



GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

Automatización del procesado y análisis de imágenes de video,
estudio del estado del arte y aplicabilidad de la inteligencia
artificial en la seguridad de las instalaciones

Autor: Iñigo Francia Alonso

Director: Emilio Manuel Domínguez Adán

Agosto 2024, Madrid

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
'Automatización del procesado y análisis de imágenes de video, estudio del estado del arte
y aplicabilidad de la inteligencia artificial en la seguridad de las instalaciones'
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el
curso académico 2023/24 es de mi autoría, original e inédito y
no ha sido presentado con anterioridad a otros efectos.
El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido
tomada de otros documentos está debidamente referenciada.



Fdo.: Iñigo Francia Alonso

Fecha: 28/08/2024

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Emilio Manuel Domínguez Adan

Fecha: 28/08/2024



GRADO EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES

Automatización del procesado y análisis de imágenes de video,
estudio del estado del arte y aplicabilidad de la inteligencia
artificial en la seguridad de las instalaciones

Autor: Iñigo Francia Alonso

Director: Emilio Manuel Domínguez Adán

Agosto 2024, Madrid

AUTOMATIZACIÓN DEL PROCESADO Y ANÁLISIS DE IMÁGENES DE VIDEO, ESTUDIO DEL ESTADO DEL ARTE Y APLICABILIDAD DE LA INTELIGENCIA ARTIFICIAL EN LA SEGURIDAD DE LAS INSTALACIONES

Autor: Francia Alonso, Iñigo

Director: Domínguez Adan, Emilio Manuel

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

En este proyecto, se ha estudiado la industria del análisis de video, principalmente ligada a los servicios de videovigilancia. Se ha analizado en detalle el estado del arte y las tecnologías empleadas actualmente, así como los próximos avances previstos. El análisis ha profundizado en la revolución del análisis de video por medio de la incorporación de inteligencia artificial en las tareas de procesado de video y se ha considerado tanto el impacto socioeconómico como las principales limitaciones.

Palabras clave: Análisis de video, automatización, video analítica, machine learning, videovigilancia, seguridad, frame, pixel.

1. Introducción

El trabajo persigue entender la industria del análisis de video y los avances que la impulsarán en el futuro. La motivación del trabajo nace de la realidad de la industria y como las tecnologías empleadas están anticuadas en comparación con muchas otras tecnologías que se utilizan en industrias similares, principalmente enfocándonos en la conducción autónoma.

La investigación de las tecnologías empleadas en la actualidad tiene un componente teórico y práctico por medio del desarrollo de un programa de prueba que permite conocer las principales limitaciones. También se analiza la tecnología existente en cuestión de machine learning en general y como está comenzando a aplicarse en el procesado de imágenes de video de forma preliminar, principalmente debido a que los avances aun no son suficientes como para justificar el alto coste.

Además, se evalúa la existencia de grandes limitaciones para la aplicación “mass market” lo cual limita que el procesado de video se automatice por completo. El estudio también abarca la conducción autónoma y como esta tecnología, paralela al análisis de video, que ha sido financiada por empresas con gran capacidad económica no se emplea aun en la práctica y sigue utilizándose en pruebas piloto, esto nos da una indicación del largo proceso que aún queda para que realmente el análisis de video sea un proceso completamente automatizado.

Por último, se realiza un estudio del impacto socioeconómico de la implementación de la automatización del procesado de video. Se abarcan las ventajas operativas y económicas que el procesado de video automatizado tendría en la sociedad en general y en la industria de la seguridad y la videovigilancia en particular. Para esto se determina cuáles son los sectores donde el análisis de video tiene una gran importancia más allá del de la seguridad, como, por ejemplo, la administración del tráfico o la supervisión de operaciones comerciales. Además, se evalúa el impacto que tendría la completa implantación tanto desde la perspectiva empresarial como desde la perspectiva del usuario.

2. Estado del Arte

El grado de automatización de la analítica de video actual es limitado, la mayoría de los programas de análisis de video actuales tan solo son capaces de extraer las características principales de los datos de un video y detectar los sucesos que ocurren en estos videos por medio de la detección de objetos y de la detección de los cambios entre los frames. El programa de prueba que se ha desarrollado en este TFG para comprender las principales limitaciones y posibilidades de avance trabaja empleando esta tecnología por medio de la cual se evalúa un frame contra el siguiente frame de ese video, observando la variación pixel a pixel, de este modo si ocurre algo en el video que sea significativo (ej. Aparición de una persona en la imagen) seremos capaces de detectarlo ya que un gran número de los pixels cambiaran de un frame al siguiente. Esta tecnología ya está completamente implantada y el siguiente gran reto de la industria es la capacidad de determinar que está ocurriendo en el video que se está analizando.

La principal línea de trabajo actual está centrada en desarrollar la capacidad de reconocer los movimientos de los seres humanos y en base a esto determinar qué clase de evento está ocurriendo en el video y tomar una decisión en consecuencia. Para ello existen distintas técnicas siendo las principales: pose-based method, tracking-based method y spatio-temporal method, estas técnicas basan el reconocimiento de los seres humanos en una serie de características que determinan la apariencia en los videos, siendo las principales las articulaciones humanas, la posición esquelética erguida y las distintas velocidades a las que se suelen realizar los movimientos. También existen líneas de desarrollo que pretenden dotar a los programas de la capacidad de detectar a seres humanos por medio de machine learning.

La realidad es que en la actualidad hay sistemas de análisis de video verdaderamente efectivos, pero en ningún caso son autónomos y saben identificar con precisión que es lo que ocurre en el video, siendo necesaria la participación humana en última instancia. Por otro lado, también se está trabajando para solucionar el otro gran problema asociado al análisis de video, el excesivo tamaño de los datos, para ello se está trabajando con la computación frontera con el objetivo de realizar un análisis en cascada e ir filtrando el video para únicamente realizar un análisis más profundo y computacionalmente complejo de aquellas partes del video que previamente se hayan determinado de interés.

3. Aplicabilidad de la IA

La inteligencia artificial y más en concreto la rama de machine learning modificará por completo el análisis de video y la forma en la que este se lleva a cabo. Es cuestión de tiempo que existan modelos que, entrenados con grandes bancos de videos, sean capaces de identificar con exactitud lo que sucede en un video que están analizando y de, en base a esta identificación, tomar una decisión sobre cómo actuar. Aun así, la realidad, es que estamos lejos de que esto suceda y que sea aplicable a gran escala, principalmente por los costes y la complejidad de esta tarea.

El principal cuello de botella que debe ser solventado está relacionado con el entrenamiento de los modelos. Para poder entrenar un modelo de machine learning de forma efectiva se necesitan bibliotecas extensas de video con el que alimentar el modelo para que este, a través de la red neuronal que utilice, sea capaz de sacar reglas que le permitan identificar sucesos en el video que analiza. Pero no solo es necesario que estas bibliotecas sean extensas y presenten una gran variedad de videos, también es de gran importancia que estos videos estén correctamente etiquetados para que el modelo pueda sacar conclusiones y que durante su entrenamiento asocie las reglas que cree para la detección de determinados videos con sucesos que deben generar una alarma. Por otro lado, hay otra gran limitación asociada a la forma en la que se realiza el análisis de video actual, que sigue un proceso descentralizado en el que no existe colaboración entre las distintas cámaras de una red, lo cual hace que el análisis que realizaría la IA este sesgado ya que carece de la visión holística que tendría un ser humano al observar todas las cámaras. Para alcanzar un análisis de video con IA exitoso es necesario centralizar el proceso de análisis y esto aumentaría significativamente el coste computacional que como ya hemos mencionado ya es excesivo actualmente y supone una gran limitación.

El coche autónomo es en la actualidad la gran esperanza para el desarrollo de las principales tecnologías necesarias para un análisis de video completamente automatizado, principalmente gracias a que las empresas del sector son capaces de financiar las inversiones necesarias. El gran retorno en términos económicos que la completa implementación del coche autónomo tendría, significa que los principales avances vendrán muy probablemente desde esta industria y serán modificados para las tareas más generales del análisis de video como la videovigilancia.

4. Impacto Socio – Económico

El empleo de análisis de video automatizado es una cuestión de tiempo, pendiente de que las limitaciones de tipo técnico sean resueltas a un coste asumible y de que las compañías dispongan de los recursos necesarios. El cambio en el modelo de operaciones de muchas industrias será completamente disruptivo. En particular en los servicios de vigilancia y seguridad, dando lugar a un modelo de empresas completamente diferente en tamaño y estructura.

Las empresas del sector de la vigilancia reducirán drásticamente su necesidad de equipos humanos, reduciendo significativamente los costes. Además, habrá una mejora significativa de los servicios ofrecidos reduciendo los errores y teniendo más cobertura. Esta creación de valor se repartirá entre los usuarios, por medio de una democratización de los servicios de

vigilancia con una clara reducción de los costes, y las empresas, que verán una evidente mejoría en sus márgenes. En el terreno social el principal impacto será la significativa reducción del empleo en el sector, muy intensivo actualmente, pero con niveles salariales ajustados. El perfil del empleado evolucionará hacia perfiles de mayor valor añadido.

A pesar de las indudables ventajas del análisis de video automatizado y de que en el futuro transformará por completo la industria, todavía se está lejos de su utilización. Las empresas que deberán promover estos avances como, por ejemplo, Securitas o Prosegur tienen limitados recursos financieros para afrontar de forma directa la inversión necesaria en el desarrollo de modelos de IA y de afrontar la propia inversión necesaria en una transformación del modelo empresarial tan disruptiva como la que acontecerá. La reducción de la inversión o la disponibilidad de modelos consolidados y accesibles en actividades relacionadas como puede ser la conducción autónoma en la que se están haciendo inversiones masivas, ayudará notablemente a hacer asumible la inversión en la automatización del análisis de video.

5. Conclusiones

El futuro del análisis de video pasa por la automatización, pero para ello deben de superarse las limitaciones existentes en la actualidad. Estas limitaciones están divididas en la rama técnica y la rama económica. Técnicamente se debe resolver la problemática del coste computacional y de procesamiento, ya que actualmente el volumen de computación requerido es insostenible y, por otro lado, también es necesario invertir significantes recursos en el entramiento de los modelos de ML. En cuestiones económicas es evidente que los avances serán absolutamente disruptivos en todas aquellas industrias que se vean permeadas por la automatización del análisis de video. Aun así, es una realidad que las empresas del sector de la seguridad y la videovigilancia no poseen los recursos necesarios ni el acceso a la tecnología para liderar estos avances. Por ello, es previsible que el camino hacia la completa automatización tengo como catalizadores a otras industrias que, sí que tienen las capacidades de inversión y de desarrollo de las principales tecnologías necesarias, destacando el sector de la conducción autónoma.

VIDEOANALISIS AUTOMATION, STATE OF THE ART AND IMPLICATIONS OF AI IMPACT IN THE SECURITY OF INSTALLATIONS

Author: Francia Alonso, Iñigo

Director: Domínguez Adan, Emilio Manuel

Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

Video analysis, mainly linked to video surveillance services is in the path towards automation. The state of the art has been thoroughly analyzed together with the technologies employed nowadays in a professional environment. The main driver towards automatization will be AI and it will have a deep impact not only on those industries that are tightly linked to video analytics such as security of installations, but also on society in general.

Keywords: Video analysis, automatization, video analytics, machine learning, video surveillance, security, frame, pixel

1. Introduction

The goal is to understand the industry that surrounds video analysis and to determine the drivers towards the future of this industry. The motivation comes from how the development of this industry is subpar when compared to its peers, as an example we can see the huge difference in how developed this industry is when compared to autonomous driving.

Investigation of the state of the art has been carried in a bidimensional way, both through a theoretical understanding of the current technology being used and by developing a trial program to further understand the main limitations of current technology. The analysis scope also covers the technology available in terms of machine learning and how it is already being used for video analysis although it is not yet used in a “Mass market” manner since the costs are very high. The development of autonomous vehicles is used across the project as a reference, due to the similarities in the technologies employed.

Also, the socio-economic impact of the total implementation of automatized video analysis is analyzed, covering the main advantages of this upcoming technology both to society in general and to the video surveillance and security industry in particular.

2. State of the Art

Nowadays there is limited automation in video surveillance, most programs are only able of obtaining the main characteristics of a video and of determining that an event is occurring in the video by detecting the objects that appear on the image and by comparing the changes in

between frames through the variation of each pixel. This is in line with what is done by the trial program which is used as support for the project and that allows us to acknowledge the main limitations of this technology. The next step for the industry is the ability to determine exactly what is occurring in a video rather than only being able to determine that something is occurring.

The main developing workstream is the ability to recognize the movements of human beings and based on this determine what kind of event is occurring in the video and decide accordingly on how to act. For this purpose, there are different techniques, the main ones being: pose-based method, tracking-based method and spatio-temporal method. These techniques base the recognition of human beings on a series of characteristics very particular to humans, the main ones being the human joints, the upright skeletal position and the different speeds at which humans usually perform different movements. There are also lines of development that aim to provide programs with the ability to detect humans in videos through machine learning.

The reality is that nowadays there are truly effective video analysis systems but none of them are fully autonomous or are able of identifying with precision what is happening in a video. Human participation is still required for all tasks. At the same time, the other main caveat towards full automatization is the excessive size of data needed for these tasks. To this end, work is being done via edge computing with the aim of performing analysis in different steps. The aim is to use lighter analysis for general video and only perform deeper analysis of those parts of the video that has been determined to be of interest.

3. AI Applicability

Artificial Intelligence and Machine Learning will completely transform video analysis and the way in which it is currently carried out. It is only a matter of time before models trained on large banks of videos will be able to identify with precision y what is going on in a video they are analyzing and, based on this identification, be able to take decisions on how to act. Despite this, the reality is that there is a long path ahead until this is achieved and even further away to the mass market implementation due to the costs and complexity of the task.

The main bottleneck to address is related to the training of ML models. To train this models effectively extensive video libraries are needed so that the neural network is capable of extracting rules that allow it to identify events in the video it analyses. These libraries not only need to be extensive and have variety amongst the videos but also it is of great importance that these videos are correctly labelled so that the model can draw conclusions and that during the training it is able to associate rules inferred from a determined set of videos to events that should become synonym of generating an alarm. The other key barrier towards the implementation of AI is the decentralized process that current video analysis has. To achieve successful video analysis with AI, it is necessary to centralize the analysis process to gain a view of the “big picture” and this would result in an increase in the processing requirements, which, as already mentioned are one of the main limitations towards automatized video analysis.

The autonomous vehicle industry is currently the great hope for the development of the core technologies needed for fully automated video analytics, mainly because the companies in the sector can finance the necessary investments and the economic returns support it. The breakthroughs achieved by this industry will undoubtedly be extrapolated for the more general tasks of video analytics, in particular, video surveillance.

4. Socio-economic Impact

The use of automated video analytics is a matter of time pending that the technical constraints are resolved at an affordable cost and that companies possess the necessary resources to implement them. The change in the operating model of many industries will be disruptive. Particularly in surveillance and security services which will lead to a completely different business model both in terms of size and structure.

Businesses in the surveillance sector will drastically reduce their need for human teams which will reduce costs significantly. Additionally, the service offered will improve via errors minimization and by extensive coverage. The value created by automation is to be split between users, who will experience a price reduction, and companies, which will achieve a clear improvement in their margins. On the social side, the main impact will be a reduction in the levels of employment of the sector, which is currently very intensive but with tight salaries. The employee profile will shift towards higher value-added profiles.

Despite the numerous advantages associated with automated video analytics and how it will completely transform the industry of video surveillance, nowadays it is still far from becoming a disruptive change. The companies that are positioned to promote these advances, such as Securitas or Prosegur, for example, have limited financial resources and can't face the required investments to develop the necessary AI models whilst facing at the same time the intrinsic cost of a transformation as disruptive as the one that will take place. Time will allow for a reduction of necessary investment as the availability of consolidated and accessible models from related activities such as autonomous driving, in which massive investments are being made, will increase. This will significantly help to make the investment in video analytics automation affordable.

5. Conclusions

The future of video analysis lies in automation, but for this to happen there are several limitations to overcome. These limitations can be divided into technical and economic limitations. Technically, the problem of computational and processing requirements must be addressed since the current requirements of computational volume are unsustainable and, also, it is necessary to invest significant resources in the training of ML models. In economic terms, the developments will be disruptive, and many industries will be permeated by this impact. Still, it is a reality that companies in the security and video surveillance sector do not have the necessary resources or access to technology to lead the developments. Therefore, it is foreseeable that the path towards full automation will be catalyzed by the parallel development in other industries, highlighting the autonomous vehicle industry, that employ similar technologies to those required for completely automated video analysis.

