REVISTA
BITS de Ciencia
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

Nº 8 / Segundo Semestre 2012

Multimedia

- **Benjamin Bustos:** Búsqueda por contenido multimedia
- **Patricio Poblete:** 25 años de NIC Chile
- **Juan Álvarez, Claudio Gutiérrez:** El Primer computador universitario en Chile

REVISTA
BITS de Ciencia
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
UNIVERSIDAD DE CHILE

Comité Editorial:
Nelson Baloian, profesor.
Claudio Gutiérrez, profesor.
Alejandro Hevia, profesor.
Gonzalo Navarro, profesor.
Sergio Ochoa, profesor.

Editor General
Pablo Barceló

Editora Periodística
Ana Gabriela Martínez A.

Periodista
Karin Riquelme D.

Diseño y Diagramación
Sociedad Publisiga Ltda.

Imagen Portada:

Resultado obtenido con un método propuesto por José Saavedra en su tesis doctoral "Image Descriptions for Sketch Based Image Retrieval" usando el benchmark propuesto por: Eitz, M., Hildebrand, K., Boubekeur, T., and Alexa, M. 2011. Sketch-based image retrieval: Benchmark and bag-of-features descriptors. IEEE Transactions on Visualization and Computer Graphics 17, 11, 1624-1636.

Fotografías:
DCC
Comunicaciones FCFM
Andrés Caro
Jorge Baier

Dirección
Departamento de Ciencias de la Computación
Avda. Blanco Encalada 2120, 3er. piso
Santiago, Chile.
837-0459 Santiago
www.dcc.uchile.cl
Teléfono: (56-2) 2 9780652
Fax: (56-2) 2 6895531
revista@dcc.uchile.cl

Revista Bits de Ciencia del Departamento de Ciencias de la Computación de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile se encuentra bajo Licencia Creative Commons Atribución-NoComercial-CompartirIgual 3.0 Chile. Basada en una obra en www.dcc.uchile.cl

Revista Bits de Ciencia N°8
ISSN 0718-8005 (versión impresa)

www.dcc.uchile.cl/revista
ISSN 0717-8013 (versión en línea)

CONTENIDOS

EDITORIAL

Estimados lectores:

Este octavo número de la Revista Bits de Ciencia del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile, gira en torno a dos ejes principales. El primero, está conformado por dos aniversarios muy importantes y, el segundo, por el tema de Multimedia, que configura el tema central de esta edición.

Partamos por los aniversarios. En la sección Investigación Destacada, Juan Álvarez y Claudio Gutiérrez, ambos académicos del DCC, nos presentan el artículo "El primer computador universitario en Chile: el hogar desde donde salió y se repartió la luz". Este artículo celebra los cincuenta años de la llegada del "Lorenz" –el primer computador digital universitario– a la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. El segundo aniversario que celebramos es el de los 25 años de NIC Chile, institución que nació en el DCC y que a pesar de contar hoy con una más que vibrante vida propia, sigue intensamente ligado a nuestro Departamento. Para conmemorar este acontecimiento, Patricio Poblete, Director del NIC y académico del DCC, nos presenta el artículo "25 años de NIC Chile", en la sección Computación y Sociedad.

Queremos recalcar que esta sección, también incluye un interesantísimo artículo de Yadran Eterovic, académico y Director del DCC de la PUC, titulado "Treinta años del DCC de la PUC: una visión muy personal".

Concentrémonos ahora en el tema central de este número: Multimedia. O, en otras palabras, el estudio de datos digitales en formatos diversos (por ejemplo, imágenes, video, audio, etc.). No cabe duda que la importancia de la Multimedia se ha incrementado dramáticamente por la alta disponibilidad de dispositivos que generan este tipo de información. Consecuentemente también han aumentado los desafíos teóricos y prácticos relacionados a su estudio. Eso nos motivó a realizar un número dedicado al tema. Como podrán ver en los artículos presentados en la Revista, existe una gran cantidad de investigación de alto impacto sobre el tema en el país, lo que es muy gratificante.

En la sección central "Multimedia" contamos con los artículos: (1) "Búsqueda por contenidos multimedia" de Benjamin Bustos. (2) "El caso de Orand y Chequemático: investigación y transferencia tecnológica en la industria chilena" de Mauricio Palma, Juan Manuel Barrios y Paola Cornejo. (3) "Ciencia computacional aplicada a la biomedicina" de Nancy Hitschfeld, Jorge Jara y Steffen Härtel. (4) "Procesando más allá de lo visible" de Domingo Mery. (5) "Reconocimiento visual con técnicas de aprendizaje de máquinas" de Pablo Espinace, Billy Peralta y Álvaro Soto. (6) "HTML5: ¿nuevas perspectivas o déjà vu?" de Nelson Baloian, y (7) "La TV en la Universidad de Chile" de Nicolás Beltrán.

También Juan Manuel Barrios nos presenta el Survey "Búsquedas por contenido en imágenes y videos", y realizamos una entrevista al profesor Daniel Gatica, de la École Polytechnique Fédérale de Lausanne, uno de los expertos mundiales en Multimedia.

Finalmente, Nelsón Baloian, Sergio Ochoa y José A. Pino nos presentan su grupo de investigación "Collaborative Applications Research Laboratory (CARL)".

Agradecemos a todos los autores por su enorme esfuerzo en la preparación de los artículos y a nuestros lectores por su fidelidad a la Revista. Esperamos sinceramente que disfruten el presente número.

Pablo Barceló
Editor Revista Bits de Ciencia

El primer computador universitario en Chile: "El hogar desde donde salió y se repartió la luz"*

Computador ER-56 de la Escuela de Ingeniería de la Universidad de Chile.

"La difusión de los conocimientos supone uno o más hogares, de donde salga y se reparta la luz, que, extendiéndose progresivamente sobre los espacios intermedios, penetre al fin las capas extremas" (Andrés Bello, Inauguración de la Universidad de Chile, 1843).

Juan Álvarez
Master of Mathematics (Computer Science), University of Waterloo; Ingeniero de Ejecución en Procesamiento de la Información, Universidad de Chile.
jalvarez@dcc.uchile.cl

Claudio Gutiérrez
Ph.D. Computer Science, Wesleyan University; Magíster en Lógica Matemática, Pontificia Universidad Católica de Chile; Licenciatura en Matemáticas, Universidad de Chile.
cgutierr@dcc.uchile.cl

El primer computador digital para aplicaciones científicas y de ingeniería se instaló en Chile en 1962: un ER-56 Standard Elektrik Lorenz ("Lorenzo") de fabricación alemana, adquirido por la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile. Fue utilizado en docencia, investigación científica y tecnológica, y en proyectos de ingeniería de instituciones estatales y empresas privadas. Este hecho instaló en el imaginario de la sociedad chilena la automatización y los computadores. Un lustro de intensa utilización sentó las bases del futuro desarrollo de la disciplina de la computación en el país.

INTRODUCCIÓN

¿Cómo se produce el encuentro entre los computadores digitales –el fundamento tecnológico de una nueva era– y la población chilena? Hasta antes del año 1960, el tema de los computadores había estado recluido a estrechos círculos. Por una parte, académicos, principalmente de ingeniería eléctrica, que ya experimentaban con computadores analógicos. Por otra parte, ingenieros que necesitaban simulación y cálculo intensivo (numérico, de estructuras, de redes). Todos ellos se habían acercado al tema por medio de revistas, contactos y visitas a

* Versión ampliada del paper presentado en el II SHIALC (Simposio de Historia de la Informática en América Latina y el Caribe), realizado en Medellín, Colombia, en octubre de 2012.

laboratorios internacionales. Adicionalmente, algunos científicos teóricos en las áreas de matemáticas y física, interesados en los fundamentos de la ciencia, comenzaban a explorar las bases teóricas del nuevo objeto tecnológico.

Los años 1961 y 1962 marcaron el inicio material de la computación en Chile con la llegada de sendos computadores digitales al país [1]. En diciembre de 1961 se instaló en la Aduana en Valparaíso el primer computador digital en Chile, un IBM 1401, destinado al procesamiento administrativo. El año 1962 llegó el primer computador de orientación científica, un Standard Elektrik Lorenz ER-56, de fabricación alemana, a la Facultad de Ciencias Físicas y Matemáticas ("Escuela de Ingeniería") de la Universidad de Chile.

El impacto cultural y mediático que produjeron estos dos primeros computadores fue disímil. El IBM 1401 de la Aduana fue incorporado como una versión mejorada del procesamiento tradicional con máquinas Hollerith, y por ello su llegada pasó casi inadvertida para la población y la prensa [2]. No ocurrió lo mismo con el computador universitario. La puesta en funcionamiento del ER-56 puso en el imaginario ilustrado la percepción de que se estaba cruzando un umbral hacia una revolución en los procesos industriales. La prensa dedicó sus mejores páginas a exaltar el "nuevo cerebro electrónico", y su puesta en funcionamiento gatilló en el mundo académico una febril actividad docente y de difusión en torno a él.

¿Por qué esta diferencia de apreciación entre el IBM 1401 de la Aduana y el ER 56 de la Universidad? En este artículo mostramos que la adquisición del ER-56 fue parte de un proceso de incorporación de nuevas ideas y tecnologías a la sociedad chilena. Su arribo fue precedido de una reflexión académica sobre los usos de la computación en la ciencia y la ingeniería. En el Consejo de la Facultad de Ciencias Físicas y Matemáticas hubo conciencia que se estaba ante una tecnología que iba a cambiar el mundo. La sesión del Consejo de 1959, que reproduce con más fidelidad esa discusión, refleja las preocupaciones sobre las necesidades que podría cubrir y los alcances de esta nueva tecnología.

La actividad posterior en torno al ER-56 confirmó esta apreciación. El computador ER-56 no se utilizó en procesos administrativos, como el IBM 1401 de la Aduana, sino que se convirtió en un centro desde donde se difundió esta nueva "luz" a otras disciplinas y organizaciones.

Junto con el ER-56 se desarrolló un Centro de Computación y un conjunto de programas, cursos y seminarios para formación sistemática de personal. Comenzaron también las primeras "investigaciones" gatilladas por la necesidad de desarrollar compiladores para lenguajes particulares. Estos antecedentes permiten considerar que el ER-56 fue el "hogar de donde se irradió la luz" de la computación digital a otros sectores de la sociedad. No es casualidad que se igualara la importancia de esta actividad con el impacto que había producido la energía nuclear y se hablara de una "segunda revolución industrial".

Este artículo presenta la etapa previa, el arribo, la instalación, y las actividades generadas en torno al computador ER-56. La investigación se basa en diferentes fuentes: entrevistas orales individuales y grupales; actas de sesiones de Consejo; memorias de título; prensa escrita; revistas científicas y profesionales; apuntes, manuales y textos contemporáneos.

EL AMBIENTE PREVIO (1958-1961)

Computación Analógica

La computación en la Universidad de Chile comenzó en 1958 en la sección de Computadores y Servomecanismos del Instituto de Investigaciones y Ensayes Eléctricos (IIEE), predecesor del Departamento de Electricidad. La sección fue creada por el profesor Guillermo González y posteriormente se transformó en el Laboratorio de Computadores y Control Automático.

Inicialmente se trabajó e investigó en computación analógica para apoyar la solución de problemas de ingeniería. De hecho, en 1958 se armó el computador analógico Heathkit. Posteriormente, se dispuso del computador analógico Applied Dynamics AD 2-64PB con tableros para realizar y mantener los programas. Adicionalmente, se contó con el computador analógico EAI modelo TR-20 [3]. Un buen resumen del estado del arte en esta área se encuentra en la memoria de título de Walter Brokering y Herbert Ohrnad: "El computador análogo electrónico y su aplicación a la resolución de problemas de ingeniería" [4].

Anticipándose al advenimiento de la computación digital, a comienzos de los sesenta, el grupo de investigación diseñó un primer computador digital experimental (COMEX) y construyó una memoria de núcleos magnéticos [3]. En este marco de experiencias se decidió la adquisición del primer computador digital para aplicaciones científicas y académicas.

Centro de Computación

En la sesión del Consejo de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, del 2 de julio de

> El primer computador digital para aplicaciones científicas y de ingeniería se instaló en Chile en 1962: un ER-56 Standard Elektrik Lorenz ("Lorenzo") de fabricación alemana, adquirido por la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.

1959, el Decano Carlos Mori dio cuenta *"de la idea que existe en la Universidad de traer un equipo computador que podría ser usado tanto para la casa Central y para las Facultades, como para las Instituciones Estatales y Particulares"* [5]. En la tabla de la sesión, el físico Carlos Martinoya señaló:

La necesidad creciente de entregar a los profesionales que forma en las escuelas de su dependencia métodos eficaces de cálculo que puedan ayudarles en la aplicación de sus conocimientos a problemas prácticos.

No obstante queda por realizar una labor importante: ella se relaciona primero con las necesidades de computación en los trabajos de los institutos tecnológicos y de investigación, dependientes de la Facultad y, segundo, con las necesidades de computación y procesamiento de información en entidades técnicas y administrativas nacionales.

Concluyó proponiendo el siguiente proyecto de acuerdo que se aprobó por unanimidad:

Acordar la creación en la Facultad, mediante la cooperación de los Institutos de Física y Matemáticas, de Investigación y Ensayes Eléctricos y Ensayes de Materiales, de un Centro de Computación que atenderá las necesidades científicas y tecnológicas de la Universidad en estas materias.

