UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE CUENCA

CARRERA DE COMPUTACIÓN

DISEÑO Y DESARROLLO DE UN SISTEMA PROTOTIPO DE VIDEOVIGILANCIA EMPLEANDO VISIÓN POR COMPUTADOR Y APRENDIZAJE AUTOMÁTICO

Trabajo de titulación previo a la obtención del título de Ingeniero en Ciencias de la Computación

AUTOR: PEDRO JOSÉ ORTIZ SOLIS

TUTOR: ING. VLADIMIR ESPARTACO ROBLES BYKBAEV, Ph.D.

Cuenca - Ecuador

2022

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Pedro Jose Ortiz Solis con documento de identificación N° 0105223986 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 8 de marzo del 2022.

Atentamente,

Pedro José Ortiz Solis
0105223986

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Yo, Pedro José Ortiz Solis con documento de identificación Nº 0105223986, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Proyecto técnico: "Diseño y desarrollo de un sistema prototipo de videovigilancia empleando visión por computador y aprendizaje automático", el cual ha sido desarrollado para optar por el título de: Ingeniero en Ciencias de la Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 8 de marzo del 2022.

Atentamente,

_____

Pedro José Ortiz Solis

0105223986

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Vladimir Espartaco Robles Bykbaev con documento de identificación N° 0300991817 docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: DISEÑO Y DESARROLLO DE UN SISTEMA PROTOTIPO DE VIDEOVIGILANCIA EMPLEANDO VISIÓN POR COMPUTADOR Y APRENDIZAJE AUTOMÁTICO, realizado por Pedro José Ortiz Solis con documento de identificación N° 0105223986, obteniendo como resultado final el trabajo de titulación bajo la opción Proyecto técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 8 de marzo del 2022.

Atentamente,

Vladimir Espartaco Robles Bykbaev, Ph.D
0300991817

DEDICATORIA Y AGRADECIMIENTO

Este proyecto está principalmente dedicado a:

A mis padres quienes con tanto esfuerzo y dedicación me han dado una gran oportunidad de alcanzar una meta más de muchas que aún quedan por delante y por siempre inculcarme la fortaleza para siempre seguir adelante y nunca rendirme.

A mi hermano quien siempre me ha demostrado mucha confianza, apoyo y un gran ejemplo en muchos aspectos, dejándome valiosas enseñanzas en toda esta trayectoria.

A mis amigos y compañeros quienes me han acompañado por toda o gran parte de la carrera y han evidenciado ser muy grandes personas y estoy seguro que también serán muy buenos y honoríficos profesionales.

Finalmente y aunque sean realmente muy pocos, a mis docentes quienes a lo largo de la carrera me han ilustrando con sus conocimientos de la mejor manera, mostrándose a si mismos como profesionales extraordinarios y preocupados por el desempeño en su labor, especialmente al docente tutor del presente proyecto, a quien considero un gran ejemplo a seguir tanto en un ámbito profesional como personal.

Mi profundo agradecimiento a todas las personas previamente mencionadas, además a todo el personal docente y administrativo de la Universidad Politécnica Salesiana que hizo posible todo este proceso.

RESUMEN

El propósito del presente proyecto es analizar la factibilidad de implementar técnicas actuales de visión por computador y herramientas de aprendizaje automático a un sistema prototipo de videovigilancia, considerando los costos de monitoreo y mantenimiento de los sistemas de seguridad modernos, además, estos no podrían mantenerse supervisados todo el tiempo, demostrando así ciertas vulnerabilidades a la seguridad de un establecimiento; por lo que sería de gran utilidad aprovechar la visión e inteligencia artificial en un sistema que se mantenga operativo y alerta todo el tiempo.

La investigación que se ha llevado a cabo para el proyecto consta de una clasificación en base a precisión de tres diferentes métodos de detección de personas o peatones, lo cual sería el principal objetivo de un proceso de videovigilancia. Para medir la evaluación de cada propuesta se ha comparado la precisión de detección de cada uno.

Con un método ya establecido de detección de personas, se procedió a recolectar datos para un dataset binario, unos capturados por el mismo sistema y otros extraídos de fuentes externas, los cuales fueron cuidadosamente analizados, pre-procesados y usados para entrenar una red neuronal convolucional con el objetivo de conseguir que el computador pueda distinguir, en base a características, una clase de la otra.

Para el entrenamiento se utilizó la técnica *k-fold* la cual divide de forma dinámica y aleatoria el dataset en partes para entrenamiento, prueba y validación respectivamente, repitiendo 5 veces el procedimiento, obteniendo una exactitud por encima del 70%.

Palabras clave: visión por computador, redes neuronales, aprendizaje automático, videovigilancia, inteligencia artificial

ABSTRACT

The main purpose of this project is to analyze the feasibility of using current computer vision techniques and machine learning tools within a video surveillance system prototype, being aware of the monitoring and maintenance costs of modern security systems, besides they couldn't be kept supervised all time, exposing with this some vulnerabilities to the secure level of a specific place; for this reason, it will be very useful to take advantage of artificial intelligence and computer vision in a system that can keep operative and stay alert all time.

The research that has been carried out for this project consists of a classification between three different methods for pedestrian or people detection, which would be the main objective in a video surveillance process. To measure the accuracy of each method, a comparison of detection accuracy of each method has been performed.

With a set detection method, a binary class dataset has been collected, images captured by the same detection method and others from external sources have been carefully analyzed, pre-processed, and used to train a convolutional neural network with the purpose of allowing the computer to tell, based on characteristics, which class belongs to each detection performed.

For the training process, a *k-fold* technique has been used, which consists in a dynamic and random division of the dataset in training, test and validation respectively, repeating 5 times this process and obtaining an accuracy above 70%.

Key words: computer vision, neural networks, machine learning, video surveillance, artificial intelligence

ÍNDICE DE CONTENIDO

INTRODUCCIÓN

Con el rápido avance de la tecnología, la videovigilancia se ha vuelto cada vez más accesible en términos económicos, paralelamente, con las mejoras en el rendimiento de las computadoras ordinarias, se podría mejorar considerablemente la seguridad, fiabilidad y accesibilidad económica, permitiendo automatizar la detección temprana de eventos sospechosos y registro en video de los mismos.

