Conteo automatizado de tráfico en Bahía Blanca basado en videos

Martín C. De Meio Reggiani[1,2] y Valentina Viego[1,3]

[1] Instituto de Investigaciones Económicas y Sociales del Sur (CONICET-UNS), Bahía Blanca, Buenos Aires, Argentina
[2] Departamento de Economía, Universidad Torcuato Di Tella, Ciudad Autónoma de Buenos Aires, Argentina
[3] Departamento de Economía, Universidad Nacional del Sur, Bahía Blanca, Buenos Aires, Argentina

Abstract. Los conteos de tráfico suelen ser insumo de otros sistemas de planificación y gestión del tráfico vehicular, como la gestión de semáforos, simulaciones de desplazamientos, elaboración de matrices origen-destino. El objetivo de este trabajo es presentar los avances de la implementación de un sistema de conteo de tránsito vehicular automatizado basado en videos que aprovecha equipamiento existente usualmente utilizado para otros fines. En particular, se aplicaron los algoritmos YOLO v4 y en DeepSORT a una muestra de 85 cámaras de vigilancia, distribuidas en toda el área urbana de Bahía Blanca, General Daniel Cerri e Ingeniero White. Los registros presentaron ciertas dificultades técnicas que han sido clasificadas para su corrección en ambos algoritmos. Los resultados del algoritmo preliminar muestran algunos problemas derivados de los microcortes en las filmaciones. Se espera que la experiencia obtenida en este experimento sirva como guía para futuros desarrollos de algoritmos de rastreo que sean robustos a problemas del mundo real.

Keywords: Conteo vehicular, YOLO v4, DeepSORT.

1 Introducción

Bahía Blanca y las localidades aledañas de General Daniel Cerri e Ingeniero White constituyen un área urbana de 115 km2, ubicada al sur de la provincia de Buenos Aires con una población de 310 mil habitantes (según proyecciones del INDEC basadas en el Censo de Población y Vivienda de 2010). La extensión espacial de la zona residencial generó nuevos barrios de ingresos altos, medios y bajos, aunque con distinto ritmo y volumen de hogares involucrados.

Esta ampliación de la "mancha" urbana tuvo lugar con una estructura de transporte público (buses) con una menor cantidad de líneas y recorridos casi sin variaciones. Desde 2012 operan en la ciudad 17 líneas de colectivos; el recorrido de cada línea ha sido bastante estable a pesar de los cambios espaciales de la distribución de la población. Por este motivo, entre otros, ha aumentado el uso del automóvil particular en los barrios de ingresos medios y altos, y de motovehículos en los barrios de ingresos bajos, en tanto las líneas de transporte público no llegan o lo hacen con bajos niveles de servicio a los distintos barrios. Ambos fenómenos han provocado la congestión de ciertas arterias y horas del día, y aumento de la accidentalidad vial [1].

El trazado de recorridos y la frecuencia de servicio son centrales en la determinación del volumen de pasajeros, además del costo monetario del viaje. La matriz origen-destino (MOD) suele ser el insumo principal en la planificación racional del transporte público. Esta herramienta permite cuantificar los viajes realizados entre distintos puntos del territorio y analizar el sistema de transporte urbano. La información necesaria para construir la matriz suele ser voluminosa y costosa (relevamientos en hogares, datos de movilidad de telefonía móvil, etc). Por este motivo, suelen emplearse herramientas de observación indirecta, como los conteos vehiculares para obtener las MOD. El desarrollo de métodos de inteligencia artificial y aprendizaje basado en máquinas (*machine learning*) ha abierto un enorme potencial para el perfeccionamiento de los modelos de planificación urbana y toma de decisiones en transporte público.

Actualmente existen herramientas computacionales que podrían colaborar en la toma de decisiones relacionadas con la gestión y ordenamiento del tránsito en las ciudades. Por ejemplo, métodos basados en inteligencia artificial para el conteo automatizado de vehículos y los simuladores de flujos de tránsito, que permiten anticipar la reacción de la movilidad urbana ante la aparición de nuevas líneas o cambios de recorridos de buses, colocación de nuevos semáforos o variación de tiempos de espera, implantación de carriles de uso exclusivo, nuevas infraestructuras de transporte (tren eléctrico, etc), entre otras. En Argentina el uso de esas herramientas por parte de los reguladores del transporte urbano de pasajeros es aún incipiente. Un proyecto que aplique los recientes desarrollos de métodos de aprendizaje automatizado

a problemáticas locales concretas representa un avance tanto para las áreas de transferencia del sector científico como para las esferas de decisión de políticas públicas ligadas a la movilidad. El presente trabajo tiene como objetivo presentar los avances de la implementación de un sistema de conteo de tránsito vehicular automatizado. Ello podría, a futuro, convertirse en un insumo para generar simulaciones de tráfico urbano y obtener matrices origen-destino para Bahía Blanca en forma periódica. Se espera, además, que la experiencia obtenida en este experimento sirva como como guía para futuros desarrollos de algoritmos de rastreo que sean robustos a problemas del mundo real.

