

MINERÍA DE DATOS PARA LA DETERMINACIÓN DE VARIABLES DE TRÁNSITO MEDIANTE LA APLICACIÓN DE MONITOREO SATELITAL

GARCÍA, Francisco_J.; ESCOBAR, Diego_A.; VÁSQUEZ, Luis_Ricardo. Universidad Nacional de Colombia. Sede Manizales. Facultad de Ingeniería y Arquitectura. Departamento de Ingeniería Civil.
fjgarciaor@bt.unal.edu.co daescobarga@bt.unal.edu.co lrvasquezv@unal.edu.co

RESUMEN

En el presente artículo se explora el espectro de las posibilidades que ofrece la minería de datos orientada a la obtención de diferentes variables del tránsito, según las múltiples bases de datos que se vienen generando como consecuencia del monitoreo satelital, aplicado a diferentes actividades comerciales, de seguridad e investigación, entre otras.

Cada vez es más frecuente la aparición de flotas completas de vehículos con sistema de posicionamiento global (GPS) instalados, procedente de las compañías que hoy en día utilizan el monitoreo con fines diferentes al seguimiento de variables del tránsito, como son el control de flota con fines operacionales, servicios de taxi, sistemas de seguridad, sistemas de transporte público colectivo, entre otros, y cuyo propósito es prestar una ayuda a la explotación del parque automotor, controlando diferentes aspectos, según la orientación o la necesidad del empresario.

De esta forma las bases de datos generadas pueden ser utilizadas para obtener subproductos relacionados con diferentes variables del tránsito como son las velocidades de operación, la tasa de paradas, zonas de parqueo, origen – destino del viaje, número de viajes diarios, kilómetros recorridos diarios, entre otras muchas posibilidades. En este artículo se propone una metodología para la obtención de diferentes variables y se presenta un caso específico de aplicación de la misma.

Palabras claves: *Minería de datos, Monitoreo, Transporte, GPS, Tránsito.*

ABSTRACT

This paper explores the spectrum of possibilities that data mining offer when used to obtain transport variables, according to the multiple databases generated from satellite monitoring, which are applied, among others, to commercial, security and research activities.

It is becoming more and more frequent to find entire fleets of vehicles with installed GPS systems, from companies that today use monitoring for purposes other than transport. Applications such as fleet control for operational purposes, taxi services, security systems, and public transport systems, among others, aim at supporting the operation of the vehicle fleet, controlling the different aspects, in line with the entrepreneur needs or orientation.

Thus the generated databases can be used to obtain byproducts related to different transport variables such as operation speeds, stop rates, parking zones, trip schedule, number of daily trips, and daily mileage, among many others. This paper proposes a methodology to obtain different variables and presents a specific case where this methodology can be applied.

Palabras claves: *Data Mining, Monitoring , Transport, GPS, Traffic.*

1. INTRODUCCIÓN

En los últimos tiempos se ha popularizado el monitoreo satelital de vehículos mediante el GPS, permitiendo obtener la posición, el recorrido y otras características de un número masivo de vehículos. Las primeras aplicaciones comerciales de dicho monitoreo estuvieron relacionadas con sistemas de alarma y recuperación de vehículos robados. Hoy, el monitoreo satelital tiene múltiples aplicaciones con un amplio rango de opciones, que van desde la simple captura de información de campo, como por ejemplo, desde el mapeo de una carretera, hasta el control en tiempo real de dispositivos instalados en vehículos que, por ejemplo, permiten abrir las puertas de su coche si usted dejó las llaves en el interior.

Una mirada general permite ver como el monitoreo abarca campos de acción cada vez más amplios, siendo hoy posible, monitorear en tiempo real aspectos relacionados con el seguimiento o abandono de una ruta (de reparto de mercancías o de un vehículo de transporte publico), si se detiene y abre una puerta en un lugar no programado, el peso de la carga que lleva, la velocidad a la que se desplaza, la velocidad máxima alcanzada, el vehículo disponible más cercano de un lugar específico (para el caso de un servicio de taxi o de ambulancia, por ejemplo), la evaluación de rutas alternativas para llegar de un sitio a otro dado algún evento de tránsito, apagar el motor del vehículo ante una acción de robo, localizar un sitio donde se desea adquirir un servicio según el lugar donde se encuentra el interesado. Finalmente, también se viene trabajando el monitoreo de personas con fines académicos, operativos y de seguridad como, por ejemplo, reos con libertad condicional que deben permanecer en una cierta área.