Índice de la Memoria

1	Introducción.....	5
1.1	Objetivos del proyecto.....	5
1.2	Motivación.....	6
1.3	Metodología.....	7
2	Estado de la cuestión	9
2.1	Estado de la cuestión de la analítica de video	9
2.2	Estado de la cuestión del Machine Learning.....	12
2.3	Limitaciones actuales al empleo generalizado de IA en análisis y procesado de video	17
2.3.1	Datos de Entrenamiento Limitados	18
2.3.2	Proceso descentralizado y colaboración limitada entre cámaras.....	20
2.4	Empleo de IA en procesado de video actualmente	20
2.5	Vinculación con el desarrollo y consolidación del coche autónomo.....	22
3	Análisis de recursos	27
4	Desarrollo de un entorno propio.....	31
4.1	Finalidad del Modelo.....	31
4.2	Desarrollo del Programa.....	31
4.2.1	Modelos de prueba	32
4.2.2	Modelo Inicial	32
4.2.3	Introducción de una máscara y reducción de coste computacional.....	35

4.2.4	Ajuste del tamaño de imagen, selección de área de exclusión simplificada e impresión de mensajes con imagen	37
4.2.5	Selección del parámetro en función de la calidad de la imagen.....	41
4.3	Aprendizajes y limitaciones del programa	42
5	Impacto Socio – Económico.....	43
5.1	Ventajas operativas y económicas de la utilización de análisis de video automatizado.....	43
5.2	Sectores en las que la aplicación de análisis de video automatizado aporta un valor añadido muy relevante.....	47
5.2.1	Algunas consideraciones de interés sobre el sector de vigilancia y seguridad 49	
5.3	Requisitos para la implantación generalizada (mass-market) de la automatización del visionado.....	59
5.4	Análisis de rentabilidad coste-beneficio.....	61
5.4.1	Costes operativos recurrentes (OPEX):.....	61
5.4.2	Incremento de inversión (CAPEX).....	62
5.4.3	Mejora de la propuesta de valor e impacto en precios para los clientes.....	64
5.4.4	Impacto de la rentabilidad y tamaño de la industria.....	64
6	Conclusiones y trabajos futuros.....	66
6.1	Conclusiones.....	66
6.2	Trabajos futuros – ADE	67
7	Glosario	69
8	Bibliografía.....	72
9	Anexo I: Objetivos de desarrollo sostenible.....	80

Tabla de Ilustraciones

ILUSTRACIÓN 1. RECONOCIMIENTO DE LA FISIONOMÍA HUMANA EMPLEANDO EL POSE-BASED METHOD (LI ET AL., 2019).....	11
ILUSTRACIÓN 2. EVOLUCIÓN DE LA PROGRAMACIÓN TRADICIONAL AL MACHINE LEARNING (MORONEY, 2020).....	13
ILUSTRACIÓN 3. ESTRUCTURA GENERAL DE UNA RED NEURONAL (NIELSEN, 2019).....	16
ILUSTRACIÓN 4. SELECCIÓN DE VIDEOS DE LA BASE DE DATOS “MOMENTS IN TIME DATASET” (MONFORT ET AL., 2019)	19
ILUSTRACIÓN 5. SELECCIÓN DE FRAMES A ESTUDIAR OBVIANDO EL FONDO DE LA IMAGEN (NGUYEN ET AL., 2020)	22
ILUSTRACIÓN 6. COMPARATIVA DE SEGURIDAD ENTRE VEHÍCULOS TESLA Y EL MERCADO AMERICANO GENERAL (TESLA INC., 2024).....	25
ILUSTRACIÓN 7. EJEMPLO DE SELECCIÓN DE ZONA DE EXCLUSIÓN EN EL VIDEO (FRANCIA, 2024).....	36
ILUSTRACIÓN 8. ESCALA DE GRISES; SIENDO 0 EL NEGRO Y 255 EL BLANCO	36
ILUSTRACIÓN 9. EJEMPLO DEL MÉTODO DE SELECCIÓN DE UNA ZONA DE EXCLUSIÓN (FRANCIA, 2024).....	38
ILUSTRACIÓN 10. IMAGEN INICIAL DE UN VIDEO ANALIZADO CON EL PROGRAMA DE PRUEBA (FRANCIA, 2024)	39
ILUSTRACIÓN 11. EJEMPLO DE DETECCIÓN DE UN SUCESO EN EL PROGRAMA DE PRUEBA (FRANCIA, 2024)	40
ILUSTRACIÓN 12. MENSAJE POR PANTALLA PARA INICIALIZAR EL PRINCIPAL PARÁMETRO DEL PROGRAMA DE PRUEBA (FRANCIA, 2024).....	41

Tabla de Ilustraciones

ILUSTRACIÓN 13. RANKING DE EMPRESAS DE SERVICIOS DE SEGURIDAD PRIVADA (ELECONOMISTA, 2024).....	52
ILUSTRACIÓN 14. PRINCIPALES MAGNITUDES FINANCIERAS DEL GRUPO SECURITAS (SECURITAS, 2023)	53
ILUSTRACIÓN 15. PRINCIPALES MAGNITUDES FINANCIERAS DEL GRUPO PROSEGUR (PROSEGUR, 2023).....	54
ILUSTRACIÓN 16. GRÁFICO DE LAS 5 FUERZAS DE PORTER (PORTER, 1979)	55
ILUSTRACIÓN 17. ODS CON LOS QUE SE ALINEA EL TRABAJO (NACIONES UNIDAS, 2024c) ..	80

1 Introducción

1.1 Objetivos del proyecto

El proyecto pretende ofrecer una imagen representativa sobre las capacidades actuales en analítica de video y los avances que pueden implementarse gracias a las tecnologías más modernas, destacando los avances que vendrán de la mano de las ramas de la inteligencia artificial conocidas como “machine learning” y “generative AI”. Para conseguirlo la primera tarea es comprender a fondo la tecnología actual empleada en la industria de la analítica de video con especial hincapié en el uso de la denominada Inteligencia Artificial, en adelante IA, y posteriormente es necesario estudiar los márgenes de mejora en la implementación de esta tecnología, así como las limitaciones existentes. Este conocimiento se basará en un estudio del estado del arte de las tecnologías aplicables y se apoyará en la construcción de un entorno automatizado propio para conocerla de primera mano. Con este trabajo también se pretende abrir la puerta a dos estudios futuros, un análisis económico del ahorro que implicaría el procesado automático de imágenes y un análisis económico y comercial de industrias donde el procesado de imagen automatizado sería de gran utilidad.

El estado de la cuestión en procesado de imágenes de vídeo, dependiente de la tecnología y recursos disponibles, determina cómo de autónomo y efectivo es el proceso. Para conocer la realidad actual se ha trabajado en tres (3) sentidos, el primero, más teórico, consiste en un estudio de la literatura y el know-how disponible, a continuación, se han analizado los recursos que emplean las empresas más punteras en este sector y por último se ha comparado de primera mano las distintas técnicas y formatos de análisis de video. Todo esto nos da un entendimiento sobre lo que buscan las empresas y las limitaciones que tienen los sistemas actuales. Todo este proceso también nos identifica las posibilidades del empleo de IA en el procesado de imágenes de video y nos da una visión de lo que se podría lograr en el medio plazo en procesos e industrias que se pudiesen ver beneficiadas si se implantasen estas tecnologías.

Por otra parte, como toda tecnología joven, existen gran número de problemas que son especialmente visibles para el empleo de esta evolución tecnológica. Mediante el entorno automatizado propio se persigue comprender los errores habituales o los procesos que resultan especialmente tediosos en el empleo actual de video análisis para así identificar que procesos deben mejorarse y detectar donde sería más útil una mayor automatización.

Por último, el desarrollo de avances tecnológicos relevantes y, más aún, la consolidación de una tecnología como estándar en un sector suelen requerir de fuertes inversiones en el desarrollo de esta y en la transformación que hay que realizar de las operaciones del sector para utilizarla como estándar y de un claro interés empresarial por impulsar la transformación que esta tecnología supone. En este sentido se ha profundizado en las implicaciones económicas y sociales que aportaría el análisis automatizado de video a diversos sectores y, en especial, en el de la seguridad y vigilancia privada de instalaciones para comprender la capacidad e interés de este sector por impulsar, primero el desarrollo y consolidación de la tecnología y, posteriormente, la implantación de la misma.

1.2 Motivación

El proyecto surge de la importancia de poder implementar por completo la inferencia en el procesado de imágenes de video, puesto que considero que supondría un grandísimo avance y tendría un elevadísimo número de aplicaciones una vez la tecnología estuviese desarrollada y consolidada. Además, siendo la IA el tema más candente en la actualidad y considerando las grandísimas inversiones que las empresas están destinando a ella considero que el marco temporal no puede ser más favorable e interesante.

La construcción de un entorno automatizado permitirá observar y actuar sobre la que acontece en los espacios analizados e identificar a que otros campos se podría ampliar la tecnología, a la vez que comprender donde están las mayores complejidades a solventar para el desarrollo generalizado.

Por otro lado, hay un gran componente de la motivación derivada del ámbito más personal del alumno. Recientemente he experimentado de primera mano la utilización de

videograbadores de última generación equipados con un sistema de análisis de video que permite detectar en tiempo real posibles intrusiones como parte de un sistema de seguridad. Observando su funcionamiento me ha sorprendido como existen falsas alarmas que realmente no son lógicas y que demuestran el margen de desarrollo que existe. Algunas de estas falsas alarmas pueden darse por fenómenos muy comunes como lluvia, pájaros cerca de las cámaras o luces. Esto me ha llevado a querer profundizar en que tecnología se emplea en la actualidad, especialmente al considerar que hay multitud de oportunidades y de herramientas disponibles, como pueden ser los sistemas de conducción autónoma, que con ligeras adaptaciones podrían ser aplicadas a este campo.

A la vez, el enfoque de este trabajo tiene otro aspecto muy positivo en el ámbito educacional ya que me permitirá investigar temas de la más absoluta actualidad destacando la Inteligencia Artificial y la programación con Python. Estos temas no son abarcados en profundidad durante el grado de ingeniería industrial, pero a la vez será inevitable cruzarse con ellos durante la carrera de un ingeniero industrial especializado en la rama de organización industrial. Asimismo, lo veo un proyecto a largo plazo ya que al ser un tema tan amplio y en un momento álgido de desarrollo e innovación que resulta prácticamente inabarcable siempre podré continuar aprendiendo y desarrollando este trabajo.

Por último, como alumno del doble grado en Ingeniería de Tecnologías Industriales y ADE mi enfoque incluye no solo la tecnología en sí misma, sino también como esa tecnología impacta en la actividad empresarial y qué es necesario para que esa innovación tecnológica se traduzca en una transformación o evolución empresarial. De esta forma, valoro la opción de que este TFG desemboque en el TFG que tendré que completar para el grado de ADE. Puede evolucionar en un plan de negocio basado en la aplicación de IA en sistemas de video, en un análisis de los ahorros asociados a emplear un análisis de video automatizado e, incluso, en una posible idea de negocio en el que se venda esta tecnología a compañías.

1.3 Metodología

El proyecto se ha desarrollado con una metodología centrada en cuatro pilares: recopilación de información acerca del problema y de la tecnología actual que puede ser empleada,

análisis de esta información con el fin de ver su practicidad y aplicabilidad para abordar el problema a resolver, trabajo del alumno en el desarrollo del proyecto y fijación de los siguientes pasos y por último reuniones bimensuales alumno-director para iterar acerca de los avances realizados.

La investigación se ha realizado usando los distintos recursos de los que se dispone como son las bibliotecas de código de acceso público o los artículos científicos que analizan la aplicación de la inteligencia artificial en los ficheros de video. En esta fase también se ha estudiado que tecnologías usan las compañías líderes del mercado, como pueden ser los videograbadores de la marca comercial “*Paradox*”, especialmente con el fin de ver qué ventajas tienen respecto al resto de las tecnologías disponibles.

Tras filtrar la información recabada se ha priorizado el análisis de su aplicación al procesado de imágenes automático apoyado en inteligencia artificial. Este análisis también ha tenido en cuenta la complejidad computacional y el coste para evitar el uso de tecnologías excesivamente complejas que supongan “*matar moscas a cañonazos*”.

Todo este trabajo continuo se ha basado en un proceso de mejora continua para ir descartando las líneas de trabajo erróneas o menos eficientes y focalizarse en aquellos avances más importantes y sobre los que es más interesante incidir e investigar. Para ello, el alumno y el director se han reunido con una frecuencia mínima de dos reuniones mensuales en las que director y alumno entablábamos discusiones sobre el tema y se comentaban los resultados obtenidos.

2 Estado de la cuestión

2.1 Estado de la cuestión de la analítica de video

Las cámaras de videovigilancia son empleadas en prácticamente todos los lugares en los que los humanos desarrollan actividades y cumplen multitud de funciones como la detección de incidentes, el control del tráfico, la detección y el recuento de objetos, el monitoreo de líquidos o incluso para detectar fugas de gas. Estas cámaras habitualmente trabajan entre los 15 y los 30 FPS con algunas que llegan a los 60 FPS (Frames Per Second), es decir, las cámaras de menor calidad sacan 15 imágenes todos los segundos, en otras palabras, después de un día de grabación esta cámara ha generado 1,296,000 imágenes. Solo en España se estima que hay más de 900,000 cámaras de videovigilancia (Continoux, 2020), lo cual supone que en total se capturan alrededor de 22 millones de horas de grabación todos los días. Por otro lado, está demostrado que un humano cuando observa en vivo una pantalla durante más de 20 minutos puede llegar a no darse cuenta del 95% de lo que acontece en las imágenes que está vigilando (Green, 1999). Todo esto pone en relieve la problemática que acarrearán los sistemas de CCTV (Circuitos Cerrados de Televisión) donde el tamaño de los datos es prácticamente imposible de gestionar y, definitivamente, es imposible de abarcar sin cierta automatización. En este contexto entran en juego los sistemas de analítica de video que permiten automatizar el procesamiento de estas imágenes. (Kulbacki et al., 2023).

Los sistemas de grabación activos complementan las grabaciones con información que infieren del video, en esto consisten la analítica de video, en tener un programa capaz de detectar sucesos en un video. En la actualidad, los sistemas de análisis de video inteligente más punteros no solo son capaces de detectar anomalías en el video, sino que pueden clasificarlas, detectar cuando hay humanos presentes en el video e incluso determinar si suponen una amenaza estas acciones.

En la actualidad los sistemas IVA (“Intelligent Video Analytics”) más avanzados en videovigilancia siguen los siguientes pasos en su funcionamiento:

1. Captación de la imagen.

2. Pre-procesado: se estabiliza la imagen y se aplican filtros que resaltan ciertas características
3. Extracción de las características principales: se extraen las características más relevantes de los datos de video
4. Detección de objetos: el programa detecta los objetos que aparecen en el video.
5. Seguimiento de los objetos: se hace un seguimiento de estos objetos en el tiempo
6. Reconocimiento de humanos (HAR – Human Action Recognition): el programa evalúa si hay movimiento humano en el video
7. Detección de eventos: se detectan los distintos eventos que involucran actividad humana
8. Toma de decisión: el programa analiza los eventos y decide si se debe generar una alerta en el caso de determinar la existencia de eventos no comunes
9. Emisión de una alarma: el programa genera una alarma que se envía al usuario

Los primeros cuatro pasos ya no suponen un problema y son desarrollados eficazmente por virtualmente todos los programas de análisis de video. El primer gran obstáculo se encuentra en el seguimiento de los objetos detectados a través de la red de cámaras interconectadas, los programas actuales no son capaces de hacer un seguimiento de un objeto efectivo si el objeto está en un espacio cerrado o si la imagen del objeto es obstruida y desaparece en algún instante. Se prevé que los principales avances en detección de objetos sean capaces de interpretar las imágenes en 3 dimensiones y con ello entender cuando los objetos desaparecen y ser capaces de encontrar patrones en el seguimiento de estos objetos. (Kulbacki et al., 2023)

La otra gran tarea pendiente de la video analítica es en cuestión de detección de humanos, esta difícil tarea es vital para que los sistemas de video analítica sean realmente efectivos ya que la mayoría de las aplicaciones de la analítica de video son únicamente posibles con una detección de humanos efectiva. Actualmente existen cuatro formas de afrontar las tareas de reconocimiento de humanos (HAR methods – “Human Action Recognition” methods): pose-based methods, tracking-based methods, spatio-temporal-based methods y, por último, deep learning-based methods.

El **pose-based method**, método de detección de poses, detecta la presencia de los seres humanos gracias a ser capaz de reconocer la fisionomía del cuerpo. En particular es capaz de identificar la presencia de articulaciones humanas en imágenes y videos. La detección de poses analiza las características más diferenciales de los seres humanos entre las que destaca la posición esquelética erguida y la posición de las diferentes articulaciones, así como la variación angular que estas permiten. También analiza los contornos y volúmenes de las figuras presentes en el video. Este método está limitado por la necesidad de requerir múltiples vistas de cámara distintas para una triangulación precisa, también es muy inefectiva cuando existe una oclusión parcial de la imagen. (Li et al, 2019)

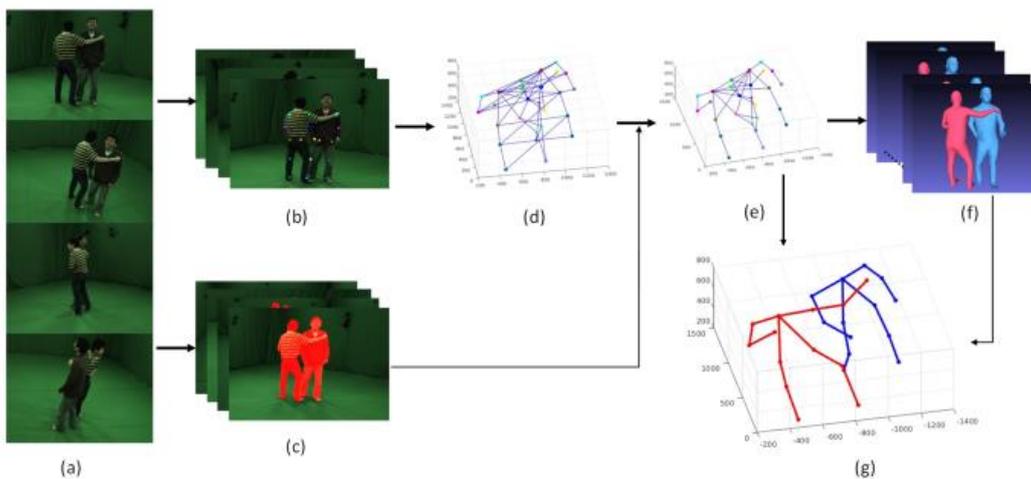


ILUSTRACIÓN 1. RECONOCIMIENTO DE LA FISIONOMÍA HUMANA EMPLEANDO EL POSE-BASED METHOD (LI ET AL., 2019)

El **tracking-based method** realiza un seguimiento del movimiento y detecta las características de movimiento de los seres humanos, para ello analiza la velocidad de los movimientos, la cadencia, las trayectorias e incluso la aceleración y compara estas con los movimientos habituales de los seres humanos, siendo incluso capaz de categorizar los movimientos. (Kulbacki et al., 2023)

El **spatio-temporal method** es una combinación de los dos anteriores, inicialmente identifica la potencial presencia de seres humanos en la imagen basándose en características espaciales como la postura y posteriormente realiza un seguimiento de las figuras que

potencialmente son humanos y estudia sus movimientos en el tiempo, de nuevo analizando velocidad, trayectoria de los movimientos, etc. Este método es más robusto que los anteriores y es menos sensible a oclusiones de la imagen gracias a que se realiza un análisis combinado.

Por último, diferenciamos los métodos basados en “deep learning” que funcionan gracias a redes neuronales capaces de detectar la presencia humana, gracias a que han sido entrenadas con grandes bancos de imágenes que en ocasiones incluyen presencia humana, estas imágenes además son modificadas de distintas formas (recortadas, giradas, etc.) para conseguir mayor diversidad en el entrenamiento y que la red neuronal haya sido expuesta a imágenes mucho más variadas entre sí.

Aun existiendo en la actualidad sistemas de análisis de video muy efectivos, el tamaño excesivo de los datos sigue suponiendo un grandísimo problema. Uno de los frentes para reducir este problema y dinamizar la tarea de procesado de video es la computación frontera, esta consiste en acercar la computación y el almacenamiento de los datos a la generación de esos datos, con ellos se consigue reducir la latencia y ahorrar ancho de banda. La computación frontera (Edge Computing) hace posible la analítica en tiempo real y también posibilita la implementación de video análisis en cascada, en el cual el primer análisis es realizado por las propias cámaras, posteriormente aquellas partes del video que la cámara identifique como potencialmente interesantes son las únicas que son transferidas a un ordenador remoto para realizar una analítica más precisa que determine si el suceso realmente es de interés y si debe de emitirse una alarma. (Zhang, Sun, Wu, & Zhong, 2019)

2.2 Estado de la cuestión del Machine Learning

El **Machine Learning** es una rama de la Inteligencia Artificial enfocada en conseguir entrenar ordenadores/máquinas para que estos sean capaces de llevar a cabo acciones de forma exitosa sin haber recibido instrucciones explícitas para ello. Desde un prisma sencillo, el objetivo es que las máquinas aprendan del mismo modo que lo hacen los humanos, es decir a base de consumir información y de prueba y error.

