por ejemplo, imágenes de videovigilancia.

Las normas biométricas describen las reglas generales, las directivas y las características relativas a los datos biométricos de entrada, como las imágenes faciales. El primer paso para poder reconocer un rostro es la localización de este. La detección de la posición se realiza en tres etapas: La primera es la reducción del impacto de los factores de interferencia con el uso de la ecualización del histograma y la reducción del ruido. La siguiente etapa es la determinación de las áreas con alta probabilidad en las que puede situarse o puede ser colocado un rostro. En una etapa posterior se realiza la verificación de las áreas previamente seleccionadas. Por último, se detecta y marca el rostro. Algunas técnicas para poder realizar esto se encuentran dentro del estado del arte de este artículo, donde se especifican en qué tipo de factores un rostro no puede ser detectado, por ejemplo, si los ojos se encuentran de manera oculta. La baja resolución es un factor que afecta en

gran parte tanto el reconocimiento como la detección de un rostro, pero estos no impiden que estas funciones puedan realizarse, esto es posible incluso en imágenes con una resolución de 21x21 píxeles, esto significa que las personas pueden ser reconocidas desde una gran distancia.

Selective Refinement Network for High Performance Face Detection (Chi et al., 2019)
Este artículo presenta un novedoso detector de rostros de una sola toma, denominada Red de Refinamiento Selectivo (SRN), que introduce novedosas operaciones de clasificación y regresión de dos pasos de forma selectiva en un detector de rostros basado en anclajes para reducir los falsos positivos y mejorar la precisión de la localización simultáneamente.

La SRN consta de dos módulos: el módulo de clasificación selectiva en dos pasos (STC) y el módulo de regresión selectiva en dos pasos (STR). El STC se divide en seis niveles de una pirámide, los 3 niveles superiores realizan una regresión más precisa de las ubicaciones de los cuadros delimitadores de los rostros, mientras que los niveles inferiores prestan más atención a la tarea de clasificación. El STC tiene como objetivo filtrar la mayoría de las anclas negativas simples de las capas de detección de bajo nivel para reducir el espacio de búsqueda para el clasificador posterior. Además, se diseñó un bloque de mejora del RFE para proporcionar un campo receptivo más diverso, que ayuda a capturar mejor las caras en algunas poses extremas.

Todos los modelos fueron entrenados con el dataset Wider Face, el cual se dividió en el 40% para entrenamiento, validación el 10% y de prueba el 50%. Dividido en tres niveles de dificultad: Fácil, Medio y Difícil basados en la tasa de detección de EdgeBox.

Los experimentos realizados con los conjuntos de datos AFW, PASCAL face, FDDB y WIDER FACE demuestran que SRN alcanza el rendimiento de detección más avanzado.

A continuación, la Tabla 3.1 muestra una síntesis de los artículos revisados.

Tabla 3.1 Resumen de artículos del estado del arte

Artículo	Objetivo	Técnicas/ Herramientas	Conjuntos	Resultados
Técnicas de súper resolución				
(Yue et al., 2016)	Ofrecer una revisión de la SR desde el punto de vista de las técnicas y las aplicaciones.	Clasificación de aplicaciones de la SR.	-	Se necesitan métodos más avanzados, adaptables, rápidos y con una amplia aplicabilidad.

Tabla 3.1 Resumen de artículos del estado del arte

Artículo	Objetivo	Técnicas/ Herramientas	Conjuntos	Resultados
Técnicas de súper resolución				
(Sha et al., 2012)	Conocer la SR y las diferentes formas y métodos en que puede trabajarse.	Clasificación de técnicas de SR.	-	La mayoría de los enfoques utilizan una secuencia de imágenes de baja resolución para extraer una imagen de SR.
(Chen et al., 2019)	Crear un algoritmo de SR de una sola imagen basado en el modelo de escala y características de deformación.	Algoritmo de SR de una imagen basado en el modelo de escala y características de deformación.	-	Imágenes con un nivel más alto de mejora, hasta de 80% evaluado por medio de SSIM y PSNR mediante el método propuesto, en comparación con los algoritmos existentes.
(Farooq et al., 2019)	Método de aprendizaje de transferencia de estilo para enfrentar SR utilizando conjuntos de LR y HR que comparten propiedades importantes.	Alinear conjuntos de datos de HR y LR para crear una sola imagen de entrada en algoritmos de SR.	Repositorio propio para experimentación.	Imágenes de SR con un nivel mayor de mejora de hasta 90% en nitidez comparado con el estado del arte de este trabajo.
(Ochoa Domínguez et al., 2020)	Presentar un estudio comparativo de cuatro métodos recientes del estado del arte de SR.	SR mediante aprendizaje profundo.	Set5, Set14, Urban100 y BSD100.	Se demuestra que las arquitecturas más profundas obtienen mejores resultados de PSNR y SSIM, hasta de 85%.
(Bhanu et al., 2012)	Realizar un algoritmo que sea capaz de entrenarse para interpolar imágenes.	Diferenciación de la señal multidimensional discreta.	-	Alto rendimiento en las evaluaciones subjetivas como en las cuantitativas, mediante SSIM y PSNR, hasta de 80%.
(Honda et al., 2018)	Crear un sistema que sea capaz de extraer características a través de filtros.	Algoritmo Lanczos	-	Obtiene una imagen de SR en menor tiempo comparado con el estado del arte, y se permite una mejor reconstrucción inicial.
(Álvarez-Ramos et al., 2018)	Proponer una nueva técnica de SR.	Características wavelet y la representación dispersa.	INRIA	Rendimiento superior en términos de criterios objetivos hasta de 85%, así como en la percepción subjetiva a través del sistema visual humano.

Tabla 3.1 Resumen de artículos del estado del arte

Artículo	Objetivo	Técnicas/ Herramientas	Conjuntos	Resultados
Técnicas de súper resolución				
(Yu et al., 2018)	Proponer un método que incorpora información estructural de rostros en el proceso de súper resolución.	Una red neuronal evolutiva (CNN) multitarea.	TDAE, CelebA, Menpo	Resuelve los rostros en diferentes distorsiones hasta un 75% de componentes faciales en las imágenes de entrada.
(Cao et al., 2021)	Proponer un método de reconstrucción de imágenes faciales.	Transformada wavelet y una red generativa adversarial de súper resolución.	The Muct Face	El método logra restauración de SR de imágenes faciales de baja resolución y cumple con precisión de reconocimiento facial en un 50%.
(Alkanhal et al., 2020)	Mejorar las imágenes de rostros capturadas de videos de vigilancia.	Un sistema de SR basada en aprendizaje profundo.	CelebA y QMUL-SurvFace.	PSNR del 7% y SSIM del 3%. La tasa de reconocimiento facial en un 45.7%.
(Zhou & Süsstrunk, 2019)	Proponer una red de SR de modelado de kernel.	Red de súper resolución de modelado de kernel que incorpore el desenfoque de una imagen de baja resolución.	Repositorio propio para experimentación.	Eficacia del enfoque de súper resolución en fotografías con núcleos de desenfoque desconocidos en un 75%.
(Han et al., 2020)	Proponer un módulo de SR de artefactos y enfoque y un extractor de características de dos flujos.	Un módulo de SR de artefactos, enfoque y un extractor de características de dos flujos.	FaceForensics++ y DeepfakeTIMIT	El método logra la restauración de SR de imágenes faciales de baja resolución y cumple con los requisitos de precisión de reconocimiento facial en un 55.6%.
(R. Chen et al., 2018)	Analizar la SR para problemáticas como la degradación, el desenfoque, los ruidos y la deformación geométrica de una imagen.	Una red de memoria persistente.	DIV2K 2018.	Los modelos propuestos logran un mejor rendimiento que el modelo EDSR en métricas SSIM de 5%. Logrando una mejora y mitigando la degradación de la imagen.

Tabla 3.1 Resumen de artículos del estado del arte

Artículo	Objetivo	Técnicas/ Herramientas	Conjuntos	Resultados
Técnicas de mejoramiento de imágenes				
(Aditya Acharya & A Venkat Giri, 2020)	Presentar una técnica de corrección Gamma que complementa las limitaciones del esquema de corrección Gamma convencional.	El factor gamma se varía localmente en función de la intensidad media local de una zona.	Repositorio propio para experimentación.	Proporciona, un rendimiento superior en un 37.5% a esquemas de mejora de la imagen existentes.
(Sousa et al., 2016)	Presentar una técnica de resonancia estocástica con ruido blanco Gaussiano para el mejoramiento de contraste y calidad visual de imágenes acústicas de sonar de barrido lateral.	Ruido estocástico y ruido blanco Gaussiano.	Repositorio propio para experimentación.	El contraste y la calidad de la imagen de salida se evaluó basándose en el contraste RMS, mejorando la calidad de la imagen en un 7%.
(Higaki et al., 2019)	Discutir técnicas para mejorar la calidad de las imágenes de diagnóstico por tomografía y resonancia magnética.	Aprendizaje profundo.	Repositorio propio para experimentación.	La SR aumenta la capacidad de diagnóstico al mejorar la resolución de las imágenes en un 4% de SSIM.
(Varonin et al., 2018)	Presentar un nuevo algoritmo de mejora de imágenes térmica basado en el procesamiento de la imagen en el dominio de la frecuencia.	Transformada logarítmica local y global de la transformación del histograma con ecualización espacial.	Repositorio propio para experimentación.	La imagen resultante del método propuesto es la más colorida y con menos ruido en PSNR en un 6% y en SSIM 3%.

Tabla 3.1 Resumen de artículos del estado del arte

Artículo	Objetivo	Técnicas/ Herramientas	Conjuntos	Resultados
Técnicas de mejoramiento de imágenes				
(Singh et al., 2017)	Presentar un nuevo método que utiliza métodos de mejora local y global en la misma imagen.	Métodos de mejora local y global en la misma imagen.	Repositorio propio para experimentación.	Este método funciona bien en la mayoría de las imágenes oscuras en un 4% de SSIM.
(Ueda & Suetake, 2019)	Presentar una nueva técnica para mejorar imágenes en color preservando el tono y no generar el problema de la gama de colores que comúnmente presentan muchas imágenes.	Especificación del histograma.	Repositorio propio para experimentación.	Los resultados demuestran que, en intensidad del contraste y la calidad de la imagen, se tuvo mejoras comparadas de forma objetiva con SSIM 4% y PSNR 7% con el estado del arte de este trabajo.
Detección de rostros				
(Adimoolam et al., 2022)	Rastrear objetos por medio de una selección de características y sustracción de fondo en sistemas de videovigilancia.	Detección de objetos mediante múltiples cámaras de videovigilancia por medio de múltiples canales.	CIFAR y datasets en tiempo real de videovigilancia.	Precisión del 98% comparado con métodos en el estado del arte.
(Menaka & Yogameena, 2021)	Centrarse en la eliminación del desenfoque en videos de videovigilancia de baja resolución para detección de rostros.	Algoritmo de des convolución ciega.	Chokepoint, OWN ATM1 2 y3, IISC Bangalore, Non-ATM 1 SC Face, ETH, ILIDS-VID, y PRID 2011.	Aumento significativo en un 35% de accuracy en la tasa de detección de rostros para su ámbito de aplicación de perfil y rostros frontales.