2. METODOLOGÍA DEL ANÁLISIS

a- Definición del objetivo

Este nace como el resultado de una necesidad proveniente de un problema específico, o puede ser que se posea la información y se desee explorar alternativas de su uso. Como el espectro de posibilidades es tan amplio, se deberá sintetizar en algunas variables a determinar, direccionando los esfuerzos a la obtención de los insumos que permitan obtener dichas variables, determinando de esta forma los alcances y limitaciones enmarcados por las características de la información fuente.

b- La base de datos

El segundo paso consiste en obtener la base de datos, bien sea existente o tomada a partir de un programa estructurado para la toma de datos de campo, instalando el dispositivo (GPS) en los vehículos objeto de la prueba.

No se abordan en este documento los aspectos relacionados con el diseño de la muestra, por ser este un tema específico del trabajo que se pretenda hacer, sin embargo es un asunto que debe ser considerado al momento de

formular y ejecutar el proyecto. Es también necesario recordar que una de las propiedades de la minería de datos, es lo relacionado con la gran cantidad de información con la que se puede disponer (base de datos con información de meses o años), pero que puede ser muy específica de un grupo en particular por lo que los resultados generalmente deberán estar enmarcados a la población a la que pertenece la información de origen.

Bien sea que se obtenga la información recolectada en el proyecto o información existente de alguna base de datos obtenida con anterioridad, se tendrán cuatro grupos de datos fundamentales:

- La identificación del dato: Código ID del dato
- La posición del punto: Coordenadas x, y, z
- Tiempo del dato: La fecha y hora del dato
- La precisión del equipo Horizontal y vertical

En la identificación del dato se puede considerar aspectos relacionados con el código en el proyecto del GPS, el vehículo en el que está instalado, número secuencial del dato, entre otros.

La posición del dato corresponde a las coordenadas que le localizan en el espacio y estas pueden tener formatos diferentes para los sistemas de coordenadas, con diferentes orígenes y unidades. Normalmente pueden ser la longitud, latitud y altura sobre el nivel del mar, éstos pueden estar en metros, kilómetros o grados.

El tiempo del dato define la fecha en un formato específico como día-mes-año, e incluye la hora-minutos-segundos-décimas o centésimas de segundo. La frecuencia del dato generalmente es un parámetro que se configura en el equipo y puede ser incluso cada segundo o cuando sucede un evento según como se programe el equipo.

Por último, la precisión del equipo es una característica propia, que se define al momento de su adquisición en función del precio, la tecnología y las necesidades del proyecto, donde generalmente interviene la marca del equipo, la capacidad de almacenamiento, los accesorios, entre otras muchas características. En todo caso la precisión está directamente asociada a la cantidad de satélites con los que trabaja para determinar el dato.

c- Base cartográfica

Si bien el manejo y procesamiento de la información se puede hacer con una adecuada codificación, es indudable que la eficiencia es mucho mayor cuando se utiliza un sistema de información geográfico (SIG), y es mejor aun si dicho SIG está orientado al manejo de datos de tránsito y transporte. Cualquiera sea el caso deberá existir una base cartográfica actualizada que permita la localización espacial de la información. En algunos casos, cuando no se dispone de la base cartográfica adecuada, su proceso de generación será la primera fase del proyecto, donde se requiere hacer una importante inversión de recursos, siendo este producto por sí mismo de gran importancia, dado que sus cualidades permiten múltiples usos y aplicaciones.