El termino Machine Learning se remonta a 1950 y fue asociado a un programa que desarrollo Arthur Samuel que aprendió de forma autónoma a jugar a las damas, siendo esto el comienzo del aprendizaje de las maquinas. Durante los siguientes años, y hasta finales de 1980, hubo pocos avances en este campo. Es ya en 1990 cuando se inventan métodos capaces de realizar de forma efectiva tareas de clasificación empleando redes neuronales. En la primera década del siglo XXI se comenzaron a desarrollar sistemas mucho más capaces y se introdujo por primera vez el concepto de “Deep learning”. Por último, en los últimos años los avances han demostrado el poder de las redes neuronales y del aprendizaje de máquinas reforzado, llevando a maquinas a ser claramente superiores a los seres humanos, un buen ejemplo de esto fue AlphaGo, que batió al mejor jugador del mundo de Go. (Al-Sahaf et al. 2021)

La programación tradicional descrita de la forma más sencilla posible es un conjunto de reglas que el programador determina y que son expresadas en un lenguaje de programación, estas reglas actúan en base a unos datos que se introducen y el resultado son las respuestas. Por ello, existe una clara limitación, solo puede automatizar procesos de los que se puedan extraer unas reglas que los rigen. Aplicando esta definición tan sencilla también es posible comprender la base detrás del Machine Learning. En este caso, no queremos encontrar las respuestas o el resultado, queremos encontrar la conexión entre los datos y estas respuestas, es decir, nuestro objetivo será determinar las reglas. El siguiente diagrama refleja la gran diferencia conceptual al más alto nivel entre la programación tradicional y Machine Learning. (Moroney, 2020)

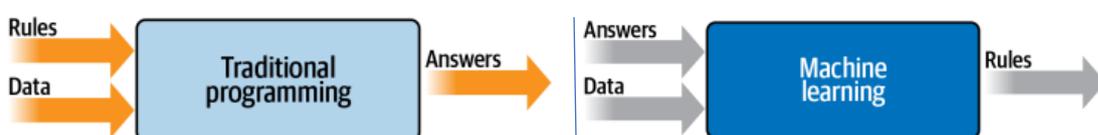


ILUSTRACIÓN 2. EVOLUCIÓN DE LA PROGRAMACIÓN TRADICIONAL AL MACHINE LEARNING
(MORONEY, 2020)

Para que todo esto sea posible existen tres pilares en Machine Learning.

El primer pilar son **los datos**, es decir, la información que consume la maquina y a partir de la cual desarrolla patrones y es capaz de aprender cómo debe actuar. La efectividad del proceso de aprendizaje está directamente ligado a la cantidad de datos y sobre todo a la calidad de estos, esta calidad se mide en base a como de amplio es el espectro de posibles situaciones que los datos cubren.

El segundo pilar son **los algoritmos**, que se definen como las reglas e instrucciones que permiten distinguir los inputs detectando que es lo que distinguen a unos datos de otros y con estos algoritmos se crea tercer pilar, **el modelo**.

Este es el último pilar, coloquialmente se dice que este ha de ser entrenado, esto significa crear un aprendizaje a base de inferencia que permita que el modelo en ultima estancia sea capaz de establecer las reglas que conectan datos y resultados. Una vez el modelo es capaz de esto, se consigue que introduciendo nuevos datos este modelo pueda predecir el resultado.

Los modelos habitualmente empleados son denominados Redes Neuronales, esto se debe a que su estructura y funcionamiento llega a asemejarse a aquellos de las neuronas biológicas. Las redes neuronales se definen por dos características principales, el número de capas de la red neuronal y el número de neuronas, también llamadas nodos, que tiene cada capa. Dentro de esta configuración existe una capa de entrada y una capa de salida y entre medias existen las denominadas capas ocultas. El número de nodos que tiene la capa de entrada generalmente será el número de posibles inputs y el número de nodos de la salida será igual a los posibles resultados.

Un ejemplo sencillo para entender una red neuronal asociada al análisis de imágenes sería el análisis de una imagen de 28x28 pixels en escala de grises que presenta un dígito del 0 al 9.

En este análisis tendríamos 784 (28x28) nodos en la capa de entrada y cada uno de ellos tomaría un valor entre 0 y 1 y representaría uno de los pixeles en la imagen. Por ejemplo, si el píxel que el nodo de entrada representa es completamente negro tomaría el valor 0 y si este píxel es completamente blanco tomaría el valor de 1.

Por otro lado, en la capa de salida tendríamos 10 nodos, cada uno representando una de las 10 posibles salidas. Estos nodos, después de aplicarse las conexiones de la red neuronal adquirirían un valor entre 0 y 1 y aquel que presente el valor más alto es el que la maquina predice que será.

Entre medias existen las denominadas capas ocultas que conectan la entrada con la salida y son las que realizan el aprendizaje para conectar unos determinados valores de entrada en la red con unos valores en la salida. En estas capas intermedias se realizan asociaciones que permiten dar peso a las capas de salida, por ejemplo en esta situación una de las neuronas de una capa intermedia puede basar su valor del 0 al 1 en sí existen formas redondeadas cerradas en la parte superior de la imagen, cuanto más seguro este de que las haya más cercano será su valor a 1 y esta neurona dará peso en la salida al 8 y al 9 que son los dos dígitos que presentan esta característica, otra asociación similar que haría otra de las neuronas de la capa intermedia es medir las líneas rectas que hay en la imagen y sobre ponderaría valores como el 1, el 4 y el 7 que son los dígitos que presentan líneas rectas. (Nielsen, 2019)

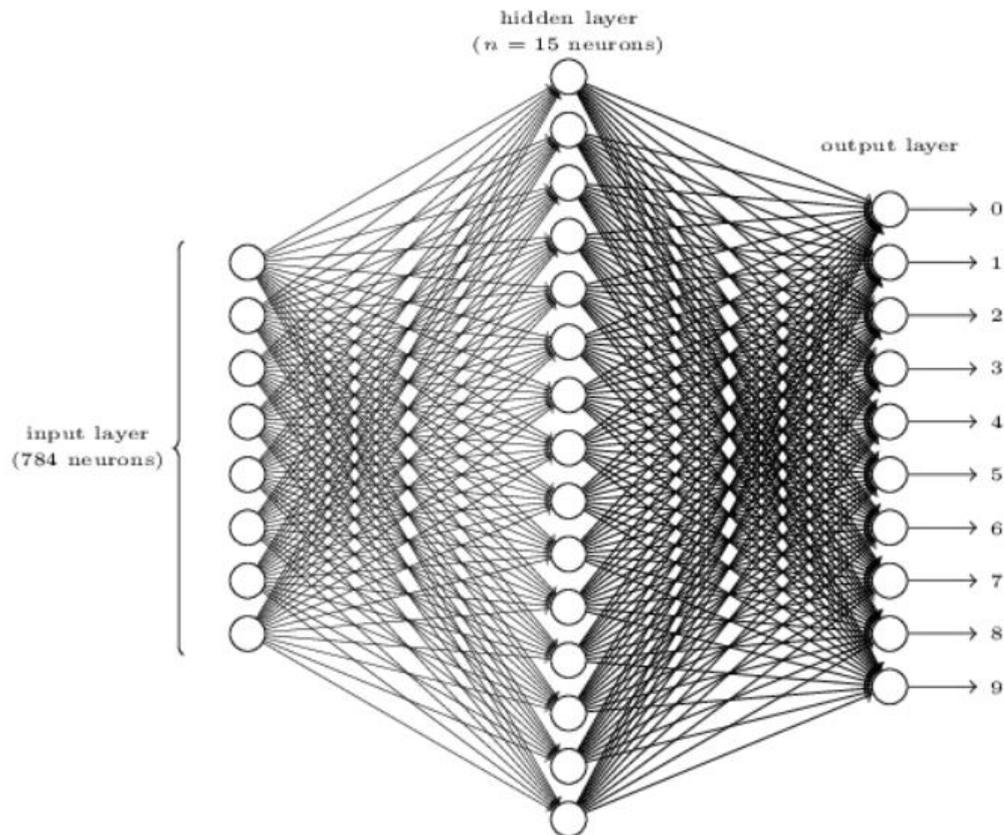


ILUSTRACIÓN 3. ESTRUCTURA GENERAL DE UNA RED NEURONAL (NIELSEN, 2019)

En estas capas intermedias es donde se realiza el aprendizaje, a medida que el modelo va siendo entrenado se realizan conexiones entre las neuronas de las distintas capas y se llega a conexiones como las previamente mencionadas en las que una neurona cuyo valor es mayor si hay líneas rectas estará conectada a las neuronas asociadas a los dígitos con líneas rectas.

Durante el proceso de entramiento inicialmente el modelo no tiene ninguna información acerca de la relación que existe entre estos valores y por ello comienza creando conexiones aleatorias. El resultado de estas conexiones se comprueba contra el resultado real asociado a esos datos y se crea lo denominado función de pérdida, sencillamente una función que mide la discrepancia entre la predicción del modelo para unos datos determinados y el valor real. Con esta función de pérdida es con lo que el optimizador del modelo trabaja para

conseguir que la siguiente función de pérdida tenga un valor menor, es decir que el modelo prediga valores más cercanos a la realidad.

Hay muchos tipos de optimizadores y se usan unos u otros en función del objetivo, los más comunes son los optimizadores SGD (Stochastic Gradient Descent) que buscan minimizar la función de pérdida por medio de su derivada. Este optimizador es sencillo y su coste computacional es muy bajo ya que no requiere ir almacenando los valores durante todo el proceso de iteración, el único input que necesita es la función de pérdida anterior.

El último concepto importante del proceso de aprendizaje de las redes neuronales son las determinadas épocas, estas se pueden definir como el número de vueltas que el modelo da sobre unos mismos datos para aprender, es decir, cuanto queremos que el modelo entrene con los datos que le hemos proporcionado. Si tenemos un set de 10 datos y queremos entrenar el modelo con 30 épocas, en total el modelo correría 300 veces, 30 para cada uno de los datos. Un modelo funcionando correctamente debe ir reduciendo la función de pérdida época tras época, por lo que cada vez, de media para el global de los 10 datos, el resultado será más cercano al real.

2.3 Limitaciones actuales al empleo generalizado de IA en análisis y procesado de video

La IA sin duda es la protagonista en el futuro del análisis de video, pero en la actualidad su uso generalizado para las tareas de procesado de imágenes y análisis de video es limitado y hay ciertos cuellos de botella que deben de solucionarse para desbloquear el total potencial de esta tecnología aplicada a la industria del análisis de video. Las principales limitaciones vienen de la mano del alto número de datos de entrenamiento necesarios para ser capaces de entrenar las soluciones de análisis de video y en concreto de datos de entrenamiento correctamente etiquetado para que la máquina asocie correctamente las imágenes con las que está siendo entrenado con un “output” concreto. También existe otra limitación asociada a la descentralización del proceso y la colaboración limitada entre las cámaras que conforman un sistema de video análisis, un análisis centralizado que evaluase de forma holística la

información recogida por diversas cámaras sería mucho más efectivo y tendría una información contextualizada que permitiría una toma de decisiones robusta.

2.3.1 Datos de Entrenamiento Limitados

La base más elemental del Machine Learning son los datos utilizados en el proceso de entrenamiento. Estos datos determinan la exposición del sistema, por ello el ideal es que los datos con los que se alimenta el entrenamiento sean variados y no solo incluyan situaciones que se consideran “normales” sino que el conjunto de datos de entrenamiento incluya situaciones que se considerarían anómalas e imprevistas, de forma que el sistema sea capaz de adaptarse a estas situaciones y detectarlas.

La exposición durante el entrenamiento determina la capacidad que tendrá la IA para reconocer patrones que no se alineen con la “normalidad” definida durante su entrenamiento. Además, en un entorno dinámico, donde los comportamientos y eventos anómalos cambian con el tiempo, el entrenamiento predefinido supone una gran limitación ya que la IA no sería capaz de adaptarse a nuevas situaciones que previamente no hayan sido consideradas por los datos de entrenamiento y por tanto su eficacia en la detección se reduce drásticamente. Por ello, se persigue que la exposición no se limite, es decir que continúe en el tiempo y por ello se persigue implementar un entrenamiento circular en el que el aprendizaje sea continuo, es decir, la IA continúa aprendiendo del video que está analizando e inmediatamente comienza a aplicar este nuevo aprendizaje.

Además, para que el entrenamiento sea exitoso es vital el correcto etiquetado de los datos con los que se realiza el entrenamiento. La precisión del etiquetado determina de forma directa la precisión del análisis, cuando los datos están mal etiquetados el rendimiento del modelo de IA se ve directamente comprometido y sus decisiones serán ineficaces ya que están siendo construidas con una premisa falsa. (Ghozia et al., 2020). Actualmente este etiquetado se realiza de forma manual, es decir, es una tarea humana la de filtrar los videos con los que se entrena la IA y determinar aquellos instantes con momentos “anómalos”, por tanto, esto supone que en la actualidad la disposición de datos óptimos es uno de los

principales cuellos de botella de cara a ser capaces de analizar video a través de IA. (Hansen et al., 2021).

En la actualidad se está trabajando con el fin de solucionar este gran cuello de botella creando grandes bases de datos compuestas de videos cortos (en el entorno de los 3 segundos) con contenido de todo tipo (incluyendo personas, animales, objetos e incluso fenómenos naturales como lluvia). Estos videos son etiquetados y clasificados a través de diversas dimensiones, por ejemplo, una de las principales bases de datos de este tipo que ya están disponibles para entrenar modelos es “Moments in Time Dataset” que categoriza más de 1 millón de videos a través de 339 clases distintas (Monfort et al., 2019).

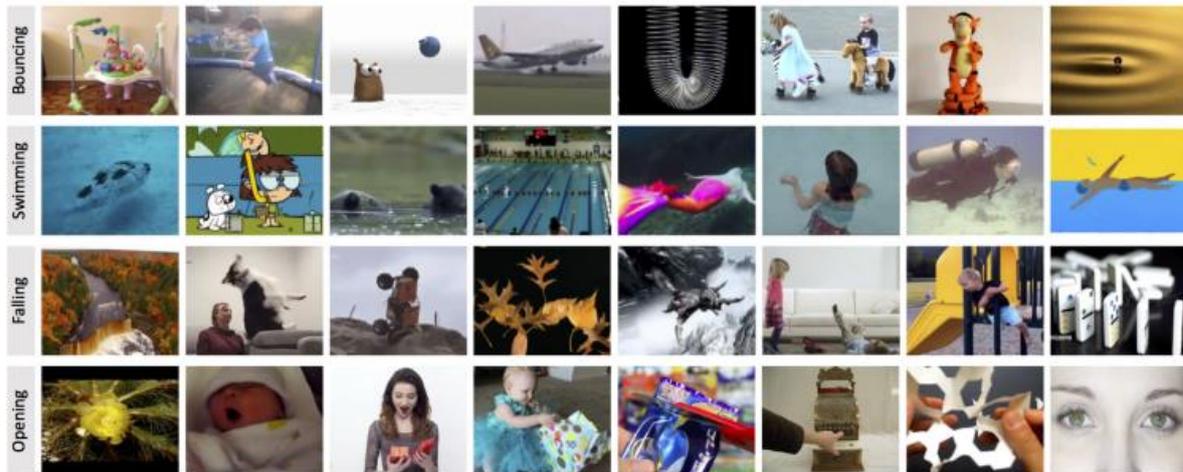


ILUSTRACIÓN 4. SELECCIÓN DE VIDEOS DE LA BASE DE DATOS “MOMENTS IN TIME DATASET”
(MONFORT ET AL., 2019)

Además, para maximizar el uso posible de estas bases de datos y que el modelo reconozca las situaciones en distintos entornos, los videos base son alterados por medio de distintas técnicas como por ejemplo haciendo zoom, recortando partes de la imagen, cambiando los colores o cambiando el ángulo, de este modo los modelos no solo se entrenan con las imágenes base, sino que se incrementa exponencialmente la cantidad de datos con los que se alimentan debido al gran número de posibles alteraciones. Esto además tiene un sentido muy lógico y es que las cámaras no siempre captaran un ángulo perfecto de los sucesos o los sucesos en su totalidad por tanto hay que entrenar los modelos con videos que no sean

perfectos (ej. Videos que no muestran la totalidad de una persona, sino que recortan la persona por la mitad).

2.3.2 Proceso descentralizado y colaboración limitada entre cámaras

En los sistemas de videovigilancia de grandes instalaciones y otras aplicaciones que requieren múltiples perspectivas, como puede ser el monitoreo del tráfico, la colaboración entre cámaras es la base para un análisis coherente. Si las cámaras trabajan de forma aislada el análisis no es preciso y no se realiza un análisis similar a aquel que realizaría un ser humano que tiene visión sobre los distintos ángulos de las distintas cámaras. La robustez y efectividad del análisis de video automatizado pasa por conseguir que el análisis sea global y valore en conjunto todo lo observado, en vez de que el análisis sea individual para cada imagen ya que esto limita claramente el resultado. (Yi et al., 2021)

2.4 Empleo de IA en procesado de video actualmente

El uso de Inteligencia Artificial para la automatización del análisis de video es ya una realidad y existen multitud de programas que han sido entrenados para ser capaces de identificar eventos extraños en situaciones específicas o identificar de forma específica objetos o personas. El principal campo de aplicación en la actualidad es en seguridad, concretamente en sistemas de videovigilancia, siendo capaces de detectar amenazas potenciales, reduciendo la dependencia de intervención humana y mejorando la seguridad general, especialmente en entornos difíciles de abarcar como aeropuertos o centros urbanos.

Actualmente el avance de la aplicación de IA en procesado de video está limitado por la capacidad computacional y de almacenamiento requerido, por ello, los principales avances que están siendo aplicados en la práctica son una combinación de las tecnologías previas y IA. La idea consiste en hacer un primer análisis utilizando métodos computacionales automatizados que analicen la imagen y evalúen cuando en esta suceden fenómenos, sin ser capaces de determinar qué clase de fenómenos son, simplemente evaluarán un frame contra el siguiente y determinarán si ha habido variaciones significativas entre los frames, lo cual sería indicativo de que ha habido algún fenómeno a estudiar en la imagen. A continuación,

aislaran únicamente lo que es nuevo y ha de ser estudiado e ignoraran todo aquello que sea el fondo del video y que no es de interés. Con esto se consigue que únicamente se haga un análisis más profundo de aquellos segmentos del video de interés, es decir aquellas partes donde ha habido cambios significativos en el video (ej. Aparición de un vehículo en la imagen). Posteriormente se aplicaría el análisis más profundo a través de IA únicamente en las áreas donde se determina que puede haber ocurrido un suceso a estudiar para determinar las características de este suceso y si se debe generar una alarma. Esta metodología que utiliza un filtro antes de realizar el análisis con IA ha demostrado ser de gran eficiencia y en pruebas preliminares se ha observado un ahorro de hasta un 80% en capacidad de almacenaje y capacidad computacional. (Nguyen et al., 2020)



a) Imagen estática de la grabación



b) Aparición de un vehículo en la imagen



c) Selección de frames donde se detecta movimiento y que serán estudiados en profundidad

ILUSTRACIÓN 5. SELECCIÓN DE FRAMES A ESTUDIAR OBLVIANDO EL FONDO DE LA IMAGEN

(NGUYEN ET AL., 2020)

2.5 Vinculación con el desarrollo y consolidación del coche autónomo

El desarrollo y consolidación de una tecnología disruptiva, como sería la aplicación generalizada de análisis de video automatizado, como estándar en una industria de consumo masivo requiere fundamentalmente de dos elementos:

- De la existencia de una tecnología disruptiva suficientemente desarrollada y consolidada que ofrece ventajas relevantes, en términos de valor añadido para el cliente y/o de eficiencia en costes, como para modificar los procesos de trabajo en una industria y sustituir los anteriores; en el caso que nos ocupa, la sustitución de visionado, análisis y actuación manual humana por uno modelo automatizado.
- De la capacidad y rentabilidad en términos de ROI (“Return on Investment”) de realizar las inversiones necesarias para disponer de la tecnología y para modificar los procesos, aspecto que requiere igualmente una elevada inversión en tiempo, recursos humanos e inversión económica.