Tabla 3.1 Resumen de artículos del estado del arte

Artículo	Objetivo	Técnicas/ Herramientas	Conjuntos	Resultados
Detección de rostros				
(X. Chen et al., 2019)	Detección de rostros basado en características haar, empleando el escalado de la imagen.	Características de tipo haar	INRIA	Reducción del espacio de búsqueda y aumento de accuracy en un 45.7%.
(Cardenas et al., 2019)	Proponer un nuevo modelo para la detección de rostros en videos de baja resolución.	Morfología de la parte superior del cuerpo de las personas, haciendo uso de aprendizaje profundo.	CAVIAR y UCSP	Aumento de precisión con baja tasa de falsos positivos y un 39% de precisión en Caviar y 32% en UCSP.
(Ranjan et al., 2019)	Realizar un sistema con capacidad de detectar rostros y reconocerlos.	Deep Pyramid Single Shot Face.	Repositorio propio para experimentación.	Accuracy en un 65% en el conjunto de datos Pascal Faces y en los conjuntos de datos WIDER, UFFDD y FDDB.
(Zheng et al., 2010)	Estudiar la relación entre la resolución y la tasa de detección automática de rostros con la técnica seleccionada.	Transformada de censo modificada.	-	Alto rendimiento de accuracy en 35% en el funcionamiento del sistema en comparación con la técnica original.
(Zhou et al., 2018)	Comparar los detectores de rostro más avanzados junto con sus protocolos de diseño.	Comparación de detectores de rostros.	-	Métodos tradicionales y de aprendizaje profundo no son suficientemente robustos para muestras de prueba borrosas.
(Kumar et al., 2019)	Estudiar la importancia de la detección de rostros.	Histogramas y estadísticas.	-	Diversas técnicas ocupan distintos factores para trabajar.

Tabla 3.1 Resumen de artículos del estado del arte

Artículo	Objetivo	Técnicas/ Herramientas	Conjuntos	Resultados
Detección de rostros				
(Sun et al., 2018)	Realizar un nuevo esquema de trabajo mediante aprendizaje profundo.	FDDB (Face Detection Data Set Y Benchmark).	-	Realiza tareas de detección de objetos a una velocidad competitiva comparado con el estado del arte.
(Marciniak et al., 2021)	Analizar la fiabilidad de un sistema a partir de imágenes de baja resolución.	Reglas generales y directivas.	-	Las imágenes con baja calidad afectan la detección y reconocimiento facial pero no impiden trabajar con ellas.
(Chi et al., 2019)	Presentar un detector de rostros de una sola toma, denominada Red de Refinamiento Selectivo.	Aprendizaje profundo con operaciones de clasificación y regresión de dos pasos de forma selectiva en un detector de rostros.	AFW, PASCAL face, FDDB y WIDER FACE	La Red de Refinamiento Selectivo. alcanza el rendimiento de detección más avanzado en un accuracy de 85% comparado con el estado del arte.

Conclusiones del capítulo

Las categorías revisadas en el estado del arte fueron detección de rostros, súper resolución y mejoramiento de imágenes. En cada una de ellas, se buscó las técnicas más recientes y utilizadas, esto con el objetivo de conocer diversas herramientas para el desarrollo del sistema.

En la categoría de detección de rostros se analizó literatura específica en el ámbito de videovigilancia, en la cual los autores comentan las problemáticas que se tienen en localizar un rostro frente a los problemas de escala, iluminación o baja resolución de las imágenes. Por eso, cada autor aborda diversas problemáticas para su solución. Las más comunes son las problemáticas con la iluminación y contraste, la pérdida de calidad de las imágenes debido a la compresión que sufren los archivos al guardarse, los rostros rotados debido a la posición de las personas, diversos ángulos que pierden la visibilidad de características faciales que dificultan un reconocimiento facial, entre otras.

La categoría de súper resolución ofrece varias técnicas para convertir imágenes de baja resolución a alta resolución. Actualmente la mayoría utiliza aprendizaje profundo, y también relatan como pueden trabajarse con los modelos pre-entrenados sin la necesidad de ocupar una red neuronal desde cero. Dentro de esta categoría los modelos más utilizados son EDSR, ESPCN, FSRCNN y LapSRN, los cuales demuestran tener un desempeño superior al 90% en el nivel de mejora de las imágenes.

Por último, el mejoramiento de imágenes ofrece una variedad de técnicas para tratar diversos aspectos, en su mayoría resuelven el mejoramiento de contraste por medio de convolución. La técnica más utilizada reporta ser el mejoramiento de imagen por medio de ecualización del histograma con una mejora en el contraste e iluminación de las imágenes.

Capítulo 4 Diseño del sistema

En este capítulo se describe el análisis y diseño del proceso necesario para el desarrollo del sistema de mejoramiento de imágenes para sistemas de videovigilancia de baja calidad: este proceso incluye: el funcionamiento del sistema y las herramientas utilizadas (software y equipo) para el desarrollo del sistema.

4.1 Selección del Ambiente del sistema

Para el desarrollo del sistema se seleccionó el lenguaje de programación Python, por su amplio repertorio de herramientas y librerías con las que permite trabajar. Una de esas librerías es OpenCV. Algunas de las ventajas con las que se cuenta al trabajar con Python son (Rivera & Zambrano., 2022):

- Facilidad en su aprendizaje.
- Sistema de código abierto.
- Multiplataforma.

Así como Python proporciona ventajas al trabajar con él, la librería OpenCV, la cual puede definirse como una librería de software de código abierto, ofrece ventajas en cuanto a eficiencia computacional y un amplio enfoque en el área de visión artificial. Algunas otras ventajas son:

- Comprende algoritmos que son capaces de procesar imágenes por visión artificial en varios niveles.
- Se constituye por muchos clasificadores estadísticos y métodos de agrupación.
- Puede almacenar videos e imágenes.

4.2 Funcionamiento del sistema

La finalidad del sistema es mejorar imágenes de videovigilancia de baja calidad con problemas de iluminación y contraste, la implementación de este sistema es conformado por una serie de procesos y condiciones de control como se muestra en la Figura 4.1

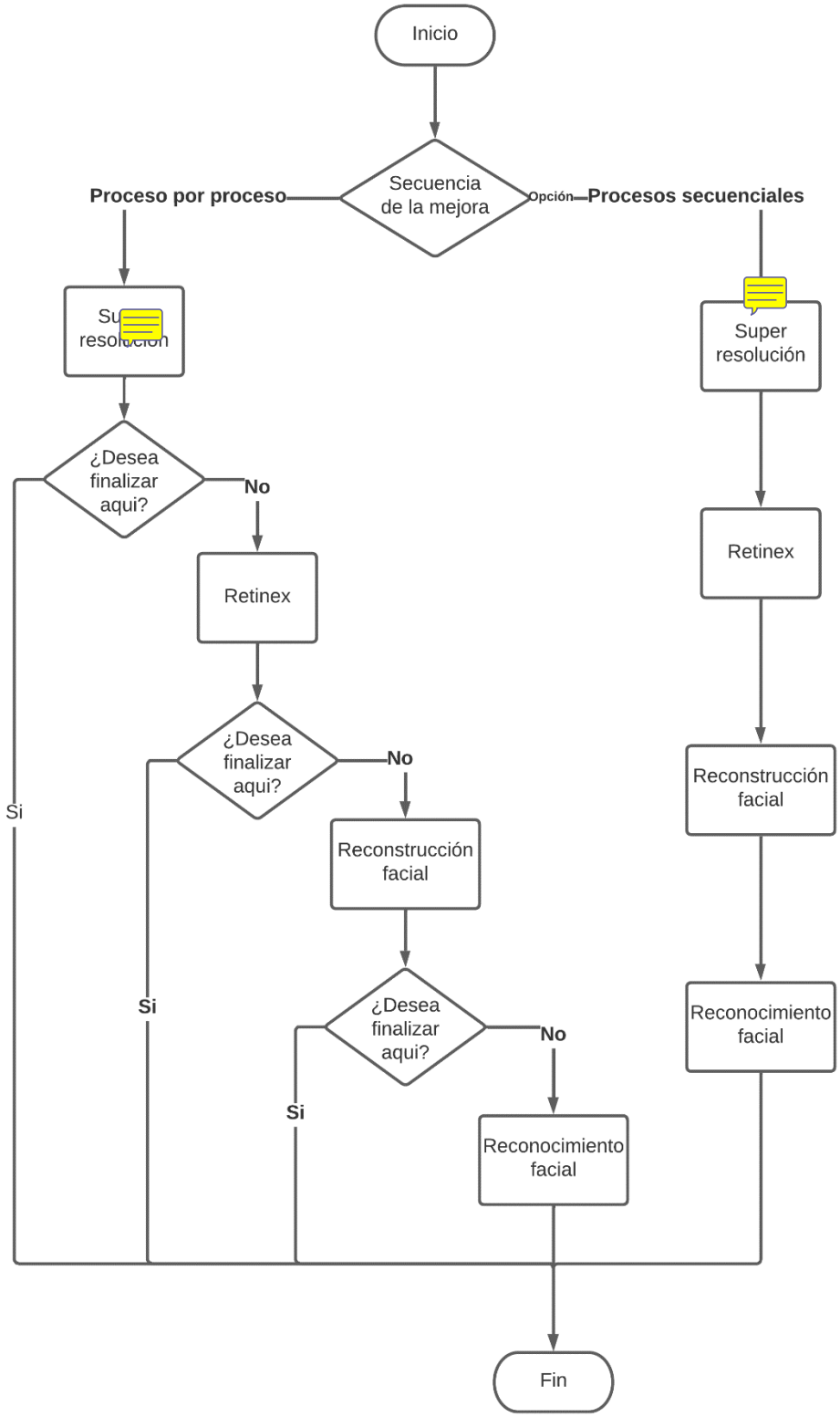

Figura 4.1 Diagrama de flujo del sistema

4.3 Diseño del sistema

Para facilidad del usuario se creó una pequeña interfaz con ayuda de la biblioteca Tkinter 8.5 (Shipman, 2013) que se tiene instalada por defecto al instalar Python. La interfaz se integra de 7 botones de operación para el usuario como se muestra en la Figura B1, súper resolución que ejecuta el algoritmo seleccionado y descrito en esta tesis, el algoritmo de mejora de iluminación Retinex, Reconstrucción del rostro detectado en la imagen con el algoritmo GFPGAN y, también el reconocimiento facial para comprobar que el mejoramiento de la imagen no se ha visto afectado, por último está el botón de todos los procesos, el cual se encarga de ejecutar todos los procesos en secuencia. Ver Figura 4.2.