d- Corrección y validación

El siguiente paso consiste en corregir y validar la información. Se debe corregir la posición del punto puesto que usualmente sus coordenadas están desplazadas y por tanto se debe realizar un post proceso de la información según un factor de corrección procedente de una estación cercana. Validar la información consiste en verificar la estructura de la base de datos, el sistema de coordenadas, el tiempo entre dato y dato, la secuencia de datos, el formato de los archivos, la precisión prevista, y los demás que se consideren necesarios. De los datos anteriores, la secuencia de datos puede ser especialmente compleja, dado que algunos equipos, cuando temporalmente pierden señal, entregan información incoherente, teniéndose como resultado una secuencia grande de puntos

correctos, mezclados con algunos pocos incorrectos, siendo estos de difícil localización y corrección por el gran tamaño de las bases de datos que se manejan.

e- Procesamiento de la información

El procesamiento deberá enfocarse a la obtención de las variables ya definidas, esto puede requerir diferentes orientaciones según la variable que se desee estudiar. No obstante, en la mayoría de los casos se requerirá localizar la información en la base cartográfica y asociarla a la misma. En este sentido, como se dijo anteriormente, será más eficiente el proceso si la información se maneja en un SIG orientado al tránsito y transporte.

Usualmente la información puede requerir que sea procesada a nivel de intersecciones (nodos), tramos de vía (arcos), áreas geográficas (barrios, comunas, vecindarios, zonas de análisis de transporte – ZAT), siendo este proceso asociativo de la información, posiblemente la parte más larga y dispendiosa de la etapa del procesamiento.

En esta etapa, vincular la información a los nodos, a los arcos y a las ZAT, permitirá que en forma directa el procesamiento de la información se haga mediante las herramientas de cálculo que posee el SIG, siendo de gran utilidad en su análisis y en la presentación gráfica de resultados.

f- Análisis de resultados

Se debe tener en cuenta que existen datos referentes con los cuales se logran comparar nuestros resultados y que pueden proceder de estudios anteriores. En algunos casos son sorprendentes los resultados obtenidos, presentando cambios significativos respecto a los existentes en otras fuentes, por lo que es necesario analizar detenidamente las posibles razones de estos cambios.

Sin embargo, las bases de datos de donde se procesa la información llegan a ser tan grandes que permiten hacer análisis en una variada gama de combinaciones de tiempo y espacio, pudiéndose determinar valores de condiciones o características sobre las cuales no se tienen valores que sirvan como puntos de referencia y que permitieran hacer comparaciones respecto a los valores tradicionales, lo que se traduce en aportes significativos de la investigación.

3. EJEMPLO PRACTICO DE APLICACIÓN

A continuación se desarrolla un ejemplo práctico de aplicación, realizado en una ciudad intermedia colombiana, donde se generó una base de datos de unos 12 millones de registros, obtenidos del monitoreo planificado de una amplia gama de vehículos, en un extenso periodo de tiempo. En este caso se trabajan las variables velocidad y tasa de paradas, sin embargo, esta misma base puede servir para determinar variables como parqueo, origen y destino de los viajes, viajes promedio diarios, horas de uso del vehículo, determinación de horas pico, recorrido promedio diario, entre otras variables. Finalmente, las variables analizadas pueden ser utilizadas en estudios de accesibilidad al territorio, modelación del tránsito, coordinación de semáforos y costos de operación vehicular, por citar solo algunos ejemplos.

3.1 OBJETIVO.

Como resultado de la investigación financiada por la Dirección de Investigaciones de Manizales (DIMA), titulada: “Determinación de algunas características operativas del tránsito sobre las vías con fines de modelación, mediante la utilización de GPS”, se logró obtener una gran base de datos (mas de 12 millones de registros) de movimiento real de vehículos sobre la red vial de la ciudad de Manizales que puede ser usada con múltiples

propósitos y la cual consistió en la instalación de equipos GPS de alta precisión, de forma simultánea en diferentes tipos de vehículos, durante largos periodos de tiempo.

Se definió como objetivo, la determinación de las velocidades sobre cada uno de los más de diez mil arcos o segmentos de vía, que componen la red vial de la ciudad de Manizales, a partir de las velocidades registradas sobre cada uno de ellos.

