

Colombia en 35 terabytes anuales

Australia inventó los cubos de datos de imágenes de satélite y los puso a disposición de la comunidad como código abierto. Aunque generalmente tienen un uso ambiental, las posibilidades de aplicación son variadas.

Analizar la cobertura vegetal del país, de sus ríos y lagunas, el comportamiento de las inundaciones, de los incendios, los cambios en los páramos y un sinnúmero de posibilidades relacionadas con el ambiente ha sido posible desde que, en 1972, el satélite Landsat 1 transmitió las primeras fotografías satelitales de la tierra y en particular de nuestro territorio.

Desde entonces, cada foto ha sido analizada individualmente por distintas entidades meteorológicas colombianas. Para obtener el material de trabajo adecuado, un experto debe escogerlas y corregirlas aplicando diferentes algoritmos y parámetros. En general, este trabajo ha sido manual, demorado y no siempre confiable; en esta labor un funcionario puede emplear hasta el 70 % de su tiempo y todavía no ha empezado la interpretación.

“Cada persona trabaja de manera independiente, con lo cual es factible la duplicación de tareas y de costos, además de conclusiones distintas sobre una misma imagen y un mismo tema”, señaló Germán Bravo, profesor del Departamento de Ingeniería de Sistemas y Computación (DISC).

Esta era la situación hasta que en el 2013 los australianos desarrollaron el concepto de los cubos de datos, conformados por una serie de imágenes de satélite superpuestas que contienen la historia física de un territorio. La implementación realizada fue publicada como código abierto.

Apropiar esa tecnología para manejar hasta 70 terabytes de información al año, llevarla a la infraestructura del Instituto de

Hidrología, Meteorología y Estudios Ambientales de Colombia (Ideam) y escribir los algoritmos adecuados para que funcione con los datos recogidos por satélites sobre Colombia fue el desafío que afrontó el grupo de Comunicaciones y Tecnología (Comit) del Departamento de Ingeniería de Sistemas y Computación. El Ministerio de Ambiente y Desarrollo Sostenible lanzó el Cubo de Datos del Ideam —como se conoce— el 17 de febrero del 2017 y el profesor Germán Bravo lo presentó en el segundo foro de Ingeniería de la Información.

El proyecto debía, además, estimar los requerimientos en tiempo de procesamiento de imágenes, en carga de red y en espacio de almacenamiento, problemas tratados en *big data*. Además de Los Andes y del Ideam, participaron Commonwealth Scientific and Industrial Research Organization (Csiro), Committee on Earth Observation Satellites (Ceos), National



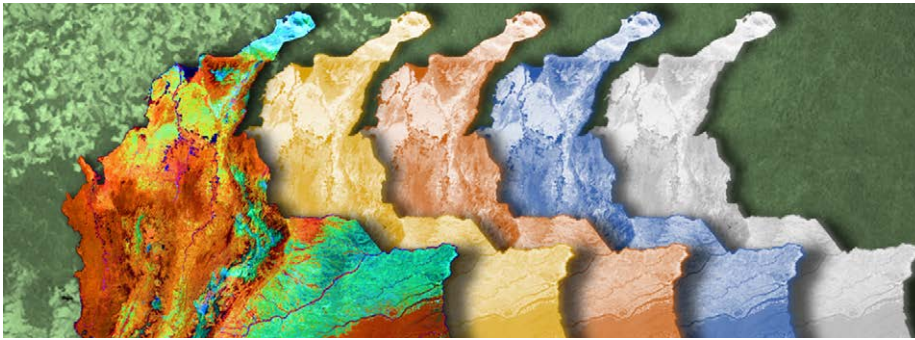
El profesor Germán Bravo explicó que la tecnología de los cubos de datos de imágenes puede aplicarse a diferentes problemáticas de *big data* que involucren fotografías.

Aeronautics and Space Administration (Nasa) y la Universidad de la Salle.

Antes de explicar en qué consistió, Germán Bravo habló de las características de las fotografías de satélite y su evolución. En las primeras, tomadas por satélites con órbitas a 700 kilómetros de altura, un píxel representaba un área de 30 metros de lado. Hoy el satélite Ikonos puede obtener fotografías en las que cada píxel representa un área de 0.80 metros de lado. Desde el punto de vista espectral, los primeros satélites disponían de sensores para dos anchos de banda, mientras que los de hoy pueden tener 15 o más sensores, incluidos los que toman de imágenes de radar. En el aspecto temporal también ha habido cambios: el período de los satélites antiguos (el tiempo en el cual un satélite pasa por exactamente el mismo sitio) era de 16 días, lo que permitía tener 23 imágenes de Colombia completa anualmente; ahora, con Sentinel, se logran cada 5 días.