En línea con el análisis que se realizara más adelante en el apartado 5.2 sobre “Sectores en las que la aplicación de análisis de video automatizado aporta un valor añadido muy relevante” y las propias conclusiones finales del apartado 6, los principales sectores privados en los que se utilizaría el análisis de video automatizado son la vigilancia y seguridad privada de instalaciones y, en segunda instancia y en menor medida, la utilización de estos sistemas para el seguimiento y supervisión de operaciones en “retail” y servicios de consumo. Ambos sectores son muy relevantes, pero, en cualquier caso, limitados en tamaño en comparación con otros sectores por lo que la inversión y la transformación completa que esto requiere hace que, con el estado actual de la tecnología comentado anteriormente, no este todavía al nivel necesario para un avance rápido dadas las cuantiosas inversiones requeridas.

En este sentido, la experiencia dice que salvo en temas de defensa, aeroespacial y sectores muy concretos, las organizaciones públicas, en cuyo marco se incluiría todos los aspectos relacionados con la supervisión y seguridad pública, no son las que acometen transformaciones disruptivas, sino que, una vez asentadas en el sector privado, se van poco a poco utilizando en el sector público. En definitiva, la iniciativa e inversión necesaria es previsible que, en el caso de la video vigilancia, tenga que venir del sector privado cuyas empresas tienen un tamaño y una capacidad limitada, como analizaremos posteriormente.

El análisis automatizado de video tiene una relación muy clara en términos tecnológicos con el desarrollo del coche autónomo puesto que es una de las partes de este. El coche autónomo

en definitiva sustituye a un humano en el análisis del contexto de la conducción, tanto en relación con el coche propio como con el entorno, y la toma y ejecuta una decisión al respecto de que debe de hacerse en la conducción del coche propio. La tecnología necesaria para la consolidación del coche autónomo es evidentemente más compleja que el análisis de video automatizado presentando las siguientes diferencias más relevantes:

- Los sensores y cámaras del coche están en movimiento y a gran velocidad.
- La decisión y actuación en caso de un evento extraordinario tienen que ser inmediata y puede tener gravísimas consecuencias en términos de pérdidas materiales y humanas y, en cualquier caso, hay un proceso de decisión continuo. Hay adicionalmente aquí, un aspecto regulatorio muy relevante.
- La decisión es mucho más compleja al poder intervenir varios objetos a la vez (incluyendo personas) y el coche propio en la previsión de cuál sería la mejor solución.
- A diferencia de los sistemas de video análisis actuales, los coches autónomos disponen de una capacidad local de computación muy relevante que les permite poder realizar una parte, o todo, el análisis y la decisión de forma local que acelera el proceso y simplifica la necesidad de computación en la nube y necesidad de comunicación con una latencia inmensa.

Independientemente de la simplicidad y brevedad de las reflexiones anteriores, es evidente que la complejidad del desarrollo y consolidación de tecnologías de conducción autónoma es mucho más complejo y sofisticado que las de análisis de video automatizado y que el avance en el desarrollo de metodologías para el coche autónomo, en gran parte, podría servir para avanzar en el análisis de video automatizado.

En este sentido, el coche autónomo es uno de los temas más candentes en el desarrollo tecnológico industrial de modo masivo en los últimos años y ha puesto de manifiesto que, aunque los modelos teóricos y las soluciones existentes funcionan y están operando actualmente con un alcance moderadamente limitado, su aplicación generalizada fuera de

entorno relativamente controlados se está demorando mucho más de lo esperado. Sirva como soporte de esto, que en España, por ejemplo, solo está legalmente autorizada la conducción autónoma si hay un conductor supervisándola y solo en autopistas o en aparcamientos. Sirva también como ejemplo, la secuencia de retrasos de TESLA en la disponibilidad del coche con “autopilot” completo, el denominado Nivel 5 SAE que supone la condición autónoma total (TESLA lo está entrenando como una red neuronal), que Elon Musk, fundador de TESLA, fijó inicialmente para 2018 y que actualmente, agosto de 2024, sigue siendo una versión Beta disponible básicamente para USA y requiriendo supervisión humana (Lee, 2021). En relación con esta última apreciación, hay que destacar también que la actual versión del autopilot de TESLA tiene un índice de accidentes que es, según el informe de seguridad de TESLA de mayo de 2024, unas 11 veces inferior en un Tesla con el Autopilot activado y unas 1.4 veces inferior sin el Autopilot activado con respecto a la media de un coche con conducción estándar en Estados Unidos (Tesla Inc., 2024)

- Tesla vehicles using Autopilot technology
- Tesla vehicles not using Autopilot technology
- United States average

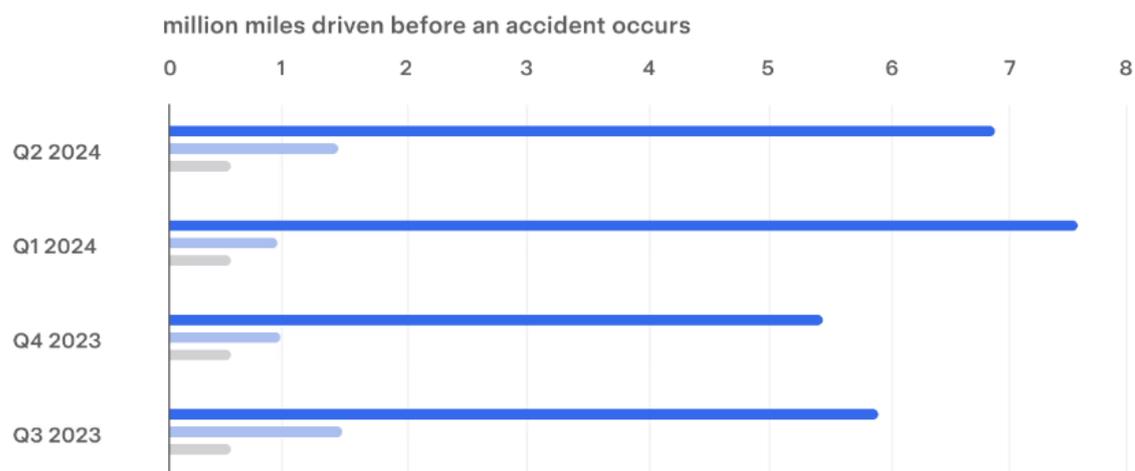


ILUSTRACIÓN 6. COMPARATIVA DE SEGURIDAD ENTRE VEHÍCULOS TESLA Y EL MERCADO AMERICANO GENERAL (TESLA INC., 2024)

En relación con la reflexión anterior, el elemento diferencial del coche autónomo respecto del análisis de video automatizado, es que su consecución de forma generalizada, dado el volumen económico que se mueve, tendría un retorno gigantesco en términos económicos, de ahorro de costes por no requerir conductor y poder operar 24x7, y en términos sociales, reducción de accidentes, y , además, que la capacidad de inversión de las grandes empresas automovilísticas es inmensa comparada con la capacidad de las empresas de seguridad. Sirva de referencia en cuanto a la capacidad de inversión que el valor de mercado del grupo SECURITAS AB, una de la mayores empresas mundiales de vigilancia y seguridad y cuya filial es líder en España, está actualmente en el entorno de 6.000 millones de euros frente a TESLA cuyo valor de mercado es actualmente cercano a los 650.000 millones de euros, más de 100 veces más, y junto a TESLA, Daimler, Volkswagen-Audi, BMW, General Motors, KIA, Hyundai, Toyota, Tata, ... están también invirtiendo grandes cantidades en el desarrollo de los modelos de conducción autónoma. Volveremos posteriormente a la influencia de la capacidad de inversión para el desarrollo y consolidación del análisis automatizado de video.

3 Análisis de recursos

Una solución tecnológica exitosa comienza en la elección de las herramientas y recursos con los que esta se construye. Esta elección requiere de estrategia, no se trata de elegir los mejores recursos para cada tarea, el verdadero reto es conseguir un balance y conseguir que la elección de recursos este hecha a medida de la solución que se pretende conseguir. Esta elección debe hacerse valorando múltiples dimensiones entre las que se encuentra el rendimiento de la herramienta, la facilidad de uso, el coste computacional y económico, la flexibilidad que ofrece y el soporte.

La librería o biblioteca de código es el primer recurso que debe seleccionarse de cara al desarrollo de un programa de software como el que se plantea en este trabajo y que permite identificar las limitaciones del procesado de video. Estas bibliotecas recogen código, funciones y recursos predefinidos para que cualquier persona los utilice y distribuya. Para este trabajo se evaluó cual era la biblioteca idónea para la tarea a realizar además de considerarse la compatibilidad con el lenguaje de programación Python que es el que se ha escogido para el desarrollo del programa debido a la simplicidad y legibilidad del código gracias a una sintaxis simple y clara, las extensas bibliotecas estándar que el propio lenguaje incluye, como por ejemplo NumPy que proporciona la infraestructura necesaria para la computación numérica en Python, y también por ser uno de los lenguajes que cuenta con una comunidad de desarrolladores más grande y activa en el mundo de la programación, lo cual se traduce en abundante documentación y soporte.

De cara a este trabajo se evaluaron las siguientes 3 bibliotecas graficas: SimpleCV, Mahotas y OpenCV

- **Simple CV**: biblioteca de código abierto para la visión por computadora, enfocada en simplificar el proceso de desarrollo de aplicaciones que requieren análisis de videos. Supone una simplificación de la biblioteca OpenCV y permite realizar tareas complejas de procesamiento de video con pocas líneas de código. (SimpleCv, 2023)
 - o Ventajas:

- Facilidad de uso y sintaxis simple: la biblioteca está diseñada para usuarios que buscan una curva de aprendizaje suave y que les permita una implementación rápida de soluciones de visión por computadora.
- Rapidez: debido a su simplicidad SimpleCV es ideal para construir prototipos de aplicaciones de análisis de video de forma rápida y no requiere dedicarle mucho tiempo a la configuración o depuración del código.
- Integración: SimpleCV está diseñada para ser muy versátil y trabajar en conjunto con la gran mayoría de las herramientas del ecosistema Python, como NumPy, lo cual amplía la funcionalidad y permite análisis más avanzados.
- Desventajas:
 - Limitaciones en funcionalidades avanzadas: su simplicidad implica que carece de muchas funcionalidades necesarias para aplicaciones de análisis de video más complejas como el procesamiento de video en tiempo real o la incorporación de algoritmos de inteligencia artificial al proceso de imágenes.
 - Comunidad y soporte limitado: la comunidad de usuarios y desarrolladores que utilizan SimpleCV como una herramienta profesional es limitada y esto dificulta la búsqueda de soporte y de soluciones a problemas específicos.
 - Rendimiento inferior: al comparar SimpleCV con otras bibliotecas más avanzadas se observa que es menos eficiente en términos de velocidad de procesamiento.
- **Mahotas:** biblioteca de Python especializada en procesamiento de imágenes, escrita principalmente en C++ para garantizar un rendimiento rápido, principalmente utilizada para el análisis de imágenes estáticas siendo especialmente útil para la segmentación y detección de imágenes.
 - Ventajas:

- Rendimiento: Mahotas, gracias a su implementación en C++ se caracteriza por ser rápida en operaciones específicas de procesamiento de imágenes.
- Fácil integración: utiliza una API sencilla y muy intuitiva lo cual facilita su uso en proyectos sencillos de análisis de imágenes.
- Desventajas:
 - Especialización: muy enfocada en tareas concretas de análisis de imágenes como la segmentación y las transformaciones morfológicas y no ofrece tantas funciones como OpenCV en áreas específicas como el procesamiento de video.
 - Menor popularidad: en comparación con OpenCV tiene una popularidad mucho menor lo cual dificulta la búsqueda de recursos y soporte y supone que se actualiza con menor frecuencia en comparación con OpenCV.
- **OpenCV**: es la principal biblioteca de software de código abierto especializada en el procesamiento de imágenes y visión por computadora. Escrita en C++ con interfaces para Python y otros lenguajes de programación destaca por la amplia gama de funciones que ofrece para la captura, procesamiento y análisis de imágenes y videos (OpenCV, 2024).
 - Ventajas:
 - Amplitud: ofrece una gama extremadamente amplia de funciones que abarcan todo el espectro del procesamiento de imágenes y videos, desde tareas simples de filtrado de imágenes hasta algoritmos muy complejos para determinar la visión de un robot.
 - Popularidad: la comunidad de desarrolladores que utilizan OpenCV a diario es extremadamente amplia y esto democratiza el acceso a tutoriales, documentación y soporte en línea.
 - Interoperabilidad: OpenCV es plenamente compatible con otras bibliotecas comúnmente utilizadas en paralelo como puede ser

TensorFlow o Keras y también es compatible con un gran número de lenguajes de programación.

- Desventajas:
 - Curva de aprendizaje: OpenCV es una biblioteca más extensa y profunda lo cual la hace más compleja de aprender y utilizar.
 - Rendimiento: el rendimiento de la biblioteca es más limitado y algunas operaciones tienden a ser menos eficientes en comparación con Mahotas.
 - Sobrecarga de funciones: para ciertos proyectos pequeños o muy específicos OpenCV puede ser excesiva y presentar un gran número de funciones que no serían utilizadas.

Evaluando las tres principales bibliotecas de código se considera que, para la finalidad de este modelo, que realmente es comprender cuáles son las limitaciones en procesamiento de video y por donde se deberá avanzar en el futuro, lo idóneo es emplear **OpenCV** como biblioteca de código ya que será esta la que se utilizaría en un entorno profesional y también es la biblioteca que ofrece el mayor número de funciones.

Por último, se escoge como editor de código el editor Visual Studio Code, desarrollado por Microsoft y que se ha convertido en una de las herramientas más populares entre los desarrolladores. VS Code se caracteriza por ser un editor ligero y versátil además de ser compatible con las principales plataformas (Windows, macOS y Linux). Además, aun siendo un programa gratuito tiene soporte constante de Microsoft por medio de actualizaciones y mejoras recurrentes. También es destacable la interfaz de VS Code que es completamente personalizable e intuitiva pudiéndola adaptar a las necesidades en función del software que se esté desarrollando.

4 Desarrollo de un entorno propio

4.1 Finalidad del Modelo

El proyecto está apoyado en el desarrollo de un entorno automatizado capaz de analizar video desarrollado íntegramente por el alumno. Este entorno automatizado está construido con el editor de código Visual Studio Code, el lenguaje Python y la biblioteca de código OpenCV.

Con el entorno automatizado de análisis de video se pretende conseguir una visión práctica de cómo funcionan estos modelos, tener una herramienta para comparar las limitaciones y experimentar con distintas formas de enfocar el análisis de video. Este entorno automatizado ha permitido realizar experimentos comparando análisis de video en color vs en escala de grises, estudiar la variación de coste computacional en función de los fps o de los pixels de la imagen.

Por otro lado, el trabajo práctico ha servido como aprendizaje para el alumno en el uso de las distintas tecnologías que se emplean en la industria del procesado de imágenes. Actualmente se diferencian dos enfoques distintos a la analítica de video, que muchas veces son complementados para conseguir la mayor efectividad posible. Existe una analítica de video más tradicional que se basa en comparar frame a frame un video, se estudia la composición de cada frame y las variaciones entre ellos. Por otro lado, existe la analítica de video apoyada en IA, esta emplea machine learning y es capaz de detectar estos sucesos al conocer sus características (por ejemplo, es capaz de identificar la figura de una persona en una imagen y de distinguirla de la de un gato). Como hemos mencionado previamente, en la actualidad los principales avances pasan por la combinación de estas dos metodologías utilizando en primera instancia la analítica tradicional para filtrar los sucesos de interés y posteriormente realizando una analítica por medio de IA que categorice el suceso.

4.2 Desarrollo del Programa

El modelo se ha ido desarrollando durante el transcurso del proyecto y se ha ido apoyando en el estudio realizado simultáneamente sobre la tecnología de vanguardia. El desarrollo del

modelo tiene dos patas, la solución de problemas y fallos que se han ido detectando a medida que se desarrollaba y la inclusión de nuevas capacidades en él.

4.2.1 Modelos de prueba

Los primeros programas tenían como única finalidad conocer el entorno y comenzar a utilizar las funciones más básicas de Python relacionadas con el tratamiento de videos. Estos primeros programas únicamente eran capaces de obtener sencillas métricas de los videos como son: la cantidad de frames del video, el brillo del video y la duración. Posteriormente estos programas sacaban otras métricas derivadas como los frames por segundo del video y los devolvían por pantalla.

Estos programas sencillos y las métricas que aportaban son de gran valor ya que realmente el desarrollo de un modelo efectivo no es posible si no se conocen al detalle los aspectos más básicos del tratamiento de videos. (Francia, 2024)

4.2.2 Modelo Inicial

La primera versión del modelo final convierte el video a escala de grises para minimizar el coste computacional, ya que se observa que el tiempo de carga del video es claramente inferior. A continuación, analiza el video frame a frame y detecta sucesos relevantes en aquellos instantes en los que hay variaciones en la imagen. Al utilizarlo con videos reales el modelo peca de hipersensibilidad y detecta sucesos donde no existen y simplemente hay algo de movimiento como puede ser el movimiento de una hoja.

El código inicialmente importa las bibliotecas necesarias:

- ``Cv2``: biblioteca de OpenCV centrada en procesamiento de imágenes y videos.
- ``Os``: biblioteca para interactuar con el sistema operativo, nos permite comprobar la existencia del archivo de video.
- ``Numpy``: biblioteca para manejo de matrices.

A continuación, se indica la ruta del archivo de video y se comprueba su existencia. También se define la ruta donde se grabará el nuevo video, es decir, el video a analizar una vez este convertido a escala de grises.

En este mismo nivel se inicializan los siguientes parámetros:

- `Difference_threshold`: este es el principal parámetro del programa y establece el umbral a partir del cual se considera que ha habido un “cambio significativo entre un frame y el siguiente. En este programa el parámetro se modifica en función de la calidad del video de forma manual, iterando.
- `Counter`: este parámetro es un contador que inicializamos a 0 en el que se acumula el número de cambios significativos consecutivos que se detectan, es decir, cuantas veces seguidas se detecta una variación superior a la definida por el parámetro `Difference_threshold`.
- `Significant_change`: especifica el número de cambios significativos consecutivos necesarios para que consideremos que está habiendo un “suceso” en el video.
- `Frame_number`: lo inicializamos a cero y nos sirve para llevar la cuenta de frames que llevamos procesados.

A continuación, a través medio de la función `cv2.videocapture` se lee el video que se le ha indicado anteriormente y se comprueba que este tiene contenido, de no ser así saldrá un mensaje por pantalla indicando que no se ha leído ningún video.

El siguiente paso consiste en un bucle `While` que lee cada frame y los convierte uno a uno a gris y los graba en un nuevo archive con la función `out.write`.

```
while True:
    # Captura frame a frame y devuelve true si el frame se ha leído correctamente
    ret, frame = cap.read()

    if not ret:
```

```
print("Can't receive frame (stream end?). Exiting ...")
break

# Convierte el frame a gris
gray_frame = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)

# Graba los frames en escala de grises
out.write(gray_frame)
```

A continuación, se inicia el procesado de video releyendo el nuevo video que hemos generado en una escala de grises. Por medio de un bucle While con la misma condición que el anterior leemos cada fotograma y usando la función `cv2.absdiff` se calcula la diferencia absoluta entre el frame actual y el anterior, con este dato se comprueba si esta diferencia supera el umbral que previamente hemos definido, si esta diferencia es significativa se incrementa el contador y de lo contrario este se resetea. Por último, comprobamos siempre si el contador supera el parámetro “Significant_change” lo cual nos indicia si durante los últimos X frames ha habido cambios significativos. De superarse, nos imprimirá un aviso por pantalla y reiniciará el contador. Este proceso se repite hasta que son analizados todos los “frames” del video, una vez finalizado se liberan recursos y se muestra por pantalla que el proceso se ha finalizado.