2 Metodología

Algunos métodos de conteo automatizado de tráfico suelen utilizar dos herramientas de aprendizaje profundo. Por un lado, algoritmos de reconocimiento de objetos combinados con algoritmos de seguimiento de objetos múltiples. La detección de objetos en la primera etapa involucra 3 operaciones: clasificación, localización y delimitación. A su vez, los algoritmos pueden basarse en métodos de clasificación o de regresión. Los segundos son más eficientes en tiempo de cómputo a expensas de cierta pérdida de precisión respecto de los primeros. El algoritmo YOLO (*You Only Look Once*), desarrollado originalmente por Redmon et al [2], corresponde a esta última clase. YOLO requiere conocer lo que se va a detectar en base a descriptores de centroides, ancho, altura, clase de objeto y probabilidad de localización en un contorno dado. Desde su desarrollo ha atravesado diferentes mejoras. En particular, la versión 4 propuesta por Bochkovskiy et al [3] ha mejorado la velocidad de entrenamiento y robustez de detección.

Respecto de los métodos de rastreo, se utilizan los denominados algoritmos de ordenamiento, que consisten en colocar los elementos de un vector en una secuencia con una relación de orden. Para el problema de conteo de tráfico interesan en particular los algoritmos de seguimiento de múltiples objetos (MOT), en la que se identifican varios objetos dentro de un marco y se los sigue hasta que lo abandonan mediante un mismo identificador. Uno de los algoritmos de rastreo más populares es el DeepSORT [4], que utiliza un filtro de Kalman para el seguimiento. Este filtro utiliza variables de posición inicial, forma y velocidades asociadas para predecir la próxima posición. El paso siguiente es asociar las nuevas detecciones con las predicciones mediante la minimización de un criterio de distancia (en este caso, el cuadrado de la distancia de Mahalanobis). La aparición de aprendizaje profundo en estos enfoques radica en la incapacidad del filtro de Kalman de lidiar con la oclusión. Para ello, DeepSORT introduce una medida de distancia adicional basada en el aspecto del objeto y modelado con una red neuronal convolucional.

Los algoritmos automatizados de conteo de flujo vehicular se aplicaron *off-line* a una muestra de 85 cámaras de vigilancia, distribuidas en toda el área urbana (Bahía Blanca, General Daniel Cerri e Ingeniero White). Las grabaciones de las cámaras fueron provistas por el Centro Único de Monitoreo del municipio de Bahía Blanca. La captura se realizó en octubre 2020, noviembre 2020, y enero 2021, en días y horarios diversos para controlar efectos estacionales.

El experimento se desarrolló entre noviembre de 2020 y marzo de 2021, y se utilizaron 19 computadoras personales con procesador i5 sin GPU. Para asegurar la integridad de los equipos y evitar daños por recalentamiento, se aplicó un protocolo de 48 horas de trabajo con pausas entre cada video procesado. Además, por cada fotograma procesado, se saltearon los 2 subsiguientes para acelerar el procesamiento. Teniendo en cuenta que los fotogramas por segundo de los videos oscilan en 25 FPS, la diferencia entre cada par de fotogramas tomados en forma consecutiva es casi imperceptible en términos de la ubicación de los objetos. Las pruebas de tolerancia del filtro de Kalman para obtener un *tracking* correcto indicaron que es seguro saltearse hasta 4 fotogramas en secuencias de video sin problemas técnicos.

Los algoritmos fueron implementados utilizando las arquitecturas y los parámetros pre-entrenados por Bochkovskiy. El proceso predictivo asociado a cada video demoró, en promedio, 4.8 hs. A modo de ensayo, también se procesaron algunos videos utilizando *Google Colab* con GPU, y el tiempo de predicción se redujo a un promedio de 40 min.

