

Cámaras con Inteligencia Artificial para las vías públicas de la CdMx

Adolfo Guzmán Arenas *

Resumen: La red de cámaras digitales de la Ciudad de México fortalece notablemente la seguridad de sus habitantes, a través de personas vigilantes que observan las imágenes transmitidas desde las cámaras a los centros de monitoreo (llamados C5, C2 y C2M), y detectan situaciones anómalas o actividades sospechosas que ocurren en la vasta metrópoli mexicana. Con esta importante inversión en recursos tecnológicos y humanos no sólo se apoya la administración de la seguridad de la megalópolis, sino se da un paso importante hacia la ciudad inteligente.

Es posible continuar más allá de la detección manual de eventos anómalos o atípicos. La introducción de la computación y la inteligencia artificial (especialmente la visión por computadora) en esta red hará posible que se detecten y registren más y mejor estos eventos, con mayor precisión y rapidez, siendo un valioso apoyo a su análisis manual. El presente artículo detalla cómo puede ocurrir esta transición.

* Ingeniero en Comunicaciones y Electrónica de la ESIME del IPN. Doctor en Ciencias de la Computación en el MIT (Cambridge, Mass.). Fue profesor del Dep. de Ingeniería Eléctrica del MIT; del Dep. de Inteligencia Mecánica de la Universidad de Edimburgo; del CINVESTAV del IPN, donde fundó la Maestría y Doctorado en Computación; del Instituto de Investigación en Matemáticas Aplicadas y Sistemas, de la UNAM, y de la Unidad Interdisciplinaria (UPIICSA) del IPN. También ha tenido importantes cargos en empresas nacionales y del extranjero, especializadas en cómputo. Recibió el Premio Nacional de Ciencias y Artes (1996) y la Presea “Lázaro Cárdenas” (1997). Fundó en 1996 el Centro de Investigación en Computación (CIC) del IPN y lo dirigió hasta 2002. Es miembro del Sistema Nacional de Investigadores (Investigador Nacional Emérito). Actualmente se desempeña en el CIC.

Palabras Clave: Análisis de escenas, visión por computadora, Inteligencia Artificial, seguridad, CdMx.

Cameras with Artificial Intelligence for public roads in Mexico City

Abstract: The network of digital cameras in Mexico City significantly strengthens the security of its inhabitants, through employees who watch the images transmitted from the cameras to the monitoring centers (called C5, C2 and C2M), and detect anomalous situations or suspicious activities that occur in the vast Mexican metropolis. This important investment in technological and human resources not only supports an enhanced security of the megalopolis, but it becomes, too, an important step towards the digital city.

It is feasible to continue beyond manual detection of anomalous or atypical events. The introduction of computer science and artificial intelligence (especially computer vision) in this network will make it possible to detect these events better, more of them, with greater precision and speed, thus becoming a valuable support for their manual analysis. This article details how this transition can occur.

Keywords: Scene analysis, computer vision, Artificial Intelligence, security, Mexico City.

Fecha de recepción del artículo: 29 septiembre 2023

Fecha de aceptación: 16 octubre 2023

1. INTRODUCCIÓN

El progreso de las comunicaciones digitales y la miniaturización de los circuitos integrados ha hecho posible que, en avenidas, centros públicos, organizaciones privadas, comercios, hospitales, escuelas, etc., se instalen cámaras digitales (“circuitos cerrados de televisión”, solían llamarse) para vigilar en tiempo real lo que sucede, y registrarlo para su análisis posterior. Esta vigilancia la llevan a cabo personas que escrutan en una pantalla digital

hasta dieciséis imágenes (fotogramas, llamados también cuadros, imágenes estáticas o instantáneas), y dan aviso o toman acción cuando observan algo anómalo.

Las cámaras necesitan enviar sus fotogramas a una computadora central que las almacena y también las despliega en una o más pantallas para su análisis visual. Después de cierto tiempo (dos semanas, típicamente) la información almacenada se guarda en otro medio como dato histórico, o se desecha.

Cuando la persona vigilante detecta un evento que requiere acción inmediata, es posible reaccionar rápidamente cuando ella misma puede atender la anomalía o evento atípico (la entrada de un intruso al almacén, por ejemplo), si se encuentra cerca del lugar vigilado, o cuando puede dar aviso (por radio o dispositivo móvil) a personal que se encuentra en la zona del suceso. Si no es posible atender el evento en tiempo real, aún se cuenta con la evidencia grabada para acciones posteriores.

Toda esta inversión en equipo (cámaras, instalaciones, red de transmisión digital, pantallas, computadoras), personal operativo (personas vigilantes, encargado de las computadoras) y gastos recurrentes (renta de la fibra óptica o canal de transmisión, mantenimiento del equipo) redundan en un aumento notable en la seguridad del edificio, comercio, escuela o vía pública.