La creación del Centro de Computación se concretó en agosto de 1961. El Decano Carlos Mori señaló que *"la creación de este Centro por la Universidad responde a la necesidad de introducir en el país una herramienta que ha revolucionado los conceptos vigentes en relación con la amplitud y alcance de las investigaciones y estudios de índole industrial, económico, administrativo, científico, etc."*. Después de un amplio debate en el Consejo de Facultad, se aprobó la creación del Centro por 25 votos contra 6 y 3 abstenciones [6]. La creación fue confirmada en la sesión del Consejo Superior Universitario del 13 de septiembre. Cabe señalar que dos años antes el Gerente General de Endesa, el ingeniero Raúl Sáez, propuso la creación de un "Centro Nacional de Cálculo" patrocinado por una universidad [7].

El Centro de Computación (CEC) se creó como una unidad independiente, bajo la dirección del ingeniero Santiago Friedmann, con los siguientes objetivos iniciales:

- Prestar servicio de procesamiento de datos a los Centros e Institutos de la Universidad de Chile, a las otras Universidades y a las demás instituciones que lo soliciten.
- Difundir el conocimiento de las técnicas derivadas de la operación de computadores digitales y formar el personal necesario, tanto para el Centro como para las instituciones similares del país.

Adquisición del ER-56

La selección y gestión de la compra de un computador digital fue encabezada por los académicos Joaquín Cordua, Director del IIEE, y Gastón Pesse, encargado de la Sección de Electrotecnia y Alta Tensión. La idea original fue comprar un computador IBM, pero la empresa sugirió esperar un nuevo modelo. Se decidió entonces adquirir el computador alemán ER-56 fabricado por Standard Elektrik Lorenz (SEL) [8].

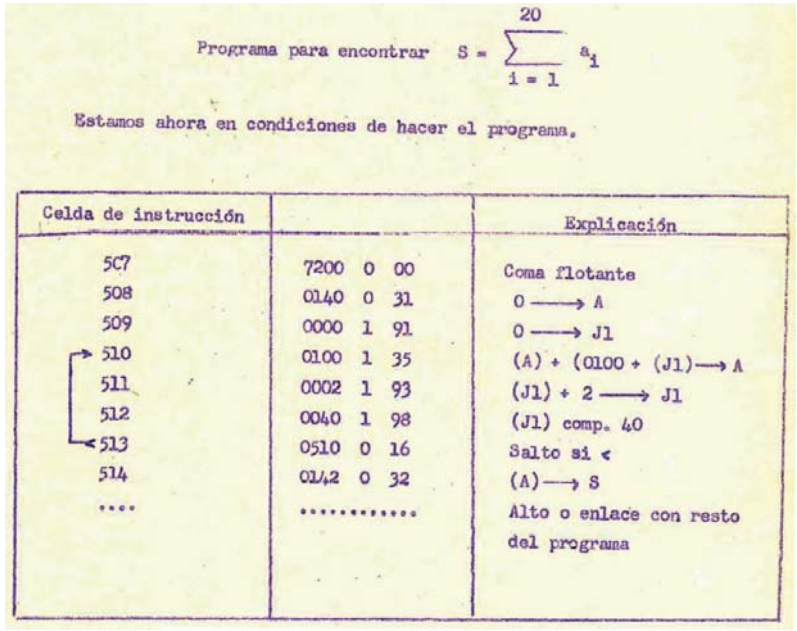
En 1961 se contrató a la firma SEL de Alemania, un crédito por la suma de DM 1.380.230 (aproximadamente 400.000 dólares) el que fue cancelado con una cuota al contado de DM 452.850 y el saldo de DM 997.380 en cinco letras iguales equivalentes cada una a DM 185.476 [9].

La sigla ER-56 corresponde a "Elektronischen Rechenanlage von 1956": Equipo Calculador Electrónico de 1956. De hecho la guía de despacho de la Aduana especificó "máquina calculadora eléctrica automática" [10]. El ER-56 se consideró pionero europeo en computadores completamente transistorizados, que reemplazaron a los computadores a tubos, mejorando la confiabilidad en un 50%. El primer computador ER-56 estuvo operativo en 1959. Su diseñador fue Karl Steinbuch (1956), quien además en 1957 acuñó el término "Informática" como un acrónimo de Información y Tecnología Automática [11].

Preparación

Una vez decidida la compra, comenzó una intensa preparación y capacitación. El profesor de ingeniería eléctrica Guillermo González impartió cursos de capacitación y preparó los primeros apuntes [12]. A modo de ejemplo, la página 22 de los apuntes contiene un programa que calcula una sumatoria matemática (Figura 1).

Figura 1



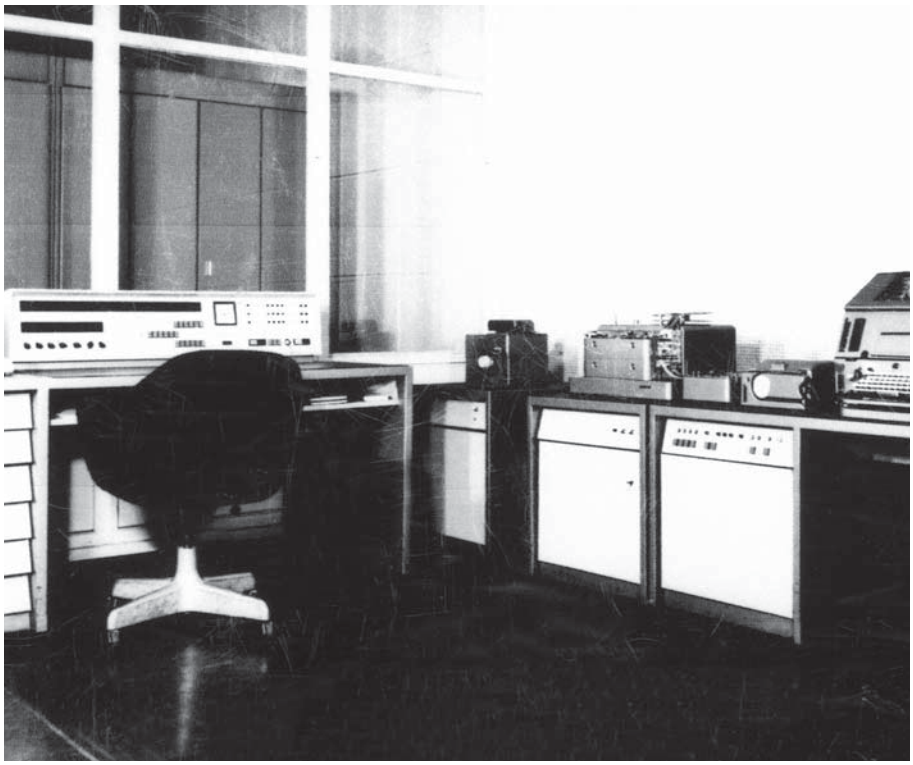
Programa para encontrar $S = \sum_{i=1}^{20} a_i$

Estamos ahora en condiciones de hacer el programa.

Celda de instrucción		Explicación
5C7	7200 0 00	Coma flotante
508	0140 0 31	0 ⟶ A
509	0000 1 91	0 ⟶ J1
→ 510	0100 1 35	(A) + (0100 + (J1)) ⟶ A
511	0002 1 93	(J1) + 2 ⟶ J1
512	0040 1 98	(J1) comp. 40
← 513	0510 0 16	Salto si <
514	0142 0 32	(A) ⟶ S
....	.....	Alto o enlace con resto del programa

Un programa del computador ER-56.

Figura 2



El computador Standard Elektrik Lorenz ER-56 (foto gentileza de W. Riesenkönig).

Por otra parte, en diciembre de 1961, se realizó por primera vez un seminario de doce sesiones dirigido a las empresas con el siguiente temario:

- El Computador Electrónico dentro de la familia de las calculadoras.
- Los componentes básicos de un computador.
- Diagrama de bloque del computador.
- Cómo se formula un problema para el computador y la entrega de datos.
- Aplicaciones sobre problemas y ejecución de sus diagramas de flujo.
- Programación para el computador electrónico.

Asistieron representantes de Ferrocarriles del Estado, Banco Crédito e Inversiones, IBM, Compañía Manufacturera de Papeles y Cartones, Molinos y Fideos Luchetti S.A., S. A. Yarur, Compañía Chilena de Electricidad, Endesa, Editora Zig-Zag, Chiprodal y Compañía de Cervecerías Unidas [13].

INSTALACIÓN Y PRESENTACIÓN DEL ER-56 (1962-1963)

Capacitación en Alemania

Para recibir la capacitación final del fabricante y preparar la llegada del ER-56 a Chile, los primeros meses de 1962 viajaron a Stuttgart José Dekovic, Guillermo González y Jean Marie de Saint Pierre [14].

José Dekovic, ingeniero investigador del CEC, viajó entre el 6 de enero y el 6 de junio de 1962 para realizar estudios relacionados con la organización, operación, explotación y puesta en marcha del Centro de Computación, primeramente en Stuttgart y luego completó su entrenamiento en el Centro Internacional de Cálculo en la Universidad de Roma.

Guillermo González, ingeniero jefe de la Sección de Computadores del IIEE, viajó el 6 de febrero de 1962 durante tres meses, y recibió capacitación en los aspectos de ingeniería de computadores digitales relacionados con el computador digital ER-56.

Jean Marie de Saint Pierre, investigador del IIEE, viajó por tres meses a partir del 11 de febrero de 1962 y recibió entrenamiento en los problemas de mantención del ER-56.

Instalación

El ER-56, que se instaló entre los meses de junio y agosto de 1962 en el subterráneo del edificio de Química, fue administrado por el Centro de Computación y su mantención técnica fue encargada al IIEE. Prontamente fue coloquialmente apodado "el Lorenz" o "Lorenzo", en alusión a su fabricante Standard Elektrik Lorenz, y al género masculino que se atribuye a los computadores en Chile.

El ER-56 se adquirió con una memoria de tres mil palabras de siete dígitos decimales, un tambor magnético de doce mil palabras, y una lectora de cinta de papel perforado. Los resultados se imprimían con un teleimpresor.

Desde el punto de vista del software, los programas se escribían en los lenguajes de máquina y Alcor (un subconjunto de Algol para el ER-46). El ER-56 funcionaba sin sistema operativo, por lo que su operación se realizaba manualmente a través del panel ("pupitre") de control (ver esquema en el Anexo 2).

Para apoyar la instalación y la operación inicial, la SEL envió al ingeniero alemán Wolfgang Riesenkönig, quien capacitó a los chilenos en Alemania y puso en funcionamiento el Centro de Computación de la Universidad de Bonn. Riesenkönig llegó el 1 de septiembre de 1962, permaneció varios meses en Chile, dictó diversas charlas y cursos en la Facultad y en otras universidades del país y publicó un apunte de programación en Algol [15].

Actividades de difusión

Una vez instalado el computador, se realizaron diversas actividades de difusión con el propósito de fomentar su uso. El 27 de agosto de 1962, Santiago Friedmann, Director del CEC, ofreció la conferencia "La era del computador se inicia en Chile – consideraciones sobre sus efectos en el

ejercicio de la ingeniería" en el Instituto de Ingenieros [16]. Su exposición abordó los siguientes temas:

- ¿Por qué fue necesario crear el computador?
- La mecanización de la inteligencia.
- Enseñando a leer y escribir (las máquinas Hollerith).
- El computador digital.
- El Centro de Computación de la Universidad de Chile.

Su charla finalizó señalando que *"el Centro de Computación está abierto al uso de sus facilidades por el público: otras universidades, empresas públicas y privadas, profesionales, etc."* en tres niveles específicos: entrenamiento, asesoría y procesamiento.

Por otra parte, el CEC organizó un seminario sobre computadores especialmente orientado hacia los profesores e investigadores de todas las universidades del país. Los participantes quedaron capacitados para formular sus problemas a un computador digital electrónico. El seminario se realizó entre el 27 de agosto y el 14 de septiembre de 1962 y fue dictado por el ingeniero Jean Marie de Saint-Pierre. Posteriormente, entre el 24 de septiembre y mediados de octubre, se realizó la segunda parte a cargo del ingeniero Manuel Quinteros, donde se resolvieron y programaron problemas específicos [17].

En febrero de 1963, en la sede del Club de la Unión, el ingeniero del CEC José Dekovic ofreció una charla para el Rotary Club que el diario La Nación tituló "Computadores electrónicos dan lugar a la segunda revolución industrial" [18]. La crónica informó que:

Caracterizando los últimos 25 años en el aspecto tecnológico, el conferenciante señaló dos hechos fundamentales: 1.- el dominio y la utilización de la energía atómica, y 2.- el gran desarrollo de la automatización, posibilitada por los computadores. El segundo de estos hechos ha dado lugar a lo que algunos llaman la Segunda Revolución Industrial.

El "cerebro electrónico" en la prensa

La presentación del ER-56 a la opinión pública se realizó en una masiva conferencia de prensa el 11 de enero de 1963. El diario La Nación tituló la noticia "Cerebro electrónico reemplazará al hombre en tareas improductivas" [19] e informó que:

La gigantesca máquina, que ocupa un gabinete de 12 por 2.30 metros y pesa 8.950 kilos, informó a los reporteros en un segundo y medio cuánto habían vendido 30 empleados de una casa comercial y cuál había sido su comisión. Santiago Friedmann, Director del Centro, dio a conocer la complicada maquinaria y señaló sus múltiples ventajas – con un funcionamiento de 7 a 8 horas diarias –explicó- su mayor ventaja es su monstruosa rapidez, puede hacer una cantidad ilimitada de operaciones y rinde 300 escudos por hora de trabajo, con lo cual amortizará en siete años su costo que es de 400.000 dólares.

El texto de la foto (ver Figura 3) titulada "Un cerebro de 8.950 kilos" señalaba que:

El computador electrónico que ocupa la Universidad de Chile en sus investigaciones científicas, hizo demostraciones ayer sobre los diferentes cálculos que puede hacer en tiempo récord para aliviar al hombre de tareas inútiles e improductivas. En el grabado, los periodistas observan mientras la máquina responde una de las preguntas que habría requerido horas de trabajo a una persona.