La ciudad de Manizales (Ver Figura 1) posee alrededor de 370000 habitantes, se encuentra ubicada en la región centro occidente de Colombia, a una altura de 2150 m.s.n.m., entre los 5.4° de latitud norte y 75.3° de Greenwich, sobre la prolongación de la cordillera andina, a una distancia de 275 km. de la capital Bogotá D.C. La conformación urbanística de la ciudad responde a la adaptación a una topografía bastante quebrada y abrupta, dado su emplazamiento geográfico, lo cual hace que ésta posea características muy particulares; ubicada en el filo de la colina, permite una apertura permanente al paisaje, a todo lo largo de su recorrido, es así como el crecimiento urbanístico se ha venido adaptando de manera orgánica a las condiciones topográficas, teniendo que Manizales se caracterizará por contar con una estructura urbana no continua, con una red de infraestructura vial con grandes limitaciones, sobre la cual, el transporte público en algunos casos particulares, transcurre sobre pendientes superiores al 18%.



Figura 1. Panorámica de la ciudad de Manizales

3.2 LA BASE DE DATOS.

La base de datos que se tomo en el GPS, se estructuro de la manera más simple posible, conformada por un código de identificación del dato, un código de identificación del GPS o Vehículo, La hora, la fecha, la latitud, la longitud y la velocidad, tal como se presenta en la Figura 2.

Los formatos que generan los GPS pueden ser muy variados, siendo comunes archivos de formato txt, xml, csv o gpx, por lo que puede requerirse un trabajo previo para llevar el archivo de este formato al de la base de datos con la que se pretende procesar.

Datmanizales						
Id_dato	Id_vehic	Hora	Fecha	Latitud	Longitud	Velocidad
1	01	105604	091007	5.064124	75.520819	
2	01	105608	091007	5.063987	75.520521	8.947713
3	01	105609	091007	5.063954	75.520448	8.901371
4	01	105613	091007	5.063826	75.520165	8.303643
5	01	105614	091007	5.063794	75.520094	8.653115
6	01	105615	091007	5.063762	75.520026	8.350345
7	01	105616	091007	5.063730	75.519958	8.350345
8	01	105617	091007	5.063700	75.519891	8.156637
9	01	105618	091007	5.063669	75.519824	8.202671
10	01	105619	091007	5.063639	75.519758	8.055356
11	01	105620	091007	5.063609	75.519693	7.954337
12	01	105621	091007	5.063579	75.519628	7.954337
13	01	105622	091007	5.063549	75.519565	7.753128
14	01	105623	091007	5.063520	75.519501	7.807080
15	01	105624	091007	5.063491	75.519438	7.706010
16	01	105625	091007	5.063462	75.519376	7.605221
17	01	105626	091007	5.063434	75.519315	7.457692
18	01	105627	091007	5.063405	75.519253	7.605221
19	01	105628	091007	5.063377	75.519193	7.356857
20	01	105629	091007	5.063349	75.519135	7.156100

Figura 2. Estructura de la base de datos

3.3 BASE CARTOGRAFICA.

La ciudad de Manizales cuenta con diferentes restitutiones cartográficas, realizadas aproximadamente cada 10 años. También se ha desarrollado por variadas entidades del estado, diferentes esfuerzos tendientes a la obtención de un Sistema de Información Geográfica (SIG) unificado para la ciudad, lográndose poseer hoy en día un sistema actualizado que permite una amplia gama de opciones de análisis. Adicionalmente, en un proceso similar, se tiene una base informática en un SIG de aplicación específica al tránsito y al transporte. En la Figura 3, se presenta una panorámica de la red vial de Manizales, en el SIG TansCAD®, sobre la cual se desarrollo la presente investigación.



Figura 3. Red vial de Manizales en TransCAD®

3.4 CORRECCION Y VALIDACION.

En la Figura 4 se muestra el recorrido de vehículo que sigue adecuadamente el lineamiento de las vías de la ciudad, lo que no siempre ocurre, como se presenta en la Figura 5, donde el recorrido se hace por el medio de la manzana. Estas situaciones se detectan más fácilmente cuando es una secuencia de datos y no un sólo punto aislado. En todo caso el proceso de corrección y validación de la información se simplifica con la ayuda del SIG.