La disponibilidad de herramientas de desarrollo de software “open source” o de código abierto ha facilitado los avances en varios campos de la tecnología, teniendo la visión artificial como uno de ellos, además, estas herramientas nos ofrecen una gran cantidad de librerías, documentación y técnicas de desarrollo para conseguir o acercarnos a nuestros objetivos.

Con la existencia de un sistema de videovigilancia, de un costo relativamente bajo y con una eficiencia muy alta, se podría incrementar en gran medida la seguridad tanto en lugares públicos como privados, y al mismo tiempo, disminuir la taza de robos, asaltos, incidentes adversos, etc. Por otro lado, el desarrollo de este tipo de soluciones abarca una serie de alternativas que pueden usarse dependiendo de que es lo que se desea obtener, existen algoritmos de detección pre-entrenados y listos para implementarse y usarse a conveniencia del proyecto en desarrollo, además ciertos algoritmos pueden ser mejorados para un uso en específico.

PROBLEMA

Año tras año el índice de delincuencia en nuestro país ha ido en aumento, según la fuente periodística GK en la publicación (Rofifah, 2021), en Ecuador, las denuncias de robos en el primer semestre de 2021 aumentaron a tal punto de compararse con todas las presentadas en 2020. Las denuncias que más van en aumento se tratan de robos de automóviles o partes de estos, domicilios, motocicletas y locales comerciales.

Se concluye que gran parte de los robos ocurren especialmente cuando cierto objeto, domicilio o local comercial carece de vigilancia o se encuentra al descuido. Se ha evidenciado la falta de seguridad y la poca importancia que se le da al tema por parte de las autoridades, además no todos tenemos la posibilidad de optar por servicios de guardias de seguridad u otras implementaciones que ofrece el sector privado.

La videovigilancia ha sido de gran utilidad para brindar una mayor seguridad en diferentes sectores sectores públicos y privados, pero no se trata solamente de tener una o varias cámaras de seguridad instaladas en los distintos sitios de interés, también es necesario que se encuentren operando y registrando video las 24 horas del día, es decir, necesitan de dispositivos con gran capacidad de almacenamiento, teniendo como limitación la calidad del video, disponibilidad de registros con cierta antigüedad, fallas en las unidades de almacenamiento o en los dispositivos que las contienen, además, el proceso de identificación de eventos de interés captados por las cámaras se vuelve costoso, demorado e impreciso.

El artículo (Babiker et al., 2018) señala que, en los sistemas de videovigilancia, existen 3 tipos principales: manual, semi-autónomo y completamente autónomo. En los sistemas de videovigilancia manuales, el manejo y análisis es realizado por un ser humano. En los sistemas semi-autónomos, la intervención del ser humano en el análisis y la toma de decisión es parcial mientras que, en los sistemas completamente autónomos, el video de entrada, análisis, procesamiento y detección de eventos de interés son completamente independientes de cualquier intervención humana. Según los autores, el proceso básico

de los sistemas de videovigilancia consiste en la extracción del fondo, lo cual se conoce también por su aplicación en la detección de movimiento. Además ellos proponen un preprocesamiento de imágenes que consisten en una serie de operaciones y técnicas de modo que sean entendibles para el computador, finalmente, con los datos preprocesados y etiquetados, entrenan una red neuronal con el objetivo de detectar 5 actividades, obteniendo resultados muy precisos. Algo similar ocurre en la propuesta de (Kakadiya et al., 2019), quienes proponen un sistema de televisión de circuito cerrado, en el cual usan la substracción del fondo, preprocesamiento de imagen y el posterior entrenamiento de una red neuronal convolucional (CNN) para la detección de armas de fuego y/o corto punzantes en lugares cerrados.

Muchas soluciones existentes utilizan la substracción del fondo, no obstante, (Wang et al., 2014) afirma que esta técnica es muy sensible a los cambios constantes de luz y tiene ciertas dificultades en el manejo de problemas en cuanto al agrupamiento y fragmentación, la mejor solución, según los autores, son los detectores basados en grandes conjuntos de datos entrenados con todo tipo de variaciones que se pueden dar.

En la propuesta de (Feng et al., 2021), la visión por computadora involucra muchas disciplinas, que incluyen no solo ciencias de la computación, neurología, óptica y matemáticas, sino también disciplinas sociales como la sociología, la filosofía y el comportamiento humano. Además insisten en la importancia de establecer y actualizar la imagen de fondo cada cierto tiempo o evento para una correcta substracción y obtención de la región de interés con mayor precisión en el seguimiento de figuras humanas.

El trabajo realizado en el artículo (Elsayed et al., 2019) muestra ser muy prometedor en cuanto a la detección de humanos y las actividades que hacen, los autores utilizan un detector pre-entrenado de personas y posteriormente utilizan múltiples máquinas de soporte vectorial (SVM) para el aprendizaje de actividades “normales” como caminar, correr, saludar, saltar, etc. Cuando el algoritmo no reconoce una actividad previamente entrenada, o la puntuación de reconocimiento es muy baja, la etiqueta como “anormal”, esto con el fin de detectar robos, destrucción de propiedad pública/privada o incluso peleas.

OBJETIVOS GENERALES Y ESPECÍFICOS

General:

Diseñar y desarrollar un sistema prototipo de videovigilancia empleando visión por computador y aprendizaje automático.

Específicos:

• **OE1.** Estudiar y conocer las principales técnicas de visión por computador y tipos de aprendizaje automático aplicados en la videovigilancia.

• **OE2.** Estudiar, obtener y preparar datos reales acerca del comportamiento humano de interés en varios entornos bajo diferentes condiciones físicas y sociales.

• **OE3.** Diseñar y entrenar un modelo de aprendizaje automático para brindar soporte en la detección de eventos de videovigilancia.