El Cubo de Datos del Ideam empleará cinco fuentes de fotografías, que son de acceso libre: Landsats 5, 7 y 8, de Estados Unidos, y Sentinel 1A y 1B, de la Unión Europea. El territorio continental de Colombia ocupa un área aproximada de 131 grados cuadrados y se cubre con 65 escenas (fotos) Landsat, cuyo contenido se ubica en una malla de 1 grado por un grado, para un total de 131 zonas, denominadas *tiles*. El conjunto de los 131 *tiles*, cada uno con todas las imágenes de satélite tomadas desde el 2000 a la fecha por cada uno de los satélites involucrados, conforman el cubo de imágenes. Se estima que almacenar estos datos ocuparía el orden de 35 terabytes anuales. Si se contempla un *backup*, se requieren equipos para almacenar 70 terabytes al año, lo que arroja un total de 1190 terabytes para los 17 años considerados en el proyecto hasta el momento.

Ahora, con la adaptación de la tecnología australiana que hizo el grupo Comit, las labores de preprocesamiento y análisis estarán estandarizadas con una serie de algoritmos que facilitan esas tareas; los resultados podrán verse incluso a los tres meses de recibidas las fotografías. Para garantizar la confiabilidad del sistema se diseñó una



Los cubos de datos están conformados por una serie de imágenes de satélite superpuestas que contienen la historia física de un territorio.

interfaz web con los roles de usuario para administrar los datos y la gestión de unidades de almacenamiento. Las imágenes resultantes contienen además la *metadata* del proceso al que fueron sometidas, de manera que estos resultados quedan completamente documentados y son replicables. Los algoritmos conocidos y calificados están disponibles en un repositorio único para manejar de manera uniforme la infor-

mación suministrada contenida en el cubo. Allí, el analista tiene acceso a diferentes versiones de cada algoritmo y puede determinar los parámetros requeridos según el objetivo de su análisis, explicó el profesor.

El proyecto se verificó con un piloto para el Sistema de Monitoreo de Bosques y Carbono de Colombia, del Ideam, que genera alertas tempranas y cuantifica la deforestación. Consistió en un minicubo de 2º x

2º, con imágenes de Landsat 7 del 2000 al 2015. “En una zona del Putumayo, utilizando el algoritmo de detección de cambio, se ve la deforestación. Con base en las medianas se calcula una máscara de “bosque-no bosque” y su evolución en el tiempo”.

Si bien el Cubo de Datos ya está en manos del Ideam, Germán Bravo aseguró que seguirán con el proyecto: “En adelante nuestro trabajo consistirá en mostrarle a la comunidad internacional que se le puede sacar más provecho al *hardware* mediante estrategias de escalamiento horizontal, pues la versión 2 del cubo de Australia aún no tiene todas las operaciones paralelizadas, es decir, no se pueden usar todos los procesadores al mismo tiempo. Nosotros podríamos trabajar en ese problema”. También es probable que se involucren en la definición de un modelo de corrección atmosférica para Colombia, pues el que se usa es de Estados Unidos y no satisface plenamente las particularidades del territorio colombiano. ■

Desarrollan aplicación para apoyar decisiones en movilidad

Con un *software* que apuntalará el análisis de información y la toma de decisiones, la Secretaría de Movilidad distrital podría actuar de manera inmediata sobre sectores críticos del tránsito, establecer qué vías requieren mantenimiento preventivo, evaluar las acciones adoptadas versus la inversión de recursos en accidentalidad, entre muchas otras acciones de Gobierno.

La profesora María del Pilar Villamil presentó este proyecto de seguridad vial desarrollado por Caoba, Centro de Excelencia y Apropiación en *Big Data* y *Data Analytics* (ver revista Foros Isis # 6, pág. 63), con el Departamento Nacional de Planeación y la Secretaría de Movilidad de Bogotá.

En su conferencia “Apoyo a decisiones

Los datos relacionados con movilidad y accidentalidad en Bogotá se emplearon para crear un instrumento que contribuirá a adoptar medidas para resolver problemas como las congestiones en la capital del país. El principal reto es integrar los datos disponibles garantizando buenos niveles de la calidad en la información.

en seguridad vial: un reto para la analítica”, la profesora del DISC aseguró que las congestiones son un asunto que incide negativamente en el producto interno bruto de la capital colombiana: “Bogotá pierde 7.000.000 de horas al año por congestiones, lo cual afecta el 2 % del PIB”. Dijo que uno de los aspectos que impacta la movilidad es la accidentalidad, que en el 90 % de los casos implica a los más vulnerables:

peatones, ciclistas, motociclistas. “En el mundo, las víctimas por esta causa son solo la mitad”.

El proyecto se motivó en comprender lo que pasa con la siniestralidad en Bogotá. El primer desafío que enfrentaron fue establecer lo que se quería hacer, con qué datos, qué decisiones apoyarán, qué personas interactuarán con esta información, con qué infraestructura se cuenta (servidores, etc.).