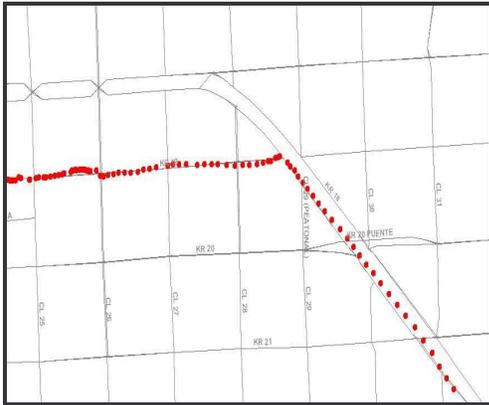


Figura 4. Recorrido bueno de un vehículo.

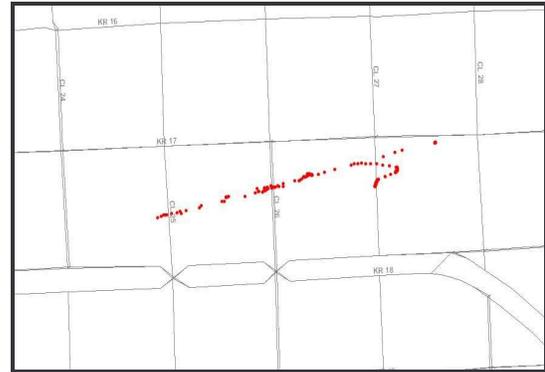


Figura 5 Recorrido erróneo de un vehículo

3.5 PROCESAMIENTO DE LA INFORMACIÓN.

En este caso, se requiere asociar todos y cada uno de los datos validados a cada arco, para lo cual se desarrolla un algoritmo que localiza el dato más cercano a un nodo (intersección), de la secuencia de cada recorrido, cruzándose esta información con la asociación de los nodos extremos de cada arco, para así definir de cada recorrido los datos que se asocian a cada arco.

Se obtiene de esta forma la estructura que se presenta en la Figura 6, adicional a las variables ya presentadas, en donde se incluye el código del arco o tramo de vía, su nodo inicial y final, la cantidad de registros que pertenecen a ese arco en ese recorrido (cant_reg), la longitud del arco (lon_arco), registro inicial y final de la secuencia de la base de datos (reg_in, reg_fin).

Datmanizales2														
	Id_datos	Id_vehic	Hora	Fecha	Latitud	Longitud	Arco	Nodo_in	Nodo_fin	Cant_reg	Lon_arc	Vel	Reg_in	Reg_fin
	247	01	151724	091007	5.068595	-75.522221	18	1034	1035	3	4.56	5.47	15438	15441
	248	01	151727	091007	5.068592	-75.522071	6590	1035	3916	10	68.83	24.78	15441	15451
	249	01	151737	091007	5.068616	-75.521553	6834	3916	4086	14	72.75	18.71	15451	15465
	250	01	151751	091007	5.068645	-75.520916	7152	4086	6673	26	71.50	9.90	15465	15491
	251	01	151817	091007	5.068702	-75.520251	6349	6673	6674	7	71.99	37.02	15491	15498
	252	01	151824	091007	5.068750	-75.519619	6849	6674	6644	253	71.92	1.02	15498	15751
	254	01	152237	091007	5.068770	-75.518958	7102	6644	6663	10	72.92	26.25	15751	15761
	255	01	152247	091007	5.068834	-75.518309	6900	6663	6609	15	72.40	17.38	15761	15776
	256	01	152302	091007	5.068864	-75.517654	6589	6609	6681	9	73.37	29.35	15776	15785
	257	01	152311	091007	5.068862	-75.516986	6459	6681	6641	9	69.72	27.89	15785	15794
	258	01	152320	091007	5.068872	-75.516368	6927	6641	6645	9	73.09	29.24	15794	15803
	259	01	152329	091007	5.068881	-75.515729	7029	6645	6638	511	73.73	0.52	15803	16314
	266	01	153200	091007	5.068887	-75.515046	6432	6638	6725	1112	69.51	0.23	16305	17417

Figura 6. Estructura de la base de datos asociada a la base cartográfica

Con la información ya estructurada de esta forma, se obtiene la velocidad segundo a segundo asociada a cada arco en cada recorrido, con lo que se puede determinar la variación de la velocidad, como se muestra en la Figura 7, en la cual se grafican 7 recorridos de los 86 disponibles, pudiéndose observar la tendencia de velocidad de cada automotor, y como uno de los vehículos (el Veh 6) se detuvo y luego continuo su recorrido.