• **OE4.** Diseñar y ejecutar un plan de evaluación que permita medir los resultados de la propuesta desarrollada.

REVISIÓN DE LA LITERATURA O FUNFAMENTOS TEÓRICOS

Visión por computador

La visión artificial o visión por computador es, en esencia, la tecnología que intenta imitar la capacidad visual y cognitiva que poseen algunos seres vivos para percibir, mediante la secuencia de imágenes, un entorno físico y actuar en respuesta a lo observado. En la actualidad, existen varios tipos de aplicaciones industriales que buscan el uso de técnicas de visión artificial, las cuales se encuentran en constante desarrollo y evolución, demostrando un acelerado avance en las últimas décadas evidenciado en la abundante cantidad de investigaciones y publicaciones académicas por parte de la comunidad científica. Esto gracias a la inmensa cantidad de contenido audiovisual que se produce a diario y el constante avance tecnológico en dispositivos electrónicos en cuanto a procesamiento y almacenamiento, además de una vasta disponibilidad de lenguajes de programación, herramientas, y librerías, según afirma (García & Caranqui, 2015).

La visión por computador involucra muchas disciplinas, que incluyen no solo ciencias de la computación, neurología, óptica y matemáticas, sino también disciplinas sociales como la sociología, la filosofía y el comportamiento humano, de acuerdo con (Feng et al., 2021).

Videovigilancia

La videovigilancia es una implementación tecnológica basada en el registro de audio y video constante en entornos específicos, esta herramienta posee amplias ventajas en lo que respecta a la vigilancia tradicional realizada por personal especializado, por ejemplo: el acortamiento de gastos, centralización del área de vigilancia, facilidad de monitorización, etc. Con la aplicación de esta tecnología también es factible monitorear un entorno más extenso, de una manera continua, permitiendo un control más profundo, el cual permite un aumento significativo en la seguridad y protección de bienes y personas, según señala (Ouahhadi Escudero, 2020).

Inteligencia artificial

La inteligencia artificial trata de imitar la inteligencia del ser humano, tanto a nivel cognitivo como intelectual. Para cumplir con este objetivo, se basa en teorías matemáticas y computacionales, logrando, de esta manera, una emulación significativa del comportamiento mediante patrones que explican determinadas conductas inteligentes, tales como: visión, razonamiento, aprendizaje, lenguaje, entre otras, siguiendo a (Regalado et al., 2019).

Aprendizaje automático

El aprendizaje automático, en inglés "machine learning", es uno de los campos claves de investigación en la inteligencia artificial, está definida como el área de estudio que se encarga de desarrollar algoritmos que permitan a las computadoras, a través de la experiencia, adquirir habilidades para las que no fueron explícitamente programadas. Por lo general, el aprendizaje automático aprende de datos estructurados de manera tabular, encontrando patrones similares entre los datos del conjunto afectado para así, tratar de ir más allá del análisis convencional, en el cual se estudian los objetos aisladamente, conforme a (Regalado et al., 2019).

Actualmente existen varios tipos de aprendizaje automático, entre los que más se destacan están: Redes Bayesianas, Modelos de Markov, Máquinas de Soporte Vectorial, Redes Neuronales, Técnicas de lógica difusa, Algoritmos genéticos, Agrupamiento y detección de outlier.

Aplicación en visión por computador y videovigilancia

La detección y reconocimiento basados en la visión es un área de investigación muy importante de la visión por computador en conjunto con otras tecnologías como la que corresponde al aprendizaje automático, cuyo aporte ha servido de gran ayuda para el desarrollo y mejora de técnicas de detección, identificación y seguimiento de objetos o

personas en video sin necesidad de la intervención humana y con una precisión muy alta en comparación con otros métodos tradicionales, de acuerdo con (Feng et al., 2021).

La videovigilancia con el uso de la visión artificial y el aprendizaje automático puede llegar a ser mucho más eficiente y accesible puesto que se reduce significativamente el margen de error, el esfuerzo humano y el almacenamiento requerido, además de aumentar el nivel de seguridad por un constante y preciso monitoreo.

MARCO METODOLÓGICO

El desarrollo de este proyecto se ha orientado hacia una iniciativa que busca aprovechar las tecnologías que tenemos a nuestra disposición y aplicarlas en sectores críticos como lo es la seguridad a nivel de sociedad; por consiguiente, se ha establecido la siguiente metodología de trabajo que también procura cumplir los objetivos propuestos.

DISEÑO DEL ESTUDIO

El estudio se dividió en dos partes fundamentales, una de carácter cuantitativo, que consiste en la investigación acerca de las metodologías usadas en la videovigilancia, principalmente en detección de peatones; y otra de carácter cualitativo, en el cual se buscó un criterio de recolección de datos que nos permitiría obtener los resultados esperados y así acercarnos a los objetivos propuestos.

La primera fase investigativa, al ser cuantitativa, nos permitió revisar datos estadísticos sobre la precisión y el uso de recursos computacionales que tiene cada método de deteccion de peatones seleccionados para **opencv** (librería *open-source* de visión artificial y aprendizaje automático), además se experimentó con estos métodos y se estableció un criterio sobre cada uno de ellos.

Al contrario de la primera, la segunda fase fue cualitativa y nos posibilitó generar un criterio de selección de datos flexible basado en la observación de casos o eventos de interés relacionados con el tema de la videovigilancia. En este procedimiento se tuvo un análisis en el material disponible en internet y se supo

OBTENCIÓN DE DATOS

Para el proceso de recolección de datos se recurrió en gran parte a las redes sociales, principalmente Youtube, Facebook, Instagram y Twitter, en donde se definió un conjunto

de videos relacionados a eventos de videovigilancia de los cuales se fueron capturando muestras según el criterio establecido en cuanto a las características fisiológicas según el comportamiento de cada individuo, tomando en cuenta sus intenciones y buscando un patrón de comportamiento. El conjunto de datos principalmente no fue tan amplio debido a la limitada disponibilidad de este tipo de material en internet, por este motivo se buscó aumentar el set de datos imitando el patrón de comportamiento analizado en los datos recolectados previamente, de este modo se aprovechó el método de detección de peatones implementado para una captación de datos automatizada, conectando una cámara IP de videovigilancia, enfocándola hacia un área de constante tránsito de peatones y guardando cada detección de forma individual.