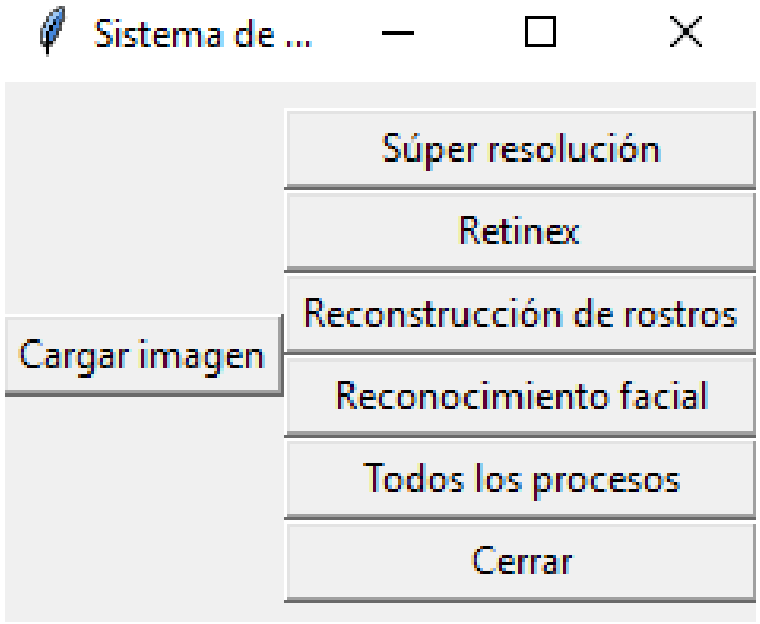

Figura 4.2 Vista principal de la interfaz

El sistema permite el ingreso de la imagen de forma individual, por lo que se asignó un botón correspondiente al nombre de la función que realiza, que es cargar una imagen.

Al presionar el botón de “Cargar imagen” se muestra la ventana que permite buscar en la computadora la imagen con que se desea trabajar (en formato .jpeg, .jpg, .png y, .bmp) como se puede apreciar en la Figura 4.3.

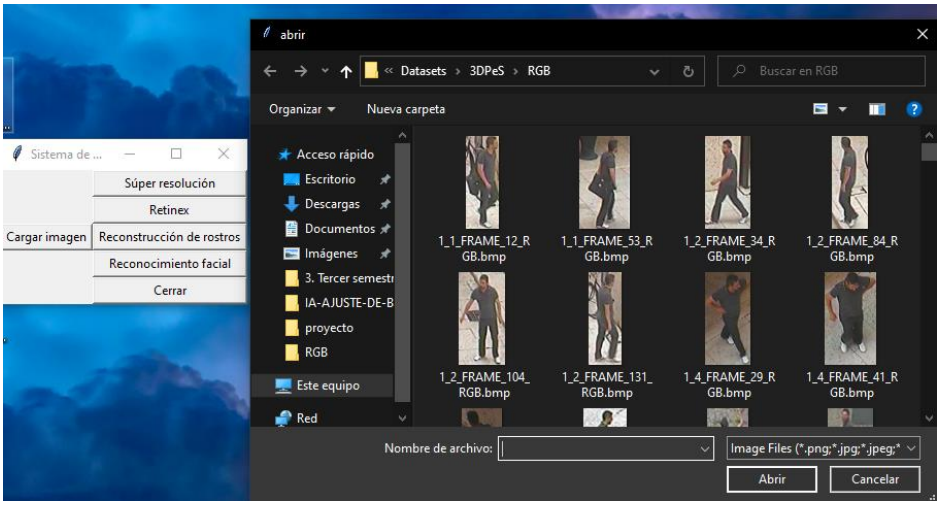

Figura 4.3 Vista de selección de una imagen a procesar.

Una vez que la imagen se haya cargado, se puede seleccionar el proceso que se desee realizar a la imagen o, si se desea se pueden aplicar todos los procesos en forma secuencial, es decir al terminar uno inmediatamente comienza otro. Una vez que se selecciona el proceso que se desea aplicar a la imagen el cuadro donde se encuentra la imagen cargada se sustituye por la imagen nueva de salida como se muestra en la Figura 4.4.

Figura 4.4 Ejemplo de resultado final de todos los procesos

4.3 Herramientas Utilizadas y Características de la Computadora

La interfaz se creó mediante la librería Tkinter 8.5 (Shipman, 2013), el sistema se creó utilizando Python 3.10.5. Para trabajar con imágenes se utilizaron las librerías de Opencv-contrib-python 4.5.5.64 mismas que permiten guardar la imagen para, posteriormente, ser evaluada.

Para el algoritmo de súper resolución solo se necesita de la librería Opencv-contrib-python 4.5.5.64 que se mencionó anteriormente.

Para la detección de rostros se utiliza MediaPipe 0.8.9.1 y para la parte de reconstrucción de rostros se utilizó el algoritmo GFPGAN (Wang et al., 2021) en su versión 1.3 que está basado en una versión anterior, pero con resultados de restauración más naturales. Las librerías requeridas por este algoritmo son torchvision, tqdm, yapf, pyyaml, scipy, basicsr, facexlib, lmdb, numpy, tb-nightly y torch.

Tanto la interfaz como el sistema se desarrollaron en una laptop Dell Inspiron 4557 con Windows 10 versión 21H1 64 bits, con procesador Intel(R) Core (TM) i7-1005G1 CPU @ 1.20GHz, disco duro de 1TB, memoria RAM 16GB.

En el anexo B, se muestran los comandos en el símbolo del sistema en Windows para la instalación de las librerías necesarias para el funcionamiento del sistema si no se tiene el ejecutable.

Capítulo 5 Experimentación y resultados

En este capítulo se presentan los datasets con los que se trabajó, así como los diferentes casos de experimentación que se realizaron.

5.1 Datasets

A continuación, se describen los datasets ocupados en la experimentación realizada, cada conjunto de datos fue seleccionado por los tamaños que presentan y su baja calidad, ya que este trabajo de tesis busca mejorar imágenes de baja calidad reales, el último dataset presentado fue seleccionado para contar con imágenes de referencia de una mejor calidad y así evaluar el nivel de mejora del trabajo de tesis presentado.

3DPeS (3D People Surveillance Dataset) (Baltieri et al., 2011): Es un conjunto de datos de videovigilancia, diseñado principalmente para la identificación de personas con un campo de visión no solapado. El conjunto es utilizado para detección de personas, el seguimiento, el análisis de acciones y el análisis de trayectorias. Permite detectar personas desde diferentes puntos de vista. Contiene alrededor de 1012 imágenes de 200 personas, estas 1012 imágenes tienen un tamaño aproximado desde 60*60 hasta 100*100 píxeles de tamaño, el área del rostro va desde 25*25 hasta 30*30 píxeles y presenta problemáticas de iluminación no controlada y rotación en diferentes ángulos de las personas. En la Figura 5.1 se observa una muestra del dataset.

Figura 5.1 Muestra del dataset (Baltieri et al 2011)

ChokePoint (Wong et al., 2011): Es un conjunto de datos de video diseñado para experimentar la identificación y verificación de personas en un entorno real de videovigilancia. Consta de tres cámaras sobre varios puntos para capturar personas caminando de forma natural, los rostros de este conjunto tienen variaciones en cuanto a iluminación, la pose, la nitidez, así como la desalineación causada por la localización y detección automática de rostros. Se cuentan con tres cámaras, es probable que una de ellas capture los rostros de manera casi frontal, pero dependerá

de la posición de las personas. El dataset contiene un aproximado de 48 secuencias de video y 64,204 imágenes de caras. El tamaño de estas imágenes es aproximadamente de 800*600 píxeles, mientras que el tamaño de los rostros va desde 40*40 hasta 200*200 píxeles, en este dataset se tiene la problemática de rotación en diferentes ángulos, iluminación en entornos controlados y brillo de forma variante en las imágenes. En la Figura 5.2 se observa una muestra del mismo.

Figura 5.2 Muestra del dataset (Wong et al 2011)

LFW (Labeled Faces in the Wild) (UMass, 2011): Es un conjunto de datos de imágenes de rostros diseñado para estudiar el reconocimiento de rostros sin restricciones. Este conjunto fue creado por la Universidad de Massachusetts con alrededor de 13,000 imágenes de diferentes personas famosas recopiladas de la web. Cada cara ha sido etiquetada con el nombre de la persona, se cuenta con 1680 identidades con dos o más fotos por cada una. Según investigadores de diferentes modelos de reconocimiento facial, las imágenes produjeron resultados superiores para la mayoría de los algoritmos de reconocimiento de rostros.

El tamaño de estas imágenes es de 250*250 píxeles, todas las imágenes de este dataset fueron reescaladas a ese tamaño. Estas imágenes no presentan problemáticas, ya que, es utilizado con la finalidad de probar algoritmos de reconocimiento facial. Todos los rostros de este dataset se encuentran totalmente de frente. En la Figura 5.3 se observa una muestra del dataset.

Figura 5.3 Muestra del dataset (UMass, 2011)

5.2 Casos de experimentación

La sección se integra de cinco casos de prueba.

Caso 1: Algoritmos de Súper Resolución. El objetivo de esta prueba es comparar de manera objetiva (con métricas de calidad) los cuatro algoritmos de súper resolución seleccionados, buscando determinar el algoritmo más eficiente para mejorar una imagen considerando los criterios de brillo, contraste y calidad.

Caso 2: Localización de rostros. El objetivo de esta prueba es comparar el desempeño de cuatro herramientas de detección de rostros en entornos de videovigilancia, con dos conjuntos de imágenes. Es decir, con fotografías que presentan diferente calidad, tamaño, escala, iluminación, rotación y presencia de lentes graduados. Estas herramientas son evaluadas con las métricas de clasificación. Esta prueba se divide en dos partes, ver sección 5.2.2.

Caso 3: Algoritmo de reconstrucción de rostros. Este caso presenta los resultados de la implementación del algoritmo de reconstrucción de rostros seleccionado, el objetivo de esta experimentación es evaluar el desempeño del algoritmo al aplicarse a las imágenes de calidad baja y media, así como conocer y analizar los puntos que toma en cuenta en los rostros de las personas para reconstruir la imagen.

Caso 4: Reconocimiento de personas. El objetivo de este caso es analizar qué tanto afecta la metodología propuesta en esta tesis a las imágenes para la tarea de identificación de las personas. Para ello, se utiliza el conjunto de datos para reconocimiento facial, imágenes a las cuales se le aplica un proceso para bajar la calidad de las mismas. La evaluación se realiza mediante las métricas clásicas de clasificación ya mencionadas en la sección 3.6.

Caso 5: Reconocimiento facial con las imágenes finales del sistema. el objetivo de esta experimentación es comprobar que, al mejorar la imagen, la identidad de la persona no se ve afectada.

En las siguientes secciones se detallan los casos mencionados.

