

VISIÓN POR ORDENADOR

CURSO 2024-2025

TRABAJO FINAL



Autores:

María Carreño Nin de Cardona

Raquel Fernández Esquinas

Madrid, 10 de Enero de 2025

Índice de Contenidos

1. INTRODUCCIÓN.....	3
2. METODOLOGÍA	4
2.1. CALIBRACIÓN DE LA CÁMARA.....	4
2.2. SECUENCIA DE TRANSFORMACIÓN DE LA IMAGEN.....	4
2.3. SISTEMA DE SEGURIDAD: DETECCIÓN DE PATRONES Y EXTRACCIÓN DE INFORMACIÓN	5
2.4. SISTEMA PROPUESTO: TRACKER, AMPLIACIONES Y SALIDA DE VÍDEO.....	5
2.5. DIAGRAMA DE BLOQUES DEL SISTEMA.....	6
3. RESULTADOS Y FUTUROS DESARROLLOS	7
3.1. RESULTADOS.....	7
3.2. FUTUROS DESARROLLOS.....	7

1. INTRODUCCIÓN

Este documento presenta el trabajo realizado por María Carreño y Raquel Fernández, alumnas de la Escuela Técnica Superior de Ingeniería (ICAI) de la Universidad Pontificia Comillas. El trabajo lleva por título “Trabajo Final” y ha sido desarrollado en la asignatura “Visión por Ordenador” del tercer curso del grado iMAT.

Este proyecto, elaborado con gran entusiasmo, tiene como objetivo el desarrollo de un sistema de visión por ordenador que incluya tecnologías avanzadas para la detección, extracción y seguimiento de patrones en tiempo real. El sistema está orientado a su aplicación en entornos de seguridad, donde su función principal es identificar y rastrear objetos o personas de interés mediante imágenes capturadas por una cámara de video. La clave del proyecto se basa en la capacidad para detectar patrones complejos, extraer características relevantes y realizar un seguimiento preciso de los mismos. En este caso, los patrones a detectar son los movimientos de las manos en el famoso juego “piedra, papel o tijera” y la identificación de rostros, que luego se utilizarán para aplicar filtros. A medida que el sistema va identificando estos patrones y objetos, genera salidas visuales en tiempo real, enriquecidas con información procesada, lo que facilita su interpretación y análisis.

El proyecto se apoya en algoritmos especializados en visión por ordenador para la detección y seguimiento de objetos. Se emplean técnicas avanzadas como el filtrado de imágenes, la segmentación, la detección de bordes y la identificación de rostros, entre otras, para garantizar un rendimiento óptimo en las tareas de procesamiento visual.

La estructura del trabajo se organiza en varias etapas. En la primera parte, se desarrolla un sistema de seguridad basado en la detección de patrones, en el cual la persona frente a la cámara debe realizar una secuencia con las manos que represente “piedra”, “tijera” y “papel” en ese orden, para poder continuar. En la versión extendida de este sistema, el desafío aumenta, ya que, en lugar de solo realizar la secuencia, la persona debe ganar a la máquina en el juego de “piedra, papel o tijera” para avanzar en el proceso.

Una vez completada la implementación del sistema de seguridad, se introduce el sistema propuesto: una cámara de filtros interactiva. En esta parte del proyecto, el usuario puede elegir entre diferentes filtros visuales. En la versión simple, se incluyen filtros como blanco y negro, normal y beige. La versión extendida ofrece una mayor variedad, con filtros adicionales en colores como rosa, rojo, verde y azul, permitiendo así una experiencia visual más dinámica y personalizada.

Este trabajo no solo permite explorar diferentes aspectos de la visión por ordenador, sino también implementar soluciones prácticas e innovadoras para la interacción entre humanos y sistemas automatizados en un contexto de seguridad.

