Herramientas para la detección y seguimiento de personas a partir de cámaras de seguridad

Leonardo D. Dominguez[1,2], Alejandro J. Perez[1],
Aldo J. Rubiales[1,3], Juan P. D'Amato[1,2], Rosana Barbuzza[1,3].

[1] PLADEMA, Universidad Nacional del Centro de la Provincia de Bueno Aires,
[2] Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET
[3] Comisión de Investigaciones Científicas, CICPBA

Resumen La inseguridad es un problema que afecta en mayor o menor medida a todas las ciudades del mundo. Las ciudades más informatizadas hacen uso de la video-vigilancia para combatirla, montando en muchos de los casos centros de monitoreo con cientos de cámaras. En su mayoria, estos centros cuentan con grupos de personas para realizar la tarea de observación, sin embargo, la velocidad de cómputo actual nos da la posibilidad de automatizar muchas de sus tareas diarias. En este trabajo, se presenta una plataforma de análisis de video que se está desarrollando en la UNCPBA para facilitar el seguimiento de una persona a través de diferentes cámaras, utilizando técnicas de proyección que convierten los puntos detectados desde las diferentes cámaras a un único espacio georeferenciado. Se presenta una discusión de los algoritmos utilizados para el seguimiento, algunos problemas propios que se suceden en este tipo de sistemas y los resultados preliminares obtenidos.

Keywords: Videovigilancia, Tracking multi-camara, Seguridad

1. Introducción

En los últimos tiempos, la seguridad se está convirtiendo en uno de los ejes fundamentales de la sociedad. La gran preocupación que generan los altos índices de violencia e inseguridad actuales [8] en países como Argentina, motivan el estudio de nuevas tecnologías que permitan vivir con mayor tranquilidad a los habitantes, sobre todo, de las grandes ciudades, en donde estas lamentables situaciones son más frecuentes. Según [12], entre el año 2000 y 2008, la tasa de población penitenciaria en América Latina, creció un 42 %, y según las tendencias en los buscadores WEB, en los últimos 10 años el interés de las personas en términos como "Robo", "Inseguridad o Insecurity" se encuentran en constante aumento. En contrapartida, países como Estados Unidos, cuentan con agencias de investigación para la defensa de sus habitantes. En particular, la agencia DARPA, cuenta con más de 60 proyectos activos, de los cuales nos interesa mencionar a Satellite Remote Listening System y a Combat Zones That See, proyectos que tienen como objetivo registrar todo lo que se mueva mediante un sistema de videocámaras [14]. Otro caso paradigmático es el de Reino Unido,

que según [12],en el informe [4] se estimaba que en Londres un individuo podía ser registrado en un día normal por aproximadamente 300 cámaras. En general, los sistemas actuales están orientados en su mayoría como herramientas de monitorización, también conocidos como CCTV (Closed Circuit Television), que guardan las imagenes en un dispositivo de almacenamiento para su futura visualización o para ser utilizadas como evidencia. En sintonía, los propios fabricantes de cámaras de vigilancia ofrecen herramientas simples, que se limitan a conectar y grabar lo que las cámaras observan. Sin embargo, pensar en un sistema con ciertas automatizaciones, que sea capaz de vigilar un área, emitir y registrar alarmas, clasificar y contar personas e incluso poder seguir sus movimientos si es necesario, nos hace presuponer que será más ventajoso para los operadores, pues les permitirá focalizar su atención y estar alerta ante un evento importante pudiendo así seguir el protocolo que corresponda. Con un propósito similar, se ha iniciado el proyecto [16], el cual pretende a partir de una extensa red de sensores y dispositivos inalámbricos, controlar grandes áreas metropolitanas de forma inteligente. Sin lugar a dudas, con al auge de la tecnología en cuanto a captura de imágenes y video, es posible adquirir cámaras con buena definición a un bajo costo, lo cual permite que una herramienta como esta pueda ser utilizada no sólo en ámbitos de seguridad, sino también en lugares concurridos que requieran analizar los patrones de movilidad diaria, los comercios podrían establecer estrategias de marketing con mayores argumentos [15].