Por su parte, la portada del diario El Mercurio informó con el título "Cerebro electrónico adquirido por la Universidad piensa y memoriza" y el siguiente resumen [20]:

El computador digital es capaz de jugar ajedrez; resuelve en apenas un segundo más de mil quinientas multiplicaciones. "Piensa" mediante fórmulas lógico-matemáticas que le son proporcionadas junto a los datos que procesará. "Recuerda" gracias a un archivo maestro que, para ulteriores operaciones, sólo requiere de datos complementarios y variables. Funciona a temperatura moderada, pues "se le calienta la cabeza", e indica su dolencia.

El texto de la foto (ver Figura 4) de la primera página del diario indicaba que:

El 18 de septiembre de 1810 fue un día martes, según la contestación que en fracciones de segundo dio el reluciente aparato que se observa en la fotografía. Esta es una máquina electrónica capaz de jugar ajedrez, formular diagnósticos médicos y resolver todo problema susceptible de ser

Figura 3

Foto del ER-56 en diario La Nación, enero de 1963.

Figura 4

Foto del ER-56 en diario El Mercurio, enero de 1963.

planteado en términos matemáticos. Fue adquirido por el Centro de Computación de la Universidad de Chile.

Por su parte, la revista Ercilla tituló la noticia "El cerebro Mecánico de la U" con el siguiente y ambicioso resumen [21]:

Sabe ejecutar y componer música, hacer acertadamente diagnósticos médicos, resolver todo tipo de problemas lógicos, jugar ajedrez, y matemáticos (sic), a la velocidad de 2.000 operaciones por segundo y posee la memoria más eficaz y amplia que es dado imaginar.

Días después una columna editorial del diario La Nación titulada "Competencia de cerebros", era más moderada expresando en el último párrafo [22]:

Magnífica la técnica, y los robots, y cuanto ingenio salga para aligerar las tareas estólidas. Pero, por favor, no aplaudamos más los cerebros electrónicos que los hechos de materia orgánica. Ellos cuando se destruyan nunca se convertirán en flores o en gusanos como los nuestros.

Tres meses después, con motivo de las elecciones municipales de abril de 1963, la revista de sátira política Topaze publicó caricaturas del Presidente Alessandri y del futuro Presidente Frei (ver Figura 5) y entrevistó al "cerebro electrónico" que falló al hacer los escrutinios [23]:

O sea, caballeros, que reconozco hidalgamente que no me la puedo – terminó confidenciándonos el cerebro electrónico-. Y por favor, nada de fotos, porque siempre salgo movido –nos dijo al despedirnos mientras nos palmoteaba cariñosamente la espalda.

La situación se produjo por una falla en la lectora de cinta debido a una pequeña desviación en su célula fotoeléctrica [10].

UTILIZACIÓN DEL LORENZO (1962-1966)

Proyectos

Inicialmente el ER-56 fue utilizado por Endesa (procesamiento de información hidrológica y estudios de energía), la Dirección de Riego (análisis de los recursos de agua), la Dirección de Vialidad (cálculo de cubicaciones de los movimientos de tierra), y Enap (interpretación de mediciones de terreno para determinar yacimientos) [16].

En la Facultad de Ciencias Físicas y Matemáticas el computador fue utilizado en proyectos de investigación de centros e institutos, en algunos casos en colaboración con empresas públicas o privadas. Sus resultados se reflejaron en las memorias de titulación de al menos cuatro ingenieros eléctricos, siete industriales y once civiles según se observa en Anexo 1.

El Centro de Computación prestó servicios a centros, laboratorios e institutos: Laboratorio de Estructuras, Departamento de Exploraciones de Geofísica, IDIEM, Instituto de Física y Matemáticas (departamentos de Biofísica y Cristalografía), IIEE, Centro de Radiación Cósmica, Centro de Geodesia y Cátedra de Química Teórica. Por otra parte, se prestó servicios a reparticiones externas: Instituto Forestal (FAO), CORFO, UTFSM, SONAP, ENDESA, Ministerio de Obras Públicas (Direcciones de Riego y Vialidad) y Municipalidad de Santiago (Intendencia) [17].

Tarifas

Después de un intenso uso inicial, a partir del 1 de enero de 1963 se fijaron las siguientes tarifas por los servicios [24]:

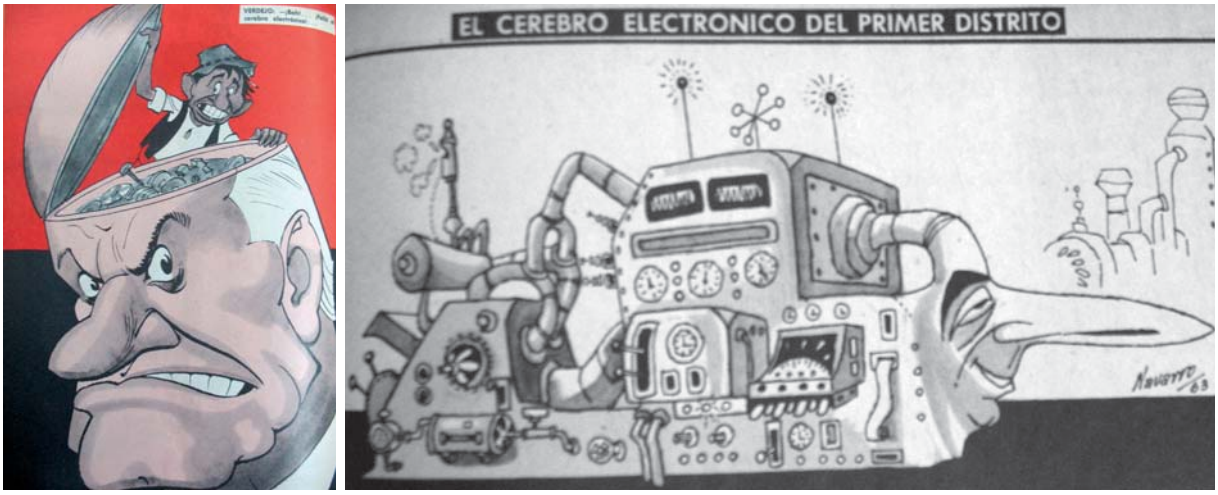
- Tarifa base por servicio del computador digital electrónico, por minuto (con un cargo mínimo de 2 minutos): E°5,00.
- Tarifa base por elaboración de cinta perforada por palabra (dato hasta siete caracteres), en exceso sobre 200 palabras por minuto de computación: E°0,002.
- Tarifa base por impresión de resultados, la hora: E°3,00.

Los trabajos que realice el Centro de Computación, para atender necesidades de la docencia universitaria, serán gratuitas.

Las tarifas antes señaladas, serán rebajadas en un 60% en los trabajos que el Centro realice en beneficio de la investigación universitaria, cuando sean financiados con fondos universitarios, y en un 50% cuando se trate de pruebas de programas. En el caso de prueba de subrutina, el trabajo que realice el Centro será gratuito por el tiempo que se haya convenido y el exceso sobre ese tiempo se cancelará con un 50% de descuento sobre la tarifa base.

Los costos de programación, asesoría, docencia, etc., serán convenidos entre el usuario y el Director del Centro de Computación.

Figura 5



Caricaturas de Alessandri y Frei Montalva en Revista Topaze, abril de 1963.

Perfeccionamiento

La utilización intensiva del ER-56 gatilló la necesidad de perfeccionamiento de los ingenieros del Centro de Computación [25]. Santiago Friedmann viajó en marzo de 1963 por un año a realizar estudios sobre Computación Digital en la Universidad de California, Berkeley, a través de una beca de la Comisión Fulbright. En enero de 1964, Jean Marie de Saint Pierre del IIEE, haciendo uso de una beca del Gobierno francés, viajó a la Universidad de París a realizar estudios sobre computadores. Posteriormente, en septiembre de 1964, Hugo Segovia viajó al Centro de Computación de la Universidad de California para desarrollar un plan de estudios e investigación. En el mismo año, Manuel Quinteros y Pedro Taborga viajaron becados al MIT.

Docencia

La labor docente relacionada con el computador consistió en seminarios de divulgación técnica generales y para empresas específicas, capacitación para investigadores, y cursos de cálculo numérico, de programación y de estadística [17]. Por su parte, el Departamento de Estudios Generales de la Escuela de Economía creó la cátedra de "Elementos de Computación Electrónica" que se impartió en el séptimo semestre de su carrera[26].

La labor docente más importante comenzó en marzo de 1966 cuando la Facultad introdujo un curso semestral de "Computación y Cálculo Numérico" en el segundo año de las carreras de Ingeniería. El curso se orientaba a la comunicación hombre-máquina a través de diversos lenguajes para el cálculo numérico y no numérico en el computador ER-56: lenguaje de máquina, Algol y Lisp. Para apoyar el curso se escribió un texto de siete capítulos cuyo editor fue Efraín Friedmann. Los cuatro primeros se refieren a los computadores y a los fundamentos de programación y fueron escritos por Adriana Kardonsky y Víctor Sánchez. El capítulo cinco presenta el lenguaje Algol y su autor es Víctor Sánchez. Los capítulos seis y siete, de "Análisis Numérico" y de "Programación no Numérica" (usando LISP), fueron escritos por Manuel Quinteros y Hugo Segovia respectivamente. En el Anexo 3 se presenta la tabla de contenidos del libro, que permite apreciar el estado del arte de la enseñanza de la computación después de tres años de experiencia del uso del ER-56 [27].

Carrera de "computación"

En el Consejo Universitario del 30 de diciembre de 1964 el Decano de la facultad de Ciencias Físicas y Matemáticas Enrique D'Etigny propone un Plan de Estudios en Computación [28]:

El Decano Señor D'Etigny expresa que el Departamento de Matemáticas de la Facultad de Ciencias Físicas y Matemáticas, a raíz de la instalación en Chile de las Máquinas Computadoras especializadas, que en este momento suman 10 y hay 3 o 4 que deben llegar el próximo año, estimó que era necesario formar matemáticos que pudieran estudiar los problemas que deben ser planteados a estas máquinas.

La Facultad, sin embargo, se anticipó a esta idea y tiene, desde hace 3 años, un grupo de trabajo en matemáticas aplicadas, en realidad, en computación y, desde hace 2 años, funciona un computador de cierta importancia. Ahora desea ofrecer, como una carrera profesional adicional, a los alumnos de Ingeniería, la posibilidad de dedicarse a este campo. En especial, se ha tenido en cuenta que para abastecer normalmente de problemas a una máquina como ésta se requiere un equipo de 6 personas, lo que supone, en este momento, 70 personas, que, sin embargo, no han sido formadas.

Después de un intenso debate, y con fuerte oposición de varios decanos, la propuesta fue postergada. Posteriormente, en la sesión del 13 de enero de 1965, mediante la gestión del Rector Eugenio González, se aprobó finalmente un Plan de Estudios distinto para una carrera de Ingeniería Matemática de cinco años de duración.

Desarrollo de lenguajes de programación

La programación del computador se realizó inicialmente en los lenguajes de máquina (programación "directa") y Alcor apoyándose en la asesoría y publicaciones del Centro

[29,30]. Para superar la incomodidad de la programación en lenguaje de máquina, el ingeniero Fernando Vildósola desarrolló Adrela, un pequeño lenguaje ensamblador. Posteriormente desarrolló Ladrea, que traducía programas desde el lenguaje de máquina al lenguaje Adrela [31]. Más tarde implementó Slap (un lenguaje simple para la programación automática) [32]. Años después, el ingeniero electricista Herbert Plett implementó el lenguaje TNP [33].

Evolución del Centro de Computación

Con la reforma estructural de la Facultad en 1964, a partir del 1 de enero de 1965 el Centro de Computación junto con el Centro de Matemáticas conformaron el Departamento de Matemáticas dirigido por el ingeniero Efraín Friedmann [17]. Para entonces el Centro de Computación contaba con el siguiente personal:

- Ocho investigadores de jornada completa.
- Cuatro investigadores auxiliares con media jornada (estudiantes de Ingeniería).
- Seis programadores de jornada completa.
- Siete programadores ayudantes con ¼ de jornada.
- Un técnico de mantención del equipo.
- Dos operadores de consola (del panel de control del computador).
- Un operador de teleimpresores.
- Una bibliotecaria.
- Dos secretarias.
- Un portero y un chofer.

Los ocho investigadores de jornada completa, con una excepción, eran ingenieros egresados de la Universidad de Chile:

- Efraín Friedmann, Ingeniero Civil y Eléctrico con estudios de ingeniería nuclear.
- Santiago Friedmann, Ingeniero Civil.
- Manuel Quinteros, Ingeniero Civil.
- Francisco Radó, Ingeniero Civil.
- Víctor Sánchez, Ingeniero Mecánico Industrial (Universidad Técnica del Estado).
- Héctor Hugo Segovia, Ingeniero Civil Industrial.
- Pedro Taborga, Ingeniero Civil Industrial.
- Germán Torres, Ingeniero Civil.

Las nuevas y crecientes necesidades docentes y de servicio aconsejaron una ampliación del ER-56. Lamentablemente, la fábrica SEL cerró y "Lorenzo" fue uno de los pocos equipos que se contruyeron. El CEC buscó una alternativa, que se concretó en diciembre de 1966 con la adquisición e instalación de un IBM/360-40, un computador de "tercera generación" (tecnología de circuitos integrados) y de "propósito general" (para aplicaciones científicas y administrativas). Consecuentemente, el intenso uso del ER-56, de alrededor de ocho horas diarias utilizando cintas de papel perforado, fue rápidamente desplazado por el nuevo computador que utilizaba tarjetas perforadas, discos y cintas magnéticas. A mediados de los setenta terminó sus días como una reliquia utilizada con fines didácticos y para demostraciones (como por ejemplo la interpretación musical del "vuelo del moscardón") hasta ser lentamente desmantelado [34].