3 Conjunto de datos

El sistema de monitoreo urbano cuenta con 450 cámaras que registran el tránsito las 24 hs. Aunque el potencial de cobertura de la base de datos pareciera promisorio, existe una relación costo-beneficio entre la cobertura y la capacidad de grabación y procesamiento de las cámaras. Bajo esta restricción, se seleccionaron 132 cámaras, a partir de las cuales se obtendrían registros de una hora de duración en hora pico y no pico, y en días laborales y no laborales (sábados, domingos y feriados). Las horas pico

comprenden el horario entre las 7 y las 9 hs., y entre las 17 y 20 hs. Luego, las horas no pico comprenden las grabaciones entre las 14 y las 16, así como las grabaciones nocturnas.

Se recolectaron 3796 videos que, en principio, deberían ser equivalentes a las horas grabadas. Sin embargo, los registros presentaron ciertas dificultades técnicas. Antes de procesar los videos, se hizo una pre-selección atendiendo a una serie de fallas presentes, clasificadas en las siguientes categorías:

1. Microcortes: el 42% de los videos recibidos (1577 casos sobre el total de las grabaciones) debieron ser descartados ante la severidad de los microcortes o, incluso, debido al contenido nulo del video. La saturación de la red inalámbrica que comunica las cámaras con el Centro de Monitoreo es una posible causa de este problema. Si bien la distribución de las anomalías es bastante uniforme en días y horas, las grabaciones efectuadas los domingos y en proximidad de la noche fueron las más afectadas.

2. Iluminación: otro factor adverso ha sido la iluminación en horas pico y nocturna. La determinación del número de pasajeros en hora pico es vital para la cuantificación de la demanda. Sin embargo, este horario coincide con la salida y la puesta del sol. En ciertas ubicaciones, las cámaras sufren debido al haz de luz que se proyecta en el lente. Por otra parte, la medición de la circulación nocturna es necesario para cuantificar la demanda en horas valle. Sin embargo, ciertas zonas de la ciudad carecen de la iluminación adecuada para delinear los contornos de los vehículos. Mientras que el número de cámaras perjudicadas por la luz solar ascendió al 11% de los videos registrados en hora pico (244 casos sobre 2231 grabaciones en los rangos horarios de 7-8 hs. y 17-18 hs.), las cámaras afectadas por poca luminosidad alcanzaron el 58% de las filmaciones en horas de la noche (255 casos sobre 459 grabaciones nocturnas).

3. Ángulo: una cuestión que limita el alcance completo es el enfoque de las cámaras. Para poder contemplar dirección y cantidad de los vehículos, es necesario que las cámaras apunten a todas las bocacalles en la intersección que se desea medir. Normalmente eso no sucede, por lo que la cuenta es una estimación del flujo total en el punto de medición e impide obtener el conteo de los giros. Al mismo tiempo, un grupo reducido de cámaras se encuentran situadas en altura, con un enfoque que ha dificultado la detección de los vehículos en función del conjunto de entrenamiento utilizado. En total, la cantidad de videos en esta categoría alcanzó el 15% de los videos recibidos (568 casos sobre el total de las grabaciones).

4. Movimiento: debido a que la principal función de las cámaras es el control y la vigilancia, es normal que algunas cámaras cambien de posición por algunos minutos, o incluso su enfoque se altere durante la mayor parte del tiempo grabado. Sólo el 2% de los videos recibidos (81 casos sobre el total de las grabaciones) han sido descartados por este motivo.

5. Suciedad, falta de nitidez y obstáculos: algunas cámaras son antiguas, o presentan suciedad en el lente que requiere cierto mantenimiento. Algunas otras presentan obstáculos en el medio de la imagen, tales como cables o postes, que representan una dificultad adicional para DeepSORT. No obstante, estos videos suman sólo el 1% del total recibido (41 casos sobre el total recibido).

Una vez que se preseleccionaron los videos, el número final a procesar descendió a 1365. Mediante el reentrenamiento de YOLO, utilizando *Data Augmentation*, probablemente puedan reinsertarse los casos con problemas de luminosidad o las capturas en altura.

4 Resultados preliminares

El algoritmo preliminar, con los parámetros de las versiones originales de YOLO v4 y DeepSORT, mostró resultados dispares. La **Tabla 1** muestra la diferencia entre el conteo automatizado y el manual de 3 cámaras testigo.