APLICACIÓN DE APRENDIZAJE AUTOMÁTICO

El modelo de aprendizaje utilizado en este proyecto consiste en una red neuronal convolucional compuesta por 4 capas, esta red fue entrenada, validada y probada por medio de un set de datos binarios, obtenidos en el paso anterior, los cuales fueron preparados y pre-procesados para alimentar dicho modelo.

La técnica que se usó para este procedimiento es conocida como *k-fold*, la cual realiza una división aleatoria del set de datos por grupos, los cuales se van intercabiando en cada iteración, logrando así un aprovechamiento más amplio sobre los datos obtenidos.

DISEÑO DE LA INTERFAZ GRÁFICA DEL PROTOTIPO

Para la implementación de la interfaz gráfica de usuario, se usó un dispositivo NVIDIA® Jetson™ TX2, la cual posee una CPU 64-bit Quad-Core A57 Complex y una GPU con arquitectura NVIDIA Pascal™.

Este prototipo fue estructurado con las funciones básicas para registrar, listar y eliminar cámaras IP de videovigilancia, a las cuales se acceden y monitorizan a través del

protocolo RTSP ó *Real Time Streaming Protocol*, y se realizan las operaciones y detecciones mediante los métodos seleccionados, todo esto ejecutado con los recursos computacionales de dicho dispositivo que fue previamente actualizado y configurado con las herramientas necesarias para un correcto funcionamiento y aprovechamiento de hardware.

PLAN DE EVALUACIÓN:

Para el proceso de evaluación, se ha optado por un área en donde, además de presentar gran afluencia de personas, tambien muestre gran variedad de características visuales y/o movimientos corporales que permita cualificar el comportamiento del modelo de aprendizaje automático, para ello hemos seleccionado un local comercial, en donde podemos encontrar varias personas interactuando con los productos a su alcance al mismo tiempo, de este modo, se colocó una cámara IP de videovigilancia y se ejecutó el proceso de detección y captura de personas.

Con los datos ya guardados se procedió a importarlos dentro del entorno en donde se encuentra el modelo de aprendizaje automático previamente entrenado, se ejecutó una predicción con cada uno de los datos, removiendo las capturas borrosas o sin ninguna apariencia de una figura humana dentro, ya sea por la calidad de la camara o por problemas de conección. Además se agregaron capturas de individuos realizando actividades delictivas tomadas de grabaciones reales y se las agregó al mismo directorio para medir el comportamiento del modelo de forma cualitativa y cuantitativa.

RESULTADOS

MÉTODOS DE DETECCIÓN DE PEATONES

A lo largo del desarrollo de este proyecto se analizaron tres técnicas de visión por computador capaces de detectar peatones o personas, siendo el tercero el seleccionado en este proyecto:

a. Histograma de Gradientes Orientados (*HOG*)
b. Estimación de Poses (*Pose Estimation*)
c. Segmentación de instancias (*Instance Segmentation*)

Histograma de Gradientes Orientados (HOG)

El primer método que se investigó y se lo llevó a la práctica fue el llamado Histograma de gradientes orientados o *HOG* por sus siglas en inglés, este método consiste en evaluar una representación de la imagen en una cuadrícula densa con histogramas individuales, completamente normalizados y orientados según su respectivo gradiente, en otras palabras, el algoritmo busca encontrar apariencias en los objetos denotando la intensidad de los gradientes o dirección de los bordes, y con la ayuda de una máquina de soporte vectorial o *SVM* realizar la detección de peatones dentro de dicha imagen, según señala (Byeon & Kwak, 2017).

Al llevar este método a la práctica, mostró un rápido rendimiento debido a su poco consumo de recursos computacionales, pero gran ineficiencia con respecto a los cambios de iluminación, entornos con amplia variedad de objetos y más aun con el uso de cámaras IP, las cuales muestran cierta latencia en el tiempo de comunicación de paquetes por red local o *LAN*. En consecuencia, se le calculó una precisión de aproximadamente 10%, por lo que no fue útil para llegar al objetivo propuesto.

Estimación de Poses (*Pose Estimation*)

El segundo método candidato a usarse en el proyecto fue el de estimación de poses, el cual se basa en la detección y localización de puntos clave que describen un objeto en específico mediante la aplicación de redes neuronales profundas, en este caso se optó por la estimación de poses humanas, en donde se requiere detectar y localizar principalmente las articulaciones del cuerpo humano para realizar un seguimiento mediante una conexión de dichos puntos, según señala (Gupta, 2018).

Una vez ya implementado en un ambiente de pruebas, pudimos observar que este método no sería de mucha utilidad en la videovigilancia, o al menos no en su estado actual. Pues al momento en que se realizan las detecciones, el modelo llega a confundir muchos objetos dentro de la imagen como partes humanas, ya sea por baja resolución, ruido o incluso por características del propio entorno. Para este método no hubo la necesidad de calcular una precisión, pues rápidamente fue descartado al no mostrar mayor utilidad para el proyecto.

Segmentación de instancias (*Instance Segmentation*)

El tercer método seleccionado y finalmente implementado en el proyecto, fue el de segmentación de instancias "*Mask R-CNN*", el cual consiste en la detección de todos los objetos dentro de una imagen y además segmentar o seccionar con bastante exactitud el área que comprende cada instancia dentro de la misma imagen, esto se consigue gracias al uso de una arquitectura múltiple: una red convolucional que resulta fundamental para la extracción de características; y otra red frontal que realiza la clasificación y regresión para el reconocimiento de cada región de interés, según señalan los autores en su artículo (He et al., 2020).

Este método resultó sumamente útil para el proyecto, puesto que, a diferencia de los anteriores, es capaz de reconocer y clasificar gran cantidad de objetos y con una precisión bastante alta, lo que le hace una excelente opción para la aplicación en videovigilancia.