2. METODOLOGÍA

2.1. Calibración de la Cámara

La calibración de la cámara es el paso inicial y crucial para asegurar que el sistema funcione con la máxima precisión. Para este proceso comenzamos con la captura de varias imágenes de un patrón conocido, el tablero de ajedrez, utilizando la cámara que íbamos a usar posteriormente. Una vez obtenidas estas imágenes, utilizamos algoritmos avanzados de la librería OpenCV para calcular los parámetros intrínsecos y extrínsecos de la cámara, que incluyen la focal, el centro de la imagen y las distorsiones provocadas por la lente. Posteriormente, realizamos una corrección de las distorsiones radial y tangencial, causadas por la curvatura de la lente y la posición del sensor, respectivamente. Este paso es esencial para eliminar los errores ópticos y garantizar la obtención de imágenes rectificadas, con una geometría más precisa y fiel a la realidad.

2.2. Secuencia de Transformación de la Imagen

En el proyecto hemos realizado una secuencia compleja de transformación de la imagen, la cual comienza con la captura de video en tiempo real, utilizando la librería Picamera2. El primer paso en el proceso de transformación es la detección de gestos de la mano y las características faciales a través de la integración de la librería Mediapipe. Mediapipe procesa la imagen capturada, detectando la posición de la mano o la cara del usuario y evaluando sus movimientos o gestos, como los de la mano para determinar si se muestra piedra, papel o tijeras. Este es un proceso clave que convierte las imágenes crudas de la cámara en datos que se pueden utilizar para interactuar con el usuario en un contexto de juego, y es el primer punto de transformación de la imagen.

Posteriormente, las imágenes capturadas por la cámara se someten a más transformaciones, dependiendo de la elección del usuario. Si se ha identificado un gesto o se ha seleccionado un filtro, la imagen experimenta un segundo proceso de transformación. Los filtros faciales (como gafas, bigote o sombrero) se aplican mediante el uso de imágenes de superposición con un canal alfa, lo que permite colocar los elementos gráficos en posiciones específicas de la cara del usuario, basándose en puntos de referencia faciales detectados previamente. Además, el programa ofrece la opción de cambiar el estilo visual de la imagen mediante filtros de color (blanco y negro, beige, verde, rosa, etc.), alterando la tonalidad general de la imagen con efectos de mezcla de imágenes. La imagen se convierte a escala de grises para el efecto blanco y negro (`cv2.cvtColor`), o se combina con un filtro personalizado para aplicar tonos cálidos (beige) utilizando `cv2.addWeighted`.

Por último, la imagen puede ser capturada y guardada en el dispositivo con un filtro aplicado, o bien ser visualizada en tiempo real, ajustándose continuamente a las interacciones del usuario. Este flujo de transformación, desde la captura de video hasta la aplicación de filtros y efectos visuales, permite una experiencia interactiva dinámica, ofreciendo un alto grado de personalización de la imagen en función de las decisiones del usuario.

2.3. Sistema de Seguridad: Detección de patrones y Extracción de Información

Al iniciar el programa, se debe completar un sistema de seguridad para poder acceder al sistema propuesto. En la versión simple, este sistema de seguridad se basa en la realización de un patrón específico por parte del usuario frente a la cámara. El usuario debe realizar, en orden secuencial, los gestos correspondientes a "piedra", "tijera" y "papel" con su mano para poder avanzar en el proceso. Para lograr esto, hemos implementado un sistema de detección de patrones utilizando algoritmos avanzados de visión por ordenador. El sistema analiza los puntos característicos de la mano, como las articulaciones de los dedos y la posición relativa entre ellos, utilizando técnicas de procesamiento de imágenes y el modelo de detección de manos de la librería MediaPipe. Esto permite al algoritmo identificar de manera precisa la configuración de los dedos, interpretando correctamente el gesto realizado. Una vez que el patrón se completa correctamente en la secuencia indicada, el sistema permite continuar con el siguiente paso del proceso.