Entendiendo que la seguridad es un tema muy importante, en este trabajo, presentaremos una plataforma distribuida para la automatización de tareas de detección y seguimiento de personas u objetos en diferentes cámaras de video. Para esta tarea es necesario unificar los diferentes puntos de vista de cada cámara, para llevarlo a un único espacio geo-referenciado. Con este objetivo, se han desarrollado herramientas de calibración para obtener las transformaciones que permitan ubicar las detecciones en un mapa. Uno de los algoritmos principales es el que realiza la detección del movimiento, análogo a la substracción de fondo (SF) y detección de objetos. Lo innovador de la plataforma es que se encuentra preparada para poder contar con diferentes algoritmos de manera sencilla y su arquitectura distribuida le permite escalar para soportar muchas cámaras sin afectar el rendimiento global.

El trabajo se presenta de la siguiente forma. En la sección 2, se presentan las herramientas existentes con propósitos similares y un resumen del estado de arte. Posteriormente, en la sección 3 se mencionan dos de los detectores utilizados y cómo se mapean al plano. En la sección 4 se presentan algunos resultados preliminares de los recorridos de una persona junto con las problemáticas encontradas, y en la sección 5 se presentan las conclusiones y los trabajos que quedan pendientes para un futuro.

2. Estado del Arte

Actualmente existen una gran cantidad de sistemas de video vigilancia y en lo que respecta al seguimiento de personas hay mucha bibliografía reciente y es

un tema muy estudiado en estos últimos años, en parte gracias al avance del poder de cómputo disponible en los procesadores actuales.

Trabajos como [5], detallan los algorítmos más comunes para la detección y seguimiento de objetos, y en particular en [7], se realiza una comparación interesante entre los distintos detectores, en donde se concluye que la combinación de detectores puede ser muy útil para disminuir la tasa de falsos positivos mientras se mantiene el tiempo y la tasa de verdaderos positivos.

En el mercado, se destacan soluciones como Blue Iris a un costo de U$S60 con soporte hasta 64 cámaras o EyeLine a un costo de U$S250 sin límites de conectividad. Por otra parte se destacan soluciones Open Source como ZoneMinder o Ispy muy evolucionadas aunque todas se limitan a funcionar en un solo equipo. Las características de hardware que recomiendan todas las soluciones son: Intel Core i7, 8 gb de RAM o más, disco SSD, placa aceleradora Nvidia para la decodificación por hardware y Windows 8 o superior.

2.1. Legislación

El esfuerzo de las autoridades por brindar mayor seguridad para sus ciudadanos, puede caer en un reto legal. Hoy en día existen debates sobre la vulnerabilidad de la privacidad por el aumento excesivo de sistemas de vigilancia. En Argentina, todavía nos debemos un gran debate para este tema. Según el artículo [3] publicado en la 44 JAIIO 2015, el 87,5 % de las provincias no cuentan con una regulación adecuada o completa, y sus autores destacan la gran disparidad en el plazo de almacenamiento establecido de las imágenes previo a su destrucción, que va desde los 30 días en Santa Fe, a los 2 años en Corrientes y San Luis. Además, señalan que la mayoría de las regulaciones no se adaptan a los principios establecidos por la disposición 10/2015 de la DNPDP[1].

3. Detección y seguimiento de personas

3.1. Detección de movimiento

El primer paso para el seguimiento de personas, es poder determinar si hubo o no movimiento a partir de una serie de imágenes, lo cual se encuentra muy estudiado. Ya existen múltiples soluciones para poder calcular la variación entre imágenes consecutivas [11][13]. Los métodos varían en complejidad computacional y eficacia del resultado. Cuando las cámaras son exteriores, se suceden otros factores que además varían dependiendo del momento del día, la nubosidad y la época del año. Para contrarrestar estos efectos, se suelen utilizar ventanas de movimiento como se menciona en [10], cuyo objetivo consiste en aplicar una operación lógica, para procesar únicamente los movimientos que se detecten dentro de sus dimensiones y descartar todos aquellos que se produzcan por fuera. En general, la salida de este paso es un conjunto de píxeles (agrupados o no) que han sido marcados como representativos de objetos en movimiento.