Este primer programa funciona con bastante imprecisión, el primer gran problema está en que hay zonas del video que por lo que se está grabando o por la distancia que alcanza la cámara están en constante movimiento y esto nos provoca que no haya un umbral en el cual se detecten solo los cambios que son de nuestro interés. Por ello uno de los avances necesarios es establecer zonas de exclusión en los videos (por ejemplo, se excluirán zonas con mucha vegetación que al ser movidas por el viento disparan el aviso de suceso). Otro de los grandes problemas es el coste computacional de ir analizando todos los frames, lleva mucho tiempo lo cual no es eficiente y no soluciona uno de los objetivos que es la detección veloz de los sucesos. Por otro lado, existen otro problema a solucionar, como la necesidad

Desarrollo de un entorno propio

de adaptar el programa manualmente a cada video/cámara (el programa podría conectarse a una cámara en vivo sustituyendo la dirección del archivo por la dirección de hardware de la cámara). Esto se debe a que dependiendo de la calidad de la imagen y de lo que se graba los umbrales han de ser distintos para conseguir resultados efectivos.

4.2.3 Introducción de una máscara y reducción de coste computacional

Se adapta el programa para introducir una máscara que permita ocultar del proceso una zona en la cual consideremos que hay demasiado movimiento y que haga poco efectivo el programa. También se reduce el coste computacional al analizar únicamente una tercera parte de los frames (más adelante se estudia si esto reduce el tiempo de procesado a una tercera parte o si no es directamente proporcional).

Para la máscara se plantea una “zona de exclusión”, para ello se introducen los valores de la zona que se quiere excluir en el formato de coordenadas (x, y, x + w, y + h). Una vez determinado el área se fija que toda esa área sea negra durante todo el video de forma que no habrá cambios en esa zona entre los frames. Para ello, se fija que los píxeles de esa zona sean de valor 0 que es el otorgado al negro en la escala de grises, siendo el blanco el valor 255.



ILUSTRACIÓN 7. EJEMPLO DE SELECCIÓN DE ZONA DE EXCLUSIÓN EN EL VIDEO (FRANCIA, 2024)

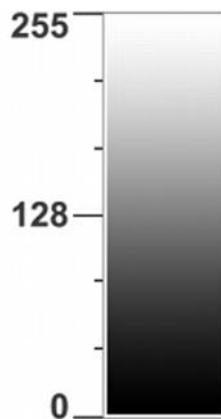


ILUSTRACIÓN 8. ESCALA DE GRISES; SIENDO 0 EL NEGRO Y 255 EL BLANCO

Para reducir el coste computacional, como se menciona previamente se reduce el número de frames que el programa analiza. Para esta labor se establecen dos nuevos parámetros:

- `Current_skip`: inicializado a 0, sirve para indicar en que posición de skip estamos, cada vez que se lee un frame se sube en una unidad su valor. Únicamente se

comparará la diferencia entre frames cuando el valor de `current_skip` coincida con el siguiente parámetro que fijaremos nosotros (`frame_skip`). Una vez esto ocurre, el valor de `current_skip` se reinicializa a 0.

- `Frame_skip`: este parámetro lo fijaremos y determinará cada cuantos frames se realiza la comparación de la imagen.

El funcionamiento es el siguiente: `current_skip` tendrá un valor de 0, y cada vez que el programa lea un frame del video sumará una unidad (leer no tiene prácticamente coste computacional y es necesario para pasar al siguiente frame). Al alcanzar `current_skip` el valor de `frame_skip` se comparará ese frame con el siguiente frame en el que esta condición suceda. En la práctica para un valor de `frame_skip` de 2 tenemos el siguiente bucle:

- A. `Current_skip` vale 2, y como coincide con `frame_skip` se graba el frame y el valor de cada pixel. A continuación, reiniciamos `current_skip` a 0.
- B. Se lee el siguiente frame sin grabar ni analizar nada y `current_skip` pasa a valer 1.
- C. Del mismo modo se lee un frame más sin analizar y `current_skip` pasa a valer 2.
- D. Al coincidir el valor de `current_skip` y `frame_skip` se analiza este frame con el del paso 1. Una vez hecho esto se reinicia de nuevo `current_skip` a 0.

En esta nueva versión también se mueve `current_skip` a un valor muy inferior a 0 siempre que se detecte una alarma, de modo que una vez detectado movimiento deje de analizar durante el número de frames que le indiquemos. Esto también sirve para reducir el coste computacional. (Francia, 2024)

4.2.4 Ajuste del tamaño de imagen, selección de área de exclusión simplificada e impresión de mensajes con imagen

El programa anterior dejaba en evidencia ciertos problemas. El primero de ellos es que al variar el video que se está analizando el programa tenía problemas con la dimensión, por ello en este nuevo programa se redimensiona el video entero a un ancho objetivo de 640 pixeles manteniendo la relación de aspecto. Esto supone un coste computacional, pero es

Desarrollo de un entorno propio

necesario para asegurar que en todos los videos se analiza correctamente el global de la imagen.

Otro de los problemas claros que se aprecian en el programa anterior es que la selección del área a excluir es demasiado manual y requiere un análisis de la imagen. Este proceso se automatiza en el nuevo código mediante una nueva función: `select_roi`. Esta nueva función nos permite seleccionar un ROI (región of interest), esto se hace nada más arrancar el programa por medio de una imagen que se imprime en pantalla en la cual se puede seleccionar un recuadro que pasará a ser la ROI. Este recuadro que introducimos es convertido en unas coordenadas (las que antes teníamos que introducir manualmente) y estas coordenadas son reseteadas al valor 0 de la escala de grises durante todo el programa, consiguiendo así que los cambios que ocurran en esa área sean obviados.



ILUSTRACIÓN 9. EJEMPLO DEL MÉTODO DE SELECCIÓN DE UNA ZONA DE EXCLUSIÓN
(FRANCIA, 2024)

Desarrollo de un entorno propio

Como se aprecia en la imagen, el programa nos permite seleccionar la zona de exclusión y el mismo es el que saca las coordenadas de esa zona señalada para posteriormente ponerla a negro.

El otro gran cambio de esta versión respecto a la anterior es en los mensajes que se muestran una vez se detecta una alarma/suceso. Esta nueva versión imprime en pantalla una imagen del frame en el cual se decide que está habiendo un cambio significativo, es decir, aquel donde la diferencia acumulada entre fotogramas excede el umbral de sensibilidad preestablecido (el parámetro `Difference_threshold`) después de también detectar cambios consecutivos en los últimos `X-1` fotogramas siendo `X` el valor que le demos al parámetro `Significant_change`. También se nos imprime simultáneamente en pantalla el instante de video que representa ese frame y que numero de frame es. Esto nos permite decidir rápidamente si el suceso detectado es de interés o es un suceso para descartar como, por ejemplo, un gato. Si se decide que es descartable con tan solo presionar enter el programa sigue analizando y así se maximiza la eficiencia y no se pierde tiempo en ese suceso.



ILUSTRACIÓN 10. IMAGEN INICIAL DE UN VIDEO ANALIZADO CON EL PROGRAMA DE PRUEBA
(FRANCIA, 2024)



ILUSTRACIÓN 11. EJEMPLO DE DETECCIÓN DE UN SUCESO EN EL PROGRAMA DE PRUEBA
(FRANCIA, 2024)

Si se comparan las dos imágenes, siendo la primera la imagen estática del video y la segunda la impresión por pantalla de un suceso se puede observar como en la segunda imagen en el cuadrante superior derecho hay un vehículo que no debería estar y que al aparecer por pantalla salta la alarma y por tanto se imprime la imagen por pantalla.

4.2.4.1 Selección de múltiples zonas de interés

Al utilizar esta versión se evidencia que la selección de zonas de interés que posteriormente excluiríamos del análisis no cubre el 100% del problema. En la mayoría de los videos se evidencia que hay más de una región que no contienen información útil o que introducen demasiado ruido en el análisis como para que nos merezca la pena considerarlas. Por esto, la selección que permite el programa anterior de un único rectángulo es insuficiente ya que o bien solo cubre una de estas áreas o, si pretendemos cubrir todas con él, ocupa demasiado del espacio de la imagen limitando mucho la sensibilidad de análisis.

Para solventar este problema evidente se modifica el código para permitir una selección de múltiples zonas de interés que posteriormente serán excluidas del análisis. Con ello se incrementa la flexibilidad al definir de forma más precisas las zonas a ignorar lo cual resulta en una mejora en precisión y relevancia del análisis. Consecuentemente se ha ajustado el código para que todas estas zonas seleccionadas sean excluidas de forma simultánea lo cual nos permite reducir el coste computacional que es tan importante en el análisis de video.

4.2.5 Selección del parámetro en función de la calidad de la imagen

El siguiente avance que se persigue es adaptar el programa a la calidad de cada cámara / video, para ello, el principal parámetro que determina la sensibilidad con la que se detectan los sucesos “difference_threshold” ahora se introducirá por pantalla de modo que el programa se puede utilizar exitosamente para videos de diferentes calidades. El usuario únicamente tendrá que determinar que parámetro se ajusta mejor a las características del video, una vez haga esto ya podrá analizar indefinidamente el video. Como referencia, para un video de seguridad en un domicilio con una cámara de alta resolución este parámetro debe situarse en el entorno de 1000 y por el contrario para una cámara en una central fotovoltaica con mucho ruido en el análisis y un gran campo de visión el programa funciona de forma efectiva fijando el parámetro en el entorno de 12,000.

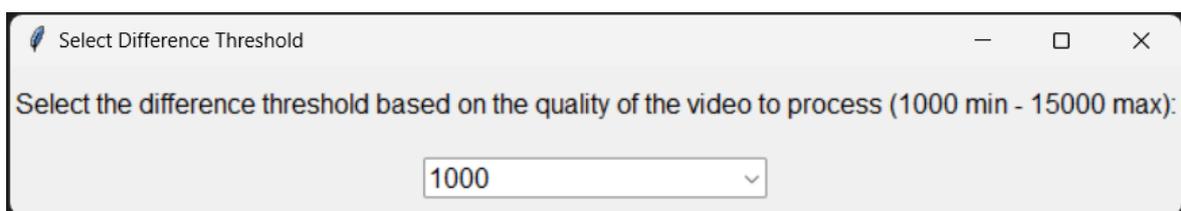


ILUSTRACIÓN 12. MENSAJE POR PANTALLA PARA INICIALIZAR EL PRINCIPAL PARÁMETRO DEL PROGRAMA DE PRUEBA (FRANCIA, 2024)

Para completar el programa también se ha implementado una muestra por pantalla de un video de 5 segundos del instante en el que se detecta un suceso. Junto con este video se indica el momento temporal en el video en el que ocurre el suceso lo cual permite al usuario identificarlo rápidamente. (Francia, 2024)

4.3 Aprendizajes y limitaciones del programa

El desarrollo del programa en paralelo con el estudio de la situación actual del análisis de video ha permitido identificar de primera mano las grandes problemáticas que envuelven al análisis de video y que concuerdan con los grandes problemas que existen en la utilización de analítica de video a nivel profesional. Durante el desarrollo del programa también se han realizado multitud de aprendizajes a través de los fallos que se iban detectando.

El principal obstáculo en la automatización del análisis de video es la gran variación que existe entre grabaciones, no solo debido a variaciones en la calidad de la imagen y la transmisión de esta, sino también debido al contenido de la imagen. No es lo mismo detectar sucesos en una grabación en el interior de una nave donde en principio no debería de haber movimiento que detectar sucesos en una grabación de una planta fotovoltaica con una cámara que esta grabando a distancias mayores a los 100 metros y con multitud de factores externos que imposibilitan la tarea como son la lluvia, la fauna (desde insectos a pájaros) o los cambios veloces de iluminación. Actualmente, similarmente al sencillo programa de prueba que se muestra en este trabajo, la gran mayoría de programas de video análisis tienen multitud de parámetros que han de ser adaptados a la grabación (cámara) concreta que debe analizar para conseguir la sensibilidad perfecta y solo detectar eventos relevantes. El futuro pasa por software que sea capaz de ir auto adaptando estos parámetros a medida que va acumulando horas de grabación para ir afinando y terminar detectando únicamente eventos relevantes, sin necesidad de ser preprogramado para esa cámara. Esto simplificaría y democratizaría la analítica de video al ser el propio programa el que se adapta a la grabación sin necesidad de adaptaciones manuales.

También hemos observado de primera mano la gran necesidad de capacidad de procesamiento de la maquina donde se realiza el análisis de video. Esta mayor capacidad de procesamiento se consigue gracias a un mejor procesador, más núcleos (CPU), más memoria RAM o una tarjeta gráfica (GPU) más potente. Todas estas características que son maximizadas en dispositivos de video análisis. Las características de un ordenador portátil común de 16 GB de RAM y un procesador Intel Core i5 con una tarjeta gráfica integrada son muy limitadas para llevar a cabo estas tareas.

Durante el desarrollo del programa otros aspectos que se han podido observar son la importancia de las zonas de exclusión para que solo se realice un análisis de las zonas que son de interés y obviar en el análisis zonas que introducen demasiado ruido, como puede ser una zona de vegetación que debido al viento complicaría mucho las tareas. Por otro lado, es esencial que los programas sean eficientes en cuestión de coste computacional, para ello durante el desarrollo de los programas esto debe de ser uno de los principales factores a tener en cuenta. (Francia, 2024)

5 Impacto Socio – Económico

Las ventajas que aporta la introducción de sistemas de análisis de video automatizados, en la medida en que la tecnología se desarrolle suficientemente y las limitaciones continúen en vías de resolución o minimización como se ha indicado anteriormente, son de tal envergadura que, sin duda, se avanzará de forma progresiva y acelerada hacia su utilización masiva en los próximos años suponiendo una revolución en la forma de trabajar en distintos sectores cuya actividad está, en una parte relevante, basada en el análisis de lo que sucede en un determinado espacio físico.

Las limitaciones a su desarrollo existentes (capacidad de proceso de datos, capacidad y velocidad de comunicación (intercomunicación), tecnología de aprendizaje del modelo, ...) para disponer de forma estable de la tecnología necesaria para los sistemas de video automatizados están en vías de eliminación y/o reducción de forma progresiva y esto permitirá el desarrollo masivo de estas soluciones empezando en los sectores de actividad donde aporten mayor valor añadido mejorando la oferta de servicio y/o aporten una mayor optimización costes.

A continuación, se realiza un primer análisis de las ventajas y la aplicabilidad en algunos de los sectores concretos donde su utilización es más clara y también se realiza con posterioridad un análisis de las limitaciones o restricciones actuales y del modelo de rentabilidad de la inversión.

5.1 Ventajas operativas y económicas de la utilización de análisis de video automatizado

El análisis de video automatizado, en definitiva, sustituye al visionado de la imagen por una persona, la reflexión/análisis sobre lo que se ve en el visionado y sus implicaciones y la toma, en primera instancia, de una decisión al respecto. Para eventos concretos pueden establecerse visionados y análisis posteriores que podrían seguir desarrollándose de forma

manual personalizada, de forma que aquellos eventos críticos pueden mantener, en la medida en que sea conveniente o necesario, una revisión manual.

Una vez desarrollado el sistema y realizado el aprendizaje con suficiente madurez, la fiabilidad continuaría mejorando a la vez que se avanza en el aprendizaje. El visionado, el análisis de las implicaciones y la toma de decisiones se realizaría de forma automatizada sin necesidad de tener un equipo humano específico, aportando las siguientes ventajas:

- Ausencia de errores humanos y garantía de respuesta, con una inmediatez muy superior a los parámetros temporales establecidos en caso de atención humana, al 100% de los visionados.
- Resolución de la problemática de recursos suficientes implícita en los servicios 24x7 permanentes que requieren de una gestión operativa muy compleja y con múltiples cambios constantes para optimizar la disposición de los recursos humanos necesarios; considerando bajas laborales, ausencias no previstas, vacaciones, festivos, renuncias, formación, Adicionalmente, en términos generales, en estas actividades hay un alto componente de estacionalidad y de horarios pico y valle que complican significativamente la gestión y optimización de los equipos humanos necesarios.
- Simplificación y optimización organizativa de la gestión de operaciones en empresas de alto volumen de personas realizando tareas manuales que requiere una estructura complementaria de niveles de supervisión (jefes de turno, área managers, supervisores, ...) y de departamentos centrales, como RRHH, de elevada dimensión cubriendo especialmente: selección, formación, administración de personal y nómina, prevención de riesgos laborales, relaciones laborales, ...
- Eliminación del riesgo de impacto en la continuidad del servicio de conflictos laborales que no son extraños en entornos de elevado volumen de equipos humanos y retribución limitada.

- Más rápida y sencilla escalabilidad en la medida en que se necesite más capacidad puesto que es básicamente una cuestión técnica de extensión de capacidad de procesamiento de datos y de comunicación, no requiriendo selección, formación,
- En la medida en que los modelos se optimicen, el coste descenderá de forma progresiva a niveles que permitirán utilizar sistemas de análisis por imágenes en actividades en las cuales actualmente no es factible utilizarlos por su elevado coste si son realizados por personas.

La automatización generalizada supondrá un cambio disruptivo en la forma de operar de las empresas y organizaciones que realizan estas actividades con la consiguiente reordenación de competitividad y rentabilidad de unas y otras empresas y la aparición de nuevos entrantes y la discontinuación de operaciones de operadores actuales.

Una mejora de competitividad tan disruptiva como la que se producirá por la automatización provocará también una reducción muy significativa del precio para los consumidores y usuarios. El enorme incremento de márgenes permitirá la bajada decidida de precios y se producirá una presión directa en los precios por posibles nuevos entrantes atraídos por el aumento de márgenes basándose en un elemental análisis aplicando el modelo de M Porter, como más adelante analizaremos con más detenimiento.

Hay muchos ejemplos de sectores enteros que han sido completamente transformados por avances tecnológicos altamente disruptivos como puede ser el fax, que había supuesto en su momento una extraordinaria transformación sobre el télex, siendo sustituido por el escáner y el mail o el mundo de la fotografía en su evolución de un mundo de film expuesto al objetivo y posteriormente revelado a un mundo digital que provocaron en muy poco tiempo un cambio completo de modelo de operaciones y hubo empresas ganadoras y perdedoras que discontinuaron su actividad. Otro sector con ciertas similitudes que se ha transformado drásticamente aunque de una forma más lenta y todavía en proceso de concluir, es la automatización de operaciones en el sector bancario y asegurador, donde multitud de tareas requerían presencia física del cliente y del empleado y actualmente se han automatizado de forma que no se requiere presencia física y muchas operaciones se pueden automatizar o

programar, habiéndose producido en este caso, a la vez una optimización de costes y márgenes que ha resultado en menores precios (comisiones) para los clientes, una mejora de las operaciones, más rápidas, sencillas y consistentes, y, también una notable reducción de equipos (incluyendo sucursales) en las empresas.

En el terreno sociolaboral, una transformación tan disruptiva provocará también un cambio muy significativo en las características y dimensión de los recursos humanos utilizados en el visionado manual de imágenes. El efecto será, lógicamente, una drástica reducción de recursos directos e indirectos, los directamente dedicados al visionado y los necesarios para gestionar organizaciones tan intensivas en personal y la creación de un número menor de posiciones de mayor especialización relacionadas con la instalación de equipos en las áreas objeto de análisis y toda la infraestructura y sistemas necesarios para la automatización y su supervisión.