1.1 La red de cámaras de la CdMx

El gobierno de la ciudad de México ha instalado más de 63 mil cámaras de vigilancia en la vía pública (Sistema Tecnológico de Videovigilancia STVs y Mi C911e), con la finalidad de prevenir y alertar inmediatamente a las autoridades de seguridad y de emergencias capitalinas sobre cualquier situación de riesgo [<https://tinyurl.com/CamarasCdMx>].¹

Estas cámaras se comunican vía fibra óptica con el Centro de Comando, Control, Cómputo Comunicaciones y Contacto Ciudadano (C5), un edificio donde las personas

¹ Ligas como ésta señalan la página en internet donde yace el documento con la información citada, más información adicional. Se puede consultar y descargar simplemente pulsando la liga con el botón izquierdo del ratón.

vigilantes monitorean manualmente (a ojo) las cámaras, a través de pantallas digitales. Las cámaras se ubican en los Tótems del programa MiC911e y de los Sistemas Tecnológicos de Videovigilancia (STVs), mejor conocidos como los postes del C5.

El C5 ofrece los servicios de Video Monitoreo, Servicio de Atención de llamadas de emergencia 9-1-1 CDMX, Denuncia Anónima 089 y Difusión de la Alerta Sísmica. Todos los servicios del C5 operan las 24 horas, 365 días al año [<https://tinyurl.com/C5CdMx>].

Además, están los Centros de Comando y Control [C2, <https://tinyurl.com/C2deCdMx>], estaciones regionales encargadas del monitoreo de las cámaras de videovigilancia por zonas en la Ciudad de México (Centro, Norte, Centro Histórico). En los C2 operan personas vigilantes que monitorean manualmente (a ojo) las cámaras. Todos los C2 tienen conexión directa con el C5 para la coordinación y canalización adecuada de los incidentes.

También cuenta la CdMx con C2 móviles [C2M, <https://tinyurl.com/C2mCdMx>], vehículos con cámaras desplegadas que permiten el monitoreo en lugares de difícil acceso y el envío de imágenes en todo momento al C5. También cuenta con dos helicópteros no tripulados con videocámaras integradas, (MUAV por sus siglas en inglés), los cuales pueden monitorear zonas más remotas en eventos u operativos especiales, ya que tienen un alcance de sobrevuelo de hasta mil metros de altura.

1.2 Un paso más hacia la ciudad inteligente

- Los avances de la computación, el aprendizaje mecánico y la inteligencia artificial hacen posible que mucho del análisis de los fotogramas, actualmente llevado a cabo por personas vigilantes entrenadas para este fin, se complementa con software y aplicaciones informáticas que los examinen. Se apoya y mejora de esta manera la labor manual llevada a cabo. Por ejemplo, la computadora puede colocarle un recuadro rojo a un fotograma que considera atípico o anómalo, y sonar una pequeña alarma, para llamar la atención de la operadora vigilante, quien tomaría la acción final.

Con el análisis por computadora se facilitan y extienden las tareas de las cámaras:

- Con el análisis por computadora se facilitan y extienden las tareas de las cámaras: -) detección de personas; -) búsqueda de personas; -) detección de multitudes y tumultos; -) análisis de multitudes; -) reidentificación de personas; -) análisis de conducta, -) detección de situaciones adversas (árboles caídos, bloqueo de calles, basura en vía pública, charcos e inundaciones, autos abandonados...); -) seguridad; -) movimientos ilegales de vehículos (vuelta prohibida, cruce en rojo, ..., en intersecciones vehiculares (ver sección 3.1); -) seguimiento de personas o vehículos, siendo esta última tarea útil en edificios y lugares controlados. Consiste en asignar una identidad única a cada persona o vehículo en el vídeo e ir generando su trayectoria conforme se mueve a través del campo de visión de una videocámara, pudiendo extenderse este problema a más de una videocámara donde su identificador único y trayectoria deben continuar a través de cada una. Esta tarea se conoce como seguimiento en multicámaras. [<https://tinyurl.com/SeguimVariasCamaras>].

2. SOLUCIONES PROPUESTAS

El problema por resolver es cómo apoyar y complementar la labor manual de análisis de imágenes, para extender las tareas que la red de cámaras puede realizar, y mejorar así la seguridad en la urbe. ¿Cómo abordar la mayor cantidad de tareas posibles, mencionadas en la sección 1.2? Propongo usar aprendizaje mecánico, visión por computadora e inteligencia artificial como parte de la solución.

La suposición principal es que cada cámara observa siempre el mismo campo visual o escena (v. gr., la misma región de la calle). No se mueve ni cambia (zoom) su campo visual.