5.1.1 Caso 1: Algoritmos de súper resolución

Para esta experimentación se utilizaron 168 imágenes de la base de datos 3DPes. Sólo se analizaron las imágenes que presentaban una posición frontal de las personas. Dichas imágenes fueron recortadas y solamente se trabajó con la región del rostro con la finalidad de conocer de forma visual y cuantitativa la eficiencia de los algoritmos seleccionados. En la Tabla 5.1 se presentan los resultados obtenidos. Los mejores resultados están marcados en color azul.

Tabla 5.1 Resultado de las métricas de evaluación de mejora de la imagen

Algoritmo	SSIM	NCC	NIQE	BIQI
EDSR	0.164	0.288	0.486	0.278
FSRCNN	0.146	0.182	0.362	0.369
ESPCN	0.277	0.188	0.278	0.302
LapSRN	**0.278**	**0.363**	**0.538**	**0.481**

Como se observa el algoritmo con mejores resultados fue LapSRN, el cual obtuvo valores ligeramente superiores que los otros algoritmos, demostrando que su pirámide Laplaciana ayuda de gran forma a la mejora de imágenes. Este modelo predice de forma progresiva los residuos de alta frecuencia de manera gruesa a fina y al sustituir a la interpolación bicúbica predefinida en capas convolucionales, la red se optimiza con una función de pérdida robusta.

De acuerdo con las métricas, LapSRN es el mejor algoritmo, sin embargo, visualmente esta mejora hace visibles algunas líneas y manchas imperceptibles en las imágenes originales, lo que provoca problemas en la detección de rostros, ya que se detectan menos rostros de los que existen dentro de las imágenes, así como también se presentan problemas en la reconstrucción de rostros, estos problemas se presentan en los casos de experimentación siguientes.

Por ello, se realizó una evaluación visual y se observó que el algoritmo EDSR (cuyos valores en las métricas son inferiores) mejora la calidad de la imagen, pero a un nivel ideal, es decir, sin modificar su contenido. Es importante señalar que esta evaluación se hizo de manera cualitativa.

5.1.2 Caso 2: Localización de rostros

En esta experimentación se comparan las herramientas de localización de rostros descritas en el marco teórico en el conjunto de datos 3DPes que se integra de imágenes de baja calidad y la base de datos Chokepoint que tiene fotografías de calidad media a baja. Para las pruebas se consideraron las siguientes variantes:

- Primero, se comparan las 4 herramientas de localización y la mejor se evalúa con el detector de rostros que viene incluido con el algoritmo de reconstrucción de rostros.
- Se realiza la prueba con las imágenes originales y con las imágenes resultantes de aplicarles el algoritmo de súper resolución.

- Se analizan imágenes donde es visible todos los componentes del rostro y posteriormente imágenes que contienen rostros ocluidos debido a la rotación, sombras, etc.

A continuación, se presenta la experimentación realizada para conocer la mejor herramienta de detección de rostros.

Caso 2.1: Evaluación de las herramientas con el dataset 3DPes.

Como ya se mencionó, la base de datos 3DPeS tiene un total de 1,013 imágenes. Se dividió en dos partes: 1) En 168 imágenes que presentan rostro cuyos puntos clave son visibles y 2) 845 imágenes donde las personas se encuentran de espaldas o el rostro tiene un grado de mayor rotación, se presentan sombras y provoca que el rostro esté ocluido de manera parcial o total. Se considera que las imágenes analizadas son de baja calidad debido a que su tamaño va desde 80x80 hasta 100x100 píxeles y el tamaño de los rostros está entre 25x25 hasta 30x32 píxeles.

Primero se ingresaron las imágenes sin realizar ninguna mejora, esto con la finalidad de conocer el comportamiento de cada herramienta. En la Figura 5.4 se muestra un ejemplo de los resultados obtenidos de esta detección.

Posteriormente, se aplica el algoritmo EDSR de súper resolución a las imágenes y las fotografías mejoradas ingresan a las herramientas de localización de rostros. En la Figura 5.5, se muestra un ejemplo de los resultados obtenidos de esta detección. En la Tabla 5.2 se muestran los resultados obtenidos.

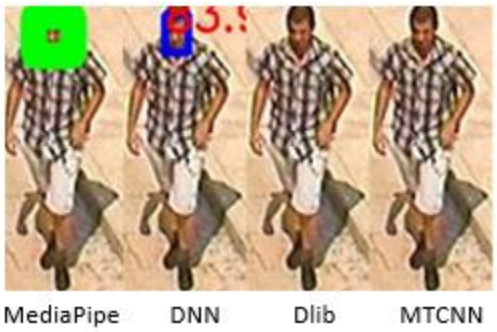

Figura 5.4 Detección de rostros sin mejora a la imagen

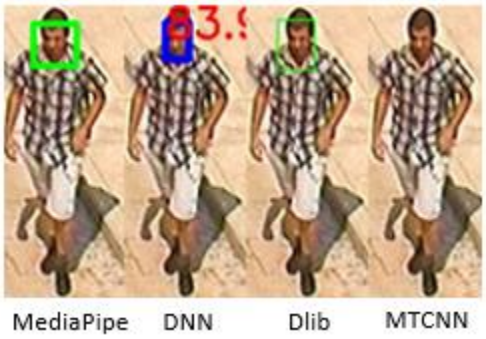

Figura 5.5 Detección de rostros con mejora de súper resolución a la imagen

Tabla 5.2 Detección de rostros en el 3DPeS, con las imágenes y después de aplicarles el algoritmo de súper resolución EDSR

Algoritmo	Rostros detectados	
	Imágenes originales	Imágenes + Súper resolución
MTCNN	1	1
Dlib	0	13
MediaPipe	**61**	**63**
DNN	25	31

Como se observa en la segunda columna de la Tabla 5.2, MediaPipe y DNN fueron los únicos capaces de detectar algunos rostros. MTCNN y Dlib tienen un pobre resultado debido al tamaño pequeño de los rostros. La aplicación del algoritmo de súper resolución ayuda en la definición del rostro, pero su contribución no logra que las herramientas analizadas mejoren de manera sustancial sus resultados. Y como se observa en la tercera columna, los resultados se mantienen, siendo MediaPipe el mejor detector de rostros.

Con el objetivo de conocer la capacidad de detección de cada herramienta ante factores más complejos como la rotación, presencia de sombras y oclusión de algunos componentes del rostro, se realizó un experimento con las 845 imágenes restantes del conjunto de imágenes ya mejoradas con el algoritmo de súper resolución. Los resultados se observan en la Tabla 5.3.

Tabla 5.3 Detección de rostros con rotación en diferentes ángulos con mejora en las imágenes

Algoritmo	Rostros detectados	
	Imágenes originales	Imágenes + SR
MTCNN	0	1
Dlib	0	1
MediaPipe	**120**	**140**
DNN	13	15

Es importante hacer notar que en estas 845 imágenes es altamente compleja la localización del rostro, porque como ya se comentó, las imágenes presentan oclusión debido a la rotación,

sombras, uso de aditamentos como lentes y el área del rostro es muy pequeña. Sin embargo, MediaPipe es capaz de localizar 140 rostros.

Con la finalidad de conocer de forma cuantitativa el desempeño de cada herramienta implementada, cada modelo se evaluó bajo las métricas de detección, como se observa en la Tabla 5.4. Para el cálculo se tomaron en cuenta las 168 imágenes donde se observa de frente a las personas con los puntos clave que toma en cuenta cada detector y estas imágenes fueron probadas con súper resolución y sin súper resolución.

Tabla 5.4 Resultados de las métricas de clasificación con las herramientas de detección de rostros

Algoritmo	Accuracy	Recall	Precision	F1 Score
Sin mejora de súper resolución				
Dlib	0	0	0	0
MTCNN	0.005	0.005	1	0.011
DNN	0.148	0.149	0.961	0.257
MediaPipe	**0.363**	**0.365**	**0.983**	**0.532**
Con mejora de súper resolución				
Dlib	0.077	0.077	1	0.143
MTCNN	0.005	0.005	0.5	0.011
DNN	0.184	0.184	1	0.310
MediaPipe	**0.375**	**0.375**	**1**	**0.545**

Como se observa, la herramienta con mejores resultados fue MediaPipe al tener una tasa de detección mayor a las otras herramientas, demostrando que, sus 6 puntos clave son de ayuda en una baja resolución o rotación en el que se encuentre el rostro. Por lo tanto, MediaPipe será utilizada en el caso 2.2.

Caso 2.2: Localización el rostro comparando MediaPipe y GFPGAN con el conjunto de datos 3DPes.

En esta experimentación, se consideran como técnicas de mejoramiento a los algoritmos de Super resolución, retinex multiescala y la reconstrucción del rostro con GFPGAN. En la Tabla 5.5 se presentan las combinaciones realizadas usando la herramienta MediaPipe, estos mismos conceptos son utilizados en la Tabla 4.5 con las combinaciones con el detector de rostros del algoritmo GFPGAN, donde:

- **GFPGAN** = Generative Facial Prior-Generative Adversarial Network.
- **Súper resolución** = Técnica EDSR (Enhanced Deep Super-Resolution Network).
- **Retinex** = Algoritmo que mejora la iluminación de las imágenes.

- **ESPCN** = Técnica de súper resolución por sus siglas en inglés (Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network).

Tabla 5.5 Localización de rostros mejorados del dataset 3DPeS con MediaPipe

Combinaciones de mejoramiento de imagen	Rostros detectados	Accuracy	Recall	Precision	F1 Score
Imágenes con rostros originales y con problemas	163 de 845 (136 VP y 27 FP)	0.973	1	0.834	0.910
Combinaciones de técnicas de mejoramiento					
Súper resolución	62 de 168	0.369	0.369	1	0.541
Retinex + súper resolución	57 de 168	0.339	0.339	1	0.506
Retinex + súper resolución + GFPGAN	57 de 168	0.339	0.339	1	0.506
Retinex + GFPGAN + súper resolución	**61 de 168**	**0.363**	**0.363**	**1**	**0.532**
GFPGAN + súper resolución + retinex	25 de 168	0.148	0.148	1	0.257
Súper resolución + GFPGAN + retinex	21 de 168	0.125	0.125	1	0.222
GFPGAN (como reconstructor)					
GPFGAN	60 de 168	0.357	0.357	1	0.526
GPFGAN + retinex	48 de 168	0.285	0.285	1	0.443
Retinex + GFPGAN	58 de 168	0.345	0.345	1	0.513

Como se puede observar en la fila 6 de la Tabla 5.5 (marcada en color azul), la combinación de Retinex, GFPGAN y el algoritmo de súper resolución EDSR permite localizar una mayor cantidad de rostros en las imágenes. Lo que significa que esta combinación para mejorar la calidad de una imagen ayuda a MediaPipe con su detección, mientras que, en la última fila de la tabla, muestra los resultados de la detección de rostros exclusivamente con la herramienta de reconstrucción y retinex, sin utilizar súper resolución.

A continuación, en la Tabla 5.6, se presentan los resultados de localización utilizando el módulo de detección de rostros de GFPGAN y su combinación con las técnicas de retinex y súper resolución.