Este aspecto es, sin duda, muy relevante en la medida en que algunos de los sectores a los que más afectará esta disrupción operativa será el de la vigilancia y seguridad que es actualmente a nivel global y, específicamente en España, un sector de alto volumen de empleo, aunque con un nivel salarial relativamente en comparación con las medias nacionales, donde un vigilante auxiliar de seguridad cobra un entorno de 18.000 de salario anual fijo de acuerdo con el convenio colectivo de aplicación (Agencia Estatal Boletín Oficial del Estado (BOE), 2024).

Como conclusiones de este análisis de ventajas de la automatización del análisis de video, se producirían efectos muy positivos para la sociedad en términos de mejor calidad y cobertura de servicios y menor precio que, a su vez, permitiría abarcar actividades que actualmente no son factibles por el elevado coste.

En el terreno económico-social, el impacto en el empleo sería muy relevante en el sentido de un empleo más cualificado en determinados sectores, pero de mucho menor volumen.

5.2 Sectores en las que la aplicación de análisis de video automatizado aporta un valor añadido muy relevante

En términos generales, se produciría un cambio muy relevante en todos aquellos sectores que utilizan de forma intensiva el visionado de imágenes para el análisis y la toma de decisiones de una forma razonablemente estandarizable. Entre estos sectores se pueden destacar los siguientes:

- Seguridad y vigilancia de instalaciones específicas y hogares. Esta actividad es desarrollada por empresas de seguridad privada que, fundamentalmente, cubren las actividades de: a) vigilancia de instalaciones, b) gestión de sistemas y alarmas y c) transporte de fondos.

Según datos de Aproser (Aproser, 2022) que es la Asociación Profesional de Compañías de Servicios de Seguridad; en España existen 1.570 empresas que facturaron €4.622M incluyendo 89.000 vigilantes. Por segmentos de actividad, la facturación es: a) vigilancia de instalaciones (59,3%), b) sistemas y alarmas (34,2%) y c) transporte de fondos y otros (6,5%). La facturación se realiza en un 74% al sector privado y en un 26% al sector público.

A nivel comparativo con otros países, en España hay 1 vigilante por cada 542 habitantes mientras que a nivel europeo hay 1 vigilante por cada 281 habitantes. (Aproser, 2022). Es necesario también destacar que algunos países en entornos con mayor inseguridad que en Europa alcanzan cifras muy superiores, como puede ser el caso de México, en el que en un artículo de 15 de agosto de 2023 la revista Forbes estimaba que el sector de seguridad alcanzaba el 1,8% del PIB con 900.000 empleos directos.

- Sistemas públicos de supervisión relacionados con la seguridad pública general, el control y supervisión del tráfico, la prevención de incendios y actividades varias de vigilancia generales y específicas.

Las actividades de los cuerpos y fuerzas de seguridad del Estado, incluyendo los servicios autonómicos y locales, incluyen, bien directamente o bien indirectamente, la vigilancia de espacios públicos o privados de forma preventiva, reactiva y como labores de investigación posteriores sobre incidentes acaecidos para lo cual, en la medida disponible, utilizan como una de las principales fuentes el análisis de las imágenes captadas por organismos públicos y privados que, de acuerdo con la Ley 5/2014 de Seguridad Privada de 4 de abril, (Agencia Estatal Boletín Oficial del Estado (BOE), 2014) cuando es necesario tienen acceso, bajo ciertas condiciones, a los datos privados.

Sin disponerse de un dato exacto de agentes, dada su distribución y múltiples dependencias, se estima que hay actualmente en España unos 155.000 agentes de policía y guardia civil y unos 65.000 agentes de policía local, a lo que hay que añadir las policías autonómicas, los servicios aduaneros, los servicios de prevención de incendios y medioambientales y diversos cuerpos adicionales. En total, un volumen de agentes cercano a los 250.000.

- Supervisión de las operaciones en actividades de “retail” y atención a consumidores y clientes para mejorar la eficiencia del servicio y la calidad de este y para entender el comportamiento del consumidor y mejorar la operativa, atractividad de la exposición de productos o la experiencia de usuario (UX)

Esta actividad de supervisión y análisis es una necesidad de gestión operacional de todos los servicios y una oportunidad de mejorar la propuesta de valor desde diferentes aspectos, especialmente aquellos servicios orientados directamente al consumidor y/o en los que hay una involucración directa de equipos humanos, tales como servicios de restauración y hospedaje, distribución y comercio, sanidad, ... en los que las actividades operativas están muy protocolizadas en modelos y procesos operativos y se requiere tanto de un seguimiento de que se cumplen los protocolos como de la identificación constante de oportunidades de mejora.

La actividad de supervisión y análisis es realizada actualmente por equipos específicos de supervisión y auditoría, propios o externos, y por los propios responsables de los distintos equipos. Si bien el personal dedicado de forma específica a estas actividades es limitado, el volumen general de la actividad (los servicios representan cerca del 70% del PIB de España medido en términos de oferta) y la relevancia de mejorar este seguimiento para elevar los estándares de calidad y de eficiencia y detectar oportunidades de mejora, lo hacen también una actividad de alto interés económico. En relación con esta actividad la utilización de modelos generalizados tendrá que ser limitada puesto que las actividades son muy distintas, aunque en contraposición requieren menor volumen para el entrenamiento de los modelos al ser entornos más cerrados y limitados.

Actualmente, hay diversas empresas ofreciendo este tipo de servicios en modelos básicamente de “start-up” tecnológica. Simplemente a modo de ejemplo, se ha escogido como referencia una de ellas YEDDA.AI (Yedda.AI, 2024) que ofrece este tipo de servicios para retail, restauración, logística, hotelería y actividades de producción y en cuya web se puede observar perfectamente como se aplica la metodología de análisis de video descrita en los apartados anteriores de este estudio

5.2.1 Algunas consideraciones de interés sobre el sector de vigilancia y seguridad

Como se ha indicado anteriormente, el sector de mayor volumen es el de la seguridad y vigilancia privada y, por tanto, es sobre el que nos centraremos a la hora de seguir analizando la evolución que supondrá la utilización generalizada de sistemas de análisis de video automatizados. A efectos de soportar determinados razonamientos en relación con los obstáculos y la velocidad estimada de implantación de soluciones de análisis de video automatizado es necesario profundizar en las características de las empresas de seguridad y vigilancia y sus relaciones competitivas.

El sector de la vigilancia y seguridad de instalaciones es actualmente un sector básicamente de servicios de consumo, de elevado volumen en el que no hay una gran diferenciación entre

los servicios que dan unas y otras empresas y donde el precio, en relación con el alcance del servicio es el elemento más importante en la elección del proveedor. El conjunto del sector evoluciona históricamente por encima del PIB dado el creciente incremento de demanda de servicios para aumentar la seguridad, tanto a nivel internacional como en España, pero con un incremento en precios en línea con IPC dado que no hay grandes evoluciones en las operaciones en los últimos años y el principal coste de la empresas son los equipos humanos altamente regulados por convenios colectivos.

Las empresas que proveen el servicio requieren de escala y concentración para poder ser competitivas; cuanto más volumen tienes y más concentrado está geográficamente más capacidad de atención tienes y más optimizas los recursos tanto en la parte de instalaciones físicas y mantenimiento, como de cobertura optimizada de la central receptora de alarmas (CRAs), como de la capacidad respuesta antes eventos extraordinarios. Sin embargo, está escala puede ser solo de alcance relativamente local, lo que hace que el sector este bastante atomizado y convivan empresas de cobertura local y grandes compañías internacionales. Igualmente, son empresas muy relacionadas con la parte de instalaciones y mantenimiento de equipos y con elevadísimos volúmenes de personal que las hacen relativamente lentas en la toma de decisiones, no habiendo un know-how o tecnología diferenciadora actualmente.

A continuación, se analizará con más profundidad este sector donde es más claro el efecto enormemente disruptivo que tendría la utilización generalizada de análisis de video automatizado.

5.2.1.1 Características de las operaciones

En línea con lo indicado anteriormente, las empresas de vigilancia y seguridad tienen básicamente tres actividades: a) vigilancia de instalaciones (59,3%), b) sistemas y alarmas (34,2%) y c) transporte de fondos y otros (6,5%). Las dos primeras, que constituyen la inmensa mayoría de su facturación, están completamente integradas puesto que, al final, la actividad consiste en la instalación de unos sistemas de vigilancia y detección en localizaciones específicas y las labores de vigilancia “in situ” o a través de los sistemas de video y detectores de distinto tipo y la intervención física en caso necesario. El desarrollo

de estas actividades requiere, además de las áreas soporte habituales de cualquier empresa en función de su tamaño, las siguientes actividades básicas:

- Instalación de sistemas de videovigilancia y detección de distinto nivel de sofisticación e interconexión, en función de las necesidades y presupuesto del cliente, y su mantenimiento y evolución permanente.
- Vigilancia “in situ” en determinados casos. En la práctica, en muchos casos hay asociados servicios complementarios de conserjería cuyo coste es inferior a los de vigilancia, pero complementan ésta y ofrecen un complemento a las necesidades del cliente.
- Acceso a una CRA (Central Receptora de Alarmas) donde se monitorizan los incidentes y se deciden las actuaciones a acometer en función del análisis.
- Disponibilidad de equipos humanos de vigilancia y seguridad que, en caso de necesidad, pueden desplazarse a la localización física específica en un corto espacio de tiempo.

La facturación al cliente se instrumenta normalmente como un coste de instalación inicial (pueden utilizarse modelos de renting o de inversión), un coste mensual o anual por los servicios de los que se dispone, incluyendo mantenimiento preventivo y los servicios de CRA y costes específicos cuando es necesario hacer modificaciones de cualquier tipo a requerimiento del cliente.

El elevado volumen de vigilantes “in situ”, la necesidad de disponer de equipos de vigilantes que puedan desplazarse físicamente, en un razonablemente corto periodo de tiempo, a cualquier instalación física, la disponibilidad de equipos humanos analizando 24x7 las video imágenes en las CRA y tomando las decisiones pertinentes y la necesidad de equipos humanos (en parte pueden ser subcontratados) de instalación y mantenimiento, hacen esta industria altamente intensiva en equipos humanos. Igualmente, es una industria donde la escala es fundamental a efectos de rentabilidad económica, al requerir una presencia local cercana de vigilantes con disponibilidad de acudir a cualquier localización física y de

equipos de instalación y mantenimiento, minimizando, en ambos casos, los costes de desplazamiento.

5.2.1.2 Tipología y características de las empresas de seguridad y vigilancia en España

Una de las principales características de la industria de la seguridad y vigilancia es la fragmentación de las empresas que participan en la misma, conviviendo grandes empresas internacionales con empresas pequeñas y medianas en general de componente local.

Como se ha indicado anteriormente, según datos basados en el ejercicio 2022 de Aproser en España hay 1.570 empresas que facturaron €4.622M, es decir, una facturación media por empresa de tan solo €2,94 millones de euros, que hace que una multitud de empresas, generalmente muy locales, sean PYMEs.

Junto con estas empresas, conviven grupos multinacionales de mucha mayor envergadura. En el caso de España las empresas más relevantes del sector, sobre datos de 2022 de acuerdo con y considerando el CNAE 8010 de Servicios de Seguridad Privada (adicionalmente con menor volumen existe el CNAE 8020 de Servicios de Sistemas de Seguridad) son:

Ranking de empresas de Servicios de Seguridad Privada	
	Ventas (millones de euros)
SECURITAS DIRECT ESPAÑA SU	1022
PROSEGUR SOLUCIONES INTEGRALES DE SEGURIDAD ESPAÑA SL	793
SECURITAS SEGURIDAD ESPAÑA SAU	631
ACCIONA FACILITY SERVICES SAU	303
EULEN SEGURIDAD SA	243
ILUNION SEGURIDAD SA	169
MOVISTAR PROSEGUR ALARMAS SL	164
PROSEGUR SERVICIOS DE EFECTIVO ESPAÑA SL	149
LOOMIS SPAIN SA	146
GRUPO CONTROL EMPRESA DE SEGURIDAD SA	105

ILUSTRACIÓN 13. RANKING DE EMPRESAS DE SERVICIOS DE SEGURIDAD PRIVADA

(ELECONOMISTA, 2024)

Aunque con empresas diferentes en función de “joint-ventures” y divisiones separadas jurídicamente, los líderes claros en España son Securitas que engloba Securitas Direct, Securitas Seguridad y Loomis con una facturación de 1.797 millones de euros y Prosegur que engloba Prosegur Soluciones Integrales de Seguridad, Movistar Prosegur Alarmas y Prosegur Servicios de efectivo con una facturación de 1.106 millones de euros.

SECURITAS

Securitas pertenece al grupo Securitas AB de origen sueco, fundado en 1934 y con sede en Estocolmo y cotiza en el mercado sueco (SECU-B).

Principales magnitudes financieras y operativas 2023	
Ventas (millones de euros)	13.681
EBIT (millones de euros)	430
EBIT margin (%)	3.15
Net Financial Debt/EBITDA Ratio	2,7
Market Cap (14/08/24)	5391
Mercados con operación	44
Empleados	341.000

ILUSTRACIÓN 14. PRINCIPALES MAGNITUDES FINANCIERAS DEL GRUPO SECURITAS
(SECURITAS, 2023)

Securitas es uno de los líderes mundiales del sector y cubre todos los aspectos de la seguridad privada. Analizando en detalle las explicaciones en las presentaciones para analistas e inversores es evidente la importancia que se le da al desarrollo tecnológico y a la implantación de soluciones tecnológicas avanzadas como vía de desarrollo futuro (Securitas, 2024).

El aspecto que se tratará con posterioridad y que destaca es la limitada rentabilidad (un margen de EBIT de tan solo un 3,15% y una deuda relevante con un NFD/EBITDA de 2,7 veces), al igual que una capitalización limitada de 5.391 millones de euros.

PROSEGUR

Prosegur es de origen español y fue fundada en 1976, tiene sede en Madrid y cotiza en el mercado español (PSG-BME). Algunos de los principales datos de referencia son:

Principales magnitudes financieras y operativas 2023	
Ventas (millones de euros)	4.310
EBIT (millones de euros)	238
EBIT margin (%)	5,53
Net Financial Debt/EBITDA Ratio	2,3
Market Cap (14/08/24)	920
Mercados con operación	31
Empleados	150.000

ILUSTRACIÓN 15. PRINCIPALES MAGNITUDES FINANCIERAS DEL GRUPO PROSEGUR
(PROSEGUR, 2023)

Prosegur es igualmente uno de los líderes mundiales del sector y cubre todos los aspectos de la seguridad privada. Al igual que Securitas, analizando en detalle las explicaciones en las presentaciones para analistas e inversores es evidente la importancia que se le da al desarrollo tecnológico y a la implantación de soluciones tecnológicas avanzadas (Prosegur, 2020).

En este sentido hay que destacar especialmente el acuerdo con Telefonica (Movistar) en 2019 por el que Prosegur vendió a Movistar el 50% del negocio de alarmas por 300 millones de euros, gestionándose esta unidad de negocio desde entonces de forma conjunta. El acuerdo busca claramente mejorar las capacidades y la posición competitiva en un entorno en el que la tecnología y la intercomunicación de los dispositivos es una tendencia clara de futuro en la industria y también para Movistar es una forma de utilizar sus capacidades digitales de forma adicional y ampliar la oferta para los clientes con los distintos servicios conectados que una casa individual pueda necesitar.

Al igual que con Securitas, analizando Prosegur, que es igualmente una empresa líder y exitosa, destaca la limitada rentabilidad (un margen de EBIT de tan solo un 5,53% y una deuda relevante con un NFD/EBITDA de 2,3 veces), al igual que una capitalización limitada de 920 millones de euros.

5.2.1.3 Tipología y características de las empresas de seguridad y vigilancia a nivel internacional

Tanto Securitas como Prosegur son dos de empresas líderes mundiales y en términos generales, las conclusiones sobre ellas son válidas a nivel internacional.

La mayor empresa de vigilancia y seguridad a nivel mundial es Allied Universal con sede en USA y 770.000 empleados (incluyendo subcontrataciones) en 2023 (Allied Universal, 2022).

5.2.1.4 Aplicación del modelo de M. Porter de análisis de rentabilidad sectorial

La aplicación del conocido modelo de análisis la competitividad estratégica publicado en la de Michel Porter y que sigue siendo un modelo completamente válido para analizar la situación competitiva de una industria, ilustra de una forma muy interesante las implicaciones de la situación de sector y las limitaciones al desarrollo del análisis de video automatizado derivadas de su situación competitiva.



ILUSTRACIÓN 16. GRÁFICO DE LAS 5 FUERZAS DE PORTER (PORTER, 1979)

La aplicación del análisis de las cinco fuerzas en el contexto de la industria y las características de las empresas participantes descritas anteriormente nos define unas conclusiones en línea con lo siguiente:

- **Poder de negociación con proveedores:** Los proveedores principales básicamente son tres: a) los proveedores de equipos y sistemas (grandes compañías como proveedoras de cámaras, hardware y similar, además de capacidad de computación y almacenamiento en la nube) sobre las que hay solamente una moderada capacidad de negociación dado su enorme tamaño y que venden a otras muchas industrias de mayor tamaño, b) los empleados que, en términos generales, están altamente sindicalizados por los que la negociación colectiva es bastante rígida y limita la capacidad de negociación y c) las empresas proveedoras (outsourcing) de servicios de instalación y mantenimiento sobre las que hay un mayor poder de negociación.
- **Poder de negociación con clientes:** Para los clientes, a) empresas que requieren servicios de vigilancia y seguridad de sus instalaciones y que tienen equipos de compra especializados en un coste relevante que están permanentemente gestionando y comparando distintas alternativas y b) hogares individuales y urbanizaciones para las cuales el coste de la vigilancia y seguridad es muy relevante (en el caso de urbanizaciones suele ser el más importante) y que disponen de múltiples alternativas de servicio y es relativamente fácil el cambio de proveedor. En definitiva, la capacidad de negociación es muy limitada siendo un servicio cuyo coste y valor añadido está constantemente en revisión por los clientes y en el cual el cambio de proveedor no presenta especiales dificultades.
- **Productos sustitutivos:** En términos generales el producto sustitutivo es la rescisión del servicio que dependiendo de los niveles de seguridad de las zonas y la criticidad de las instalaciones es una alternativa de ahorro de costes para los clientes o simplemente la reducción del alcance del servicio.
- **Rivalidad y barreras de entrada y salida:**

A nivel de empresa de cobertura local, las barreras de entrada son limitadas requiriéndose fundamentalmente un volumen mínimo en la zona de operación y a nivel más global, las barreras de entrada si son relevantes y se centran en disponer de una cobertura global en el mercado y un reconocimiento como proveedor de servicios de referencia que requiere inversión constante y cuantiosa en marketing publicidad.

En relación con las barreras de salida hay tener en cuenta que éstas empresas rara vez forman parte de grupos empresariales con diversas actividades y eso limita las posibilidades de salida o reorientación empresarial, puesto que es el cese de actividad, aunque no hay inversiones de CAPEX importantes, es la única alternativa con una pérdida completa de valor.

En términos de rivalidad competitiva, nos enfrentamos, como se ha indicado anteriormente, a una industria fragmentada y en el que todas las empresas a partir de un cierto tamaño dan un servicio muy similar sin que haya diferencias relevantes para el usuario; lo que unido a que es un sector donde el coste es muy relevante para el cliente hace que exista una presión muy importante en los márgenes para mantener los clientes. En términos generales, el crecimiento de una empresa en particular tiene mucho de conseguir contratos existentes a través de un mejor precio y mantenerlos con contención de precios y constante elevación del nivel de servicio.