2.1 Solución básica. Análisis de todos los fotogramas de cada cámara

En esta solución, cada cámara envía sus fotogramas periódicamente (cinco cada segundo, por ejemplo) a una computadora asignada a ella, para su análisis. Una computadora (una PC o laptop típica) puede analizar simultáneamente los fotogramas de varias computadoras

(diez a veinte). Para las 63,000 cámaras, se necesitarían unas tres mil computadoras. Es inviable. Para acelerar el proceso de cada fotograma, se añade un filtro de parecido (ver figura 1) por software, que rápidamente compara el fotograma entrante con el fotograma anterior, y lo desecha (evita procesarlo más) si es muy parecido. Como dos fotogramas secuenciales tienen alta probabilidad de parecerse mucho, esto reduce hasta en más del 90% el trabajo de comparación detallada de la computadora. Con este truco, el número de computadoras se reduce a unas trescientas, que se pueden instalar repartidas en los siete C2s (sección 1.1) de la CdMx, correspondiendo en promedio cincuenta a cada C2. Ver figura 1.

Figura 1
Arquitectura del sistema básico

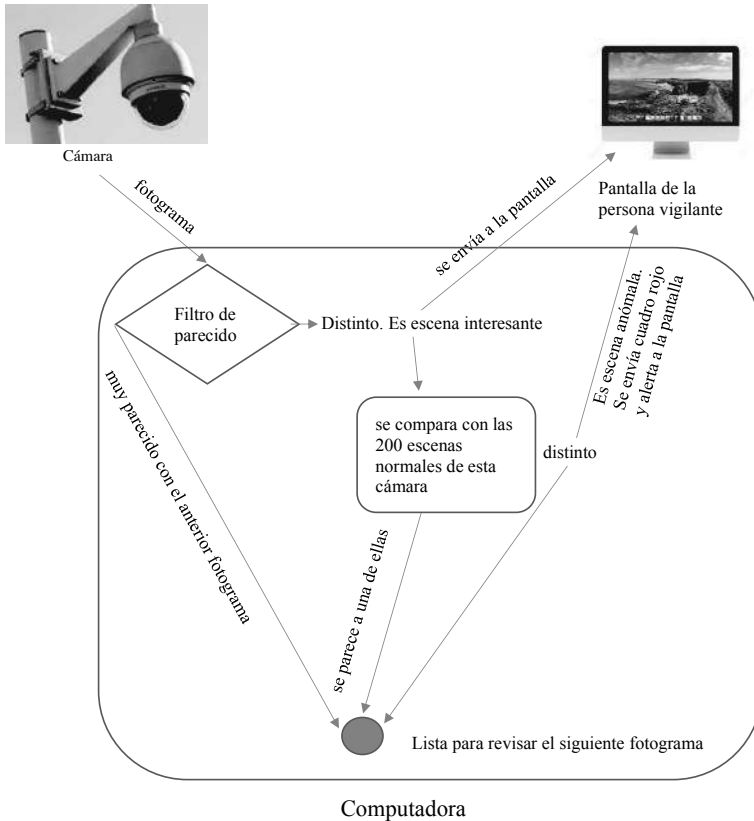


Fig. 1. Arquitectura del sistema básico. La adición a la entrada de la computadora del filtro de parecido (por software) permite que cada computadora maneje alrededor de 100 a 200 cámaras. En la figura aparece solo una cámara. Cada una de las escenas se puede reducir a un vector característico de tamaño mucho menor, un "resumen" de la escena, para reducir espacio y aumentar rapidez de procesamiento.

2.1.1 Arquitectura y funcionamiento de la solución básica

Como ya se dijo, no es posible ni conveniente que una sola computadora analice los fotogramas de todas las cámaras. Sería muy grande, formaría un cuello de botella y su falla colapsaría el análisis automático. Es mejor usar varias computadoras y colocarlas en forma descentralizada (en los C2s, por ejemplo, sección 1.1). Así, cada computadora sería responsable del análisis de un subconjunto de cámaras, “sus cámaras”.

Cuando un fotograma llega a la computadora para su análisis, ésta debe rápidamente de tectar si es “normal” o “anómalo”, y en este último caso, señalar con un cuadro rojo y una señal sonora el fotograma a la persona vigilante, para que ella tome la acción final.

Etapas de aprendizaje. La suposición principal es que cada cámara observa siempre el mismo campo visual o escena. La computadora previamente aprendió y tiene guardadas unas doscientas “escenas esperadas” correspondientes a una cámara (calle oscura, sin vehículos; calle con vehículos con luces frontales; calle con vehículos con luces rojas, alejándose; calle lluviosa; amanece en la calle; calle con polvo; ...) que son clasificadas como “normales” en esta etapa de aprendizaje mecánico.