Para la plataforma, se utilizaron las capacidades del paquete AForge (open source), con los detectores que incluye para hacer las primeras pruebas del sistema. Observando que todos informan el nivel de movimiento detectado en el video con valores entre 0 y 1, lo cual le permite al programador establecer un umbral para comparar el movimiento y definir diferentes tipos de alarmas, que es uno de los objetivos del proyecto. Hoy en día, se estan probando otros algoritmos más sofisticados de sustracción de fondo, tal como VIBE[6], pero que escapan al alcance actual del artículo.

- Two Frame Difference (TF): Es el detector más simple y rápido. Se basa en encontrar la diferencia entre dos fotogramas consecutivos.
- Simple Background Modeling (BM): En contraste con el anterior, este detector se basa en encontrar la diferencia entre el fotograma actual y un fotograma definido que representa el fondo. También existen técnicas para actualizar el fondo a medida que avanza el tiempo.

Figura 1. A izquierda, detección de un objeto con sombra. A derecha, zona de mapeo

Cuando se trabaja con cámaras en el exterior, se suceden algunas situaciones climáticas que afectan el resultado de una detección, como por ejemplo la sombra. Para minimizar el error que provoca la sombra, se propone utilizar el pixel central del objeto detectado. Otros trabajos proponen utilizar la minima coordenada "Y"del objeto detectado (se considera que es la ubicación de los pies), pero esto no siempre se corresponde en todos los puntos de observacion. Se pretende en un futuro mejorar estos algoritmos, aplicando filtros que primero eliminen la sombra, a partir del análisis de la variación de la iluminación.

3.2. Seguimiento en el mapa

Para poder establecer una relación entre las personas detectadas en cada cámara y el mapa satelital del área vigilada, es necesario calibrar cada una de las cámaras. La salidad de este proceso es una matriz homográfica, que permite mapear cada pixel en un punto geo-referenciado. Esta matriz se obtiene de

manera manual, utilizando una serie de herramientas que permiten seleccionar varios puntos de referencia desde la imagen e inferir la matriz correspondiente, denominada HS. Los puntos se eligen sobre una region rectangular y plana.

Esta matriz es luego almacenada en la Base de Datos. Este proceso es sensible al error, por lo que se repite varias veces, calculando la matriz utilizando el promedio de puntos seleccionados. En tiempo real, por cada nueva detección en el espacio de la imagen, se aplica la transformación sobre el punto mas representativo del movimiento, para obtener así un punto sobre el espacio del mapa.

En la Figura 2 se puede observar en forma simplificada el camino recorrido por el proceso de detección y mapeo de una persona en un centro de monitoreo. Primero para lo cual se utilizó la detección de objetos obtenidas en el paso anterior y luego se aplica una transformación espacial sobre un mapa.

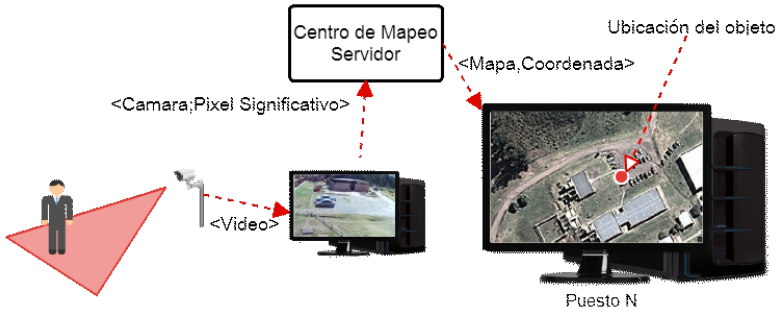

Figura 2. Recorrido de una detección hasta su visualización

3.3. Calibración

Dado que una matriz Hs relaciona puntos de dos imágenes, y teniendo en cuenta que se requiere de una diferente para calibrar cada cámara con cada mapa satelital, se implementó una herramienta que le permite al operador cargar dos imágenes (un mapa y un frame de una cámara) y establecer una referencia entre ambas a partir de un mismo cuadrilátero que pueda observarse en las mismas, de modo tal que se pueda calcular la matriz de perspectiva necesaria para adaptarlas.