En la versión extendida del sistema, el desafío se incrementa al introducir una variante del juego clásico de "piedra, papel o tijera", en el que el usuario debe competir contra la máquina. Para que el sistema reconozca y valore correctamente los gestos, hemos diseñado un algoritmo que, además de la detección de patrones, se encarga de la extracción y análisis de características más detalladas de la mano del usuario. El sistema evalúa la forma y disposición de los dedos en tiempo real, utilizando un conjunto de puntos de referencia específicos proporcionados por el modelo de MediaPipe, para poder determinar qué gesto está realizando el usuario. Posteriormente, el algoritmo compara el gesto del usuario con el de la máquina, basándose en las reglas estándar del juego. De esta manera, el sistema asegura que el usuario deba ganar a la máquina para poder continuar al siguiente paso.

2.4. Sistema Propuesto: Tracker, Ampliaciones y Salida de Vídeo

El sistema propuesto en este trabajo consiste en recrear la experiencia interactiva de las aplicaciones de redes sociales como Instagram y Snapchat, permitiendo aplicar filtros visuales y efectos en tiempo real. Este sistema tiene dos versiones principales: una simple y una extendida. En la versión simple, tras completar un sistema de seguridad basado en patrones, el usuario puede acceder a una interfaz de filtros donde es posible seleccionar entre una serie de objetos superpuestos en la imagen en vivo, como un bigote, gafas o sombrero, que se ajustan dinámicamente al rostro detectado. Además, se pueden aplicar filtros de color que alteran la visualización de la imagen, como modos en blanco y negro, beige o normal.

En la versión extendida, se añade el clásico juego de "Piedra, Papel o Tijera" como parte del sistema de seguridad, en el que, para acceder a los filtros, el usuario debe ganar al sistema. Una vez superado este reto, se amplían las opciones de filtros disponibles. Además de los objetos superpuestos y los filtros de color de la versión simple, se introducen nuevos colores, como rosa, verde, rojo y azul.

En cuanto al seguimiento facial, el sistema implementa un "tracker" que detecta y sigue continuamente el rostro del usuario en tiempo real, representado visualmente por un cuadro que delimita el área detectada. Este componente es esencial para garantizar que los filtros y objetos se posicionen correctamente y se ajusten dinámicamente a los

movimientos del usuario. El uso del tracker mejora significativamente la experiencia interactiva, proporcionando un sistema más preciso y fluido que se adapta al entorno en tiempo real.

2.5. Diagrama de Bloques del Sistema

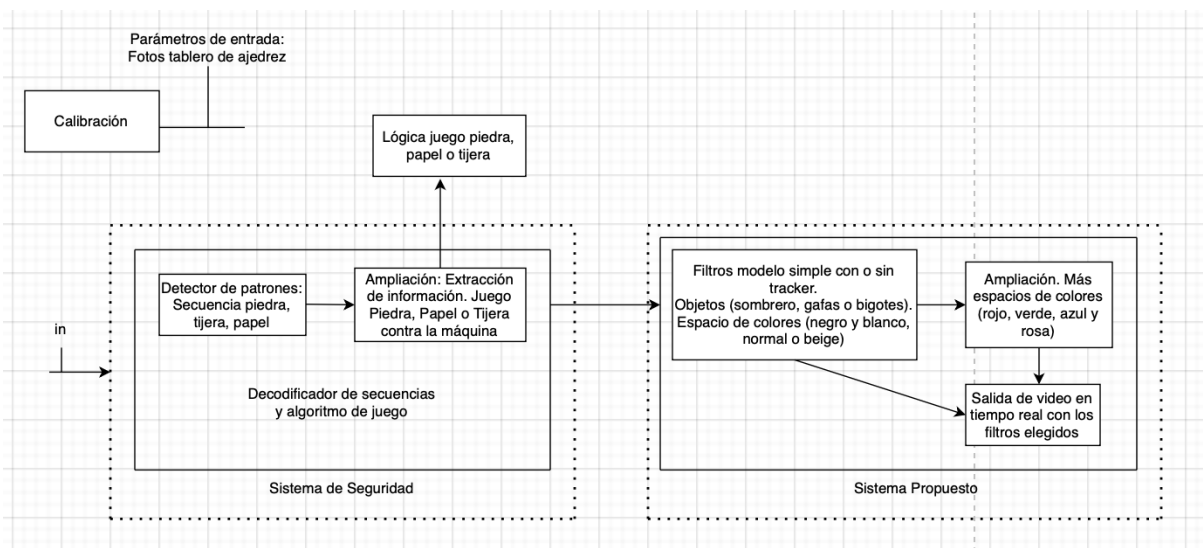
Teniendo en cuenta toda la lógica de nuestro proyecto, lo podemos resumir de forma más esquemática de la siguiente manera, donde se representan los principales bloques que forman nuestro programa.