Figura 6

Personal del Centro de Computación en un cóctel de celebración, septiembre de 1964.

CONCLUSIONES

Hemos trazado la trayectoria del primer computador universitario que llegó al país, y el ambiente que lo permitió y que a su vez se generó producto de su llegada.

El computador ER-56 estableció las relaciones entre la computación y el mundo científico e industrial en Chile, resolviendo problemas de diversas áreas de la ingeniería. Su entorno, el Centro de Computación de la Universidad de Chile, se constituyó en un "Centro Nacional de Cálculo" que brindó servicios de entrenamiento, asesoría y procesamiento al medio nacional.

El Centro hospedó el núcleo inicial, el embrión, de la investigación en computación en Chile. Fue en torno al ER-56 que académicos e ingenieros diseñaron lenguajes e implementaron compiladores. La continuidad y progreso de este quehacer gatilló posteriormente la creación de las primeras carreras profesionales y un departamento científico y académico que contribuyó importantemente a la configuración y el desarrollo posterior de la disciplina.

De los antecedentes presentados, es interesante colegir que esta empresa fue una exitosa labor colectiva, con participantes de diversas formaciones y orígenes. Muchos de ellos contribuyeron posteriomente a desarrollar el área en otras instituciones nacionales.

La actividad computacional desarrollada en torno al ER-56 no pasó inadvertida. Los principales periódicos recogieron su inauguración, y siguieron las noticias sobre sus actividades. Las organizaciones profesionales y académicas comenzaron a hablar de la computación e incorporar el tema en sus agendas. Una nueva época había nacido para el país.

AGRADECIMIENTOS

Agradecemos la valiosa colaboración de los profesores Guillermo González, Wolfgang Riesenkönig, José Dekovic, Jean Marie de Saint Pierre, Víctor Sánchez, Hugo Segovia, Fernando Vildósola, Enrique D'Etigny, Joaquín Córdua, Santiago Friedmann y Julio Zúñiga. Agradecimientos también para Gabriel Ahumada de la Biblioteca del Congreso; Carlos Adriazola del Archivo del diario La Nación; Patricia Liberona del Archivo Central Andrés Bello; Ana María Carter, Daniel Encalada y Luis Cortés de la Biblioteca de la Facultad de Economía y Negocios; Rosa Leal, Directora de la Biblioteca Central, y Luis Celis, Jefe de Recursos Humanos, de la Facultad de Ciencias Físicas y Matemáticas. Y a Nelson Baloian, Director, y Pablo Barceló, editor de la Revista Bits, y, a través de ellos, a toda la comunidad del Departamento de Ciencias de la Computación que colabora con el proyecto de Historia de la Computación en Chile. BITS

REFERENCIAS

[1] Álvarez, Juan; Gutiérrez, Claudio. "History of Computing in Chile, 1961-1982: Early years, Consolidation and Expansion". IEEE Annals of the History of Computing. Vol 34 N°3. July-September 2012.

[2] Álvarez, Juan. "El primer computador digital en Chile: Aduana de Valparaíso, diciembre de 1961". Revista Bits. DCC, FCFM, U. de Chile. 2do semestre 2011. Versión digital en: http://dcc.comopapel.com/revista_bits_de_ciencia/6/#/page/20-21.

[3] González, Guillermo. Presentación en I Taller de Historia de la Computación en Chile. Santiago. 2009.

[4] Brokering, Walter; Ohrband, Herbert. "El computador análogo electrónico y su aplicación a la resolución de problemas de ingeniería". Memoria de Ingeniero Civil Electricista. FCFM, U. de Chile. 1960.

[5] Facultad de Ciencias Físicas y Matemáticas. Acta de la sesión N° 103 del Consejo de la Facultad. 2 de julio de 1959.

[6] Facultad de Ciencias Físicas y Matemáticas. Acta de la sesión N° 122 del Consejo de la Facultad. 17 de agosto de 1961.

[7] Sáez, Raúl. "Universidad y Empresa". Boletín de la Universidad de Chile N° 1. Abril 1959.

[8] Comunicaciones FCFM. "Hombres que moldearon la historia de Beauchef". Revista FCFM N° 46. Primavera 2009.

[9] Universidad de Chile. Acta del Consejo Universitario. 27 de mayo de 1964.

[10] Prenafeta, Sergio. "La nueva era de los computadores viejos". Revista Informática. Vol.1 N° 3. Mayo 1979.

[11] Widrow, B.; Hartenstein, R.; Hecht-Nielsen,R. "1917 Karl Steinbuch 2005". IEEE Computational Intelligence Society. August 2005.

[12] González, Guillermo. "Curso de programación del computador digital ER-56". IIEE. 1961.

[13] Diario La Nación. "Computador electrónico se incorpora a administración de las empresas chilenas". 9 de diciembre de 1961.

[14] Universidad de Chile. Actas del Consejo Universitario del 14, 21 y 28 de marzo de 1962.

[15] Riesenkönig, Wolfgang. Presentación en I Taller de Historia de la Computación en Chile. Santiago. 2009.

[16] Friedmann, Santiago. "La era del computador se inicia en Chile. Consideraciones sobre sus efectos en el ejercicio de la Ingeniería". Anales del Instituto de Ingenieros. Año LXXV N° 4. Agosto-octubre 1962.

[17] U. de Chile. "La Facultad de Ciencias Físicas y Matemáticas". Noviembre 1964.

[18] Diario La Nación. "Computadores electrónicos dan lugar a la segunda revolución industrial". 8 de febrero de 1963.

[19] Diario La Nación. "Cerebro electrónico reemplazará al hombre en tareas improductivas". 12 de enero de 1963.

[20] Diario El Mercurio. "Cerebro electrónico adquirido por la Universidad piensa y memoriza". 12 de enero de 1963.

[21] Revista Ercilla. "El cerebro Mecánico de la U". N° 1443.16 de enero de 1963.

[22] Diario La Nación. "Competencia de cerebros". 13 de enero de 1963.

[23] Revista Topaze. "El cerebro electrónico". N° 1590. 12 de abril de 1963.

[24] Universidad de Chile. Actas del Consejo Universitario. 18 de enero de 1963.

[25] Universidad de Chile. Actas del Consejo Universitario. 26 de diciembre de 1962, 29 de enero de 1964, 16 de septiembre de 1964.

[26] Universidad de Chile. Actas del Consejo Universitario. 3 de abril de 1963.

[27] CEC. "Curso de Computación y cálculo numérico". CEC, FCFM, U. de Chile. 1966.

[28] Universidad de Chile. Actas del Consejo Universitario. 30 de diciembre de 1964.

[29] CEC. "Descripción del computador digital electrónico Standard Elektrik Lorenz ER-56.1". CEC. 1963

[30] Radó, Francisco. "Fundamentos de Programación (orientada al computador ER-56.1)". CEC. 1963.

[31] Vildósola, Fernando. Presentación en I Taller de Historia de la Computación en Chile. Santiago. 2009.

[32] Vildósola, Fernando. "SLAP: un lenguaje simple para la programación automática de un computador digital". CEC. 1966.

[33] Plett, Herbert. "TNP: un lenguaje versátil para la programación del computador digital ER-56". Departamento de Electricidad, FCFM, U. de Chile. 1969.

[34] Otero, Sofía. "Beauchef, la cuna del desarrollo computacional chileno". Revista FCFM. N° 43, pp. 37-39. Primavera 2008.

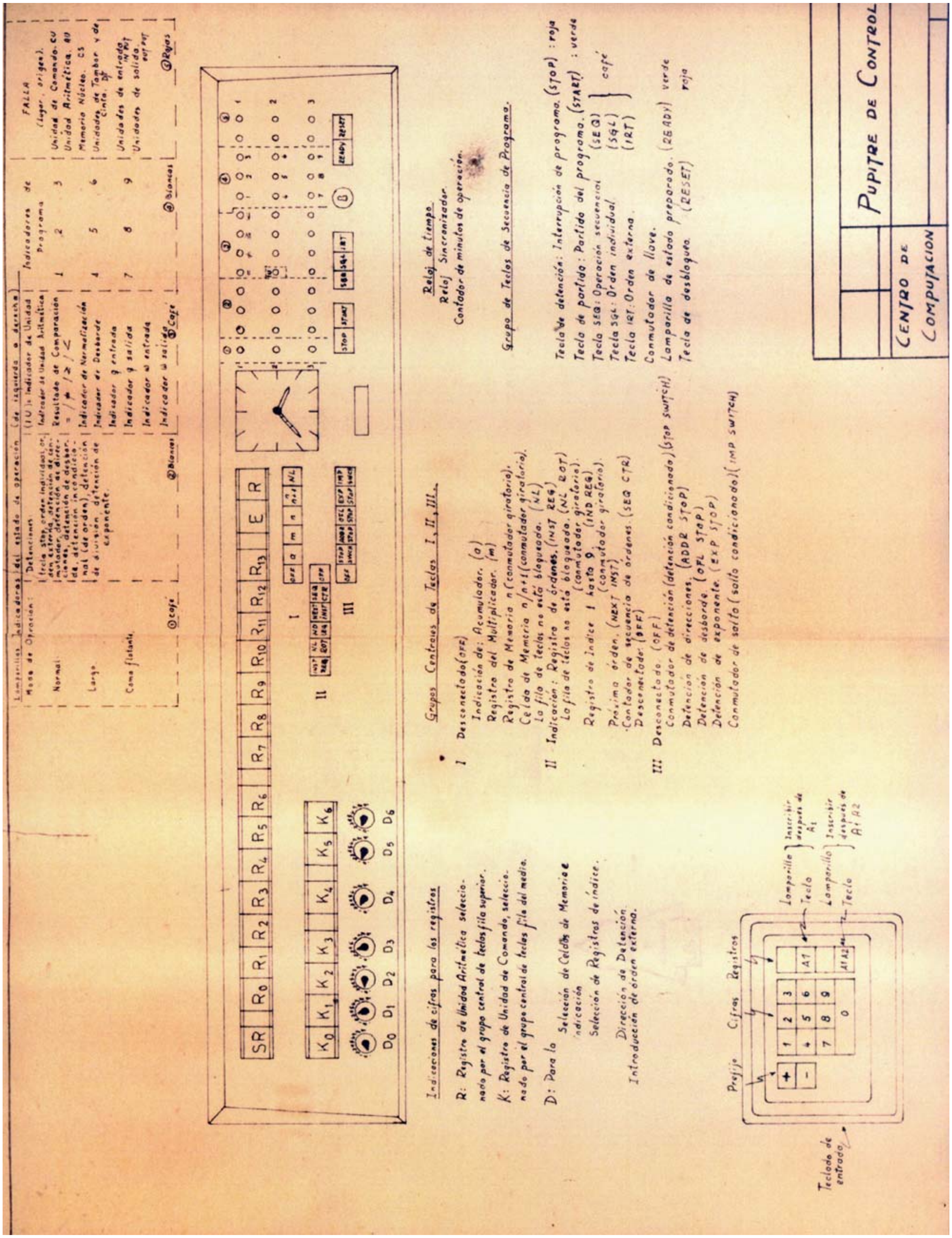
Anexo 1

Trabajos de Título de la Escuela de Ingeniería de la Universidad de Chile que utilizaron el computador ER-56.

Año	Especialidad	Ingeniero	Prof.Guía	Empresa	Título de la Memoria
62	Eléctrica	E.Gaete	E.D'Etigny	Endesa	Despacho de carga económico del sistema interconectado mediante el computador digital.
63	Eléctrica	E.Bianchi	E.Friedmann	Endesa	La previsión de las demandas de energía eléctrica a corto plazo.
63	Industrias	P.Taborga	J.Cauas	CAP	Aplicación de la teoría de esperas.
63	Civil	M.Quinteros	A.Arias		Estudio teórico del comportamiento de edificios con pisos recogidos sometidos a temblor.
63	Industrias	H.Segovia		CMPC	Aplicación de los métodos de investigación operacional al control de inventarios en la industria.
64	Industrias	J.Sutter		Endesa	Análisis comparativo de sistemas tradicionales y de trayectoria crítica para prog. y control de obras.
64	Industrias	A.Krell	J.Cauas	Enami	Aplicación de programación lineal al problema de distribución de productos minerales de la Enami.
64	Civil	F.Radó	A.Arias		Estudio teórico del comportamiento de edificios con pisos recogidos y que se deforman por flexión.
65	Civil	M.J.Bolelli		Agua Potable	Estudio del cálculo de redes de agua potable usando la computación digital.
65	Civil	C.Leighton	J.Monge		Análisis dinámico de edificios con torsión en planta.
65	Eléctrica	M.Mertens			Solución de problemas de sísmica artificial mediante el computador ER-56.
65	Civil	R.Peralta	A.Quintana		Algunos métodos de aerotriangulación.
65	Eléctrica	L.Ponce	A.Cisternas		Propagación de pulsos en esferas fluídas heterogéneas.
65	Civil	G.Torres		MOP	Diseño de superestructuras en puentes con viga metálica y losa colaborante.
65	Civil	J.Wachholtz			Análisis elástico de vigas altas.
66	Industrias	H.Donoso	J.Cauas	Endesa	Programación Matemática y economías de escala en planificación eléctrica.
66	Civil	C.Gaete			Estudio elástico de muros.
66	Civil	P.Ortigosa			Solución computacional de un emparrillado de vigas.
66	Civil	J.Poblete			Programación matemática aplicada al estudio del aprovechamiento de recursos hidraúlicos.
66	Industrias	J.Radmilovic			El método de los planos cortantes en la programación no lineal.
66	Civil	M.Sarrazín			Proyecto de losa de pruebas para un laboratorio de ensayo de estructuras.
66	Industrias	A.Scopelli			Modelo de simulación del sistema eléctrico interconectado chileno.