Etapas de producción. Al llegar un nuevo fotograma a la computadora para su análisis, se compara rápidamente con el fotograma anterior de la cámara respectiva, y enseguida pasa a reemplazar al fotograma anterior. Llamemos a esta comparación “filtro de parecido”. Si hay pocos cambios en la escena, el fotograma se ignora. “Todo sigue igual”. No se analiza más por la computadora, ni se envía a la pantalla de la persona vigilante. Por ejemplo, es de noche, y no circulan vehículos.

Cuando hay un cambio significativo, se trata de una “nueva escena” (por ejemplo, entran varios automóviles con sus luces prendidas), que atraviesa el filtro. Se le denomina escena interesante. Se envía a la pantalla de la persona vigilante, y se compara contra las escenas normales guardadas previamente. Si resulta que es normal, el análisis termina. Si no coincide con alguna escena normal,

la escena interesante se considera anómala, y se envía un cuadro rojo a la pantalla de la persona vigilante, rodeando el fotograma anómalo. También se envía una señal para que suene una pequeña alerta sonora en la bocina de su pantalla. Ver figura 1.

Etapas de reaprendizaje. Con el transcurso del tiempo, cambia el campo visual de una cámara en la vida real: se erige un nuevo edificio, se pinta una pared de verde, etc. Esta escena se señalará como “anómala”, pero la persona vigilante puede reclasificarla como “normal” (es una nueva normalidad), para que la computadora la integre a su colección de escenas normales correspondientes a esa cámara.

2.1.2 Detalles técnicos

Un detalle: las comparaciones de fotogramas hechos por el filtro (etapa de aprendizaje) y por el análisis (etapa de producción) no se hacen pixel por pixel, sería muy lento. Usamos una especie de “hash” (dispersión al azar) que consiste en reducir cada fotograma a un vector de rasgos (alrededor de 15 rasgos numéricos), y sólo comparar los vectores mediante el producto interno o producto punto de ellos. Otro detalle técnico importante: cada cámara debe entrenarse manualmente, para asignarle unas doscientas escenas “normales”.

2.1.3 Problemas con la solución básica

La solución básica requiere de unas 300 computadoras (sección 2.1). Impone una alta tasa de transmisión de fotogramas a la red de conducción de información, normalmente fibra óptica. También impone una alta tasa de procesamiento a la computadora, porque tiene que filtrar (con el filtro de parecido, sección 2.1.1) cada fotograma que llega de cada una de sus cámaras, y analizar las escenas que resulten interesantes, para averiguar si son normales o anómalas. Aun así, es una solución viable. Empero,

puede mejorarse.

2.2 Solución mejorada. La cámara envía a la computadora solo fotogramas interesantes

Se puede mejorar la solución básica si eliminamos de la computadora el filtro de parecido y de alguna manera lo colocamos directamente en la cámara, o cerca de ella, para que sólo deje pasar a la computadora escenas interesantes: los fotogramas que difieren notablemente del fotograma anterior. Esto reduce ambas tasas de trabajo considerablemente, pues por la red de transmisión sólo viajan escenas interesantes, por lo que la computadora sólo procesa escenas interesantes. El filtrado por parecido que hacía el software de la computadora, ahora lo hace el filtro añadido a la cámara, que es mucho más veloz (que el software de la computadora).

Con la solución mejorada, la computadora sólo analiza escenas interesantes. Éstas le llegan a una tasa mucho más moderada, una cada diez segundos, exagerando. Las ventajas de esta solución son: -) se necesitan menos computadoras para analizar escenas interesantes y descubrir si son anómalas (estimo que la eliminación del filtrado por parecido en la computadora permitirá que ella procese de cinco a diez veces más escenas interesantes, dividiendo el número de computadoras necesarias mediante un divisor que oscila entre cinco y diez); -) se requiere rentar líneas de transmisión de información de menor velocidad (y menor costo). Queda por definir el tipo de filtro de parecido a usar en la cámara. Las secciones 2.2.1 y 2.2.2 definen dos tipos de filtro en la cámara, siendo mucho más barato y fácil de instalar el de la sección 2.2.2.