Tabla 5.6 Combinaciones realizadas con el detector de rostros de GFPGAN

Combinaciones de mejoramiento de imagen	Rostros detectados	Accuracy	Recall	Precision	F1 Score
Imágenes con rostros no visibles y cierto grado de oclusión (sin ningún procesamiento)	0 de 845	0	0	0	0
Combinaciones de técnicas de mejoramiento					
Súper resolución - EDSR	**75 de 168**	**0.446**	**0.446**	**1**	**0.617**
Súper resolución - ESPCN	66 de 168	0.393	0.363	1	0.564
Retinex + GFPGAN + Súper resolución	48 de 168	0.286	0.286	1	0.444
Retinex + Súper resolución	52 de 168	0.310	0.310	1	0.473

Como se puede observar en la Tabla 5.6, el utilizar exclusivamente el algoritmo de súper resolución EDSR permite la detección de más rostros. Como se mencionó, este algoritmo tuvo resultados bajos en la evaluación con métricas de calidad de la imagen, mientras el algoritmo ESPCN con mayores valores en las métricas, reporta nueve rostros menos.

Caso 2.3 Localizar el rostro en una base de datos de media-baja calidad

El dataset Chokepoint tiene un total de 2,292 imágenes donde en 656 no se cuentan con la presencia de un rostro y en 1,636 se cuenta con la presencia de al menos un rostro. El tamaño de estas imágenes es de 800x600 píxeles, mientras que el tamaño de los rostros va desde 40x40 píxeles hasta 200x200.

En la Tabla 5.7 se encuentran todas las combinaciones realizadas a las imágenes utilizando la herramienta MediaPipe donde:

- **GFPGAN** = Generative Facial Prior-Generative Adversarial Network.
- **Súper resolución** = Técnica EDSR (Enhanced Deep Super-Resolution Network).
- **Retinex** = Algoritmo que mejora la iluminación de las imágenes.
- **ESPCN** = Técnica de súper resolución por sus siglas en inglés (Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network).

Para este caso, no se realizaron todas las combinaciones de mejoramiento de imagen del caso anterior, ya que, se tomaron solo las que obtuvieron mejores resultados para una nueva comparación.

Tabla 5.7 Combinaciones realizadas con el detector de rostros Mediapipe y el dataset Chokepoint

Combinaciones de mejoramiento de imágenes	Rostros detectados	Accuracy	Recall	Precision	F1 Score
Imágenes sin ningún procesamiento	971 de 1636	0.710	0.594	1	0.745
Retinex	371 de 1636	0.443	0.222	0.970	0.361
Súper resolución					
Súper resolución	1680 de 1636	0.981	1	0.974	0.987
GFPGAN					
GFPGAN	1596 de 1636	0.983	0.976	1	0.988
Retinex	1574 de 1636	0.973	0.962	1	0.981

Como se puede observar en la Tabla 5.7 con la base de datos Chokepoint, las combinaciones en las cuales se aplica súper resolución se detectan más rostros, incluso más de los que realmente existen. En muchos casos, la propuesta de localización que realiza MediaPipe se debe a que para la herramienta es suficiente ubicar la posición de los ojos para proponer la región potencial del rostro. Y en el video las personas ingresan al pasillo y salen del mismo en un espacio cercano a la cámara, lo que provoca que cuadro a cuadro se vayan perdiendo facciones del rostro que son puntos clave para ser considerado un rostro como lo son la nariz y la boca, pero MediaPipe sigue proponiendo su localización al detectar únicamente los ojos, esto ocasiona que se generen falsos positivos como se muestra en la columna de súper resolución, es decir, indican que existen más rostros de lo que se encuentran presentes en las imágenes, pese a que verdaderamente existe un rostro en la posición de los ojos, estas imágenes no son útiles para una reconstrucción facial, ya que se necesitan de todos los elementos clave de un rostro. En la Figura 5.6 se muestra un ejemplo de estas imágenes.

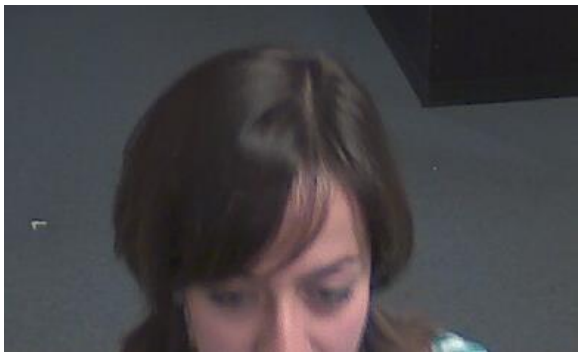

Figura 5.6 Ejemplo de falsos positivos con Mediapipe

A continuación, en la Tabla 5.8, se presentan los resultados obtenidos combinando las técnicas de mejoramiento y el detector de rostros de GFPGAN.

Tabla 5.8 Combinaciones realizadas con el detector de rostros GFPGAN y el dataset Chokepoint

Combinaciones de mejoramiento de imagen	Rostros detectados	Accuracy	Recall	Precision	F1 Score
Imágenes sin ningún procesamiento	1,596 de 1,636	0.983	0.976	1	0.988
Retinex	1,574 de 1,636	0.973	0.962	1	0.981
Súper resolución					
Súper resolución – EDSR	**1,680 de 1636**	**0.981**	**1**	**0.974**	**0.987**
Súper resolución – ESPCN	1,678 de 1636	0.982	1	0.975	0.987

Como se puede observar, los resultados en general son buenos ya que las imágenes proporcionan rostros de buen tamaño y con suficiente información. Sin embargo, se puede ver que los algoritmos de súper resolución generan nuevamente falsos positivos. En este caso, de igual forma el algoritmo GFPGAN toma los ojos como un punto clave para la detección del rostro e intenta dibujar cómo luce el rostro completo de la persona. En la Figura 5.7 se muestra un ejemplo de estos falsos positivos.

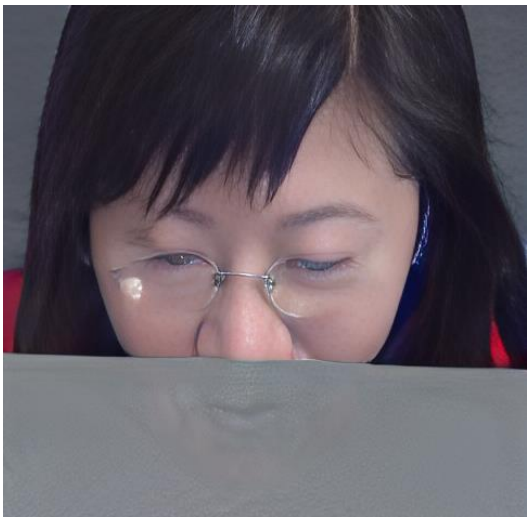

Figura 5.7 Ejemplo de falsos positivos con GFPGAN

4.1.3 Caso 3: Evaluación del algoritmo GFP-GAN en su fase de reconstrucción de rostros

Una vez que el rostro es localizado la imagen se recorta y solo se trabaja con el área del rostro. Posteriormente, se aplica el algoritmo GFP-GAN en su fase de reconstrucción. La imagen de

entrada tiene un aproximado de 100x100 píxeles en el caso de la base de datos 3DPes y 200x200 en el conjunto de datos Chokepoint, dando como imagen de salida una imagen del mismo tamaño.

La reconstrucción de la imagen de forma visual se aprecia significativamente. Sin embargo, debido a que no se cuenta con una imagen de referencia de cómo luce realmente el rostro de la persona, la reconstrucción solo puede ser evaluada de forma cualitativa y subjetiva. En la Figura 5.8, se muestra un ejemplo del resultado de la aplicación de las distintas técnicas y sus combinaciones con el conjunto de datos 3DPes.

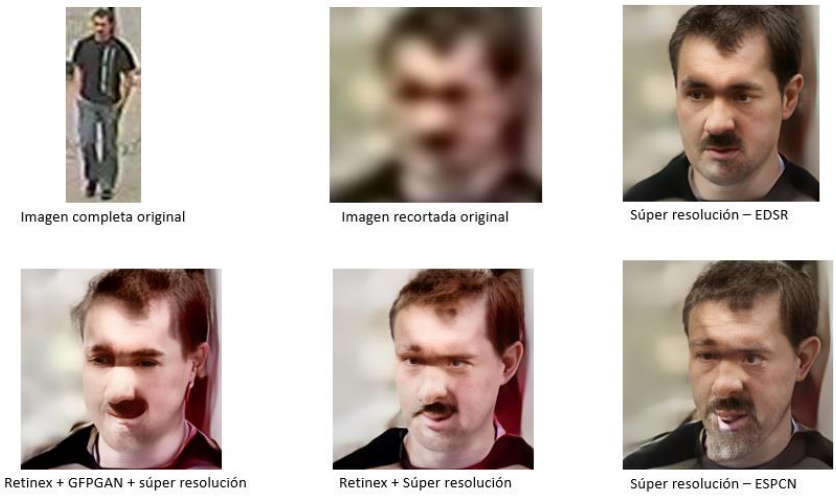

Figura 5.8 Ejemplo de reconstrucción del rostro bajo diferentes mejoramientos de imagen con el dataset 3DPeS

Como se puede observar, visualmente se observa un mayor contraste y detalle de las facciones con solo aplicar el algoritmo de súper resolución ESPCN. En la Figura 5.9 se muestran ejemplos de esta misma combinación, pero con imágenes de la base de datos Chokepoint.

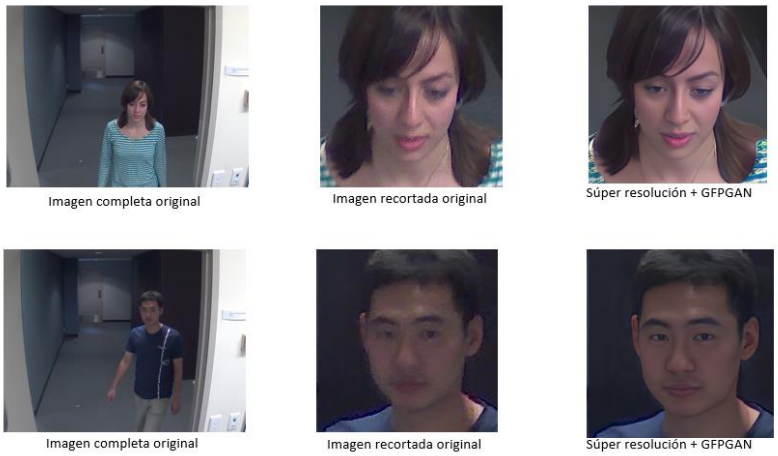

Figura 5.9 Ejemplo de reconstrucción del rostro bajo diferentes mejoramientos de imagen con el dataset Chokepoint

5.1.4 Caso 4: Reconocimiento de las personas

El objetivo del experimento tiene la finalidad de evaluar las técnicas de mejoramiento aplicadas (el algoritmo de súper resolución y retinex), y su influencia en el reconocimiento de la identidad del rostro. Además, determinar qué tan cercano es el resultado respecto a la imagen original, para ello se consideró el conjunto de datos LFW y se tomaron 220 imágenes aleatoriamente.