IMPLEMENTACIÓN DE LA INTERFÁZ GRÁFICA CON LA SEGMENTACIÓN DE INSTANCIAS

Al momento de su implementación se evidenció que la segmentación de instancias es un proceso muy pesado para el computador, el tiempo de procesamiento con CPU ralentiza de manera drástica su rendimiento, pero muestra bastante fluidez si se trabaja con GPU.

Para solventar el consumo constante e innecesario de recursos computaciones, se propuso añadir un método de detección de movimiento con el objetivo de minimizar dicho consumo, ya que en la videovigilancia pueden existir periodos bastante largos en donde las cámaras se encuentran captando video sin ningún tipo de movimiento.

La detección de movimiento se basa en la simple eliminación del fondo de cada imagen mediante una sustracción del fotograma de referencia (principalmente el primer fotograma) al fotograma actual, este método es el más conocido y simple para realizar la detección de movimiento en video, pero se vuelve ineficiente con los cambios de iluminación y movimientos permanentes de objetos dentro del área enfocada, dependiendo del entorno en donde se encuentra la cámara. Para esto se creó un temporizador que se encarga de actualizar el fotograma de referencia cada cierto número de segundos y así mejorar este proceso de detección, que además resulta ser muy eficiente y consume muy pocos recursos computacionales.

En la videovigilancia, la región de interés siempre se centrará en las personas y en sus acciones, por lo que se modificó el método de segmentación de instancias para solamente trabajar con las detecciones de personas.

Finalmente se estableció que el computador se mantiene siempre en espera de detección de movimiento. Cuando esto ocurre, se realiza un proceso de segmentación de instancias en busca de personas y se repite cada cierto número de segundos, mientras se siga detectando movimiento, si en dicho proceso detecta personas, las señala y guarda dentro de un directorio en el mismo computador, si ya no se detecta movimiento, el computador suspende el proceso de segmentación de instancias y nuevamente vuelve a la espera de movimiento.

ENTRENAMIENTO DEL MODELO DE APRENDIZAJE AUTOMÁTICO CON EL SET DE DATOS OBTENIDO

Luego de finalizar el proceso de obtención de datos, se realizó un análisis sobre los mismos y se destacó que existen ciertos elementos característicos que posibilitarían una correcta clasificación como el tiempo en el que se lleva a cabo ciertas acciones, la vestimenta, los tipos de movimientos, etc.

Este proyecto se centró en el análisis del comportamiento de cada individuo en el set de datos, y se lo relacionó con la intención que tenía o la acción que realizaba. De este modo, se propuso realizar una clasificación binaria, etiquetando los datos como normales o sospechosos, tal como se puede observar en Ilustración 1: Ejemplificación de la clasificación del set de datos.

Ilustración 1: Ejemplificación de la clasificación del set de datos.

Este criterio de clasificación consiste en la postura fisiológica de ciertos individuos frente a diferentes eventos de videovigilancia. Para un comportamiento normal, tenemos que el individuo mantiene su postura erguida en todo momento, demostrando tranquilidad en todas las acciones que lleva a cabo; en lo contrario, un comportamiento sospechoso, se caracteriza por mantener una postura encorvada con sus extremidades superiores tratando de cubrirse el rostro en todo momento y sus extremidades inferiores inusualmente más flexionadas y separadas, demostrando cierta intranquilidad y apuro en todas sus acciones.

El modelo de aprendizaje automático está implementado con *TensorFlow* y se basa en una red neuronal convolucional 2D para clasificar imágenes de personas capturadas en la segmentación de instancias, la arquitectura de dicha red consta de 4 capas de 32, 64, 128 y 512 filtros respectivamente, usando un optimizador *ADAM* con sus valores por defecto, los pesos inicializados por el esquema por defecto *Xavier GLOROT* y su función de pérdida está definida por la función de entropía cruzada binaria en conjunto con la activación sigmoidea, esto para poder obtener valores de salida en un rango de 0 a 1, los cuales representan ambas clases definidas.

Para la validación cruzada, se utilizó un esquema de validación *K-fold* en lugar de la validación tradicional, con esto obtuvimos un mayor aprovechamiento de los datos que se vieron limitados por la escasa disponibilidad y el poco acceso público en internet, además con esta técnica se minimiza el posible sobre entrenamiento. Finalmente se realizó el entrenamiento con 5 pliegues o *folds*, en donde los datos se dividieron en grupos y en cada iteración de entrenamiento se intercalaron para formar los grupos de entrenamiento y validación por 30 *epochs* o periodos en cada *fold*, este procedimiento se representa en Ilustración 2: Resultados del entrenamiento en 5 *folds*. Y en Ilustración 3: valor de pérdida y precisión durante los últimos 30 epochs.