Figura 2
Arquitectura mejorada intermedia

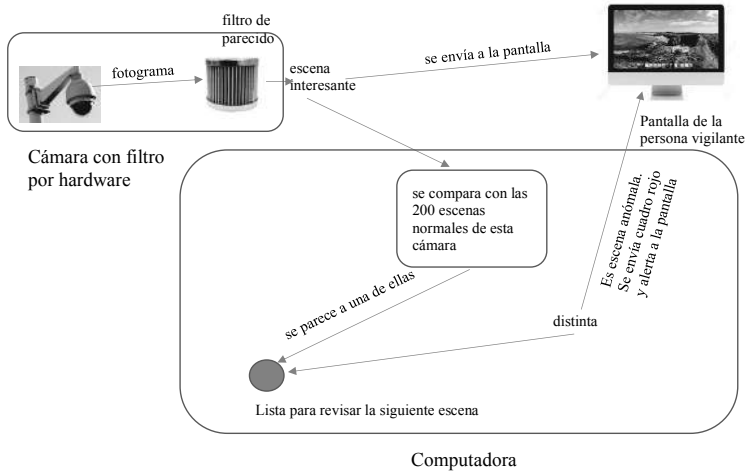


Fig. 2. Arquitectura mejorada intermedia. Ahora cada cámara está dotada de un filtro electrónico de parecido. A diferencia del esquema de la figura 1, cada computadora maneja ahora alrededor de mil cámaras, suponiendo que una cámara envía una escena interesante cada diez segundos. En un segundo la computadora recibirá cien escenas interesantes, que debe comparar para determinar si son normales o anómalas. En la figura aparece solo una cámara. Si reemplazamos el filtro por hardware por un detector de movimiento, llegamos a la arquitectura mejorada final, mucho más económica (sección 2.2.2) que la de esta figura.

2.2.1 Arquitectura y funcionamiento, solución mejorada intermedia

La figura 2 describe una nueva arquitectura. Se construye un hardware (tarjeta digital) que contiene la memoria necesaria para almacenar los cerca de doscientos vectores característicos de las escenas (figura 1) más un procesador digital de señales que hace la comparación. La cámara dotada del filtro de parecido por hardware sólo emite escenas interesantes, que se muestran en la pantalla y entran a la computadora para su análisis, una cada diez segundos, exagerando. La escena interesante se sigue comparando (usando su vector característico) con las doscientas escenas normales que la computadora almacena por cada cámara. Si resulta escena normal, no hay nada más que hacer. En caso contrario, se señala el fotograma en la pantalla mediante un cuadro rojo y una alarma sonora.

Esta arquitectura es viable, pero puede simplificarse con un filtro mucho más sencillo y económico (sección 2.2.2).

2.2.2 Solución mejorada final

La idea es muy sencilla. En vez del filtro por hardware de la sección 2.2.1, usemos un detector de movimiento, que mantiene apagada (sin enviar fotogramas) la cámara, y sólo la activa cuando se percibe movimiento.

Hay cámaras que llevan este sensor integrado. El sensor percibe el movimiento utilizando el CCD (dispositivo de carga acoplada, «charge-coupled device», que detecta la imagen originada por el campo visual) de la cámara.

2.3 Solución híbrida

También se puede tener una arquitectura híbrida, convirtiendo paulatinamente la solución básica a la solución mejorada.

2.4 Seguimiento en multicámaras

Las soluciones básicas (sección 2.1) y mejorada (sección 2.2) consideran el análisis de los fotogramas de una cámara sin tomar en cuenta lo que sucede en las cámaras vecinas. Es posible, en recintos cerrados o controlados, permitir que la computadora considere la información de una cámara para influenciar en el análisis de otra cámara. Se dice que las cámaras intercambian información (en realidad, este intercambio ocurre en la computadora). Con esto se extienden las aplicaciones posibles, y se aprovecha el hecho de que las cámaras forman una red: generalmente el recinto tiene varias cámaras y a veces sus campos visuales se traslapan. Esta comunicación es útil para detectar movimiento que abarque varias cámaras.

Cuando una persona o vehículo (“el objeto”) se mueve y sale del campo de visión de cierta cámara, sólo puede entrar en el campo de visión de pocas otras: las cámaras vecinas. Esto es, las cámaras forman una red o grafo, donde los nodos de la red son las cámaras, y las aristas conectan dos cámaras (nodos) si una persona puede ir de una cámara a otra directamente.

Una implementación reciente de este seguimiento aparece en la sección 3.2.

A veces, para seguimiento en multicámaras, el filtro de parecido se reemplaza por un filtro de movimiento, que evita que la computadora analice fotogramas estáticos.

3. TRABAJOS PREVIOS DE INTELIGENCIA ARTIFICIAL EN LA VIDEOVIGILANCIA

Por restricciones de espacio, cito sólo un par de trabajos llevados a cabo recientemente.

3.1 Detección de movimientos anómalos de vehículos

En el Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada del IPN en Querétaro, Qro., se desarrolló un trabajo que utiliza cámaras en la vía pública para detectar movimientos anómalos o prohibidos de vehículos: vuelta prohibida, cruce en rojo, etc., en intersecciones viales. Ver figura 3. [J. Salas *et al.*, <https://tinyurl.com/MovimProhibidos>, <https://tinyurl.com/MoviProhi>]. Esto es importante porque cerca del 30% de colisiones entre vehículos ocurre en intersecciones.