Para modificar la calidad de las imágenes se cambió la escala, el contraste y también se aplicaron cambios en la iluminación tener imágenes oscuras y brillantes. La Tabla 5.9 muestra los valores que se tomaron para reducir la calidad de la imagen.

Tabla 5.9 Cambios aplicados al dataset para bajar su calidad

	Contraste	Brillo	Escala	No. de imágenes
Img. con iluminación alta	1.5	25	100*100	110
Img con iluminación baja	0.9	-30	100*100	110

La Tabla 5.10 presenta en la primera fila el resultado de la comparación entre la imagen original y la imagen a la cual se le bajó su calidad. El segundo renglón muestra el resultado de similitud entre la imagen original y la imagen que paso por toda la metodología de mejora propuesta.

Tabla 5.10 Comparación entre la imagen original y la imagen modificada

Imagen	SSIM	NCC	NIQUE	BIQI
Original vs baja calidad	0.675	0.650	0.580	0.620
mejorada	0.666	0.643	0.543	0.578

Recordando, las primeras dos métricas son con referencia y las últimas dos sin referencia, todas ellas proporcionan valores en el rango de entre 0 y 1. Para SSIM entre más cerca esté del 1 existe mayor similitud entre las imágenes comparadas. En el caso de NCC de igual forma, entre mayor sea su valor más calidad se tiene en la imagen. Mientras que en las métricas sin referencia que son NIQE y BIQI entre más cercanos estén sus valores a 0 mejor calidad tiene la imagen. Estas métricas calculan puntuaciones de calidad en función de las estadísticas de la imagen, por lo que sus valores siempre serán más bajos en comparación con métricas de referencia completa, siempre y cuando la calidad de la imagen haya sido mejorada.

Como se puede observar en las métricas sin referencia (NIQUE y BIQI) los valores son bajos, lo que significa una mejora en la calidad de la imagen. Las métricas de referencia completa también

tienen valores bajos, esto se debe a que, al reconstruir el rostro de las personas, se modifican algunos rasgos presentes en las imágenes originales como arrugas, color de ojos, así como iluminación y contraste. Por lo tanto, al compararlas existen características diferentes entre cada imagen. En la Figura 5.10 se muestra un ejemplo de la imagen original, modificada y la mejorada. Visualmente se puede ver cómo la imagen reconstruida tiene mejor calidad, pero en cuestión de métricas, la editada tiene características más similares a la original.

Figura 5.10 Ejemplo del resultado final entre la imagen original, editada y reconstruida

5.1.5 Caso 5: Reconocimiento facial con las imágenes finales del sistema

El objetivo de esta prueba es evaluar si la identidad de la persona se ve alterada con la mejora de la imagen. Se utilizó el algoritmo Face Recognition de la librería de OpenCV (OpenCV, 2023), este algoritmo es fácil de usar por la optimización que posee y, con una sola imagen es suficiente para su entrenamiento.

En este caso, se cuenta con 113 identidades dentro de las 220 imágenes procesadas del dataset LFW. Primero, se entrenó y evaluó el rendimiento de este algoritmo con las imágenes originales. Posteriormente, con las imágenes editadas de baja calidad y, por último, con las imágenes mejoradas del sistema propuesto (Súper resolución con EDSR, y GFPGAN como reconstructor). A continuación, en la Tabla 5.11 se muestra un promedio en porcentaje del resultado de cada experimento.

Tabla 5.11 Resultados del reconocimiento facial

Métrica	Original	Editada	Reconstruida
Precision	96%	93%	94%

Como se puede observar, con las imágenes originales se tiene un 96% de precisión en identificar a la persona. Cuando las imágenes tienen menor calidad, su precisión disminuye a 93%, esto

debido a los cambios de iluminación que presentan las imágenes. Pero, cuando el rostro ha sido reconstruido este porcentaje sube ligeramente, alcanzando un 94%. La exactitud no alcanza al porcentaje obtenido con las imágenes originales debido a que el algoritmo de reconstrucción elimina arrugas que puedan llegar a tener las personas, hace visibles líneas o ciertos atributos del rostro, cambia el color de los ojos, marca más las sombras presentes en el área del rostro e incluso en algunas ocasiones modifica la forma de los ojos y boca. Sin embargo, los rasgos fuertes se preservan. La Figura 5.11 muestra un ejemplo de algunos de estos casos.

Figura 5.11 Ejemplo de imágenes modificadas del dataset LFW.

5.3 Análisis de los resultados

En el ámbito de la mejora de imágenes y la detección de rostros, se han implementado diversas técnicas de súper resolución con el objetivo de mejorar la calidad de las imágenes de baja resolución. Al analizar los resultados cuantitativos de diferentes algoritmos de súper resolución, se ha observado un patrón interesante. En particular, se ha identificado que en situaciones donde las imágenes de entrada poseen una baja calidad se tiene un impacto significativo en la detección de rostros, ya que las imágenes de baja calidad pueden presentar sombras, ruido y distorsiones que dificultan la identificación precisa de los rasgos faciales.

En contraste, se ha notado que, en imágenes de mayor calidad, donde no se introducen sombras ni ruido significativo, los algoritmos de súper resolución han demostrado ser más efectivos al proporcionar imágenes mejoradas y nítidas. Esta mejora en la calidad de las imágenes es crucial para garantizar una detección de rostros precisa, ya que las características faciales se vuelven más distinguibles y se reduce la posibilidad de falsos positivos o negativos en el proceso de detección, en los resultados obtenidos se observa que al trabajar con imágenes de baja calidad es más efectivo utilizar algoritmos que puedan manejar este tipo de imágenes aunque obtengan bajos resultados en métricas de mejora de imagen, como se observa en la tabla 5.1.

En relación con el proceso de mejora de imágenes en sí mismo, se ha observado que las técnicas que se centran en la corrección de la iluminación y el contraste desempeñan un papel fundamental en el éxito de los procesos posteriores. Al realzar la iluminación y el contraste de las imágenes, se proporciona una base sólida para que las técnicas de detección de rostros puedan operar de manera más efectiva. La mejora en la calidad de la imagen no solo ayuda a revelar detalles sutiles en las características faciales, sino que también contribuye a una detección más precisa y confiable. Entre los algoritmos de detección de rostros evaluados, la herramienta Mediapipe ha demostrado ser la más completa y efectiva para abordar las diversas problemáticas mencionadas anteriormente como se muestra en la tabla 5.2. Su capacidad para lidiar con imágenes de baja calidad, distancias variables y diferentes condiciones de iluminación la convierten en una herramienta valiosa para la detección precisa de rostros en una variedad de contextos.

En resumen, el análisis de técnicas de súper resolución, mejora de imágenes y detección de rostros revela que la calidad de las imágenes de entrada, así como las técnicas de mejora aplicadas, juegan un papel crítico en el rendimiento de los algoritmos de detección de rostros. El algoritmo Mediapipe se destaca como una solución integral que aborda estas consideraciones y demuestra ser altamente efectivo en la detección precisa de rostros en una amplia gama de condiciones, esto se ve reflejado en la tabla 5.3 al aumentar la cantidad de rostros detectados

debido a su mejora, en la tabla 5.4, se pueden apreciar este aumento de mejora de forma cuantitativa mediante métricas de clasificación. Mientras que, en las tablas 5.5, 5.6, 5.7 y 5.8 se observan las diferentes combinaciones y pruebas con las herramientas MediaPipe y GFPGAN.

En el caso de reconocimiento de rostros, este solo se ve afectado en su precisión debido al cambio de rasgos existentes causados por la mejora de calidad de la imagen, como lo son la disminución de arrugas, en algunos casos el cambio de color de ojos, la presencia de lentes graduados y en algunos otros casos ligeros cambios en el peinado de la persona, para evaluar el nivel de mejora se deterioró la calidad de las imágenes utilizadas y así, poder aumentar su mejora con el trabajo de tesis realizado como se muestra en la tabla 5.9. Lo cual prueba que la mejora de la imagen no afecta la identificación de la persona, esta sólo disminuye su porcentaje de precisión debido a esta mejora, como se aprecia de forma cuantitativa en la tabla 5.10 y de forma porcentual en la tabla 5.11.

Capítulo 6 Conclusiones

En este capítulo se presentan las conclusiones y objetivos logrados del trabajo de tesis presentado.

6.1 Conclusiones generales

En el presente proyecto de tesis se desarrolló un sistema que, mediante técnicas de mejoramiento de imágenes se aumenta la calidad de estas. Este sistema toma en cuenta problemáticas de iluminación y contraste en las imágenes adquiridas en un ambiente no controlado.

El presente trabajo exploró diferentes algoritmos para el mejoramiento de imágenes de videovigilancia de baja calidad, con un enfoque específico en la detección y reconstrucción de rostros. Primero, para mejorar las condiciones de las imágenes se evaluaron cuatro algoritmos de súper resolución. Es importante hacer notar que el algoritmo que obtuvo los valores más altos en las métricas también agregaba líneas y sombras que afectan a la imagen. Y el algoritmo que presentó peores resultados en términos de métricas, al combinarse con retinex, permitía localizar una mayor cantidad de rostros, este algoritmo es EDSR. En otras palabras, si bien este algoritmo puede no ser el más adecuado para el mejoramiento de imágenes de videovigilancia, destaca la importancia de elegir cuidadosamente el enfoque de súper resolución según las necesidades y características de las imágenes.

En los hallazgos se encontró que la herramienta MediaPipe, posee una gran capacidad para identificar y utilizar puntos clave del rostro, es altamente efectiva en la detección de rostros en imágenes reales de baja calidad; es decir, cuando las imágenes tienen un tamaño pequeño y tengan problemas de contraste, iluminación, presencia de sombras e incluso tengan porciones del rostro ocluidas.

Por otra parte, el algoritmo GFPGAN mostró su utilidad en la reconstrucción de áreas faciales en las imágenes, aunque con ciertas modificaciones en las facciones del rostro. Esto sugiere que, si bien puede ser una herramienta prometedora para mejorar la calidad de las imágenes, es importante considerar las posibles alteraciones que pueda introducir.

La combinación estratégica de EDSR y GFPGAN se erige como la opción preeminente para lograr mejoras sustanciales. La sinergia entre estos dos algoritmos no solo optimiza la resolución, sino que también potencia la capacidad de reconstrucción facial, ofreciendo una solución integral y efectiva. La aplicación de EDSR como primer paso en este enfoque demostró ser fundamental.