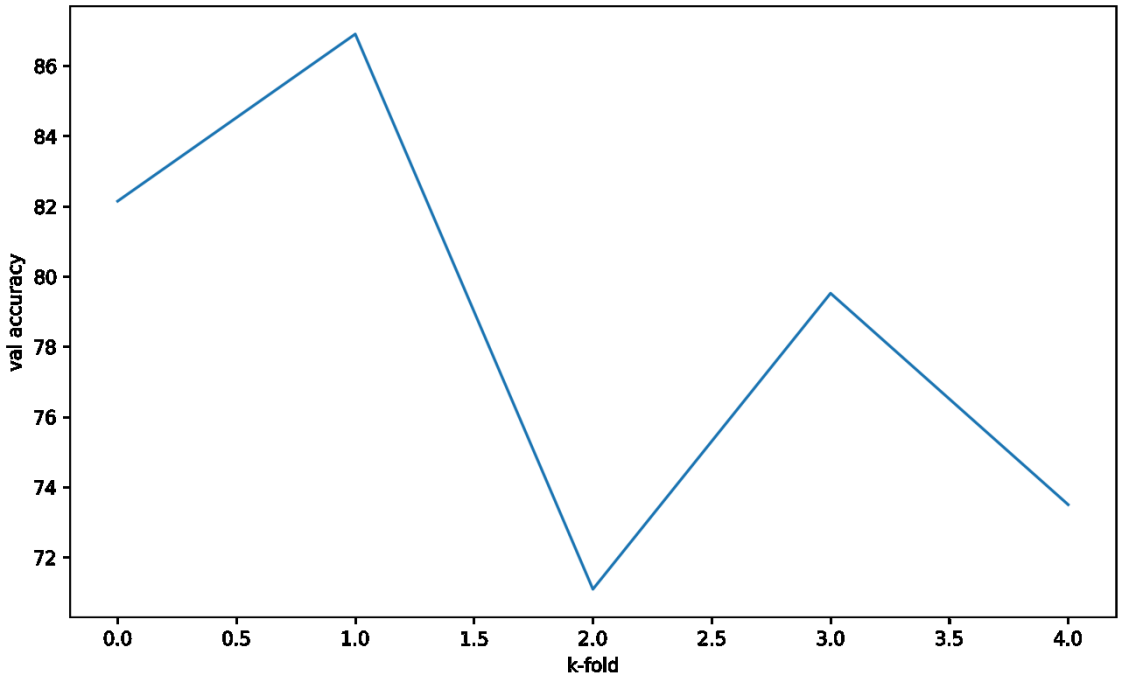

Ilustración 2: Resultados del entrenamiento en 5 folds.

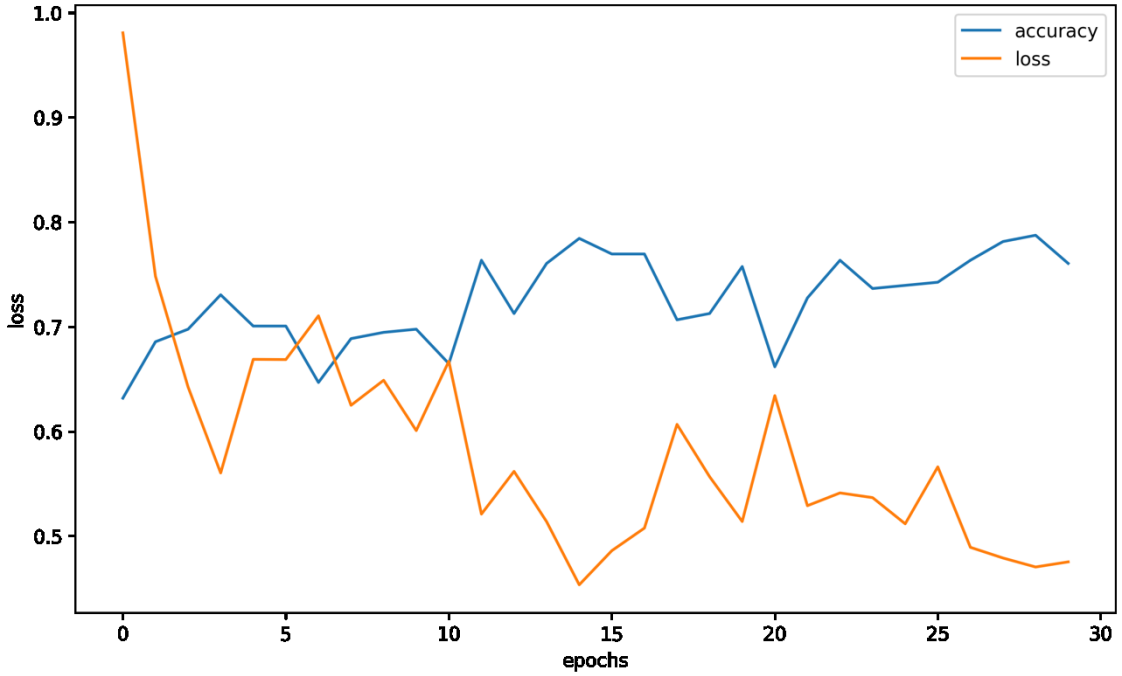

Ilustración 3: valor de pérdida y precisión durante los últimos 30 epochs.

EVALUACIÓN

Para la fase evaluativa, se instaló el prototipo en un local comercial, esto con el objetivo de capturar un gran número de personas realizando múltiples movimientos y además se agregó un conjunto pequeño de imágenes de eventos de videovigilancia reales etiquetados como sospechosos para poder evaluar el comportamiento del modelo de aprendizaje previamente entrenado, al obtener los resultados se seleccionó una muestra (que se puede apreciar en Ilustración 4: Muestra seleccionada de resultados finales) que permite describir el comportamiento que tuvo dicho modelo.