La Cámara 1 es un dispositivo de alta resolución y con una conexión inalámbrica sin fallas durante la hora registrada, mientras que la Cámara 2 es tecnológicamente más antigua y cuya conexión durante la grabación presenta microcortes. Si bien el error de conteo del primer caso es relativamente bajo, la discrepancia en la segunda cámara es extremadamente elevada. La causa de esta distorsión podría estar atribuida mayormente a la calidad de la grabación. En una filmación de buena calidad, los vehículos transitarían a lo largo de la calle, generando una sucesión continua y contigua de objetos en cada cuadro. En la Cámara 2, por el contrario, los vehículos desaparecen repetidamente y vuelven a aparecer en lugares alejados debido a los cuadros faltantes. Entonces, DeepSORT podría atribuirles una identidad distinta al no ser capaz de predecir su ubicación correctamente.

La Cámara 3, que es un dispositivo de alta definición, muestra la complejidad de este fenómeno. Aunque la calidad del video no parece diferir demasiado en apariencia, las grabaciones de la misma cámara tienen dos niveles de error distintos. El conteo muestra un nivel bajo durante las grabaciones del martes, mientras que el error sube fuertemente durante la noche del viernes y el domingo a la tarde. Una hipótesis sobre las causas de este problema es la existencia de una mayor cantidad de cuadros corruptos durante esos momentos

de la semana, que provocaría predicciones erróneas de DeepSORT. Este tipo de errores se traducen en la asignación de identidades múltiples a un mismo vehículo, cuestión que se evidencia en la sobreestimación del conteo automatizado.

Con respecto a la omisión de fotogramas para la aceleración de las predicciones, los resultados podrían haberse visto afectados significativamente si las porciones de video faltantes hubiesen sido del orden de los microsegundos. Sin embargo, los problemas técnicos en los videos implican la pérdida de secuencias de varios segundos. Luego de un microcorte pequeño que puede implicar la pérdida de decenas de fotogramas, la ubicación de un objeto que estaba en un cuadrante de la imagen aparece en otro lugar totalmente diferente. Por lo tanto, la predicción del filtro de Kalman sobre ese mismo objeto se ve afectada de igual manera que si no se hubiesen eliminado periódicamente los dos fotogramas para la aceleración de las predicciones.

Tabla 1. Test del algoritmo. Conteo automático vs. conteo manual.

Cámara	Horario	Duración (min.)	Conteo Visual	Conteo Algoritmo	Variación Absoluta	Variación Relativa
Cámara 1	Sab. 17/10 13 hs.	4:32.	20	19	-1	-5%
Cámara 2	Vie. 20/11 8 hs.	2:52	38	94	+56	+147%
Cámara 3	Mar. 17/11 7 hs.	1:25	7	7	0	0%
Cámara 3	Mar. 17/11 14 hs.	1:10	12	13	+1	+8%
Cámara 3	Vie. 20/11 24 hs.	3:52	7	29	+22	+314%
Cámara 3	Dom. 15/11 17 hs.	1:49	31	59	+ 28	+90%

5 Comentarios finales y trabajo futuro

Con el fin de estimar el flujo de tránsito vehicular en la ciudad argentina de Bahía Blanca (Argentina) aprovechando la infraestructura de seguridad urbana, se aplicaron redes neuronales para el reconocimiento y seguimiento de vehículos. En particular, se implementaron los algoritmos YOLO v4 y DeepSORT. El experimento se realizó en base a grabaciones de cámaras de vigilancia urbana distribuidas en la ciudad, las cuales contemplan horarios y días diversos para captar la intensidad vehicular en horas pico y no pico.

Los resultados del algoritmo preliminar muestran algunos problemas derivados de los microcortes en las filmaciones. Las distorsiones en los resultados son fuertes aún en videos cuyas imágenes, a primera vista, no parecen sufrir problemas. Resta identificar un mecanismo robusto ante la presencia de fallas en la continuidad del video.

Como trabajo futuro inmediato se planifica implementar aumentación de datos para incorporar aquellos videos con problemas de sol o poca luminosidad que aún conserven algún grado de identificabilidad en los vehículos. Adicionalmente, se espera incorporar las cámaras en altura al re-entrenar ambas redes neuronales con la base de datos VeRi [5], que también dispone de capturas de vehículos en ángulos que nos son los habituales. Finalmente, se implementará en la nube para paralelizar las tareas y disminuir los tiempos de cómputo.

Referencias

1. GIECOV (2017). Informe colisiones viales. Departamento de Ciencias de la Salud. UNS. Disponible en: https://www.cienciasdelasalud.uns.edu.ar/docs/repositorio/Informe%20final%202017.pdf
2. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
3. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.
4. Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In 2017 IEEE international conference on image processing (ICIP) (pp. 3645-3649). IEEE.Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
5. Liu, X., Liu, W., Mei, T., & Ma, H. (2017). Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. IEEE Transactions on Multimedia, 20(3), 645-658.