Figura 3
Movimientos anómalos de vehículos en la vía pública



Figura 3. Movimientos anómalos de vehículos en la vía pública. A la izquierda, cruce en rojo. En el centro, vuelta prohibida. A la derecha, vuelta prohibida (muy amplia). Imagen tomada del artículo citado en la referencia anterior, con autorización del autor.

El modelo de detección usado se describe en [J. Salas *et al.*, <https://tinyurl.com/Modelo-Detec>]. Para la detección, los fotogramas obtenidos se comparan contra modelos (previamente aprendidos) de las anomalías que pueden ocurrir.

3.2 Seguimiento de personas en multicámaras

Esta sección describe una implementación reciente donde se usan varias cámaras para detectar trayectorias en un recinto cerrado, concretamente en el Centro de Investigación en Computación del IPN. Cito la introducción de la disertación en su examen de grado del M. en C. César Olivares Espinosa [<https://tinyurl.com/SeguimVariasCamaras>]. Ver figura 4.

“En los últimos años ha existido una tendencia en el alza del número de videocámaras en lugares tanto públicos como privados, generando una cantidad de información tal que su análisis manual se vuelve virtualmente posible, lo que abre la necesidad de procesar y analizar esta información de forma semiautomática ayudando a las personas que vigilan estas videocámaras a facilitar sus tareas o de manera automática donde una computadora procesa toda esta información por sí misma permitiendo el análisis de grandes volúmenes de información.”

“En conjunto con el aumento de número de videocámaras y la cantidad de información generada por éstas, la tarea del análisis automático de una videocámara ha sido un tema muy investigado, pero la extensión a un sistema de multicámaras no ha tenido la misma exploración, lo que abre la posibilidad de presentar una propuesta de solución a este problema.”

“El tema principal de trabajo de esta tesis, en la cual se plantea un sistema para realizar el seguimiento de personas en un sistema multicámaras en línea, es decir, la identidad y trayectorias de cada persona se van generando al momento y éstas se mantienen aún y cuando una o más personas cambien de una cámara a otra logrando esto realizando una reidentificación de las personas.”

“El desarrollo de un sistema de seguimiento y reidentificación de personas en vídeo permite su utilización para aplicaciones de seguridad, como el tener control sobre zonas restringidas y/o el número de personas que pueden estar en una zona, vigilancia habitual de zonas para evitar siniestros, análisis de trayectorias extraídas de los vídeos para responder preguntas de negocio, comportamiento de

multitudes con motivos de seguridad, evitar siniestros, búsqueda de personas, en el caso que una persona sea dada por desaparecida y se tenga un punto inicial de búsqueda se pueden utilizar estos sistemas para agilizar su búsqueda, detección de eventos como altercados entre personas, aumento de velocidad de las mismas, trayectorias inusuales, movimientos sospechosos, entre otros.”

“El sistema propuesto realiza estas tareas de manera automática para un conjunto de vídeos provenientes de un sistema multicámaras del cual se conoce su topología.”
“Esta tarea no es directa de solucionar ya que presenta un número de problemas que lo dificultan, dentro de estos se encuentran los relacionados a las videocámaras como lo son la resolución de sus vídeos, el número de fotogramas por segundo, la posición en la que se encuentren, su ángulo de visión, nivel de ruido, las problemáticas inter-vídeo como la iluminación, variaciones de pose de cada persona, oclusión y pérdida de personas y las intra-videos relacionadas al sistema multicámaras como la variación de iluminación entre zonas, posibles cambios de escala, variación de ángulos de visión y pose, desincronización de las cámaras, distintos codificadores de vídeo, entre otros.”
“En este trabajo se plantea proponer una solución al problema del seguimiento de personas en un ambiente multicámaras con una aplicación en línea, generando identidades y trayectorias conforme avanzan los vídeos, siendo éste uno de los pocos trabajos con este enfoque y siendo ésta la mayor aportación del trabajo, proponiendo técnicas novedosas para el cambio de información entre cámaras y detección de cambio de zonas, la posibilidad de ignorar información de alta incertidumbre que pueda afectar el desempeño del sistema, la posibilidad de trabajar con vídeos mono y multicámara, y un método para la reidentificación de personas mediante imágenes. Asimismo, se propone una arquitectura modular para la resolución del problema. De esta manera se es posible actualizar módulos del mismo conforme existen nuevos con un mejor desempeño o dependiendo de la aplicación en específico.”