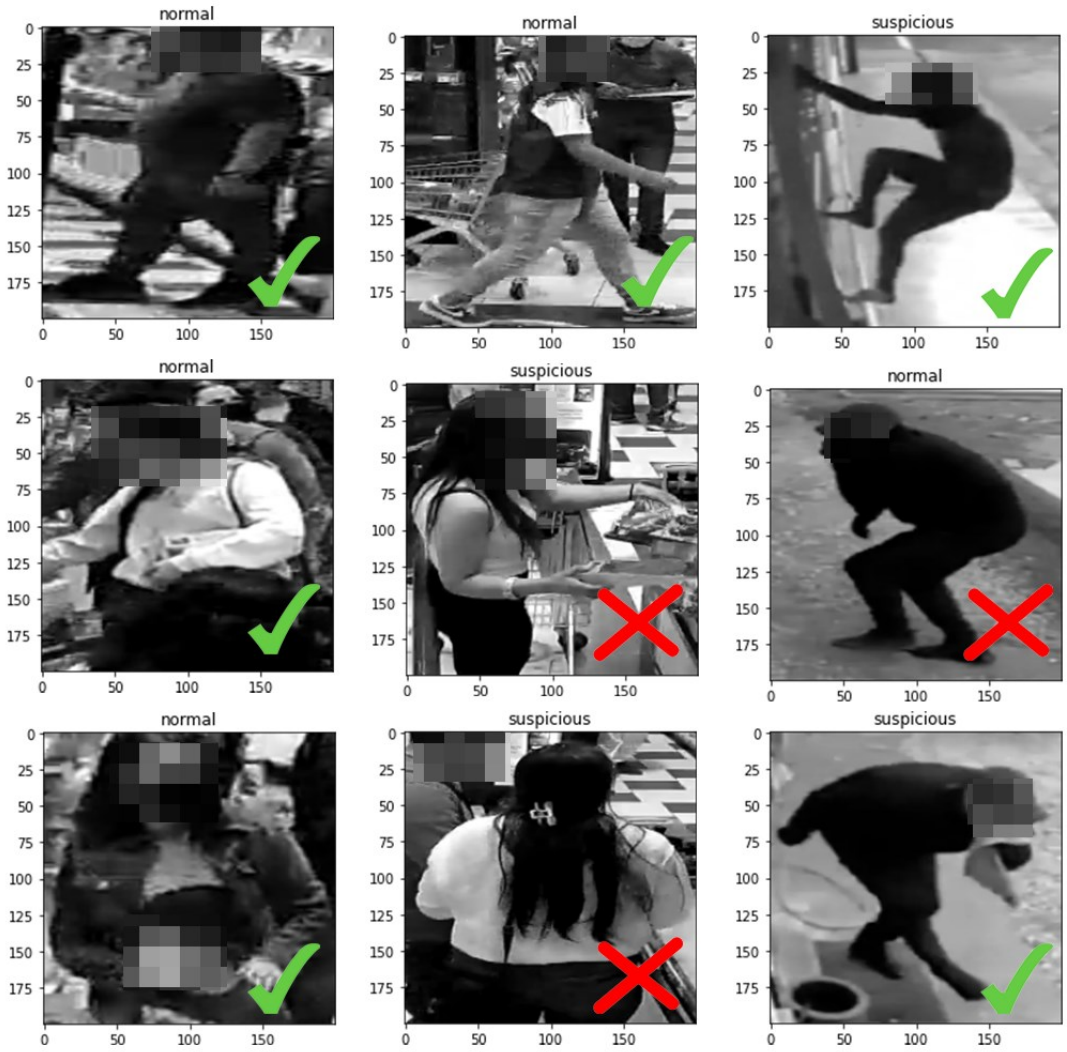

Ilustración 4: Muestra seleccionada de resultados finales

Como podemos observar, el modelo es capaz de clasificar en muchos casos de forma correcta, pero también llega a fallar en varios escenarios en donde la contextura de las personas varía o cuando realizan acciones extendiendo las extremidades como para realizar un pago o revisar un producto, de esta forma podemos ver que este modelo puede llegar a confundirse y clasificar lo sospechoso como normal y viceversa. Cabe recalcar que el set de datos no fue tan amplio debido a la escasa disponibilidad de material válido y de libre acceso, aún cuando el proceso de búsqueda y obtención de dichos datos fue prolongado.

CRONOGRAMA

- OE1. Estudiar y conocer las principales técnicas de visión por computador y tipos de aprendizaje automático aplicados en la videovigilancia.

Tabla 1: Actividades para OE1

Nro.	**Actividad**
1	Realizar una búsqueda exhaustiva de trabajos relacionados, tomando en cuenta los más recientes.
2	Efectuar una lectura comprensiva de todos los documentos académicos seleccionados, resaltando las ideas de mayor interés en cada uno.
3	Realizar un análisis comparativo de la información abstraída.
4	Documentar el resultado de todo el proceso investigativo.

- OE2. Estudiar, obtener y preparar datos reales acerca del comportamiento humano de interés en varios entornos bajo diferentes condiciones físicas y sociales.

Tabla 2: Actividades para OE2

Nro.	Actividad
1	Realizar una revisión de los eventos de interés más comunes en registros de videovigilancia.
2	Seleccionar los registros de mayor utilidad teniendo en cuenta la cantidad de características presentes.
3	Extraer información acerca del comportamiento humano en los datos seleccionados.
4	Realizar una clasificación por tipos de comportamiento.
5	Etiquetar los datos.
6	Aplicar técnicas de pre-procesamiento de imágenes a los datos obtenidos.

- OE3. Diseñar y entrenar un modelo de aprendizaje automático para brindar soporte en la detección de eventos de videovigilancia.

Tabla 3: Actividades para OE3

Nro.	Actividad
1	Determinar un algoritmo para la detección de peatones, basado en un modelo pre-entrenado.
2	Implementar el algoritmo.
3	Analizar y seleccionar una técnica de aprendizaje automático.
4	Entrenar el modelo de aprendizaje con los datos conseguidos.
5	Revisar que el proceso de aprendizaje haya sido correcto.

- OE4. Diseñar y ejecutar un plan de evaluación que permita medir los resultados de la propuesta desarrollada.

Tabla 4: Actividades para OE4

Nro.	Actividad
1	Definir un escenario de pruebas.
2	Determinar un periodo de muestreo documentado.
3	Realizar mediciones con los datos recopilados en las pruebas.
4	Documentar el resultado de todo el proceso de pruebas.