De igual manera se calcula la velocidad promedio registrada para cada uno de los 10555 arcos que conforman la red vial, asociando estos resultados a cada segmento de vía. Este procesamiento se repite teniendo en cuenta la combinación de variables como día de la semana, hora del día y tipo de vehículo, entre otras posibilidades.

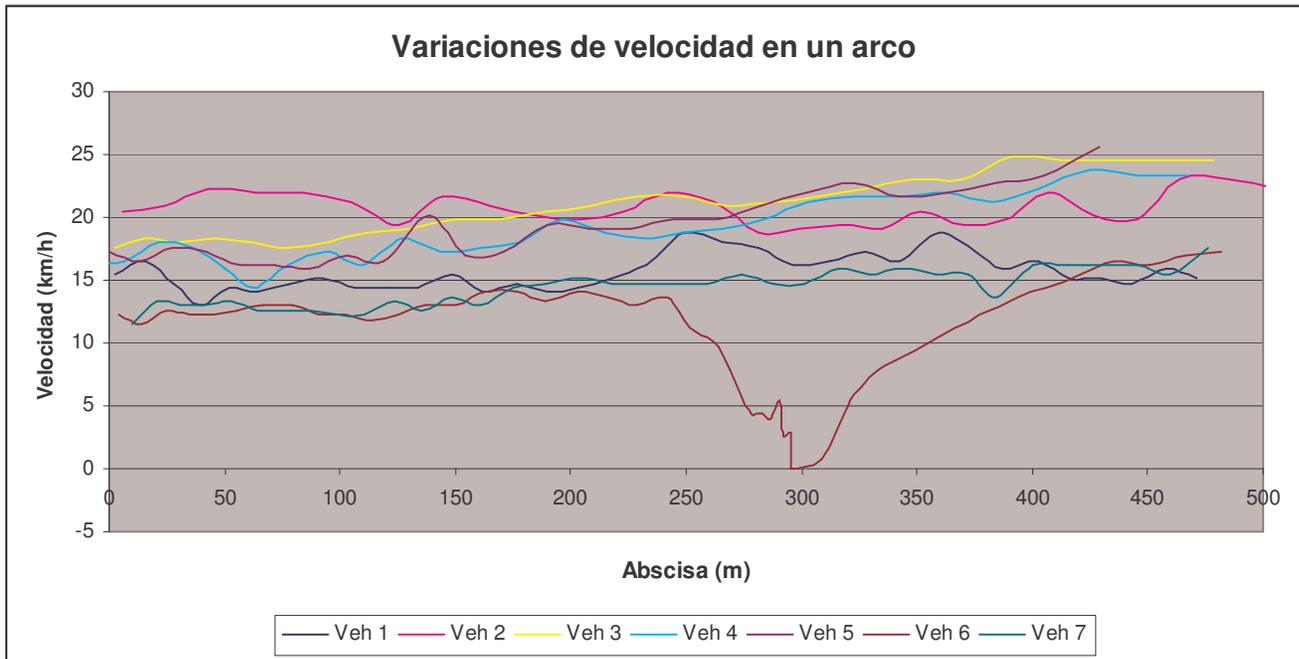


Figura 7. Variaciones de la velocidad en el arco 9869 de la base cartográfica

3.6 ANÁLISIS DE RESULTADOS.

3.6.1 ANÁLISIS DE VELOCIDAD

Con la información obtenida, se puede hacer una variada cantidad de análisis, sobretodo combinando dicha información con el SIG, generando graficas como la que se presenta en la Figura 8, en la cual se resaltan las pocas vías sobre las que se logran velocidades promedio superiores a los 45 km/h. Se encuentra también en esta figura, una muy buena correspondencia entre las altas velocidades y las vías arterias principales, sobre todo en los sectores alejados del centro de la ciudad. Es también evidente como los sitios reconocidos por los continuos periodos de congestión, impiden que se logren altas velocidades de operación.