- **Nuevos entrantes:** En la actualidad se está empezando a producir la entrada moderada vía compra o asociación de empresas de seguridad por parte de compañías más vinculadas a temas tecnológicos en línea con la tendencia expuestas en este trabajo. En España el caso más claro es la alianza Movistar-Prosegur comentada anteriormente.

Esta situación caracteriza claramente a una industria de márgenes bajos y limitada capacidad de creación de valor al estar limitada la mejora de márgenes por la competitividad y las presiones sobre los precios de los clientes.

El cambio de modelo hacia unas operaciones basadas en análisis de video automatizado y mucho menos dependiente de una fuerza laboral masiva, permitiría elevar los márgenes de forma muy significativa, a la vez que reducir los precios por los elevados ahorros, como posteriormente analizaremos, y crearía unas barreras de entrada inalcanzables para empresas medianas y pequeñas, incluso de ámbito local dando mucha más estabilidad a los precios y limitando la presión permanente en precios mediante cambios de proveedor. Se produciría una clara evolución hacia un sector mucho más atractivo para la creación de valor en términos económicos.

5.2.1.5 Conclusiones sobre la rentabilidad del sector y la situación de las empresas

A modo de conclusiones, nos encontramos ante un sector en el que hay un claro dominio de empresas multinacionales con implantación en determinados territorios geográficos que presentan relevantes economías de escala y un claro valor añadido en la propuesta de valor para los clientes, pero dentro de un contexto de fragmentación y donde hay competidores locales relevantes que son una alternativa y una presión en precios por parte de los clientes muy fuerte siendo muy difícil la diferenciación de la propuesta de valor entre empresas de las mismas características (en España Securitas y Prosegur Movistar, los dos grandes líderes, tienen una propuesta de valor completamente intercambiable).

A los efectos de entender la velocidad de desarrollo de la aplicación generalizada de análisis de video automatizado, esta situación nos lleva a tener en cuenta un elemento clave que es la muy limitada capacidad de las grandes empresas actuales para acometer, por si solas, la inversión necesaria para desarrollar los modelos y metodologías necesarias en estos momentos.

Como se ha descrito en apartados anteriores, las grandes empresas son empresas con márgenes reducidos, relevante endeudamiento y baja capitalización de mercado o valoración de la empresa, en su caso. Si se compara con la capacidad de las empresas que están actualmente liderando la inversión en el desarrollo de aplicaciones basadas en IA de cualquier naturaleza, las empresas de seguridad no están, en absoluto, en condiciones, en el corto/medio plazo, de competir en términos de inversión y esto, sin duda, actúa de freno al

desarrollo del análisis de video automatizado, independientemente de la generación de valor masiva que se producirá en su momento cuando se disponga de estas tecnologías y la transformación a positivo en términos de creación de calor que se producirá en la industria.

5.3 Requisitos para la implantación generalizada (mass-market) de la automatización del visionado

Las dificultades para la evolución de las operaciones a visionado automatizado se realicen a nivel generalizado en el mercado se centran en los siguientes aspectos cuya previsible evolución pasamos a analizar:

- **Calidad de la imagen:** El visionado automatizado requiere una calidad adecuada de la imagen, al igual que se requiere que las condiciones del entorno, tales como niebla, lluvia, oscuridad, deslumbres y similares no impidan una calidad razonable del visionado, siendo ambos aspectos totalmente superados en las cámaras y sistemas estándar que se utilizan actualmente, tal y como se ha descrito técnicamente con anterioridad.

Asimismo, como se indicaba anteriormente, la generalización del visionado automatizado traerá consigo una reducción de costes y precios y, en consecuencia, un incremento muy relevante de la necesidad de equipos que supondrá, a su vez, un abaratamiento de estos. En definitiva, no hay limitación técnica ni económica en este aspecto y el desarrollo generará a su vez una mayor facilidad en este sentido al abaratar los equipos que, en cualquier caso, continuaran mejorando de forma significativa. Es adicionalmente importante destacar que con los estándares actuales un mayor nivel de definición en términos generales no sería ni necesario ni conveniente, puesto que requiere más capacidad de comunicación y de procesado de datos.

- **Capacidad de procesamiento:** Actualmente hay capacidad y tecnología para procesar enormes volúmenes de procesado de relativamente fácil acceso e instalación

y el incremento de capacidad que es, como se ha indicado anteriormente, un elemento técnico todavía relevante de limitación es una cuestión meramente económica en términos coste-beneficio de la inversión en visionado automático versus recurso humano. Actualmente, la capacidad de procesado se incrementa algorítmicamente, a la vez que se reduce el coste de equipos por dato procesado, facilitando esta tendencia la disponibilidad de capacidad creciente a un coste progresivamente menor.

- **Capacidad de intercomunicación:** La conectividad está actualmente a un nivel muy alto de cobertura tanto de red de fibra como de comunicación móvil de última generación (5G) que permiten el envío inmediato de imágenes y la intercomunicación con una latencia muy baja. En los países desarrollados este elemento es actualmente una limitación reducida a determinadas zonas con peor cobertura que, en cualquier caso, de forma progresiva se reducen.
- **Tecnología de aprendizaje del modelo:** Como se ha analizado de manera técnica con anterioridad, existen actualmente, y avanzan continuamente, tecnologías de machine learning y de IA generativa que permiten desarrollar un modelo de análisis del visionado y respuesta totalmente automatizado que continuará mejorando hasta alcanzar una madurez muy elevada de forma rápida.
- **Capacidad de aplicaciones complementarios;** En términos generales la operativa alrededor de la toma de decisiones y actuación posterior en función del resultado del análisis del visionado automático es fácilmente automatizable también utilizando herramienta completamente de mercado tales como herramientas de CRM (Customer Relationship Management) para conexión por mail / WhatsApp /... o Chatbots con el cliente o terceros (por ejemplo vigilantes físicos, policía y similar), además de la posibilidad de inculcación humana cuando se dan circunstancias específicas
- **Aspectos legales y regulatorios:** En este sentido las limitaciones generales se reducen al cumplimiento de la normativa de protección de datos ampliamente desarrolladas y en intensa evolución a medida que la tecnología y los usos sociales

evolucionan. Dependiendo de los sectores afectados puede haber implicaciones específicas como, por ejemplo, en los temas de seguridad en la medida en que para poder pedir la intervención a las fuerzas de orden público se requiere el cumplimiento de determinados requisitos y, actualmente, no puede ser una comunicación de forma automatizada la que sea la base de una intervención, aspecto que puede resolverse con intervención humana en el último eslabón del proceso y solo en los casos en que se requiera esta intervención humana (en tanto no evolucione la regulación)

- **Capacidad de inversión y apuesta empresarial transformadora:** El desarrollo de un modelo de visionado automatizado para aplicar de forma generalizada supone no solo una elevada inversión en CAPEX, y en menor medida en OPEX, sino una apuesta empresarial por un cambio disruptivo de modelo operativo y de la propuesta de valor empresarial. En definitiva, no es solo una fuerte inversión económica sino una apuesta por liderar un cambio que sin duda se producirá.

5.4 Análisis de rentabilidad coste-beneficio

En línea con las ventajas operativas descritas en el apartado 5.1 sobre ventajas operativas y económicas de la utilización de análisis de video automatizado, el análisis de la rentabilidad coste-beneficio de la evolución del modelo de operación debería considerar lo siguiente:

5.4.1 Costes operativos recurrentes (OPEX):

Se produciría las siguientes reducciones costes operativos:

- Equipos humanos asignados específicamente a cada instalación en la medida en que el análisis automatizado permita la eliminación total o la reducción de la cobertura, bien en términos de horarios cubiertos o en términos de número de personas asignadas en cada turno horario. En la medida en que en la actualidad la situación habitual es que en cada instalación hay permanentemente, al menos, dos personas (a veces combinando personas con categoría de auxiliar y vigilantes certificados) y en todo a en gran parte del horario se podría reducir a uno (o en casos a ninguno), el ahorro sería enorme.

- Equipos humanos asignados a las CRA que se reducirían de forma drástica al ser necesaria la visualización humana solo en casos puntuales que requieran confirmación.
- La mejora de las herramientas de visualización que reduzcan el número de falsas alarmas también permitiría reducir los equipos humanos disponibles para acudir a localizaciones específicas en caso de necesidad.
- La reducción de tamaño de las compañías en cuanto a número de empleados y complejidad de gestión permitiría también generar ahorro de los costes generales de forma muy relevante, especialmente en algunos departamentos como el de RRHH, donde las tareas de selección, formación, supervisión, etc. se reducirían drásticamente.

Se produciría los siguientes aumentos de costes operativos:

- Equipos humanos y software necesario para gestionar y mantener el enorme incremento de capacidad de procesamiento de datos y la automatización de las actividades y coste de mayor volumen de almacenamiento y procesamiento de datos en la nube. Estos incrementos de costes, si bien relevantes, no serían en absoluto comparables a los ahorros identificados anteriormente y, además, tendrían un componente de alta escalabilidad por la cual incremento de volumen de actividad básicamente solo requeriría aumentos de capacidad de almacenamiento y proceso de datos (que de forma lógica se realizaría en la nube).

5.4.2 Incremento de inversión (CAPEX)

La automatización del análisis de video y de todo el proceso de toma y ejecución de decisiones requerirá lógicamente de una mayor inversión en equipamiento a dos niveles:

A nivel de empresa: donde será necesario invertir en el incremento significativo de la capacidad de proceso de datos y en el desarrollo o adquisición de herramientas de sistemas de muy diverso para integrar y completar las actividades. Estas inversiones tienen

normalmente un componente de CAPEX y de OPEX, como se ha indicado en el apartado anterior, y se pueden financiar de forma diversa acudiendo a: adquisición directa, renting, leasing o directamente servicios en la nube o SAS.

A nivel de cada instalación individual: donde será necesario, en determinados casos, aumentar la capacidad de computación local incluyendo capacidades de “Edge computing” y, en los casos de reducción de capacidad de vigilancia “in situ” que requerirá de una mayor cobertura de cámaras, grabadores y equipos de proceso que asegure que la eliminación de personal asignado directamente no supone una pérdida de cobertura. Los costes de estos equipos en la medida en que se generalicen y de acuerdo con la tendencia experimentada en los últimos años de forma continuada incrementarán su capacidad y reducirán su coste.

De forma adicional hay que considerar otro coste extraordinario, cuyas características son realmente de inversión “one-shot” que es el de **transformar las operaciones de las compañías**. Como se ha comentado con anterioridad en el trabajo, la utilización de forma generalizada de análisis de video automatizado provocará un cambio completamente disruptivo en las operaciones de las compañías, en su modelo y tamaño de organización y en las propias dinámicas competitivas de la industria. Esta transformación requerida una inversión importante en el diseño y ejecución de la transformación en general y, específicamente, en la drástica reducción y transformación de los equipos humanos.

Por último, está la **inversión necesaria en el desarrollo y entrenamiento del modelo** que requerirá de evolución permanente y que dependerá de la tecnología de desarrollo y entrenamiento seleccionada y de la capacidad de generación de los datos necesarios, en loas con las consideraciones técnica analizadas al largo del trabajo. En este caso, también es una inversión muy relevante inicualmente que tiene un componente de escalabilidad que hace que las inversiones posteriores en la mejora del modelo sean progresivamente inferiores y, en parte, realizadas con recursos sin coste de la experiencia de la propia operación ordinaria.

5.4.3 Mejora de la propuesta de valor e impacto en precios para los clientes

Las distintas mejoras comentadas a lo largo del trabajo que supondría la utilización generalizada de sistemas de análisis de video automatizado suponen una clara mejora de la propuesta de valor en la medida en que se incrementaría y mejoraría el análisis y se aseguraría una cobertura 24x7 mucho más consistente y con una drástica reducción de errores. Consecuentemente, se produciría una mejora sustancial para los clientes de los servicios recibidos.

En términos de precios, si bien la mejora de la propuesta de valor podría entenderse como la capacidad de incrementar precios, la realidad de la enorme mejora de márgenes para las empresas de seguridad y la reducción de la presencia “in situ” de equipos humanos en las instalaciones supondrá una profunda reducción de precios para los consumidores; que dependiendo de las dinámicas competitivas de la industria será mayor o menor y el balance en términos de cuanto del valor generado por la transformación se queda en las empresas o se traslada a los clientes.

Otro efecto destacable de la bajada de precios y de la mejora de la propuesta de valor será **la consolidación del mercado** en torno a las grandes empresas que serán las únicas que podrán competir, salvo aspecto de nicho muy locales, de una forma eficiente. Esto ayudará también a que la industria tenga una mayor estabilidad al reducir en parte la capacidad de presión de los clientes y reducir drásticamente la rivalidad entre competidores y la fragmentación del sector.

5.4.4 Impacto de la rentabilidad y tamaño de la industria

De acuerdo con lo planteado anteriormente, la redefinición de las dinámicas de la industria que supondrá un cambio tan disruptivo provocará un aumento de rentabilidad de las grandes empresas y un efecto de consolidación del mercado por parte de estas, aumentando su cuota de mercado. La magnitud de la mejora de rentabilidad y, por tanto, de la creación de valor dependerá de la forma en que la industria y, especialmente, las grandes empresas que sean pioneras acometan la transformación.

Por otro lado, hay un efecto adicional también claro que es aumento del volumen de actividad que permitiría la reducción de precios de forma que instalaciones que actualmente no están cubiertas por el elevado coste en relación con los beneficios posibles, puedan pasar a tener cobertura con el efecto de incremento de volumen del mercado.

El incremento del mercado por la mayor accesibilidad a estas soluciones tendrá también un impacto muy positivo en términos sociales al mejorar claramente la seguridad general del país y de la población en general y al hacer que determinados servicios públicos puedan aumentar de forma significativa su cobertura.

6 Conclusiones y trabajos futuros

6.1 Conclusiones

Basándonos en el trabajo realizado, podemos establecer dos líneas de conclusiones y una reflexión final.

En el lado técnico/tecnológico las soluciones para la utilización de análisis de video automatizado están muy avanzadas y de hecho son completamente factibles en la actualidad, aunque todavía adolecen de fundamentalmente dos problemas: a) el elevado volumen de computación que todavía requieren y b) la necesidad de invertir enormes recursos en los entrenamientos de los modelos apoyados en ML. Las actuales líneas de trabajo para la resolución de estos problemas están en vías: a) el aumento y abaratamiento de la capacidad de computación y la simplificación de los sistemas de entrenamiento de los modelos que, en parte, avanza en el tiempo de forma rápida con el avance general de capacidad de computación y b) la utilización de Edge Computing en las propias cámaras o en sistemas locales que permitan reducir el número de eventos que requieren computación en origen (en la nube) y de esta forma reducir significativamente las necesidades de computación en origen y de intercomunicación. En definitiva, se trata de aplicar con Edge Computing un esquema de Pareto en el que los análisis más sencillos son realizados de forma local y de esta forma se reduce enormemente el volumen de eventos analizados y se profundiza solo en los más relevantes o complejos de analizar.

En el lado más relacionado con la implantación en la industria de la vigilancia y seguridad de instalaciones, y en otros de los sectores que se han analizado; la implantación generalizada de análisis de video automatizado supondrá, en su momento, un cambio disruptivo completo transformado completamente el sector y las dinámicas de las empresas y de estas con los clientes, abaratando precios y mejorando coberturas y, transformado completamente las operaciones con un efecto también paralelo de reducción drástica y evolución del empleo en el sector.

A pesar de que en los próximos años se acometerá la generalización del uso de análisis de video automatizado y se transformará completamente el sector dadas las inmensas ventajas; en la actualidad el acceso a la tecnología y la inversión necesaria para hacerla realidad no es una opción todavía para las empresas del sector, ni nivel España ni a nivel global. Es necesario, que se siga avanzando técnicamente y se sigan reduciendo los costes para que las grandes empresas del sector puedan asumir la inversión necesaria en la transformación total de las operaciones.

En este sentido y tal como se ha comentado a lo largo de este trabajo, el desarrollo de técnicas de análisis de video y relacionadas en otros sectores en los que si existe la capacidad de inversión y hay un procesos muy intenso de desarrollo actualmente en marcha, **como es el caso de la conducción autónoma**, puede de forma colateral facilitar el contexto para que se acometa en la industria de la vigilancia y seguridad de instalaciones la implantación de un modelo de operaciones basado en análisis de video automatizado.

6.2 Trabajos futuros – ADE

La previsible transformación disruptiva que se producirá en el futuro en el sector de la vigilancia y seguridad de instalaciones será de una magnitud enorme para el sector y esto abre varias vías de profundización en el trabajo para entender y dimensionar algunos de los efectos que puedan producir:

- Por el lado del modelo de operaciones y su eficiencia donde se requerirá un cambio de modelo de organización con una estructura de costes completamente diferente.
- Por el lado de la propuesta de valor, en la que se evolucionara de forma muy significativa y actividades/instalaciones que actualmente no están cubiertas por no ser posible asumir el coste lo podrán estar, además de analizar el impacto en términos de evolución de la propuesta de valor y de reducción de precios para los clientes actuales.

Conclusiones y trabajos futuros

- Por el lado de las relaciones en el sector y la capacidad de este para aumentar la capacidad de creación de valor, puesto que se producirá una redefinición de la industria con una fuerte concentración empresarial y la creación de unas barreras de entrada y escalas completamente diferentes.
- Por el lado más social, puesto que se producirá una muy significativa reducción del empleo en el sector y se evolucionará también, en parte, los perfiles actuales del mismo.

En definitiva, el análisis realizado claramente muestra que el sector se transformará completa y disruptivamente en un futuro todavía por definir y el análisis de cómo y cuándo se producirá presenta un enorme interés de análisis académico anticipando la evolución futura.

7 Glosario

- **Análisis de video:** una tecnología que permite analizar de manera rápida y eficiente las imágenes de una cámara de seguridad, detectando objetos y personas.
- **Aproser:** Asociación Profesional de Compañías de Servicios de Seguridad.
- **CCAA (Cuentas Anuales):** Estados financieros anuales de una compañía completos.
- **CCTV:** Circuito Cerrado de Televisión, funciona tras la instalación de un sistema de cámaras de seguridad que permiten comprobar desde otra ubicación (en forma remota), el funcionamiento o el estado de un lugar.
- **Chatbot:** Software que interactúa con clientes de una forma preestablecida y entrenada bien oralmente o a través de texto.
- **Customer Relationship Management (CRM):** Herramienta de software para la gestión automatizada de relaciones con clientes.
- **Deep Learning, o aprendizaje profundo,:** es un método de inteligencia artificial (IA) que enseña a las computadoras a procesar datos de una manera inspirada en el cerebro humano. Los modelos de aprendizaje profundo pueden reconocer patrones complejos en imágenes, texto, sonidos y otros datos para producir conocimientos y predicciones precisos.
- **EBIT:** Earnings Before Interest and Taxes, beneficio antes de intereses e impuestos.
- **EBITDA:** Earnings Before Interest, Taxes, Depreciation and Amortization.
- **FPS (Frames o fotogramas por segundo):** el número de imágenes unidas que se necesitan para crear un segundo de vídeo.
- **Frame o fotograma:** imagen concreta dentro de una sucesión de imágenes en movimiento.
- **Gen AI (Generative AI):** modelos de inteligencia artificial diseñados para generar nuevo contenido en forma de texto escrito, audio, imágenes o videos.
- **HAR (Human Action Recognition Methods):** una rama de la ciencia e ingeniería computacional que intenta crear sistemas y técnicas capaces de reconocer y categorizar automáticamente las acciones humanas basándose en datos de sensores

- IA (AI) o Inteligencia Artificial: es un campo de la informática que se enfoca en crear sistemas que puedan realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la percepción.
- IVA (Intelligent Video Analytics): tecnología que utiliza la IA para el análisis y gestión de un alto contenido en videos.
- Latencia: Medida de la velocidad de respuesta entre dos dispositivos
- Market capitalization: Valor de mercado en un momento dado de una compañía cotizada.
- ML (Machine Learning): disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones.
- NFD: Net Financial Debt: deuda financiera neta.
- NU (Naciones Unidas): Organización mundial.
- ODS (Objetivos de Desarrollo Sostenible): constituyen un llamamiento universal a la acción para poner fin a la pobreza, proteger el planeta y mejorar las vidas y las perspectivas de las personas en todo el mundo.
- Pixels: el elemento más pequeño de una imagen reproducida digitalmente.
- Pose-based methods: métodos basados en detectar y rastrear la posición y orientación de partes del cuerpo humano en imágenes o videos.
- Python: lenguaje de programación ampliamente utilizado en las aplicaciones web, el desarrollo de software, la ciencia de datos y el machine learning (ML).
- Redes Neuronales: Son un modelo de creación cuyo sistema se basa en el funcionamiento del cerebro humano. Están formadas por diferentes nodos que funcionan como neuronas, y que transmiten señales e información entre sí.
- ROI (Región de Interés): es una parte de una imagen que se desea filtrar o procesar de alguna manera.
- ROI (Return on Investment): medida para valorar la eficiencia de una inversión.
- SAS (Software as a Service): soluciones de software cuya actividad se realiza completamente en la nube y no requieren implantaciones “on-premise” para las empresas o particulares.