Tabla 5: Cronograma de actividades con fechas

Objetivo	Actividad	Semana	Fecha Inicio	Fecha Fin	Días	Horas
OE1	Ac. 1	1	26/10/2021	30/10/2021	5	15
	Ac. 2	2	2/11/2021	8/11/2021	7	21
	Ac. 3	3	9/11/2021	13/11/2021	6	18
	Ac. 4	4	15/11/2021	18/11/2021	4	12
OE2	Ac. 1	5	19/11/2021	26/11/2021	6	18
	Ac. 2	6	27/11/2021	30/11/2021	6	18
	Ac. 3	6	1/12/2021	4/12/2021	5	15
	Ac. 4	7	6/12/2021	11/12/2021	3	12
	Ac. 5	8	13/12/2021	21/12/2021	7	21
	Ac. 6	9	22/12/2021	27/12/2021	6	18
OE3	Ac. 1	9	28/12/2021	31/12/2021	4	16
	Ac. 2	10	3/1/2022	8/1/2022	6	18
	Ac. 3	11	10/1/2022	15/1/2022	6	18
	Ac. 4	11	17/1/2022	22/1/2022	6	18
	Ac. 5	12	24/1/2022	29/1/2022	6	18
OE4	Ac. 1	12	31/1/2022	5/2/2022	6	18
	Ac. 2	13	7/2/2022	12/2/2022	6	18
	Ac. 3	14	12/2/2022	14/1/2022	2	8
	Ac. 4	15	14/2/2022	17/2/2022	1	4
Total					98	304

PRESUPUESTO

Tabla 6: Tabla de presupuesto

Denominación	Cant.	Costo Unitario	Costo Total
	Unidades	Dólares	Dólares
1. Tecnológico			
Computadora	1	1,000.00	1,000.00
Cámara de videovigilancia	2	50.00	100.00
Dispositivo Nvidia Jetson TX2	1	600.00	600.00
2. Servicios			
Servicio de transporte	1	50.00	50.00
Servicio de internet	3	28.00	168.00
3. Personal			
Estudiante/Desarrollador	304 horas (1 estudiante)	4.00 /hora	1,216.00
4. Otros			
Imprevistos	1	100.00	100.00
TOTAL			3,234.00

CONCLUSIONES

Los avances tecnológicos en el área de la visión por computador está en constante crecimiento y aplicar estos conceptos al campo de la videovigilancia es una muy buena propuesta, teniendo en cuenta las mejoras que esto nos puede traer hacia la seguridad a nivel de sociedad. Existen muchos otros criterios de aplicación, unos mejores que otros, sin embargo todos necesitan de un estudio muy extenso para lograr una propuesta factible. En este caso, se realizó una obtención de datos reales provenientes de eventos delictivos grabados y subidos a internet a través de las redes sociales, que aunque fueron muy limitados, se logró pre-procesarlos, entrenarlos con un modelo de aprendizaje y, finalmente, se comprobó que el computador es capaz de clasificar una actividad como normal o sospechosa si tiene un entrenamiento adecuado.

RECOMENDACIONES

Teniendo en cuenta lo anterior mencionado, se establecieron ciertas recomendaciones tales como:

- Diseñar un conjunto de datos colectivo, en donde todo quien se interese pueda aportar con sus propios datos e incluso con una revisión de los ya existentes.
- Solicitar la colaboración de un sistema de cámaras de videovigilancia más amplio como lo es el ECU911, para así obtener un set de datos más robusto y un modelo de aprendizaje automático mejor entrenado y más preciso.
- Buscar inversiones externas en las tecnologías aplicadas a la videovigilancia.

REFERENCIAS BIBLIOGRÁFICAS

Babiker, M., Khalifa, O. O., Htike, K. K., Hassan, A., & Zaharadeen, M. (2018). Automated daily human activity recognition for video surveillance using neural network. *2017 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2017*, *2017-Novem*(November), 1–5. https://doi.org/10.1109/ICSIMA.2017.8312024

Byeon, Y. H., & Kwak, K. C. (2017). A Performance Comparison of Pedestrian Detection Using Faster RCNN and ACF. *Proceedings - 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017*, 858–863. https://doi.org/10.1109/IIAI-AAI.2017.196

Elsayed, O. A., Mohamed Marzouky, N. A., Atef, E., & Salem, M. A. M. (2019). Abnormal Action detection in video surveillance. *Proceedings - 2019 IEEE 9th International Conference on Intelligent Computing and Information Systems, ICICIS 2019*, 118–123. https://doi.org/10.1109/ICICIS46948.2019.9014712

Feng, Z., Zhu, X., Xu, L., & Liu, Y. (2021). Research on human target detection and tracking based on artificial intelligence vision. *Proceedings of IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2021*, 1051–1054. https://doi.org/10.1109/IPEC51340.2021.9421306

García, I., & Caranqui, V. (2015). La visión artificial y los campos de aplicación. *Tierra Infinita*, *1*, 94–103. http://revistasdigitales.upec.edu.ec/index.php/tierrainfinita/article/view/76

Gupta, V. (2018). *Deep Learning based Human Pose Estimation using OpenCV*. https://learnopencv.com/deep-learning-based-human-pose-estimation-using-opencv-cpp-python/

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 386–397. https://doi.org/10.1109/TPAMI.2018.2844175

Kakadiya, R., Lemos, R., Mangalan, S., Pillai, M., & Nikam, S. (2019). AI Based Automatic Robbery/Theft Detection using Smart Surveillance in Banks. *Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019*, 201–204. https://doi.org/10.1109/ICECA.2019.8822186

Ouahhadi Escudero, F. (2020). *La videovigilancia como instrumento de control en las relaciones laborales* [Universidad Pública de Navarra]. https://hdl.handle.net/2454/37597

Regalado, B., Sonia, E., Bautista, U., & Alexis, E. (2019). *Análisis , implementación y evaluación de modelos de aprendizaje automático relacional*.

Rofifah, D. (2021). *Hubo más robos en el primer semestre de 2021 que en ese periodo en 2020*. Paper Knowledge . Toward a Media History of Documents. https://gk.city/2021/08/05/robos-primer-semestre-2021/

Wang, X., Wang, M., & Li, W. (2014). Scene-specific pedestrian detection for static video surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(2), 361–374. https://doi.org/10.1109/TPAMI.2013.124

ANEXOS

A continuación, se adjunta los enlaces al código fuente del proyecto concluido, en dos versiones, una compatible con CPU x86 (Intel & AMD); y otra compatible con CPU ARM64 – GPU NVIDIA (CUDA):

https://estliveupsedu-my.sharepoint.com/:f:/g/personal/portizs2_est_ups_edu_ec/El3zNTGypWdPg-5tas60jV8BRLSynhQd5KPsE2k7IA3Nvw?e=IeVsnB