Anexo 2

Esquema de panel de control del computador ER-56 (tomado de [29]).



Anexo 3

Tabla de contenidos del texto "Curso de Computación y Cálculo Numérico" [27].

I N D I C E

25 años de NIC Chile

Equipo de trabajo de NIC Chile.

Patricio Poblete
Ingeniero Matemático, Universidad de Chile, M.Mat. y Ph.D. in Computer Science University of Waterloo, Canadá. Director de la Escuela de Ingeniería y Ciencias, FCFM, Universidad de Chile; Director de NIC Chile. Profesor Titular, co-fundador del DCC Universidad de Chile; Director del DCC 1988-1992 y 1996-1999.
ppoblete@dcc.uchile.cl

LOS INICIOS

El Departamento de Ciencias de la Computación de la Universidad de Chile fue uno de los pioneros en la interconexión de Chile a las redes globales, dando a mediados de los años ochenta pasos tales como el envío del primer email entre dos universidades chilenas y luego el establecimiento de conexiones UUCP por vía telefónica que permitieron el envío y recepción de email internacional, y luego de las "news" de Usenet.

En estos primeros experimentos aún no se utilizaba el sistema de nombres de dominios que conocemos hoy, sino que era necesario especificar toda la ruta que debía seguir un mensaje hasta llegar a su destino. A poco andar, el mundo UUCP comenzó a utilizar los nombres de dominio que recientemente se habían introducido en Internet, lo cual puso de inmediato en evidencia que para integrarnos a ese sistema debíamos contar con un nombre de dominio para el país. Respecto de esto, no había mucha elección, debía ser .CL de acuerdo con el estándar ISO3166-1, y para nuestra Universidad elegimos "uchile.cl". Sin embargo, no había nadie aún que estuviera a cargo de llevar la lista de los dominios inscritos bajo .CL, y eso condujo de manera natural a que esa tarea la asumiéramos nosotros, lo que hemos hecho a partir de 1987.

Por cierto, no bastaba con mantener una lista, sino que dichos dominios debían "resolverse", es decir, debían traducirse a direcciones IP de manera instantánea, cosa de lo cual se encargó por varios años el servidor "uunet" en Estados Unidos, a cargo de Rick Adams, quien realizó una labor importantísima ayudando a muchos países como el nuestro a conectarse a la Red. Sólo una vez que Chile se conectó definitivamente a Internet, en 1992, pudimos traer esa función de "servidor primario" a nuestro país.

EL DESARROLLO DEL .CL

Durante los primeros diez años, el servicio de inscripción de nombres de dominio bajo .CL se brindó de manera totalmente gratuita y fue llevado a cabo como una más de las tareas del grupo de sistemas del DCC. Al término de esa primera década, había alrededor de mil nombres de dominios inscritos.

Sin embargo, a esas alturas, después de cinco años de conexión de Chile a Internet, el tema de la Red y de los nombres de dominio estaba empezando a hacerse conocido en el país y la carga que significaba encargarse de este registro de nombres estaba empezando a ser significativa. Peor aún, los primeros conflictos por nombres de dominio estaban empezando a aparecer, amenazando con volverse un problema agudo.

Todo esto condujo a que en septiembre de 1997 se estableciera una formalización del servicio, que incluyó una tarifa por la inscripción de dominios, un sistema de resolución de controversias basado en arbitraje, una reglamentación escrita para regular el funcionamiento del sistema y un nombre para identificar a todo lo anterior: NIC Chile.

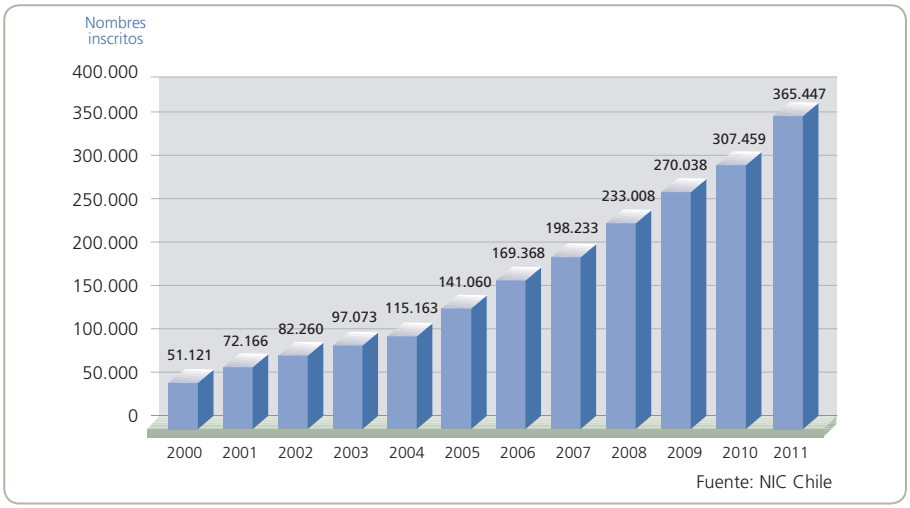
El crecimiento del registro de nombres de dominio para .CL ha sido sostenido a partir de ese momento, como lo muestra la Figura 1, y en la actualidad el total de nombres inscritos ya bordea los 400 mil.

Este crecimiento ha hecho que el número de dominios per cápita en .CL sea uno de los más altos de América Latina, superado sólo por Argentina (uno de los pocos países en el mundo en que no se cobra por registrar dominios) y por Colombia, cuyo ".co" se está comercializando (y de manera muy exitosa) como un competidor para el ".com" (Figura 2).

Nuestro país se caracteriza también por que el .CL tiene una muy alta penetración en el país. De todos los dominios inscritos desde Chile, más del 90% está bajo .CL (Figura 3).

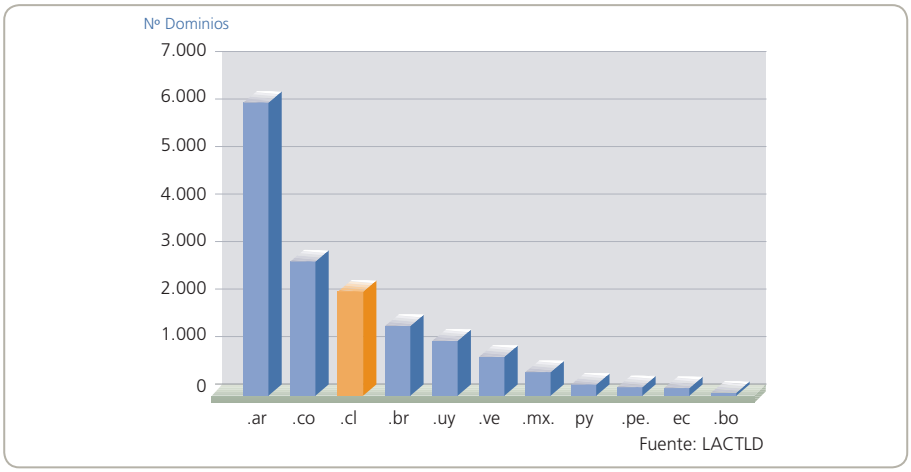
Esta situación, que podría parecer natural, dado que .CL es el nombre que identifica a Chile, es en realidad excepcional en el contexto mundial. En muchos países, la

Figura 1



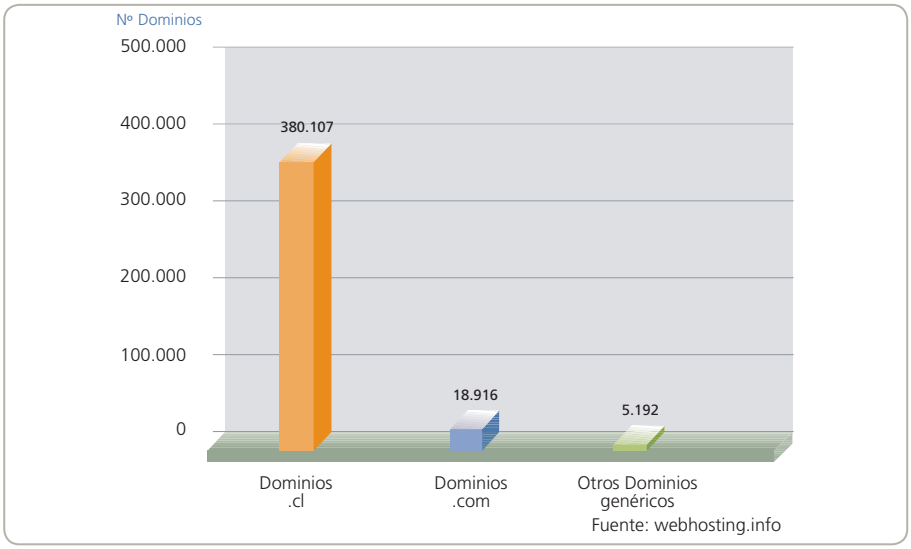
Crecimiento de nombres de dominio inscritos por año (tamaño del registro).

Figura 2



Dominios territoriales por cada 100.000 habitantes.

Figura 3



Dominios .CL versus dominios genéricos en Chile.

proporción se invierte, y la mayoría de los dominios se inscriben bajo ".com" y otros dominios genéricos.

Las razones de esta masiva adopción del .CL en Chile son variadas y no del todo conocidas, pero sin duda han influido en ello el temprano comienzo de la operación del .CL, la ausencia de trabas burocráticas para el registro de nombres y el hecho de que no se subdividiera el dominio, registrando todos los nombres directamente bajo el .CL, sin hacer uso de subclasificaciones como podrían haber sido ".com.cl", ".edu.cl", etc. En la actualidad, muchos países que en su momento adoptaron subdivisiones de ese estilo hoy están migrando a "abrir el segundo nivel" y permitir el registro directo bajo el código del país.

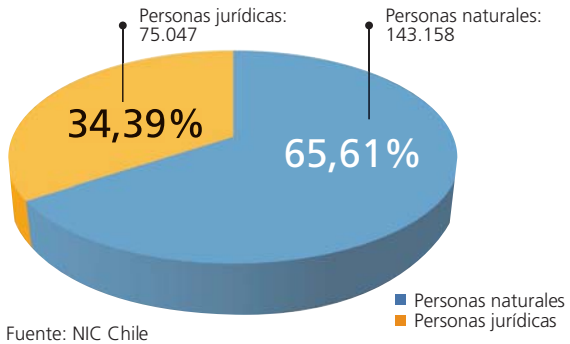
Estos dominios son mayoritariamente inscritos por personas naturales, lo cual se equilibra un poco si en lugar de contar a los titulares se cuenta a los dominios inscritos, porque las empresas tienden a registrar más dominios en promedio que las personas (Figura 4).

RESOLUCIÓN DE CONFLICTOS POR NOMBRES DE DOMINIO

Tan pronto se vio que Internet tenía un gran potencial comercial y que la identificación en la Red era clave para ganar la visibilidad necesaria, surgieron problemas tales como la ciberocupación, en que marcas registradas o nombres asociados a empresas o personas eran registrados por terceros, con la intención de obtener un provecho económico o de obstaculizar el acceso a la Red. Tales conflictos siempre pueden llevarse a los tribunales de justicia, pero la lentitud de ese proceso no se avenía con la velocidad con la cual Internet se estaba desarrollando. Para responder a este desafío, en 1997 Chile fue pionero en el mundo en establecer un sistema de resolución de conflictos por nombres de dominio basado en arbitraje. Dos años después, en la reunión de ICANN realizada en Santiago, un sistema similar llamado UDRP (Uniform Dispute Resolution Policy) fue adoptado para los dominios genéricos tales como .com, .net, etc.

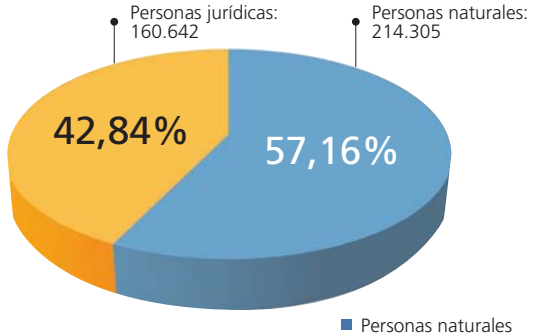
Figura 4

Cantidad y proporción de titulares de nombres de dominio que son personas naturales o jurídicas.



Fuente: NIC Chile

Cantidad y proporción de nombres de dominio asignados a personas naturales y jurídicas.