Aumentar la resolución de las imágenes de baja calidad no solo mejoró la claridad visual, sino que también proporcionó una base sólida para el proceso posterior de reconstrucción facial, ya que en términos de computación esto ayuda a mejorar su detección y posteriormente a su reconstrucción. Un paso previo a la reconstrucción facial es la localización del rostro, para la cual MediaPipe demostró los mejores resultados por su gran capacidad de detección mediante los puntos clave del rostro de una persona, ayudando así a que la reconstrucción facial pudiera identificar las facciones del rostro de la persona y dando como resultado final una imagen con mayor resolución y facciones faciales mayormente definidas.

Los resultados obtenidos subrayan la importancia de seleccionar adecuadamente los algoritmos y enfoques utilizados, considerando tanto la precisión en la detección como la preservación de las características faciales. Estos resultados pueden servir como base para futuras investigaciones y mejoras en el campo de la videovigilancia y el procesamiento digital de imágenes.

Al destacar la combinación de los algoritmos EDSR y GFPGAN, se abre la puerta a futuras investigaciones y desarrollos en la mejora de imágenes de videovigilancia de baja calidad. Esta sinergia ofrece una base sólida para la innovación continua en la optimización de la calidad visual en aplicaciones específicas de videovigilancia.

En resumen, la combinación de EDSR y GFPGAN no solo resuelve algunos de los desafíos de la videovigilancia de baja calidad, sino que establece un estándar más alto para la mejora de imágenes en este contexto. Este enfoque integral representa un paso significativo en la búsqueda de soluciones prácticas y efectivas para la seguridad visual.

En conclusión, la integración de EDSR, MediaPipe y GFPGAN representa un avance significativo en la mejora de imágenes para la videovigilancia de baja calidad. Cada componente desempeña un papel crucial en este proceso, culminando en una solución integral que supera los desafíos de iluminación, contraste y tamaño pequeño de este entorno, marcando un hito en la eficacia de la seguridad visual.

6.2 Objetivos logrados

En la Tabla 6.1 se listan los objetivos del proyecto y una breve descripción de cómo se trabajó para cumplirlos.

Tabla 6.1 Objetivos logrados

Objetivo General	**Actividades realizadas**
Desarrollar un sistema de visión artificial que realice el mejoramiento de imágenes de rostros de baja calidad.	Se diseñó y desarrolló en Python un sistema de visión artificial que, mediante el procesamiento de una imagen es capaz de mejorar su calidad. El sistema alcanza hasta un 85% de mejora en la imagen y un 94% de conservación de identidad de la persona.
Objetivos específicos	
Objetivos	**Actividades realizadas**
Revisar en el estado del arte las técnicas utilizadas para súper resolución, detección de rostros y técnicas de mejoramiento de imágenes.	Se analizaron 70 artículos y se reportan 38 artículos en el estado del arte referentes a este trabajo. En las categorías de súper resolución, detección de rostros y, técnicas de mejoramiento de imágenes.
Analizar herramientas para detectar un rostro en ambientes reales e implementar uno.	Se analizaron 20 artículos y se reportan 12 artículos en el estado del arte referentes a la categoría de detección del rostro. El análisis permitió conocer, aplicar y evaluar 4 herramientas de localización
Estudiar y seleccionar técnicas de mejoramiento de imágenes.	Se analizaron 20 artículos y se reportan 12 artículos en el estado del arte referentes a la categoría de técnicas de mejoramiento de la imagen. Se aplico retinex y se cuenta con una revisión de 25 métricas respecto a la calidad de las imágenes.
Diseñar e implementar un sistema que realice el mejoramiento de las imágenes.	Se cuenta con un sistema que realiza el mejoramiento de la calidad de las imágenes mediante los algoritmos de super resolución, retinex multiescala y un algoritmo de reconstrucción facial**.**
Evaluar el sistema de mejoramiento de imágenes con métricas seleccionadas con base al estado del arte para ello.	El sistema fue evaluado con 4 métricas clásicas en el área de clasificación Exactitud (Accuracy), Sensibilidad (Recall), Precisión, F1 score y 4 métricas que evalúan la calidad de la imagen: SSIM, NIQE, NCC y BIQI.

6.3 Aportaciones

La principal aportación de este trabajo es el desarrollo de un sistema de visión artificial que mejora la calidad de imágenes de sistemas de videovigilancia de baja calidad con problemas de escala, iluminación y borrosidad.

El sistema es capaz de mejorar la calidad de la cara, a partir de rostros de un tamaño de 20x20 píxeles.

El sistema entrega como respuesta una imagen mejorada del rostro de la persona.

Este sistema tiene la característica de poder trabajar con rostros que, presenten diferentes perspectivas o, no sea posible localizar todos los elementos del mismo en la imagen como en una cara de perfil o viendo al piso.

Al finalizar el mejoramiento en cualquiera de las tres modalidades (súper resolución, reconstrucción del rostro y localización del rostro) se guarda la imagen resultante.

El sistema puede localizar el rostro ante la presencia de:

- Lentes de lectura (graduados) y/o protección de luz azul.
- Diversa Escala del rostro.
- Cambios en la iluminación (se requiere visibilidad del rostro en la imagen analizada).
- Cambios en el contraste (se requiere visibilidad del rostro en la imagen analizada).

6.4 Trabajos futuros

Como trabajo futuro, se podría reducir el tiempo de procesamiento del sistema; es decir, que funcione en un tiempo estipulado, así como se pueden abordar otro tipo de problemáticas como lo son escala, enfoque de la imagen y presencia de alguna oclusión en el rostro de la persona.

Referencias

Acharya, A., & Giri, A. V. (2020, March). Contrast improvement using local gamma correction. In *2020 6th international conference on advanced computing and communication systems (ICACCS)* (pp. 110-114). IEEE.

Agarwal, V. (2021, 15 diciembre). Face Detection Models: Which to Use and Why? - Towards Data Science. Medium. Recuperado 29 de septiembre de 2022, de https://towardsdatascience.com/face-detection-models-which-to-use-and-why-d263e82c302c

Alkanhal, L., Alotaibi, D., Albrahim, N., Alrayes, S., Alshemali, G., & Bchir, O. (2020). Super-resolution using deep learning to support person identification in surveillance video. *International Journal of Advanced Computer Ence and Applications*, *11*(7).

Alvarez-Ramos, V., Ponomaryov, V., & Sadovnychiy, S. (2018). Image super-resolution via wavelet feature extraction and sparse representation. *Radioengineering*, *27*(2), 603.

An, L., & Bhanu, B. (2012, September). Image super-resolution by extreme learning machine. In 2012 19th IEEE international conference on image processing (pp. 2209-2212). IEEE.

Arismendi Sánchez, J.A (2021). PCNN en la Mejora de Imágenes a Color [Tesis de maestría, CENIDET]. Repositorio Académico del Centro Nacional de Investigación y Desarrollo Tecnológico.

Baltieri, D., Vezzani, R., & Cucchiara, R. (2011, December). 3dpes: 3d people dataset for surveillance and forensics. In Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding (pp. 59-64).

Barrios, J. I. (2019, June 26). La matriz de confusión y sus métricas – Inteligencia Artificial –. https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/

Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K., & Grundmann, M. (2019). Blazeface: Sub-millisecond neural face detection on mobile gpus. arXiv preprint arXiv:1907.05047

Cárdenas, R. J., Beltrán, C. A., & Gutiérrez, J. C. (2019). Small face detection using deep learning on surveillance videos. *environment*, *2*(5), 14.

Cao, M., Liu, Z., Huang, X., & Shen, Z. (2021, March). Research for Face Image Super-Resolution Reconstruction Based on Wavelet Transform and SRGAN. In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)* (Vol. 5, pp. 448-451). IEEE.

Cejudo García, M.P (2020). Desarrollo de un FrameWork para la experimentación con algoritmos de Súper Resolución [Tesis de maestría, CENIDET]. Repositorio Académico del Centro Nacional de Investigación y Desarrollo Tecnológico.

Chen, R., Qu, Y., Zeng, K., Guo, J., Li, C., & Xie, Y. (2018). Persistent memory residual network for single image super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 809-816).

Chen, Y., Wang, J., Chen, X., Zhu, M., Yang, K., Wang, Z., & Xia, R. (2019). The algorithm research of single-image super-resolution using scale model and deformation features. *IEEE Access*, *7*, 58791-58801.

Chi, C., Zhang, S., Xing, J., Lei, Z., Li, S. Z., & Zou, X. (2019, July). Selective refinement network for high performance face detection. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 8231-8238).

D.J. Jobson, Z. Rahman, and G.A. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, IEEE Transactions on Image Processing, 6 (1997), pp. 965–976. http://dx.doi.org/10.1109/83.597272.

Dlib. (2022, January 24). dlib C++ Library. http://dlib.net/

Dong, C., Loy, C. C., & Tang, X. (2016, October). Accelerating the super-resolution convolutional neural network. In European conference on computer vision (pp. 391-407). Springer, Cham.

E. Land and J. McCann, Lightness and retinex theory, Journal of the Optical Society of America, 61 (1971), pp. 1–11. http://dx.doi.org/10.1364/JOSA.61.000001.

Face Detection - mediapipe. https://google.github.io/mediapipe/solutions/face_detection

Faces in Real-Life Images workshop at the European Conference on Computer Vision 2008, run by Erik Learned-Miller, Andras Ferencz, and Frederic Jurie.

Farooq, M., Dailey, M. N., Mahmood, A., Moonrinta, J., & Ekpanyapong, M. (2021). Human face super-resolution on poor quality surveillance video footage. *Neural Computing and Applications*, *33*(20), 13505-13523.

González, R. C., & Woods, R. E. (1996). *Tratamiento digital de imágenes* (Vol. 3). Addison-Wesley New York.

Han, X., Ji, Z., & Wang, W. (2020, December). Low Resolution Facial Manipulation Detection. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)* (pp. 431-434). IEEE.

Higaki, T., Nakamura, Y., Tatsugami, F., Nakaura, T., & Awai, K. (2019). Improvement of image quality at CT and MRI using deep learning. *Japanese journal of radiology*, *37*(1), 73-80.

Honda, T., Hamamoto, T., & Sugimura, D. (2018, October). Low-light color image super-resolution using RGB/NIR sensor. In 2018 25th IEEE International Conference on Image Processing (ICIP) (pp. 56-60). IEEE.

K. B. Singh, T. V. Mahendra, R. S. Kurmvanshi and C. V. Rama Rao, "Image enhancement with the application of local and global enhancement methods for dark images," *2017 International Conference on Innovations in Electronics, Signal Processing and Communication (IESC)*, 2017, pp. 199-202, doi: 10.1109/IESPC.2017.8071892.

Kumar, A., Kaur, A., & Kumar, M. (2019). Face detection techniques: a review. Artificial Intelligence Review, 52(2), 927-948.

Lai, W. S., Huang, J. B., Ahuja, N., & Yang, M. H. (2018). Fast and accurate image super-resolution with deep laplacian pyramid networks. IEEE transactions on pattern analysis and machine intelligence, 41(11), 2599-2613.

Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144).