El proceso de la misma está dividido en 3 solapas (Figura 3), Mapa, Cámara y Resultado. En la primera, se permite cargar el mapa y en la segunda la imágen de la cámara. Luego haciendo clic sobre ambas imagenes se pueden ir definiendo los vértices del cuadrilátero. La herramienta permitirá además, definir el punto donde se posiciona la cámara y calcular las medidas del cuadrilátero (ancho y alto) y su ángulo de rotación α respecto al eje X del satélite que nos provee el mapa, con el propósito de generar no solo la matriz homográfica de la cámara,

sino una matriz que contemple todas las transformaciones necesarias entre ambas imágenes.

Figura 3. Herramienta de calibración. De izq. a der.: Mapa, Cámara y Resultado

3.4. Integración de los datos

La plataforma desarrollada está concebida para escalar a múltiples cámaras, por lo que funciona de manera distriubuida. En este sentido, basado en [9], se contempla tener varios equipos de cómputo (computadoras de escritorio o micro-PCs), todos conectados a un servidor central. Cada equipo de cómputo se conecta con una o varias cámaras IPs y se encarga del procesamiento. El servidor coordina la detección de movimiento de todas las cámaras, y realiza las operaciones de transformación de los puntos que reciba de los puestos a los mapas. El servidor registra todo en una Base de Datos, la cual luego puede ser accedida por diferentes medios. Un tercer módulo es el encargado de la visualización sobre un mapa. Este módulo puede correr en algunos de los equipos e incluso en el servidor.

Las etapas básicas de este proceso son las siguientes:

- Un equipo 1, establece una conexión con una cámara IP y aplica un algoritmo de SF a las imágenes , indicándole además su estado actual a un servidor S.
- Un equipo 2, se conecta al servidor con la funcionalidad de visualización, suscribiéndose a la recepción de puntos para un mapa que abarca una determinada región satelital.
- Cuando el equipo 1 detecta movimiento, envía al servidor S, un mensaje que contiene el identificador de la cámara procesada y las coordenadas (x,y) .
- Cuando S recibe el mensaje, verifica si algún equipo está suscripto a los mensajes del área que vigila esta cámara. En caso de existir al menos un cliente, S busca la matriz de calibración y transforma el pixel (x,y) del mensaje al espacio del mapa. En caso que se encuentre activado, aplica un pos-procesamiento a la detección para obtener informacion extra, como por ejemplo una preclasificación.

De esta manera, con este pequeño esquema de mensajes, un servidor actuará como concentrador y conocerá en todo momento qué equipos se encuentran procesando señales de video y qué equipos desean recibir las coordenadas transformadas para un mapa.

3.5. Resolución de superposición de multi-camara y falsos positivos

El área de observación de las cámaras muchas veces se superpone entre sí, por lo que un punto de movimiento puede ser informado por varios equipos a la vez. Al mismo tiempo, se pueden generar falsos positivos, producto de un movimiento de un objeto.

Para unificar estos puntos, y ser tratados como uno solo y reducir los falsos positivos, se debe aplicar un segundo procesamiento de los mismos.

Una de las propuestas que se plantea en este trabajo consiste en ordenar las capturas por intervalos de tiempo. Para que esto funcione correctamente, todos los equipos están sincronizados con un reloj único, para lo cual se utiliza la hora del Servidor. A continuación, por cada conjunto de puntos proyectados dentro de un intervalo, se los promedia (se eligió un intervalo por cada segundo), y luego se calcula la distancia euclidea del grupo con respecto al siguiente intervalo. Aquellos puntos que se encuentran a una distancia mayor de cierta tolerancia son descartados. Estos umbrales se basaron en [2] donde se expresa que en promedio una persona camina a 5 kilómetros por hora (o lo que es igual a 1,38 mts por segundo).

4. Resultados preliminares y discusiones

El seguimiento de una persona incluye dos etapas que se consideran críticas, como lo son la detección del objeto en la imagen de la cámara y el mapeo del mismo hacia el plano. Para las pruebas se utilizó una PC virtualizada de 4 núcleos y 2 GB de RAM con 4 cámaras conectadas y los videos contaban con una resolución de 640x480 pixeles. Las cámaras se ubicaron tal cual se observa en la Figura 4.