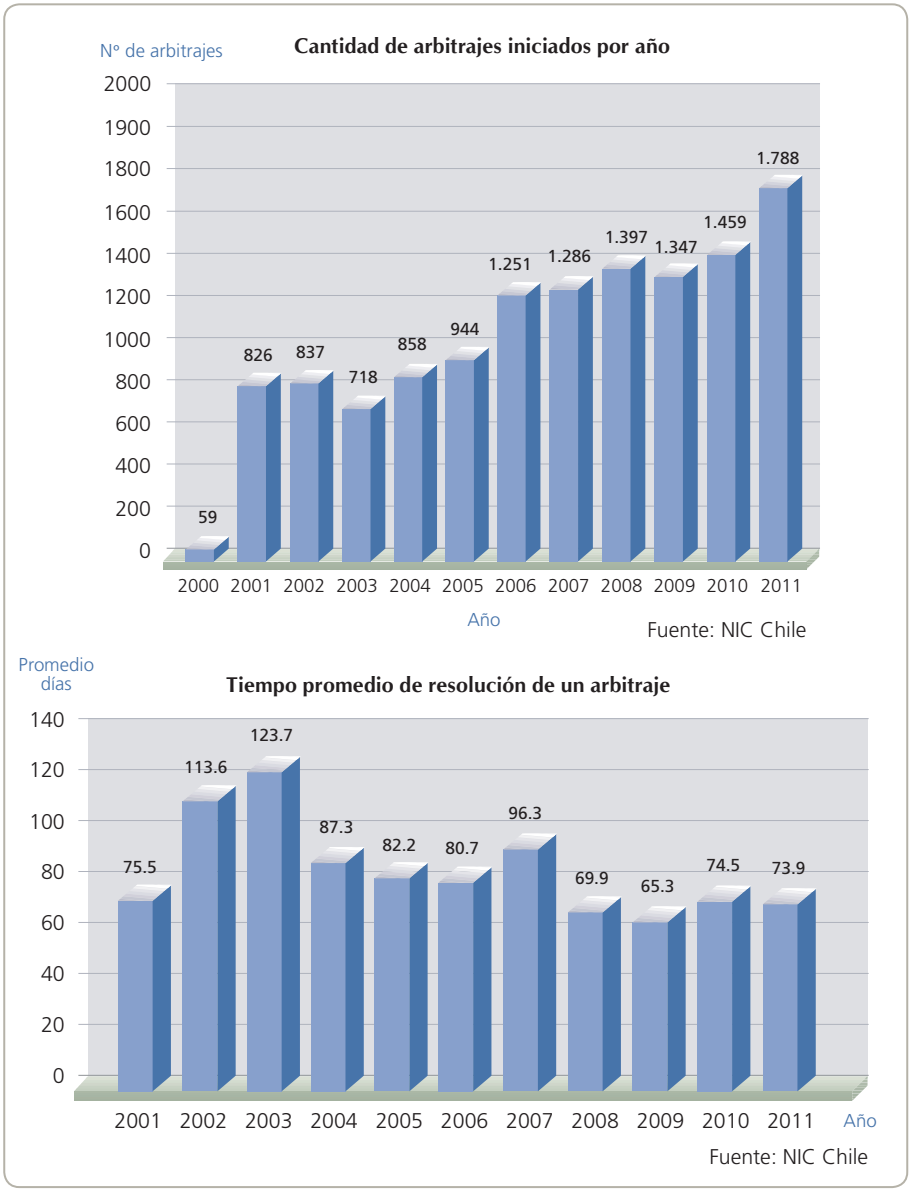
Fuente: NIC Chile

El sistema vigente en Chile se ha ido perfeccionando y robusteciendo, y en la actualidad es parte integral de la "cultura" desarrollada en torno a este tema. Las estadísticas de la Figura 5 muestran que el sistema ha logrado en general cumplir con la promesa de ser una manera eficiente de resolver estos conflictos.

En la actualidad está próximo a entrar en funcionamiento un sistema de arbitraje completamente en línea, vía web, lo cual va a mejorar aún más estas cifras (se espera disminuir la duración promedio a menos de sesenta días) y a hacer el acceso más expedito, evitando la concurrencia personal ante el árbitro.

A lo largo de sus 25 años de vida, NIC Chile ha estado permanentemente a la vanguardia en la implementación de innovaciones en sus propios sistemas, pero al mismo tiempo aportando en modernizar la infraestructura del país.

Figura 5



LA TECNOLOGÍA DETRÁS DEL .CL

A pesar de que los nombres de dominio han adquirido un innegable valor de marketing y comercial, el objetivo fundamental del sistema es permitir que estos identificadores se traduzcan a las respectivas direcciones IP, de manera confiable y eficiente. Este proceso, llamado "resolución", es crítico, porque su eventual falla podría dejar inaccesible a grandes partes de Internet. De hecho, el sistema de nombres de dominio (DNS) es uno de los pocos *"single points of failure"* en una arquitectura que en general evita tener este tipo de vulnerabilidades. Los servidores de nombres, que se encargan de implementar esta resolución, son objetivos frecuentes de ataques de denegación de servicio, y NIC Chile ha establecido una infraestructura de resolución de nombres preparada para enfrentar este tipo de amenazas, así como posibles catástrofes naturales.

En la actualidad, NIC Chile posee una red de servidores de nombres distribuida en todo el mundo, de modo que la falla de uno o incluso muchos de ellos no ponga en riesgo la continuidad del servicio. Algunos de estos servidores son operados directamente por NIC Chile, y otros son servicios contratados a empresas especializadas. El mapa de la Figura 6 muestra la ubicación de estos servidores.

Figura 6

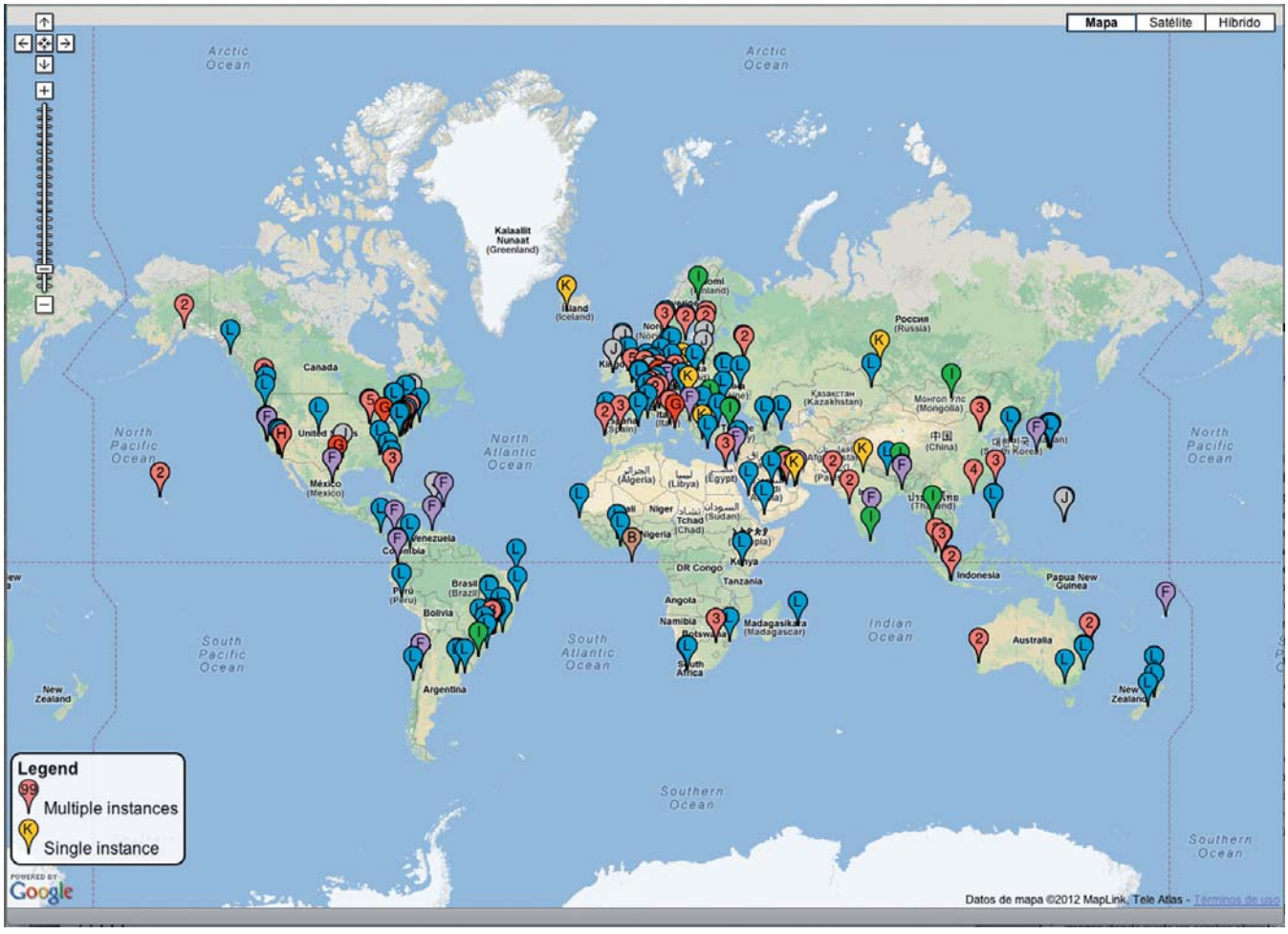
Esta arquitectura redundante también se extiende a otros componentes de la infraestructura de NIC Chile, tales como enlaces con múltiples proveedores, sitios de contingencia, etc.

Junto con robustecer su propia infraestructura, NIC Chile ha aportado a hacer que los "servidores raíz" también sean más resistentes a fallas y ataques. Estos servidores son los que permiten el inicio de los procesos de resolución de nombres; en principio eran sólo trece (identificados por letras de la A a la M), pero gracias a una tecnología llamada *"anycast"* en la actualidad poseen múltiples "espejos" ubicados en todo el mundo, lo que disminuye fuertemente la vulnerabilidad de estos servidores frente a ataques u otras amenazas. NIC Chile colabora en este esfuerzo de protección de la infraestructura global albergando espejos de los servidores F y L (Figura 7).

UNA TRAYECTORIA DE INNOVACIÓN

A lo largo de sus 25 años de vida, NIC Chile ha estado permanentemente a la vanguardia en la implementación de innovaciones en sus propios sistemas, pero al mismo tiempo aportando en modernizar la infraestructura del país. Algunos de sus aportes más significativos han sido:

Figura 7



Fuente: http://root-servers.org/map/

Factura Electrónica

NIC Chile, con sus especialistas, colaboró con el Servicio de Impuestos Internos en el diseño de la factura electrónica y fue parte del piloto inicial de dicho sistema, siendo en ese momento la única entidad del sector público que participó. La adopción de esta innovación ha sido clave para permitir la operación eficiente del registro de nombres de dominio, y miles de empresas del país se han beneficiado con esta modalidad de emisión de facturas. NIC Chile también hizo un aporte importante a la implementación de estos sistemas, al desarrollar una biblioteca de software y ponerla a disposición de la comunidad como *"open source"*.

Nombres de dominio internacionalizados (IDN)

NIC Chile fue el primer país de habla hispana en ofrecer la posibilidad de registrar nombres de dominio con letras acentuadas, eñes y "u" con diéresis.

DNSSEC

El sistema DNSSEC permite que las "zonas" dentro del DNS puedan ser firmadas criptográficamente para asegurar la autenticidad de las respuestas que entregan los servidores. De esta manera se evitan ataques del tipo *"man in the middle"*. La zona .CL está firmada con DNSSEC y en la actualidad NIC Chile está promoviendo la adopción de DNSSEC entre sus clientes, especialmente aquellos más expuestos a suplantaciones de tipo *"phishing"* u otros ilícitos.

IPv6

Las direcciones IP tradicionales, llamadas IPv4, se están agotando rápidamente, frente a lo cual se ha implementado un nuevo sistema llamado IPv6. La adopción oportuna de IPv6 es clave para permitir el desarrollo ininterrumpido de Internet y NIC Chile ha liderado este proceso en Chile, a través de iniciativas que han traído al país a los principales expertos en el tema, así como mediante proyectos público-privados orientados a promover esta tecnología.

ENTRANDO AL SEGUNDO CUARTO DE SIGLO

Cuando a mediados de los ochenta, un grupo de investigadores del Departamento de Ciencias de la Computación de la Universidad de Chile dábamos los primeros pasos para conectar nuestro país a la naciente Red global, entre el escepticismo de quienes nos escuchaban decir que el futuro estaba en ella, no nos imaginábamos, primero, que la realidad iba a superar largamente nuestras proyecciones más optimistas, ni menos que el grupo que impulsaba esta iniciativa iba a terminar convertido en lo que hoy conocemos como NIC Chile, una pieza clave para el funcionamiento y el desarrollo de la Sociedad de la Información en nuestro país.

Desde el inicio de la operación del dominio .CL, nuestro objetivo fue proveer a la comunidad un servicio de la mayor calidad, fácil de usar y a precios accesibles, para ser un socio confiable y eficiente de quienes elijan identificarse en la Red con el "punto CL". Para ello, hemos tenido más de una vez que "hacer camino al andar", cuando nos encontramos en terrenos inexplorados, pero también aprovechamos la experiencia de nuestro colegas, a través de las diversas instancias de colaboración internacional en que participamos y que, muchas veces, hemos ayudado a crear.

A medida que el uso de Internet se ha ido masificando y que más y más personas, empresas e instituciones dependen crucialmente de la Red, hemos ido robusteciendo nuestros mecanismos operativos para asegurar el funcionamiento del dominio .CL, incluso en las circunstancias más adversas. Hoy contamos con una de las redes de servidores de nombres más diversificadas a nivel mundial, y seguimos constantemente en busca de formas de optimizar este servicio.

La comunidad a la cual servimos ha ido, al mismo tiempo, integrándose más a los mecanismos de generación de las políticas que aplicamos. Es así como desde hace años opera el Consejo Nacional de Nombres de Dominio y Números IP, organismo que ha tenido un rol importante en la discusión y perfeccionamiento de todos los últimos cambios en las políticas de .CL.

Este Consejo acaba de emitir recientemente una opinión favorable a un cambio de políticas que va a permitir que .CL siga avanzando para mantener su línea de liderazgo en los años venideros. La industria mundial de nombres de dominio ha ido evolucionando y desarrollando "mejores prácticas" de las cuales nuestro registro se debe beneficiar, y en conjunto con el Consejo hemos generado un consenso en dirección a ello.

Al implementarse estas nuevas políticas, el proceso de registro de nuevos dominios en .CL se acercará significativamente a lo que ocurre en la mayoría de los TLDs (Top Level Domains). Al mismo tiempo, mediante la implementación del protocolo EPP, se han establecido las bases para la posible introducción de registradores, y el cambio anterior, sumado a la eliminación del contacto local para las inscripciones desde el extranjero, permitirán abrir .CL a los usuarios internacionales.