-
- SGD (Stochastic Gradient Descent): algoritmo de optimización comúnmente utilizado en el entrenamiento de modelos de aprendizaje automático.
 - Videgrabador: dispositivo de grabación de vídeo en formato digital.
 - VS Studio, Visual Studio Code: editor de código utilizado.

8 Bibliografía

Agencia Estatal Boletín Oficial del Estado (BOE). (2014). *Ley 5/2014, de 4 de abril, de Seguridad Privada*. Agencia Estatal Boletín Oficial del Estado (BOE). Recuperado de <https://www.boe.es/buscar/act.php?id=BOE-A-2014-3649>

Agencia Estatal Boletín Oficial del Estado (BOE). (2024). *Resolución de 19 de marzo de 2024, de la Dirección General de Trabajo, por la que se registra y publica el II Convenio colectivo de empresas de servicios auxiliares de información, recepción, control de accesos y comprobación de instalaciones*. Agencia Estatal Boletín Oficial del Estado (BOE). Recuperado de https://www.boe.es/diario_boe/txt.php?id=BOE-A-2024-6406

Algotive. (2022a). *Guía de Inteligencia Artificial Autónoma: El futuro de la IA*. Algotive. Recuperado de <https://www.algotive.ai/es-mx/blog/gu%C3%ADa-de-inteligencia-artificial-aut%C3%B3noma-el-futuro-de-la-ia>

Algotive. (2022b). *La guía definitiva de la analítica de video: La tecnología que definirá a esta década*. Algotive. Recuperado de <https://www.algotive.ai/es-mx/blog/la-guia-definitiva-de-la-analitica-de-video>

Allied Universal . (2022). *ESG Report 2022*. Allied Universal . Recuperado de https://www.aus.com/sites/default/files/2023-06/1023999840_2022-ESG-Report-FINAL.pdf

Al-Sahaf, H., Bi, Y., Chen, Q., Lensen, A., Mei, Y., Sun, Y., . . . Zhang , M. (2021). A Survey on Evolutionary Machine Learning. *Victoria University of Wellington*,. Recuperado de <https://www.tandfonline.com/doi/full/10.1080/03036758.2019.1609052>

Aproser (2022). *El sector de la Seguridad Privada en España. Datos y Cifras 2022*.

Aproser. Recuperado de <https://www.aproser.es/web/>

Badidi, E., Moumane, K., & El Ghazi, F. (s.f.). *Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review*. IEEE Access. Recuperado de

<https://ieeexplore.ieee.org/document/10198424>

Chityala, R., & Pudipeddi, S. (2021). *Image Processing and Acquisition using Python*.

CRC Press.

Continox point for fichet . (s.f.). ¿Cuántas cámaras de videovigilancia hay en España?

Continox Servi, S.A. . Recuperado de <https://www.continox.es/cuántas-cámaras-de-videovigilancia-hay-en-espana/>

Duan, L., Lou, Y., Wang , S., Gao, W., & Rui, Y. (2019). *AI-Oriented Large-Scale Video Management for Smart City: Technologies, Standards, and Beyond*. IEEE

MultiMedia, vol. 26, no. 2, pp. 8-20. Recuperado de <https://ieeexplore.ieee.org/document/8509149/authors>

El Economista. (2024). *Ranking de Empresas Españolas por facturación* . El Economista.

Recuperado de <https://ranking-empresas.eleconomista.es/>

Feng, W., Ji, D., Wang, Y., Chang, S., Ren, H., & Gan, W. (2018). *Challenges on Large Scale Surveillance Video Analysis*. China : Computer Vision Foundation.

Recuperado de

https://openaccess.thecvf.com/content_cvpr_2018_workshops/papers/w3/Feng_Challenges_on_Large_CVPR_2018_paper.pdf

- Fernandez, P., Jimenez, J., Astarloa , A., Idirin , M., & Salas , S. (s.f). *Accelerating Video Analytic Processing on Edge Intelligence*. IEEE. Recuperado de <https://ieeexplore.ieee.org/document/9970152>
- Francia I. (2024). *Código programa de prueba TFG I.F.* Obtenido de Git Hub: <https://github.com/i13fran/TFG-I.F>.
- Ghozia, A., Attiya, G., Adly, E., & El-Fishawy, N. (2020). *Intelligence Is beyond Learning: A Context-Aware Artificial Intelligent System for Video Understanding*. Hindawi. Recuperado de <https://onlinelibrary.wiley.com/doi/10.1155/2020/8813089>
- Green, M. W. (1999) *The Appropriate and Effective Use of Security Technologies in U.S. Schools, A Guide for Schools and Law Enforcement Agencies*, Sandia National Laboratories. Recuperado de <https://nij.ojp.gov/library/publications/appropriate-and-effective-use-security-technologies-us-schools>
- Guruvarreddiar, P., & Prasad, P. (2020). *Artificial Intelligence based region of interest enhanced video compression*. Intel Corporation. Recuperado de <https://ieeexplore.ieee.org/document/9105758>
- Howse, J., Prateek , J., & Beyeler , M. (2016). *OpenCV: Computer Vision Projects with Python*. Packt .
- HUTCHINSON, M., & GADEPALLY, V. (2021). *Video Action Understanding*. IEEE. vol. 9, pp. 134611-134637. Recuperado de <https://ieeexplore.ieee.org/document/9548074>
- IBM. (s.f). *¿Qué son las redes neuronales?* IBM. Recuperado de <https://www.ibm.com/es-es/topics/neural-networks>

Ibrahim, N., Maurya, P., Jafari, O., & Nagarkar, P. (2021). A Survey of performance optimization in neural network based video analytics systems. A Preprint. Recuperado de <https://arxiv.org/abs/2105.14195>

International Trade Administration. (2024). *Country commercial guides - Spain - Safety and Security*. Recuperado de <https://www.trade.gov/country-commercial-guides/spain-safety-and-security>

Iqbal Khan, A., Afzal, S., Ahmad Bhat, F., & Wani, A. (2020). *Advances in deep learning*. Springer. Recuperado de <https://link.springer.com/book/10.1007/978-981-13-6794-6>

Kulbacki, M., Segen, J., Chaczko, Z., Rozenblit, J., Kulbacki, M., Klempous, R., & Wojciechowski, K. (2023). *Intelligent Video Analytics for Human Action Recognition: The State of Knowledge*. Sensors. MDPI. Recuperado de <https://www.mdpi.com/1424-8220/23/9/4258>

Li, X., Fan, Z., Liu, Y., Li, Y., & Dai, Q. (2019). *3D Pose Detection of Closely Interactive Humans Using Multi-View Cameras*. MPDI. Recuperado de <https://www.mdpi.com/1424-8220/19/12/2831>

Lee, T. B. (2021) *Tesla Autopilot director contradicts Musk's self-driving timeline*. Arstechnica. Recuperado de <https://arstechnica.com/cars/2021/05/tesla-autopilot-director-contradicts-musks-self-driving-timeline/>

Monfort, M., Andonian, A., Zhou, B., Ramakrishnan, K., Adel Barga, S., Yan, T., . . . Oliva, A. (2019). *Moments in Time Dataset: one million videos for event understanding*. Boston: Massachusetts Institute of Technology. Recuperado de <https://arxiv.org/abs/1801.03150>

Montes, R., Melero, F., Palomares, I., Alonso, S., Chiachio, J., Chiachio, M., . . . Herrera, F. (2021). *Inteligencia Artificial y Tecnologías Digitales para los ODS*. Real Academia de Ingeniería pp. 40, 181 y 208. Recuperado de https://issuu.com/raing/docs/ia_y_tecnolog_as_digitales_para_los_ods

Moroney, L. (2020). *AI and Machine Learning for Coders*. O'Reilly Media, Inc. Recuperado de <https://www.oreilly.com/library/view/ai-and-machine/9781492078180/>

Naciones Unidas. (2024a). Objetivo 7 - Energía asequible y no contaminante. *Naciones Unidas*. Recuperado de <https://www.undp.org/es/sustainable-development-goals/energia-asequible-no-contaminante>

Naciones Unidas. (2024b). Objetivo 8 - Trabajo decente y crecimiento económico. *Naciones Unidas*. Recuperado de <https://www.undp.org/es/sustainable-development-goals/trabajo-decente-crecimiento-economico>

Naciones Unidas. (2024d). Objetivo 9 - Industria, innovación e infraestructura. *Naciones Unidas*. Recuperado de <https://www.undp.org/es/sustainable-development-goals/industria-innovacion-infraestructura>

Naciones Unidas. (2024e). Objetivo 11 - Ciudades y comunidades sostenibles *Naciones Unidas*. Recuperado de <https://www.undp.org/es/sustainable-development-goals/ciudades-comunidades-sostenibles>

Naciones Unidas . (2024c). *Objetivos de desarrollo sostenible* . Obtenido de Naciones Unidas . Recuperado de <https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/>

Naciones Unidas. (s.f.). *Objetivo 11: Lograr que las ciudades sean más inclusivas, seguras, resilientes y sostenibles* . Naciones Unidas. Recuperado de <https://www.un.org/sustainabledevelopment/es/cities/>

- Nawaratne, R., Alahakoon, D., De Silva, D., & Yu, X. (2020). *Spatiotemporal Anomaly Detection Using Deep Learning for Real-Time Video Surveillance*. IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. Recuperado de <https://ieeexplore.ieee.org/document/8820090>
- Nguyen, M., Truong, L., Tan, T., & Chien, C.-F. (2020). *Artificial intelligence based data processing algorithm for video surveillance to empower industry*. Elsevier Ltd. Recuperado de <https://www.sciencedirect.com/science/article/abs/pii/S0360835220304058>
- Nielsen, M. (2019). Chapter 1 - Using neural nets to recognize handwritten digits. *Lambda*. Recuperado de <http://neuralnetworksanddeeplearning.com/chap1.html>
- Open CV. (2023). *Open CV*. Recuperado de <https://opencv.org/>
- Parker, J. (2011). *Algorithms for Image Processing and Computer Vision*. Wiley Publishing, Inc. Recuperado de http://kiwi.bridgeport.edu/cpeg585/Algorithms_for_Image_Processing_and_Computer_Vision.pdf
- Porter, M. E. (1979). The Structure within Industries and Companies' Performance. *The Review of Economics and Statistics*, 61(2), 214–227. Recuperado de <https://doi.org/10.2307/1924589>
- Prosegur (2020). Prosegur IR Corporate 230920. *Prosegur*. Recuperado de <https://www.prosegur.com/accionistas-inversores/comunicaciones-cnmv>

- Prosegur (2023). Informe de auditoría de cuentas anuales consolidadas. Prosegur. Recuperado de <https://www.prosegur.com/dam/jcr:3fd5aca0-1aca-4264-8650-a376ed91c70c/CCAACC%202023%20PCS.pdf>
- Python. (s.f.). *The Python Standard Library* . Recuperado de <https://docs.python.org/3/library/>
- Santafe, M. (2021). Cómo la tecnología de vigilancia puede ayudar a las ciudades a alcanzar sus objetivos de sostenibilidad. *AXIS Communications* . Recuperado de <https://newsroom.axis.com/es-es/article/como-la-tecnologia-de-vigilancia-puede-ayudar-a-las-ciudades-a-alcanzar-sus-objetivos-de>
- Securitas. (2024). *Securitas invites to an Investor Day in Stockholm on March 7, 2024*. Securitas. Recuperado de https://www.securitas.com/en/newsroom/press-releases_list/securitas-capital-markets-day-march-7-2024--save-the-date/
- Securitas (2023). *2023 Annual and Sustainability report*. Securitas. Recuperado de <https://www.securitas.com/en/newsroom/regulatory-press-releases/securitas-publishes-annual-and-sustainability-report-for-2023/>
- Sengönül , E., Samet, R., Abu Al-Hajja, Q., Alqahtani, A., Alturk, B., & A. Alsulami, A. (2023). An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey. *Ke Gu. MDPI*. Recuperado de <https://www.mdpi.com/2076-3417/13/8/4956>
- Simple CV. (2023). *SimpleCV*. Recuperado de <https://github.com/sightmachine/SimpleCV>
- Stackoverflow. (2023). *2023 Developer Survey* . Recuperado de <https://survey.stackoverflow.co/2023/#overview>
- Stig Hansen, U., Landau, E., Patel, M., & Hayee, B. (2021). *Novel artificial intelligence-driven software significantly shortens the time required for annotation in computer*

vision projects. Endosc Int Open. Recuperado de
<https://pubmed.ncbi.nlm.nih.gov/33869736/>

Tesla, Inc., ed. (2024). *Tesla Vehicle Safety Report*. Tesla. Recuperado de
https://www.tesla.com/es_es/VehicleSafetyReport

Yedda.ai (2024). Use AI to measure, drive action and improve. Yedda.ai. Recuperado de
<https://www.yedda.ai/>

ZHANG, Q., SUN, H., WU, X., & ZHONG, H. (2019). Edge Video Analytics for Public Safety: A Review. *IEEE*. Recuperado de
<https://ieeexplore.ieee.org/document/8781894>

9 Anexo I: Objetivos de desarrollo sostenible

El 25 de septiembre de 2015, los líderes mundiales adoptaron un conjunto de objetivos globales para erradicar la pobreza, proteger el planeta y asegurar la prosperidad para todos como parte de una nueva agenda de desarrollo sostenible. Cada objetivo tiene metas específicas que deben alcanzarse en los próximos 15 años. Se definieron 17 Objetivos de Desarrollo Sostenible (en adelante ODS), que se puede clasificar en tres dimensiones, social, económica y medioambiental. El trabajo esta alineado con 4 de los ODS establecidos por las Naciones Unidas (NU) en 2015 con la finalidad de contribuir a un futuro más sostenible, estos objetivos orbitan alrededor de la eliminación de la pobreza, la consecución de la paz y la protección del planeta.



ILUSTRACIÓN 17. ODS CON LOS QUE SE ALINEA EL TRABAJO (NACIONES UNIDAS, 2024C)

La aplicación de modelos de inteligencia artificial en el procesado de imágenes tiene multitud de usos en diferentes campos, uno de ellos es el control de plantas de energía renovable, tanto fotovoltaicas como eólicas, que por su gran extensión como por su ubicación en zonas remotas se beneficia profundamente del control por medio de cámaras o drones. La capacidad de realizar ese control o mantenimiento robótico permitirá crear instalaciones allá donde más recursos naturales haya, aunque sea en zonas altamente remotas, como desiertos o alta mar. El entorno ayudará en tareas como pueden ser la detección de intrusiones o anomalías, además de incrementar la producción y abaratar los costes de la productora lo cual repercute en el consumidor final, con eso contribuirá al

Anexo I: Objetivos de desarrollo sostenible

séptimo objetivo de las naciones unidas de la dimensión medioambiental centrado en la consecución de una *energía asequible y no contaminante*. (Naciones Unidas, 2024a)

De la misma manera, la aplicación de la IA en el procesado de imágenes también está en línea con el octavo objetivo de dimensión económica y social: *trabajo decente y crecimiento económico*. (Naciones Unidas, 2024b) Estas tecnologías sustituirán trabajos mal remunerados y que ofrecen malas condiciones como los de vigilante de seguridad en zonas remotas, especialmente en los turnos de noche, por empleos de calidad que estarían mejor remunerados. El uso de la tecnología y procesamiento de imágenes para monitorizar actividades como la agricultura y la ganadería, sobre todo en países menos desarrollados, podría permitir liberar a la población juvenil e infantil de realiza estas labores y dedicarse más a formación. (Montes et al, 2021)

La versatilidad del procesado de imágenes también se alinea con el noveno objetivo de dimensión económica: *industria, innovación e infraestructura*. (Naciones Unidas, 2024d). El video análisis apoyado en inteligencia artificial propone un avance en seguridad en el más amplio de los sentidos, permitirá avances como poder controlar en tiempo real la cantidad de personas que están dentro de una estación o como controlar intrusiones en grandes recintos como puede ser una fábrica. Otro gran avance en infraestructura será el control del estado de esta, mediante cámaras se podría llegar a detectar si se está formando una fractura una carretera, a controlar si existe alguna obstrucción en las vías de un tren o si una zona de una ciudad tiene un tráfico excesivo. El mantenimiento en remoto predictivo semiautomático permitirá no desplazar, en ocasiones, personal al terreno con el consiguiente ahorro económico y mejor impacto medioambiental, La tecnología por sí sola ya supone innovación y un gran avance en las tecnologías de la información, pero como vemos, sirve a la vez de catalizador para que otras industrias puedan avanzar al hacerlas más eficientes. (Montes et al, 2021)

Asimismo, el procesado de imágenes puede ayudar a la consecución del decimoprimer objetivo: *ciudades y comunidades sostenibles*. Especialmente importante puede ser el mantenimiento de la seguridad pública para una ciudad inteligente (Smart Cities), la

Anexo I: Objetivos de desarrollo sostenible

videovigilancia de espacios públicos, que favorece la seguridad ciudadana. En esta misma línea, se podría favorecer este objetivo con la detección y prevención de situaciones de emergencias, que puede pasar por el rastreador de personas, siempre con las medidas de control suficientemente reguladas para no poner en riesgo la privacidad (por ejemplo puede ser una buena medida para seguimiento de aventureros a la hora de tomar decisiones de rescate en casos necesarios), de animales (mascotas que se pierden al dar un paseo por el campo) o de ganado, que puede mejorar la eficiencia y productividad, evitando robos y pérdidas. (Montes et al, 2021).

En definitiva, la IA aplicada al procesado de imágenes supone un avance para toda la población y permitirá que muchas industrias, destacando la energética en este proyecto, se apoyen en ella y ganen en eficiencia. Los recursos que se ahorrarán en este ámbito podrán ir destinados a otros campos, generarán eficiencias en todas las industrias (al abaratar el coste de los insumos) y habrá un avance muy notable en seguridad lo cual repercute en el bienestar la población. Se pueden reducir las desigualdades si se extienden estos avances a los países menos desarrollados con grandes recursos naturales.