Similar al análisis anterior, se puede evaluar la combinación de las variables antes citada, disponiendo de parámetros como día de la semana, hora del día y tipo de vehículo, lo que permite enriquecer el análisis.

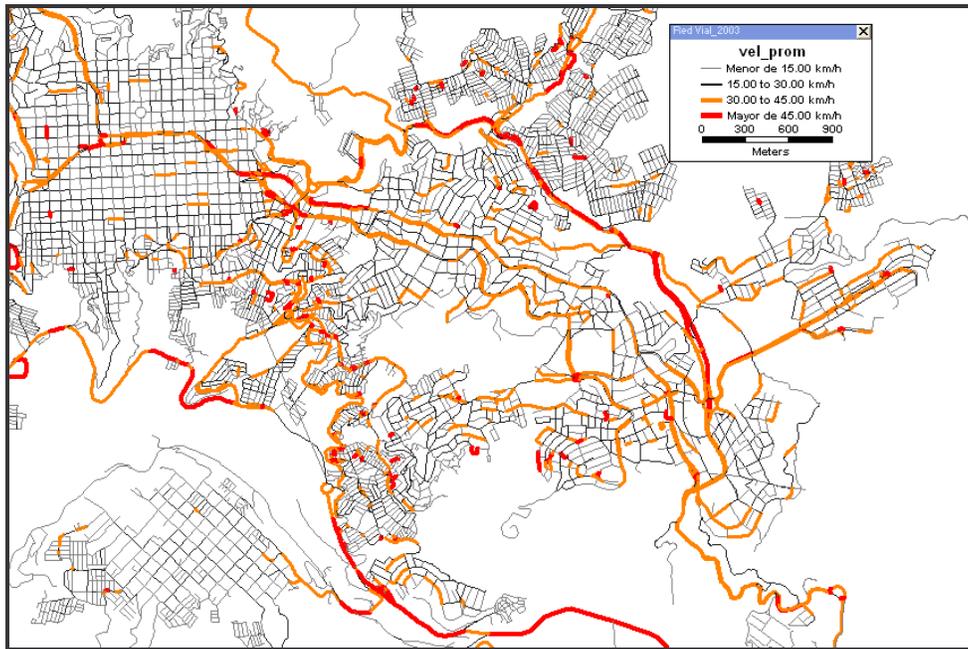


Figura 8. Mapa de velocidades de operación promedio en diferentes arcos de la red.

3.6.2 TASA DE PARADAS

Un análisis adicional que se puede realizar con los datos obtenidos mediante el monitoreo satelital, se presenta en la Figura 9, donde se observa una análisis de la tasa de parada, con rangos cada 25%, donde se destaca en color rojo las vías que refieren detenciones superiores al 75% .