Estos cambios, por otra parte, se han diseñado velando por mantener y robustecer los mecanismos de resolución de conflictos que han permitido que en .CL problemas como la ciberocupación se resuelvan de manera expedita. Es así como la esencia del sistema tradicional de disputa de nombres de dominio se mantiene, y a esto se suma la introducción de un sistema de arbitraje totalmente en línea, que hará el sistema mucho más accesible y eficiente para todos los participantes.

Al entrar al segundo cuarto de siglo, estamos proyectando a .CL hacia el futuro, aprovechando lo mejor de nuestro pasado y presente, manteniendo la línea de innovación que nos ha caracterizado. Sólo así podemos responder a la confianza que la comunidad de Internet mundial, y especialmente la de nuestro país, ha depositado en nosotros.BITS

Treinta años del DCC de la PUC: **una visión muy personal**

Cuerpo académico del DCC de la PUC.

Yadran Eterovic
Ingeniero Civil Electricista de la Universidad Católica de Chile (PUC) y M.Sc y Ph.D. en Computer Science U. of California, Los Angeles (UCLA). Sus áreas de trabajo son el desarrollo de software y la programación concurrente. Recién cumplió treinta años como profesor del Departamento de Ciencia de la Computación (DCC) de la PUC. Ha sido Director del programa INGES y coordinador docente del DCC. Actualmente es Director del DCC.
yadran@ing.puc.cl

En mi opinión, el DCC de la Universidad Católica se encuentra hoy en una muy favorable posición para seguir cumpliendo a cabalidad su función y lograr plenamente su objetivo fundacional de desarrollar (promover y realizar) docencia de pre y postgrado, investigación y extensión en la disciplina de la Computación, tanto en los aspectos científicos como tecnológicos. Este objetivo fue establecido por la Universidad, y por su Escuela de Ingeniería, en los documentos que crearon el actual Departamento en mayo de 1983, es decir, hace casi treinta años.

Si bien nos hemos desarrollado al interior de una Escuela de Ingeniería muy prestigiosa, que nos ha permitido tener excelentes estudiantes de pre y postgrado, el camino recorrido en estos treinta años no ha sido fácil. Primero, durante los ochenta, tuvimos que establecer nuestra validez como un departamento académico más (el noveno) al interior de la Escuela de Ingeniería, cuando incluso fuimos calificados de "chacrientos" al querer incluir el curso de Teoría de Autómatas y Lenguajes Formales en el currículo de nuestra especialidad. Luego, en los noventa, tuvimos que definir nuestra propia identidad, desafío que demostró ser mucho más difícil de superar que lo que cualquiera podría haberse imaginado, especialmente considerando que la mayoría de los profesores compartíamos, de alguna manera, el que habíamos obtenido nuestros doctorados en la disciplina de Ciencia de la Computación en buenas universidades de Norteamérica y Europa. Sin embargo, por lo menos para mí, lo más difícil ha sido tratar de superar la temprana partida de dos excelentes profesores y amigos, que nos dejaron cuando aún tenían mucho que contribuir en lo humano y en lo académico, y en momentos en que el Departamento pasaba por situaciones muy complicadas: Javier Pinto (agosto de 2001) y Álvaro Campos (julio de 2003); en agradecimiento a lo mucho que nos entregaron, nuestras dos salas destinadas a la docencia llevan sus nombres.

LOS INICIOS

En 1970 la Universidad crea el Centro de Ciencias de la Computación (CECICO), dependiente de la Facultad de Ingeniería, para la realización de docencia, investigación y servicios de computación, en particular, impartiendo cursos de computación para Ingeniería y otras carreras, e impartiendo la carrera de Programador de Computadores. A fines de 1981, CECICO es separado en dos: la unidad Servicios de Computación para Administración UC (SECICO) y el Departamento de Ciencia de la Computación (DCC), dependiente, en ese momento, de la Facultad de Matemáticas.

Así, el DCC, antes de establecerse en la Escuela de Ingeniería, había formado parte de la Facultad de Matemáticas por un par de años, bajo el decanato de Rolando Chuaqui. A ese Departamento llegamos varios profesores, algunos, que ya tenían formación de postgrado, con jornadas completas; otros, que aún no terminábamos nuestras carreras de pregrado, con jornadas parciales. Entre estos últimos estábamos Javier Pinto, David Fuller y yo, que estudiábamos el último año de ingeniería eléctrica. Nos hicimos cargo de la Licenciatura en Ciencias de la Computación, que la Facultad impartió por primera vez en esos años. Ocupábamos una o dos oficinas en el tercer piso del edificio Raúl Devés, piso que en ese tiempo albergaba a la Facultad de Matemáticas. Paralelamente, la Escuela de Ingeniería empezó a formar un "grupo" de profesores de computación, contratando, entre otros, a Álvaro Campos, Miguel Nussbaum e Ignacio Casas (quien había formado parte de CECICO y de la Facultad de Matemáticas), y creó la carrera de Ingeniería Civil de Industrias con mención en Ingeniería de Computación. Desconozco cuál habrá sido la relación formal, a nivel de autoridades, entre el Departamento y el grupo, pero la relación entre los profesores, la mayoría muy jóvenes, era cordial y de colaboración.

Luego, cuando el Departamento es (re) creado en 1983 al interior de la Escuela de Ingeniería, quedó integrado tanto por los profesores que proveníamos de la Facultad de Matemáticas, como por los que ya estaban en el grupo que tenía la Escuela, entiendo que sin excepciones. Éramos once profesores, y seguimos siendo once por casi veinte años. A partir de 1983 y durante los siguientes diez o doce años, la mayoría de los profesores del DCC salió a hacer sus estudios de postgrado, principalmente en universidades de Estados Unidos, Canadá, Inglaterra y Suiza. Casi heroicamente, los pocos profesores que permanecían en Chile sacaban adelante la investigación, la extensión y la docencia, tanto de la carrera de ingeniería como de la licenciatura.

Con respecto a la extensión, quiero destacar dos experiencias iniciadas en esos años bajo la dirección de Ignacio Casas y que siguen siendo exitosas más de veinte años después. La primera, llamada ecompuc y realizada en conjunto con el Centro de Extensión de la universidad, está orientada a la capacitación en aplicaciones y lenguajes de uso masivo. La segunda, realizada inicialmente en conjunto con el Departamento de Ingeniería Industrial y de Sistemas, fue la introducción del (probablemente) primer Programa de Postítulo en Gestión Informática del país, programa vespertino conocido como INGES; al poco andar, la responsabilidad del INGES quedó completamente en manos del DCC.

EL DIFÍCIL CRECIMIENTO

Es justamente cuando los profesores del DCC volvemos con nuestros doctorados, a mediados de los noventa, que se empieza a manifestar, cada vez con más fuerza, nuestra "crisis" de identidad. El DCC se había creado en una época en que la Ciencia de la Computación era toda la computación, y como tal se había consolidado en las más prestigiosas universidades del mundo. Muchos de los profesores que hemos sido parte del DCC en estos treinta años, obtuvimos nuestros doctorados en universidades extranjeras en la disciplina de Ciencia de la Computación. Sin embargo, a partir de los noventa, varios fuimos abordando otras áreas de conocimiento, distintas a las tradicionales de la Ciencia de la Computación. Esto se debió tanto al desarrollo internacional de la computación —en sus aspectos científicos y, principalmente, tecnológicos— como también a las necesidades y oportunidades que presentaba su desarrollo tecnológico en Chile.

Por ejemplo, la creciente complejidad de los sistemas de software requeridos por las organizaciones más diversas hizo del diseño y construcción de software una verdadera disciplina de ingeniería de software, la que ha contado desde fines de los noventa con sus propios principios y modelos, métodos y técnicas, y herramientas para la producción de software de calidad. También, la importancia que fueron adquiriendo ciertos sistemas de software, con acceso a grandes volúmenes de información estratégica, para la ejecución eficiente de los procesos y servicios de las empresas, fue tal que dio origen a una disciplina dedicada al diseño, análisis y evaluación de lo que llamamos *sistemas de información*.

> Si bien nos hemos desarrollado al interior de una Escuela de Ingeniería muy prestigiosa, que nos ha permitido tener excelentes estudiantes de pre y postgrado, el camino recorrido en estos treinta años no ha sido fácil.

Profesor Álvaro Campos.

Sin embargo, como no todos los profesores del DCC compartían estas ideas, durante los próximos diez años nos cuestionamos todo: si éramos un Departamento científico o uno de ingeniería; si debíamos hacer más o menos investigación; si la investigación que debíamos hacer tenía que ser más básica o más aplicada; si teníamos que hacer asesorías y de qué tipo; hasta dónde llegaba nuestra "libertad" académica para decidir en qué actividades participar, y hasta el contenido del programa curricular de nuestra especialidad. Lamentablemente, como yo lo veo, no lo supimos hacer con la altura de miras que tan importantes decisiones ameritaban. Cada uno de nosotros, unos más que otros, nos creíamos dueños de la verdad: teníamos el mismo grado de Doctor, más o menos la misma edad, la mayoría habíamos "fundado" el Departamento a comienzos de los ochenta y no nos poníamos de acuerdo en qué es la computación. Nuestras legítimas diferencias en lo académico pronto se transformaron en problemas muy profundos entre personas, que nos costó muchísimo superar.

Históricamente, no ha sido fácil definir *computación*; por lo menos, la definición no ha sido la misma a lo largo del tiempo. Una característica ampliamente reconocida de la computación es que evoluciona muy rápidamente y los desarrollos científicos son transferidos en poco tiempo a aplicaciones concretas. En mi opinión, teníamos que encontrar una definición que incluyera a la totalidad de la computación, porque ese es el sentido del término empleado en los documentos relativos a la creación del DCC. No éramos únicamente Ciencia de la Computación o únicamente ingeniería de computación, como podría suponerse considerando el nombre del Departamento o su pertenencia a la Escuela de Ingeniería.

Por lo tanto, para empezar a resolver nuestros problemas a partir de una definición de computación, recurrimos a los informes más recientes en esos momentos de la *Joint Task Force on Computing Curricula* de la IEEE-CS y ACM (CC2001). A nivel mundial, la IEEE-CS y la ACM son reconocidas como las organizaciones más importantes en computación. La CC2001 había sido creada con el propósito de desarrollar pautas curriculares de pregrado *"acordes a los últimos desarrollos de las tecnologías computacionales en la década pasada (los noventa) y que puedan mantenerse vigentes durante la próxima década"*.

De acuerdo con la CC2001, en particular con el informe *Computing Curricula 2001: A Vision of the Overview Volume* de enero de 2001, en ese momento el término "computación" (computing) incluía varias disciplinas distintas, a saber (al menos para el propósito de pautas curriculares):

- Ciencia de la Computación *(Computer Science)*.
- Ingeniería de Computación *(Computer Engineering)*.
- Ingeniería de Software *(Software Engineering)*.
- Sistemas de Información *(Information Systems)*.

Esta constatación por parte de la IEEE-CS y de la ACM corroboraba la realidad cambiante de la computación y de paso nos ayudaba a entender mejor, hasta cierto punto, lo que nos estaba pasando. Consideremos que una *joint task force* similar en 1991 había establecido como equivalentes los términos *Computer Science y Computer Engineering*, y hasta mediados de los noventa, *information systems* era considerada una disciplina distinta a *computing*, no formando parte de ésta aunque compartiendo cierta formación común. Es decir, durante los noventa la computación se fue convirtiendo en un área del saber cada vez más amplia, que para 2001 alcanzaba mucho más allá de los límites tradicionales de la Ciencia de la Computación o de la Ingeniería de Computación.

Así, en lugar de tratar de ver la computación como una única disciplina con límites claramente demarcados, a algunos nos pareció más constructivo, para los propósitos de las definiciones fundamentales que nuestro Departamento buscaba, reconocer su realidad diversa y definirla de alguna manera como la unión de las cuatro disciplinas mencionadas. También acordamos que estas disciplinas no eran totalmente independientes entre ellas. Es decir, a pesar de su rápida evolución, la computación tiene un núcleo estable de conocimientos sobre el cual se fundamentan todos los desarrollos y aplicaciones de nuevas tecnologías. Nosotros identificamos este núcleo con la Ciencia de la Computación, e identificamos los desarrollos y aplicaciones de nuevas tecnologías basadas en *software* con la

Ingeniería de Software y los Sistemas de Información, y los desarrollos y aplicaciones de nuevas tecnologías basadas en *hardware* con la Ingeniería de Computación.

Por lo tanto, y considerando que desde su creación el DCC había decidido no abordar la disciplina de Ingeniería de Computación (entendida como en el párrafo anterior), la computación para el DCC la empezamos a entender a comienzos del año 2000, y la seguimos entendiendo hoy, como aquella área del saber científico y tecnológico constituida por tres disciplinas:

- Ciencia de la Computación: estudio sistemático de procesos algorítmicos que describen y transforman información: su teoría, análisis, diseño, eficiencia, implementación y aplicación; la pregunta fundamental en esta disciplina es, ¿qué puede ser eficientemente automatizado?, en que el sentido de automatizado es mediante el uso de *software*.
- Ingeniería de Software: producción de artefactos de *software* grandes y complejos de alta calidad dentro de plazos preestablecidos con los recursos asignados; esto involucra tanto aspectos que tienen que ver con el diseño y construcción del *software* como también con métodos de gestión y control del proceso de desarrollo.
- Sistemas de Información: estudio de los procesos de creación y operación de información, y del contexto y consecuencias sociales de la manipulación de información; incluye el uso de información por individuos y grupos, especialmente en el contexto organizacional; y el impacto, implicancias y gestión de los sistemas de información.

Esta nueva manera de entender la especialidad se sumó a las dos ya existentes: Ingeniería Civil de Industrias con mención en Ingeniería de Computación, creada a comienzos de los ochenta junto con el nacimiento del Departamento, e Ingeniería Civil de Computación, creada a mediados de los noventa gracias a una iniciativa liderada por Leopoldo Bertossi.