“Este sistema se construye y pone a prueba utilizando videocámaras del primer nivel del Centro de Investigación en Computación del Instituto Politécnico Nacional, el cual se compone de cuatro videocámaras con una topología

cuadrada entre sí, de las cuales se construyeron dos conjuntos de vídeos etiquetados con un enfoque en el seguimiento de personas a través de las cámaras, siendo esta otra aportación importante del trabajo.”

A continuación, doy algunos detalles importantes de su trabajo, y remito al lector a su tesis para los detalles omitidos.

Detección de personas. Inicialmente, se detecta una persona mediante su movimiento. Para esto se usa una red neuronal convolucional llamada Yolo («You only look once», disponible como software libre), para predecir su movimiento. Así se sigue a una persona hasta que se sale del campo de visión de la cámara. Se suponen restricciones razonables: una persona tiende a desplazarse en línea recta; su movimiento es continuo, no existen saltos abruptos en su trayectoria; distintas personas tendrán vestimentas distintas, que se usan para identificar a las personas cuando dos o más personas se cruzan y la cámara no percibe a todas (oclusión).

Filtro de Kalman. La detección por Yolo tiene tres funciones de pérdida: pérdida de clasificación, que cuantifica la clasificación de cada pixel del fotograma; pérdida de localización, que cuantifica el error en el centroide de la caja de localización (Yolo encierra a las personas en un cuadrado o caja, que se puede visualizar en el fotograma); y pérdida de confianza, que mide la certidumbre de que la caja contenga un objeto (una persona) o sea el fondo (no la contenga). La pérdida general es la suma de las tres pérdidas.

Nos dice Olivares en su tesis: “Considerando un conjunto de mediciones de un sistema a las cuales les consideramos un ruido añadido a cada una de éstas se pueden expresar estas mediciones con una función de densidad de probabilidad, en donde el promedio de esta función es el valor esperado de la variable aleatoria (las mediciones), la diferencia entre el valor esperado y el valor verdadero es la exactitud de la medición, a esta diferencia también se le conoce como el sesgo y la dispersión de esta distribución es la precisión de la medida o el ruido de la medición, se utilizó el Filtro de Kalman como un sistema de predecir

una nueva medición dado un conjunto previo de estas.”

Zonas de fuga. César las llama zonas de transición. Debido a la colocación de las cámaras en el recinto, se conoce la topología (la estructura) de la red de cámaras. Eso permite señalar en el campo visual de la cámara **a** las zonas de fuga hacia la cámara **b**, la **c**, etc. Si una persona en la cámara **a** cae dentro de la zona de fuga hacia la cámara **b**, debe aparecer en la cámara **b**, o regresar a la cámara **a**.

“Memoria Global. Este módulo es el que permite la transición de identidades entre distintas cámaras, conecta a los pares de zona de transición, almacenando la información de cada persona y utilizándola para su recuperación.”

Figura 4



Figura 4. La aplicación de César Olivares detecta y sigue a una o más personas en la planta baja del Centro de Investigación en Computación, IPN. Figura tomada de la tesis de Olivares.

“Módulo de paciencia. Utilizado en algunos trabajos, este módulo mantiene durante un número de pasos fijo la información de una persona que haya desaparecido de la imagen para su probable posterior recuperación con la misma identidad.”

“Recuperación de identidades. El algoritmo utilizado para recuperar la identidad de una persona entre un conjunto de candidatos es modular el cual puede ser intercambiado sin modificar el resto del sistema de acuerdo con la necesidad.”

“Procesamiento en paralelo de los fotogramas. Al tratarse de un sistema de multicámaras en todo momento se debe procesar cada fotograma proveniente de las distintas cámaras, los cuales se hacen en paralelo aprovechando la arquitectura de las tarjetas gráficas.”

“Zonas oscuras. Al estar lidiando con vídeo estamos sujetos a las características de la video- cámara y las condiciones en que ésta se encuentra, como lo puede ser su campo de visión. Pueden existir zonas donde las detecciones puedan tener muy alta incertidumbre e inclusive sea conveniente perder dicha información. Para aliviar este inconveniente se propone el uso de Zonas Oscuras o Puntos Ciegos Artificiales, como su nombre lo dice, de forma manual definiremos zonas dentro del vídeo en donde las detecciones no serán consideradas, basándonos en la idea que no podríamos controlar la arquitectura de la zona vídeo grabada y que ésta podría tener la vista obstaculizada por un elemento físico.

De esta manera, con una exploración previa de las zonas videograbadas se pueden definir zonas en las que es preferible ignorar sus detecciones. Esto también ayuda cuando el compa ración tenga que ser únicamente por imágenes, si dos detecciones se encuentran en condiciones distintas de escala una comparación directa de las imágenes no generaría resultados con fiables; asegurándonos que las detecciones se encuentren en una escala parecida estaría brindando una comparación más fidedigna.” Los párrafos entre comillas (“”) son citas de la tesis de Olivares.