En esta configuracion, existian varias zonas de oclusión, donde la persona no podía ser observada. Se definieron zonas de interés por cada cámara para obtener el área de análisis relevante. En "blanco", se marcan los puntos por donde pasó una persona.

Para efectuar las pruebas, se aplicó un filtrado de puntos a cada conjunto, en donde se lo agrupó por segundo, promediando las coordenadas x e y. Una vez definidos los conjuntos que contienen una única marca por segundo, se calculó la medida y se promediaron las distancias obtenidas. Se aplicó un filtrado tal como se explicó en la sección 3.5. Conociendo la escala del mapa, se definieron como umbrales máximos de movimiento en un segundo entre 1.38mts y 11.04mts, que corresponden a cuatro veces la velocidad promedio al caminar [2].

Por otra parte, con el fin de comparar la eficacia de los detectores, se calcularon las distancias que existen de sus resultados hasta un recorrido de referencia.

Figura 4. Áreas observadas por las cámaras

Para medir esta distancia, se utilizó el indicador eq.1 que mide como el promedio de las mínimas distancias desde todos los puntos del recorrido original hasta los segmentos consecutivos .

$$\sum_{j=1}^{n} \frac{1}{n}(Min_{pb}(Distancia(pA, pB(j), pB(j+1)))) \tag{1}$$

donde pA ∈ PuntosOriginales ; pB ∈ PuntosDelDetector

A su vez, se aplicó un filtro de ganancia del (1 % al 99 %) a la imagen de diferencia, previo a aplicar un umbral de binarización y obtener los píxeles de movimiento del objeto de interés. En estos casos, una menor ganancia eliminaba píxeles del objeto y una mayor ganancia aumentaba la cantidad. En cualquier caso, se modificaba el centro geométrico del objeto, utilizado posteriormente en la proyección.

Se generó la tabla en la (Figura 5) para analizar estas combinaciones, donde se contabilizaron la cantidad de puntos .

Detector	Umbral = 1.38	Umbral = 2.76	Umbral = 5.52	Umbral = 11.04
B.M. con G=1	1,615	1,598	1,729	1,712
B.M. con G=10	1,610	1,497	1,837	1,765
B.M. con G=30	1,448	1,549	1,678	1,702
B.M. con G=99	1,515	1,747	1,690	1,667
T.F. con G=1	1,561	1,435	1,894	2,204
T.F. con G=10	2,087	2,182	2,357	2,243
T.F. con G=30	1,429	1,914	2,666	2,639
T.F. con G=99	1,301	1,582	2,180	2,086

Figura 5. Continuidad (Cantidad de puntos totales luego de filtrar los puntos)

donde se midió la distancia del trayecto obtenido por cada método respecto al de referencia, para ambos detectores

En todos los casos, se observó la influencia de la sombra, desplazando el centro del objeto. Se observa que cuando el umbral es menor a los 5 mts. la distancia promedio es menor (dado que se descartan falsos positivos lejanos); en otro caso, aparecen puntos que no pertenecen a la persona, por lo que las distancias aumentan.

Los recorridos resultantes antes y después del filtrado fueron los que se ven en la Figura 6

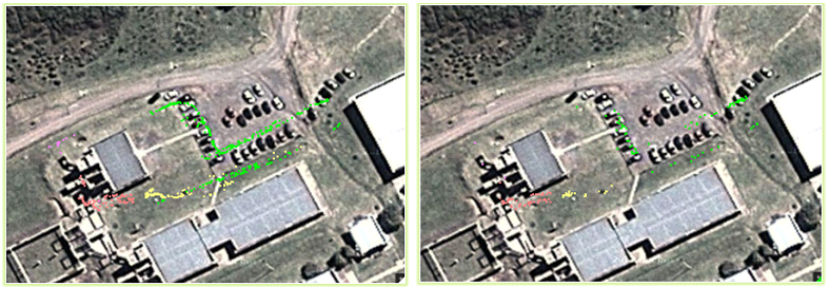

Figura 6. (izq.) Puntos obtenidos originalmente (der) Puntos luego del filtrado

Se observó que gran cantidad de muestras fueron filtradas, muchas producto de falsos movimientos; y los desplazamientos observados se han debido en su mayoría a la proyección de la sombra.