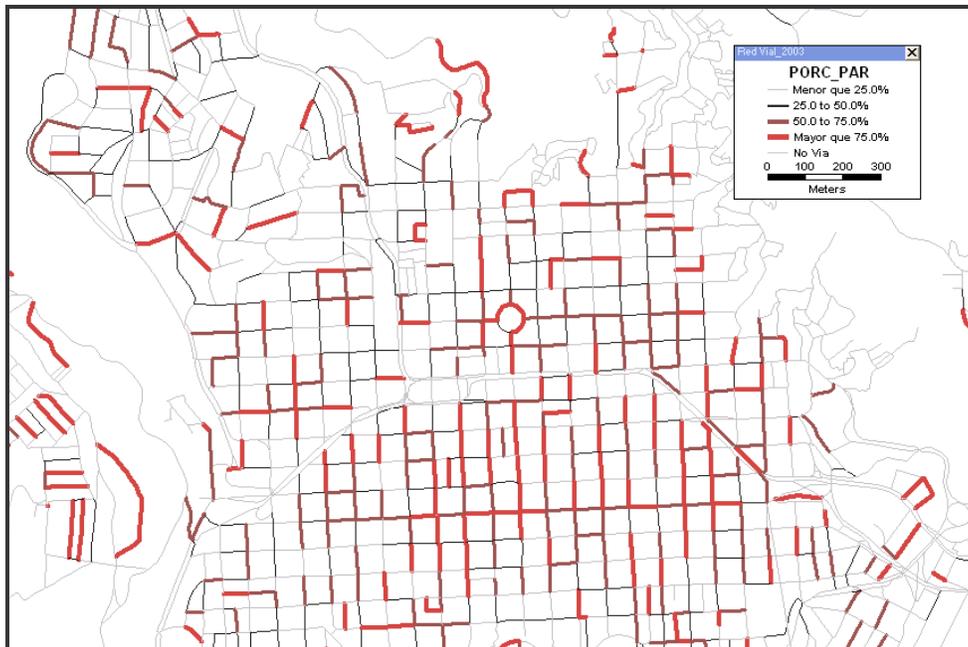


Figura 9 – Tasa de paradas por rangos

Los valores con tasas de paradas altos, son coherentes con la existencia en estos sitios de dispositivos de control tipo pare o semáforo, valores altos de tráfico vehicular y/o peatonal, carencia de coordinación entre los semáforos existentes, uso de carriles de servicio, existencia de maniobras permanentes de parqueo y paraderos del sistema de transporte público entre otros.

4. CONCLUSIONES

Es lógico pensar que el uso de las tecnologías relacionadas con el monitoreo satelital será cada vez más frecuente, y que por tanto estas aplicaciones serán de uso general. Es indudable que los beneficios serán muy altos y que estos análisis suministran información y resultados prometedores, más aún cuando la mayoría de la información podría resultar gratis, si se usa la existente, procedente de las compañías que hoy en día utilizan el monitoreo con fines diferentes a los expuestos en este documento, como son el control de flota con fines operacionales, servicios de taxi, sistemas de seguridad, sistemas de transporte público colectivo entre otras, y que hoy en día ya disponen de información almacenada de varios meses e incluso años.

Son inmensas las posibilidades de aplicación a casos prácticos, que permitirán un nivel de detalle que antes era inimaginable, como es por ejemplo, determinar para cada hora del día las velocidades de operación de los vehículos, cuando hoy en día se suele manejar por ejemplo, un sólo valor para fines de modelación de redes.

También es posible discriminar variables a nivel de vehículos cada vez más específicos como por ejemplo auto particular, camionetas, camperos, motos, taxis que generalmente son expresados como vehículo liviano. Este nivel de detalle redundara en mejores resultados cuando se pretenda crear y evaluar escenarios muy particulares, incluso en el horizonte de microsimulación de problemas específicos.

REFERENCIAS

Bullock, Ph. et al. (2005) Using GPS technology to measure on-time running of schedules bus service. University of Sydney. Journal of Public Transportation, Vol. 8, No. 1.

Michael D Fontaine, Brian L. Smith. (2007) Investigation of the performance of wireless location technology – based traffic monitoring systems. Journal of transportation Engineering.

Ministry of Transport. (2008). Technical Annex to the satellite monitoring service for vehicles of the National Highways Safety Program. Republic of Colombia.

S. Shekhar, C.-T. Lu, S. Chawla, and P. Zhang. 2000. Data mining and visualization of twin-cities traffic data. In Technical Report TR 01-015, Dept. of CSE, Univ. of Minnesota.

M. Hadjieleftheriou, G. Kollios, D. Gunopulos, and V. Tsotras. 2003. On-line discovery of dense areas in spatiotemporal databases.

Autorización y Renuncia

Los autores autorizan a LACCEI para publicar el escrito en los procedimientos de la conferencia. LACCEI o los editores no son responsables ni por el contenido ni por las implicaciones de lo que esta expresado en el escrito