Marciniak, T., Chmielewska, A., Weychan, R., Parzych, M., & Dabrowski, A. (2015). Influence of low resolution of images on reliability of face detection and recognition. Multimedia Tools and Applications, 74(12), 4329-4349.

Menaka, K., & Yogameena, B. (2021, June). Face Detection in Blurred Surveillance Videos for Crime Investigation. In *Journal of Physics: Conference Series* (Vol. 1917, No. 1, p. 012024). IOP Publishing.

Moorthy, A. K., & Bovik, A. C. (2010). A two-step framework for constructing blind image quality indices. IEEE Signal processing letters, 17(5), 513-516.

OpenCV. (2023). OpenCV. Obtenido de OpenCV: https://opencv.org/opencv-face-recognition/

Ranjan, R., Bansal, A., Zheng, J., Xu, H., Gleason, J., Lu, B., ... & Chellappa, R. (2019). A fast and accurate system for face detection, identification, and verification. IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(2), 82-96.

Rivera Itúrburu, I. G., & Zambrano Guaranda, D. F. (2022). Implementación de reconocimiento facial y visión artificial en robot nao con Python y Opencv (Bachelor's thesis).

Shah, A. J., & Gupta, S. B. (2012, December). Image super resolution-a survey. In 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking (pp. 1-6). IEEE.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1874-1883).

Shipman, J. W. (2013). Tkinter 8.5 reference: a GUI for Python. Computer, 1–118. tcc-doc@nmt.edu

Sousa, A. L., Villar, S. A., Korneta, W., Acosta, G., & Rozenfeld, A. (2016, June). Resonancia estocástica para el mejoramiento del contraste y calidad en imágenes acústicas de sonar de barrido lateral. In *2016 IEEE Biennial Congress of Argentina (ARGENCON)* (pp. 1-6). IEEE.

Sun, X., Wu, P., & Hoi, S. C. (2018). Face detection using deep learning: An improved faster RCNN approach. Neurocomputing, 299, 42-50.

Ochoa Domínguez, H. (2020). Estudio comparativo de algoritmos de súper resolución de una sola imagen basados en aprendizaje profundo. *Instituto de Ingeniería y Tecnología*.

V. Voronin, S. Tokareva, E. Semenishchev and S. Agaian, "Thermal Image Enhancement Algorithm Using Local and Global Logarithmic Transform Histogram Matching with Spatial Equalization," *2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, 2018, pp. 5-8, doi: 10.1109/SSIAI.2018.8470344.

Valderrama Cardenas, W (2019). Reconocimiento automático del rostro para verificación de identidad para evaluación en línea [Tesis de maestría, CENIDET]. Repositorio Académico del Centro Nacional de Investigación y Desarrollo Tecnológico.

Wang, X., Li, Y., Zhang, H., & Shan, Y. (2021). Towards real-world blind face restoration with generative facial prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9168-9178)

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

Wu, L., Zhang, X., Chen, H., Wang, D., & Deng, J. (2021). VP-NIQE: An opinion-unaware visual perception natural image quality evaluator. Neurocomputing, 463, 17-28.

Xu-hui, C. H. E. N., Haq, E. U., & Chengyu, Z. H. O. U. (2019, June). Face Detection and Extraction from Low Resolution Surveillance Video Using Motion Segmentation. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)* (pp. 547-552). IEEE Computer Society.

Y. Ueda and N. Suetake, "Hue-Preserving Color Image Enhancement on a Vector Space of Convex Combination Coefficients," *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 939-943, doi: 10.1109/ICIP.2019.8803035.

Y. Wong, S. Chen, S. Mau, C. Sanderson, B.C. Lovell Patch-based Probabilistic Image Quality Assessment for Face Selection and Improved Video-based Face Recognition IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops, pages 81-88. IEEE, June 2011.

Yoo, J. C., & Han, T. H. (2009). Fast normalized cross-correlation. Circuits, systems and signal processing, 28(6), 819-843.

Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., & Zhang, L. (2016). Image super-resolution: The techniques, applications, and future. Signal Processing, 128, 389-408.

Yu, X., Fernando, B., Ghanem, B., Porikli, F., & Hartley, R. (2018). Face super-resolution guided by facial component heatmaps. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 217-233).

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE signal processing letters, 23(10), 1499-1503.

Zheng, J., Ramirez, G. A., & Fuentes, O. (2010, June). Face detection in low-resolution color images. In International Conference Image Analysis and Recognition (pp. 454-463). Springer, Berlin, Heidelberg.

Zhou, R., & Susstrunk, S. (2019). Kernel modeling super-resolution on real low-resolution images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2433-2443).

Zhou, Y., Liu, D., & Huang, T. (2018, May). Survey of face detection on low-quality images. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018) (pp. 769-773). IEEE.

Anexos

Anexo A

Actividades académicas adicionales

Se presentó (de manera presencial) el artículo “Comparativa de Algoritmos de Súper resolución en Imágenes de Videovigilancia” en el marco de la 9ª jornada de ciencia y tecnología aplicada, que se celebró del 16 al 18 de noviembre de 2022 en el TecNM/CENIDET, ver Figura A.1.


EDUCACIÓN

EL TECNOLÓGICO NACIONAL DE MÉXICO
A TRAVÉS DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO

OTORGA EL PRESENTE

RECONOCIMIENTO

A

REBECCA YUZIEL ZUMAYA LANUZA
CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO

POR LA PRESENTACIÓN DEL ARTICULO:
COMPARATIVA DE ALGORITMOS DE SÚPER RESOLUCIÓN EN IMÁGENES DE VIDEOVIGILANCIA
EN EL MARCO DE LA 9ª JORNADA DE CIENCIA Y TECNOLOGÍA APLICADA, CELEBRADA
DEL 16 AL 18 DE NOVIEMBRE DE 2022, EN EL TECNM/CENIDET

CUERNAVACA, MORELOS, 16-18 DE NOVIEMBRE DE 2022

DRA. YESICA IMELDA SAAVEDRA BENÍTEZ
DIRECTORA DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO

Sello Digital:

cenidet
JORNADA DE CIENCIA Y TECNOLOGÍA APLICADA

Figura A.0.1 Reconocimiento 9ª jornada de ciencia y tecnología aplicada

Se presentó (de manera presencial) el poster “Sistema para el Mejoramiento de Imágenes para Sistemas de Videovigilancia de Baja Calidad” en el marco de la 9ª jornada de ciencia y tecnología aplicada, que se celebró del 16 al 18 de noviembre de 2022 en el TecNM/CENIDET, ver Figura A.2.

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

EL TECNOLÓGICO NACIONAL DE MÉXICO
A TRAVÉS DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO

OTORGA EL PRESENTE

RECONOCIMIENTO

A

REBECCA YUZIEL ZUMAYA LANUZA

CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO

POR LA PRESENTACIÓN DEL ARTICULO:
SISTEMA PARA EL MEJORAMIENTO DE IMÁGENES PARA SISTEMAS DE VIDEOVIGILANCIA
DE BAJA CALIDAD
EN EL MARCO DE LA 9ª JORNADA DE CIENCIA Y TECNOLOGÍA APLICADA, CELEBRADA
DEL 16 AL 18 DE NOVIEMBRE DE 2022, EN EL TECNMCENIDET

CUERNAVACA, MORELOS, 16-18 DE NOVIEMBRE DE 2022

DRA. YESICA IMELDA SAAVEDRA BENÍTEZ
DIRECTORA DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO

Sello Digital:

cenidet
Centro Nacional de Investigación y Desarrollo Tecnológico

JORNADA DE CIENCIA Y TECNOLOGÍA APLICADA

Figura A.0.2 Reconocimiento 9ª jornada de ciencia y tecnología aplicada

Se presentó la conferencia (de manera virtual) “Sistema para el Mejoramiento de Imágenes para Sistemas de Videovigilancia de Baja Calidad” en el marco del simposio internacional de ingeniería es sistemas computacionales – sistemas transversales del Instituto Tecnológico de Cuautla que se celebró el día 23 de marzo del 2023, ver Figura A.3.

EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

EL TECNOLÓGICO NACIONAL DE MÉXICO
A TRAVÉS DEL INSTITUTO TECNOLÓGICO DE CUAUTLA

OTORGA EL PRESENTE

RECONOCIMIENTO

A

REBECCA YUZIEL ZUMAYA LANUZA

POR SU VALIOSA PARTICIPACIÓN COMO PONENTE EN LA CONFERENCIA:

“PROYECTOS DE INVESTIGACIÓN DE LA LÍNEA DE INTELIGENCIA ARTIFICIAL”.

IMPARTIDA EL DÍA 29 DE MARZO DE 2023, EN EL MARCO DEL SIMPOSIO INTERNACIONAL DE INGENIERÍA EN SISTEMAS COMPUTACIONALES – SISTEMAS TRANSVERSALES.

H. H. Cuautla, Morelos a 29 de marzo de 2023

PORFIRIO ROBERTO NÁJERA MEDINA
DIRECTOR DEL INSTITUTO TECNOLÓGICO DE CUAUTLA

Figura A.0.3 Reconocimiento del Instituto Tecnológico de Cuautla

Presentación del poster “Sistema para el Mejoramiento de Imágenes para Sistemas de Videovigilancia de Baja Calidad” en la escuela de inteligencia computacional y robótica 2022, realizada en la Universidad Tecnológica Emiliano Zapata (UTEZ), los días 16 y 20 de agosto, ver Figura A.4.

Figura A.0.4 Reconocimiento escuela de inteligencia computacional y robótica 2022

Anexo B

A continuación, se muestran los comandos en el símbolo del sistema en Windows para la instalación de las librerías necesarias para el funcionamiento del sistema si no se tiene el ejecutable.

Si utiliza la versión 2.7 de Python además de la versión 3.9 sustituya el comando pip por pip3 como se muestra a continuación:

opencv-contrib-python 4.5.5.64

```
pip install opencv-contrib-python==4.5.5.64
pip3 install opencv-contrib-python==4.5.5.64
```

Verificar la versión de python

```
python –version
```

Mediapipe 0.8.9.1

```
pip install mediapipe==0.8.9.1
pip3 install mediapipe==0.8.9.1
```

Basicsr 1.3.4.0

```
pip install basicsr>=1.3.4.0
pip3 install basicsr>=1.3.4.0
```

facexlib 0.2.3

```
pip install facexlib>=0.2.3
pip3 install facexlib>=0.2.3
```

lmdb

```
pip install lmdb
pip3 install lmdb
```

numpy

```
pip install numpy<1.21
pip3 install numpy<1.21
```

pyyaml

```
pip install pyyaml
pip3 install pyyaml
```

scipy

```
pip install scipy
pip3 install scipy
```

tb-nightly

pip install tb- nightly

pip3 install tb-nightly

torch>=1.7

pip install torch>=1.7

pip3 install torch>=1.7

torchvision

pip install torchvision

pip3 install torchvision

tqdm

pip install tqdm

pip3 install tqdm

yapf

pip install vapf

pip3 install vapf