5. Conclusiones

En este trabajo se presentó una plataforma distribuida para la gestión de múltiples cámaras, con el soporte para el segumiento de personas a través de las mismas. La plataforma permite aplicar diferentes técnicas de substracción de fondo para detectar objetos de interés y luego aplicar una proyección de los puntos, para ser visualizados en un espacio georeferenciado, más amigable para el usuario. La plataforma es muy configurable, permitiendo distribuir la funcionalidad de procesamiento, visualización y segumiento en cualquier equipo.

También se comenzaron a analizar técnicas de filtrado de los datos, a fin de obtener los puntos de movimiento con mayor precisión. Para esto, se permite ingresar un recorrido de referencia (ya sea de forma manual o por GPS) y compararlo con los resultados obtenidos del análisis de imágenes. Los primeros resultados, no fueron del todo satisfactorios, ya que en un caso ideal deberían ser cercanos a 0; pero dan el pie para seguir profundizando y perfeccionando. Lo que se observó es la influencia de la sombra al momento de la captura, generando perturbaciones en las imágenes, que luego son potenciadas al ser proyectadas.

En un trabajo futuro, se pretende perfeccionar las técnicas de substracción de fondo que contemplen la aparición de sombras, a fin de obtener valores más precisos. También la idea es trabajar sobre el seguimiento de muchas personas de manera simultánea, a la vez que de integrar con otras técnicas de reconocimiento de objetos basadas en Deep-Learning; a fin de optimizar la clasificación de los objetos.

Referencias

1. Dirección nacional de protección de datos personales - disposición 10/2015. http://www.infoleg.gob.ar/infolegInternet/anexos/240000-244999/243335/norma.htm.
2. Velocidad promedio del desplazamiento humano. https://es.wikipedia.org/wiki/Kil
3. Cejas E. B. and González C. C. Estado de la normativa sobre video vigilancia en argentina y su relación con la protección de datos personales. *44 JAIIO*, 2015.
4. Watch Big Brother. The price of privacy: How local authorities spent£ 515m on cctv in four years. *A Big Brother Watch report, February*, 2012.
5. Legua C. C. Seguimiento automático de objetos en sistemas con múltiples cámaras. 2013.
6. Bouwmans D., Porikli F., B. Höferlin, and Vacavant A. *Background Modeling and Foreground Detection for Video Surveillance*. CRC Press, 2014.
7. Hall D., Nascimento J., Ribeiro P., et al. Comparison of target detection algorithms using adaptive background models. In *Visual Surveillance and Performance Evaluation of Tracking , 2nd Joint IEEE Int. Work.*, pages 113–120. IEEE, 2005.
8. Kessler G. *El sentimiento de inseguridad: sociología del temor al delito*. Siglo Veintiuno Editores, 2009.
9. Foresti G.L., Mähönen P., and Regazzoni C. S. *Multimedia video-based surveillance systems: Requirements, Issues and Solutions*, volume 573. Springer Science & Business Media, 2012.
10. Kruegle H. *CCTV Surveillance: Video practices and technology*. Butterworth-Heinemann, 2011.
11. Shaikh S. H., Saeed K., and Chaki N. *Moving Object Detection Using Background Subtraction*. Springer, 2014.
12. Dammert L., Salazar F., Montt C., and González P. Crimen e inseguridad: indicadores para las américas. *FLACSO-Chile/Banco Interamericano de Desarrollo (BID)*, 2010.
13. Piccardi M. Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on*, volume 4, pages 3099–3104. IEEE, 2004.
14. Shachtman N. Big brother gets a brain. *Village Voice*, 48(28):40, 2003.
15. Chandon P., Hutchinson J., Bradlow E., and Young S. H. Measuring the value of point-of-purchase marketing with commercial eye-tracking data. *INSEAD Business School Research Paper*, (2007/22), 2006.
16. Celtic Telecommunication Solutions. Husims - human situation monitoring system, 2012. https://www.celticplus.eu/wp-content/uploads/2014/09/HuSIMS-leaflet_lq.pdf.