Los parámetros de entrada para la calibración son un conjunto de fotos que hicimos con la cámara a un tablero de ajedrez, en el que es fácil extraer los pintos de interés y llevar a cabo dicha calibración; tras la cual obtuvimos los parámetros intrínsecos, extrínsecos y distorsiones radiales y tangenciales de la cámara.

La entrada del programa (IN en el diagrama), corresponde con la captura de imagen/video que capta la cámara para dar lugar al sistema de seguridad. El cual consiste en detección de patrones y extracción de información.

Siendo la misma entrada para el sistema propuesto, tras recibir la entrada se lleva a cabo la aplicación de filtros y transformaciones de la imagen, así como la detección facial.

Por último, la salida final del programa es un video a tiempo real donde se pueden ver los filtros elegidos por el usuario.



3. RESULTADOS Y FUTUROS DESARROLLOS

3.1. Resultados

Tras finalizar el desarrollo del proyecto y realizar diversas pruebas, podemos concluir con satisfacción que hemos logrado nuestra idea inicial, lo que nos llena de orgullo.

La calibración de la cámara, realizada en clase, nos permitió aplicar los conocimientos adquiridos en una práctica previa. Este paso fue fundamental para obtener sus parámetros, obteniendo una comprensión más precisa de su funcionamiento.

En cuanto al sistema de seguridad, nuestra idea original era implementar el juego de "piedra, papel o tijera". Sin embargo, siguiendo el consejo de nuestro profesor, decidimos comenzar con un diseño más sencillo para luego incrementar la complejidad. Así, desarrollamos inicialmente la detección de patrones con la secuencia explicada anteriormente. Una vez conseguido, nos animamos a implementar el juego completo. Conseguimos que el jugador, frente a la cámara, tuviera que ganar a la máquina para avanzar, consiguiendo así nuestro objetivo inicial.

En el sistema propuesto, queríamos incorporar una funcionalidad presente en nuestra vida diaria: filtros. Comenzamos implementando algunos filtros, y posteriormente ampliamos, añadiendo más opciones.

Además, implementamos el tracker facial, que nos permitió realizar un seguimiento preciso del rostro del usuario después de superar el sistema de seguridad. Por último, logramos que la salida del sistema ofreciera un video en tiempo real, mostrando los filtros seleccionados por el usuario

3.2. Futuros Desarrollos

En cuanto a los futuros desarrollos de este proyecto, una posible mejora sería integrar una mayor variedad de gestos y movimientos de la mano para ofrecer una interacción más rica y precisa. Actualmente, el sistema solo detecta los gestos básicos de piedra, papel y tijeras, pero se podrían incluir otros gestos que permitirían al usuario realizar acciones adicionales o incluso manipular los filtros y efectos en tiempo real. Por ejemplo, se podría implementar la detección de gestos de deslizamiento o giro de la mano para activar diferentes modos o cambiar los filtros de manera más intuitiva. Además, se podría explorar el uso de más cámaras o sensores para mejorar la precisión de la detección de gestos, especialmente en situaciones donde la cámara principal no capte bien los movimientos debido a la iluminación o el ángulo.

Otro desarrollo interesante sería la implementación de un sistema de inteligencia artificial que reconozca características faciales de forma más detallada, permitiendo personalizar los filtros basados no solo en la ubicación de los ojos y la nariz, sino también adaptándolos al tipo de rostro del usuario. Por ejemplo, un sistema que pueda detectar la forma del rostro y aplicar filtros que se ajusten mejor a la morfología facial del usuario, haciendo la experiencia aún más personalizada.