4. COMENTARIOS FINALES

4.1 Advertencia

Los cálculos expresados en la sección 2 son estimados, y sufrirán cambios (quizá significativos) al momento de implementar la solución. Están, pues, sujetos a errores, y no deben tomarse como definitivos.

4.2 Conclusiones

La adición de algoritmos de inteligencia artificial, aprendizaje mecánico y visión por computadora a la red de cámaras de la CdMx traerá beneficios importantes para la seguridad del capitalino, acercando a esa metrópoli hacia una ciudad inteligente.

Para recintos cerrados o controlados, la adición de una aplicación para seguimiento en multicámaras permite detectar movimientos de personas o vehículos, con varias ventajas.

4.3 Recomendaciones

Dada la gran inversión en equipo, personal y mantenimiento ya hecha por la CdMx en su extensa red de videocámaras, conviene hacer la transición hacia la solución básica (sección 2.1) primero, luego hacia la solución híbrida (sección 2.3), y finalmente llegar a la solución mejorada (sección 2.2).

4.4 Agradecimientos

Al Instituto Politécnico Nacional y al Consejo Nacional de Ciencia y Tecnología (ahora Conahcyt), con cuyo apoyo ha sido posible llevar a cabo los trabajos mencionados. En especial, al Dr. Joaquín Salas, al M. en C. César Olivares Espinosa, cuyos trabajos en el Politécnico han sido descritos brevemente, y al Dr. Jesús Olivares Ceja, quien codirigió la tesis de César.

4.5 Referencias

Las referencias (ligas) en este artículo fueron verificadas el 24 de septiembre de 2023. Hay trabajos, proyectos

e investigaciones adicionales con Ciencia de Datos, aprendizaje automático e Inteligencia Artificial, que se han desarrollado o están en desarrollo en el Laboratorio de Ciencia de Datos y Tecnología de Software (https://www.cic.ipn.mx/index.php/acerca_de-cdyts) del Centro de Investigación en Computación (www.cic.ipn.mx) del IPN. Muchos de los dirigidos o llevados a cabo por el autor aparecen en su blog (a-guzman.blogspot.com).

Bibliografía.

Las ligas fueron consultadas el 13 de octubre de 2023.

¿Qué es un C2 móvil? Página describiendo un C2 móvil de la CdMx. Gobierno de la CdMx. <https://tinyurl.com/C2mCdMx>

Blog del Dr. A. Guzmán, autor de este artículo. Contiene trabajos y tesis recientes desarrollados por él, o bajo su supervisión. a-guzman.blogspot.com

Canchola Magdaleno, Sandra Luz, Joaquín Salas Rodríguez, Hugo Jiménez Hernández, José Joel González Barbosa, Juan B. Hurtado Ramos (2010) A Machine-Vision System to Detect Unusual Activities Online at Vehicular Intersections. *Computación y Sistemas*, Vol. 3, No. 3. Pp. 209-220, SSN 1405-5546. <https://tinyurl.com/MoviProhi>

Jiménez Hernández, Hugo, Joaquín Salas (2009) Temporal Templates for Detecting the Trajectories of Moving Vehicles. <https://tinyurl.com/ModeloDetec>

Olivares Espinosa, César (2023). Seguimiento de personas en multicámaras. Tesis de Maestría, Centro de Investigación en Computación, Instituto Politécnico Nacional (CIC-IPN). <https://tinyurl.com/SeguimVariasCamaras>

Página con la historia del C5 de la CdMx. Gobierno de la CdMx. <https://tinyurl.com/C5CdMx>

Página describiendo las cámaras de videovigilancia del C5 de la CdMx. Gobierno de la CdMx. <https://tinyurl.com/>

com/CamarasCdMx

Página que describe el Centro de Investigación en Computación (CIC-IPN). www.cic.ipn.mx

Página que describe el Laboratorio de Ciencia de Datos y Tecnología de Software del CIC-IPN. <https://www.cic.ipn.mx/index.php/acerca-de-cdyts>

Proceedings of the Conference: Advanced Concepts for Intelligent Vision Systems, 11th International Conference, ACIVS 2009, Bordeaux, France, September 28 - October 2, 2009. DOI: 10.1007/978-3-642-04697-1_45

Salas, J., H. Jiménez, J. González and J. Hurtado, "Detecting Unusual Activities at Vehicular Intersections," Proceedings 2007 IEEE International Conference on Robotics and Automation, Rome, Italy, 2007, pp. 864-869, doi: 10.1109/ROBOT.2007.363094. <https://tinyurl.com/MovimProhibidos>