En todo caso, en estos años complejos para el DCC, también ocurrieron hechos positivos destacables: nació Solex, la primera empresa derivada de la Escuela de Ingeniería, producto de la actividad de investigación y desarrollo de Miguel Nussbaum; se doctoró el primer alumno del Programa de Doctorado de la Escuela de Ingeniería, Marcos Sepúlveda, hoy profesor nuestro, bajo la supervisión del propio Miguel; y, como acabo de mencionar, creamos la carrera de Ingeniería Civil de Computación.

Entre 2000 y 2005, aproximadamente, en el DCC fuimos resolviendo poco a poco nuestros problemas, aunque también en este período debimos lamentar el fallecimiento de dos colegas y amigos. El 2005 asumió la dirección del DCC Domingo Mery, quien en mi opinión realizó una gestión que fue clave para que los profesores del Departamento nuevamente formáramos un equipo capaz de poner los intereses y el bien del grupo en primer lugar.

HOY

Definimos nuestra misión por primera vez sólo en 1994. A partir de entonces, la hemos actualizado un par de veces y hoy es la siguiente:

> *La principal tarea del Departamento de Ciencia de la Computación es crear, difundir, extender y promover conocimientos y experiencias que estimulen la formulación de ambientes y modelos adecuados a los alumnos, para que éstos puedan convertirse en científicos, ingenieros y profesionales de excelencia, líderes en su área, tanto en lo que tiene relación con los valores como con lo profesional, a través de la entrega y comunicación de conocimientos, experiencias y habilidades que les permitan llevar a cabo dicho propósito.*
>
> *Por esto, es de vital importancia que las actividades realizadas, tanto por nuestro Departamento como por los alumnos egresados de éste, puedan aportar al conocimiento, gestión y perfeccionamiento de las disciplinas ligadas a la computación, en los distintos ámbitos en que éstas se desarrollan, y tanto dentro de nuestro país, como en el extranjero.*
>
> *Esta misión involucra compromiso y responsabilidad con la sociedad, con la Universidad Católica y con nuestro Departamento, en el marco y orientaciones entregadas por los principios y valores declarados por nuestra universidad.*
>
> *Por esto, es labor de nuestro Departamento el promover y realizar docencia, investigación y extensión en el campo de la Computación, tanto en el aspecto científico, como en el*

Jorge Baier, Cristian Ruz, Enrique Guadalupe, Álvaro Campos y Jesús Figueroa.

El profesor Álvaro Campos junto a Cristián Ruz, Martín Gutiérrez y Mauricio Vinés, y el computador IBM 1620.

tecnológico, y en lo social y empresarial; y que estas iniciativas resulten en un desarrollo concreto, que permitan mejorar sustancialmente la calidad de los procesos en los cuales la computación es parte importante de su quehacer.

El número total de lo que podríamos llamar áreas de conocimiento comprendidas en las disciplinas de Ciencia de la Computación, Ingeniería de Software y Sistemas de Información es demasiado grande y las áreas son muy diversas como para ser abordadas en su totalidad, o incluso en su mayoría, por una planta que desde comienzos del 2000 y hasta recientemente era de sólo doce o trece profesores full-time. Por otra parte, debido a la realidad nacional e internacional de las aplicaciones de la computación, nos ha parecido válido abordar áreas de conocimiento de las distintas disciplinas, complementarias entre ellas, en lugar de privilegiar las áreas de una sola disciplina.

Desde el punto de vista de la docencia, el DCC es actualmente responsable de tres programas de pregrado y titulación profesional ofrecidos por la Escuela de Ingeniería. Los tres programas comparten una formación común en computación, con cursos de programación, matemáticas discretas, estructuras de datos, arquitectura, sistemas operativos y redes, sistemas de información, ingeniería de software, bases de datos, proyecto, y gestión de proyectos (por supuesto, los programas también comparten una formación común, para todos los programas de la Escuela de Ingeniería, en matemáticas, ciencias básicas y ciencias de la ingeniería, y requieren que los alumnos cumplan el plan de formación general de la universidad).

I) Ingeniería Civil de Computación. Este programa responde a nuestra decisión original de preparar ingenieros en la disciplina de Ciencia de la Computación, e incluye cursos sobre lógica, teoría de autómatas, algoritmos, e inteligencia artificial.

II) Ingeniería Civil de Industrias con Diploma en Ingeniería de Computación. A partir de 2009, este programa responde a nuestra observación de que un número importante de nuestros egresados se desempeñan en la práctica como ingenieros de software. El programa incluye cursos sobre diseño de software y testing, además de una preparación especial en gestión de Ingeniería, consistente en cursos dictados por el departamento de Ingeniería Industrial y de Sistemas.

III) Ingeniería Civil de Industrias con Diploma en Tecnologías de Información. Este programa está orientado a preparar ingenieros en la disciplina de sistemas de información. Incluye cursos sobre modelos de procesos, gestión de operaciones e integración de tecnologías, y, al igual que el Programa anterior, incluye una preparación en gestión de Ingeniería.

En los últimos cinco años (2007 a 2011), hemos titulado más de 320 alumnos, sumando nuestras tres especialidades. Las áreas de desempeño profesional de nuestros ex alumnos son muy diversas. Algunos, que siguieron estudios de Doctorado con nosotros o en otras universidades, son profesores en universidades en Estados Unidos o aquí en Chile. Otros son ingenieros de software en empresas tales como Google, Groupon o Amazon en Estados Unidos, y otros en empresas dedicadas al desarrollo de software aquí en Chile. También hay quienes han echado a andar sus propias aventuras empresariales en nuevas iniciativas dedicadas al desarrollo y a las aplicaciones de las tecnologías de información y comunicación en distintos ámbitos productivos y de servicios. Y también están los que ocupan diversos cargos en departamentos de informática (o equivalentes) en empresas dedicadas a otros rubros o en instituciones gubernamentales.

Desde el punto de vista de la investigación —que se refleja en participación en proyectos de investigación con financiamiento Fondecyt o Fondef, publicaciones en revistas especializadas, presentaciones en conferencias internacionales, supervisión de alumnos de los programas de Magíster y Doctorado, y cursos dictados a nivel de titulación o postgrado— nuestras cifras son my buenas. En los últimos cuatro años, hemos publicado aproximadamente 80 papers en revistas ISI, y hemos graduado 12 doctores y 30 magísteres; y actualmente tenemos 36 alumnos en el Programa de Doctorado, que está acreditado por cinco años, y 39 en el de Magíster, también acreditado por cinco años.

Por otra parte, nuestro programa de postítulo vespertino, INGES, nacido, como dije, a

principios de los noventa, ha recibido 176 alumnos en total en los últimos cinco años. Además, desde 2005, tenemos un programa de magíster profesional vespertino, conocido como MTIG, cuya matrícula total en los últimos cinco años suma 180 alumnos. Si bien nuestra contribución al medio profesional del país son principalmente nuestros ingenieros, estos programas nos permiten hacer un aporte en su formación a un número significativo de profesionales que se desempeñan en la práctica en áreas informáticas, pero que no se formaron en informática, o que se formaron hace algunos años y quieren actualizar o profundizar sus conocimientos.

Finalmente, el hecho de que la computación tiene cada vez más importancia y aplicabilidad en una creciente diversidad de disciplinas, más allá de la ingeniería, se refleja plenamente en la actividad del DCC. En la actualidad, nuestros profesores lideran o participan en proyectos, ya sea docentes, de desarrollo o de investigación, junto a profesores de astronomía, agronomía, biología, comunicación, diseño, educación, letras, matemáticas, medicina y psicología.

NUESTRO FUTURO

Estoy muy optimista con respecto a lo que podemos lograr en el futuro cercano. En los próximos meses vamos a recibir a cuatro nuevos profesores, ya doctorados, con lo cual vamos a conformar una planta de 17 profesores full-time. También vamos a seguir formando ingenieros en las especialidades mencionadas más arriba, y desarrollando investigación y docencia de postgrado al mejor nivel en las seis áreas en que, en la práctica, hemos (re) organizado nuestro quehacer últimamente: inteligencia de máquinas, bases de datos y teoría, ingeniería de software, informática educativa, sistemas de información, y recuperación de información.

En particular, creo que tenemos muy buenas posibilidades de posicionarnos como un Departamento líder internacionalmente en algunas de estas áreas. Por ejemplo, en inteligencia de máquinas, el grupo GRIMA, creado en 2005 e integrado por cuatro profesores, dicta cursos de pre y postgrado, y hace investigación y consultoría especializada en robótica, visión, aprendizaje de máquinas, planificación y minería de datos. Este grupo ha graduado veinte doctores y magísteres, es claramente líder en su área en Chile y está entre los mejores de Latinoamérica. Su trabajo está siendo aplicado en áreas tales como alimentos, astronomía, biología y retail. Su propósito es convertirse en un grupo líder a nivel mundial, un partner estratégico de la industria y de los principales centros de investigación internacionales, y una fuente de científicos y profesionales especialistas reconocidos internacionalmente.

En informática educativa, el trabajo de Miguel Nussbaum es claramente líder a nivel internacional, con más de 1.800 citas en Google Scholar, 17 doctores graduados, y proyectos con los Gobiernos de Uruguay y Colombia. Este trabajo ha producido innovaciones relevantes y el propósito de Miguel ahora es potenciarlo mediante su integración con las áreas de ingeniería de software, inteligencia de máquinas y sistemas de información.

Finalmente, en bases de datos y teoría, la investigación que desarrolla Marcelo Arenas en el tema de la complejidad computacional de los lenguajes de consulta en relación a la estructura de las bases de datos —investigación que ya cuenta con casi 3.400 citas en Google Scholar y varios Best Paper Award en las más destacadas conferencias internacionales especializadas— se va a ver potenciada con la llegada al DCC de dos nuevos profesores, que fueron alumnos de Marcelo y ahora están terminando sus doctorados en universidades del Reino Unido.

Por otra parte, en cuanto a la formación de pregrado, si bien estamos manteniendo nuestras tres especialidades, a partir de 2013 vamos a ofrecer una única licenciatura —correspondiente a los primeros cuatro años— compuesta por diez cursos de Ciencia de la Computación, además de los cursos exigidos por el Plan Común de Ingeniería (matemáticas, ciencias, formación general, etc.). Estamos definiendo los contenidos de estos diez cursos a partir de la más reciente propuesta curricular conjunta de la IEEE-CS y ACM, conocida como CS2013, actualmente en su versión *strawman*. Esto nos va a permitir concentrar mejor nuestros esfuerzos docentes en el pregrado. Pero lo más importante es que en la formación de todos nuestros egresados —independiente de su título profesional posterior— vamos a asegurar un conjunto sólido de conocimientos y competencias fundamentales de Ciencia de la Computación.

Por supuesto, también tenemos desafíos, desde fortalecer y consolidar nuestra investigación en ciertas áreas, hasta ser más atractivos para más alumnos de la Escuela de Ingeniería, tanto para que sigan nuestros programas de pregrado como los de postgrado. El tema de poder titular más alumnos me parece especialmente importante, considerando la gran necesidad de ingenieros de computación (o equivalentes, especialmente ingenieros de software) que hay actualmente en Chile, en nuestra Región, y más internacionalmente, y considerando que no pareciera que esta necesidad vaya a disminuir sino, por el contrario, que todas las señales muestran que va a aumentar. BITS

El hecho de que la computación tiene cada vez más importancia y aplicabilidad en una creciente diversidad de disciplinas, más allá de la ingeniería, se refleja plenamente en la actividad del DCC.

Búsqueda por contenido multimedia

Benjamin Bustos
Profesor Asistente DCC Universidad de Chile. Doctor en Ciencias Naturales de la Universidad de Konstanz, Alemania (2006); Magíster en Computación (2002) e Ingeniero Civil en Computación, Universidad de Chile (2001). Áreas de investigación: Bases de Datos Multimedia, Búsqueda por Similitud e Indexamiento de Bases de Datos Espaciales y Métricas.
bebustos@dcc.uchile.cl

En la actualidad nos encontramos en medio de un diluvio de datos digitales multimedia. Recientemente se ha reportado [Quora11] que cada día se suben aproximadamente 200 millones de fotos a Facebook. En el caso del video, se estima que para el año 2016 se transmitirán a través de Internet en un mes el equivalente a seis millones de años de video, y que en cada segundo se transmitirá un promedio de 1.2 millones de minutos de video. Asimismo, se estima que para el mismo año el tráfico de video en Internet corresponderá al 55% del tráfico total en la Red [Cisco12]. Todo este diluvio digital está siendo impulsado por la alta disponibilidad de aparatos productores de contenido multimedia (como cámaras digitales y de video, smartphones o tablets) y la necesidad de compartir dicha información, facilitado por el boom de las redes sociales.

Las aplicaciones derivadas de la producción y uso de datos multimedia son muy variadas, por ejemplo en áreas como la biometría, la bioinformática, la industria manufacturera, la industria del entretenimiento, los servicios financieros, la medicina, la química, la biología, y un sinfín de otras áreas. Las problemáticas asociadas a Multimedia son asimismo muy amplias: cómo acceder a colecciones masivas de datos multimedia, cómo publicarlos y distribuirlos, cómo interactuar con ellos, cómo garantizar su autenticidad, cómo sacarles el máximo provecho en nuestro quehacer diario y en las distintas tareas productivas, etc. Surge entonces la necesidad de realizar investigación en una amplia gama de dominios distintos para resolver todas las problemáticas mencionadas.

TIPOS DE DATOS MULTIMEDIA

Los datos multimedia se pueden dividir en distintos tipos. Dentro de los más comunes se encuentran:

- **Imágenes.** Conceptualmente es una matriz donde se almacena la información del color representado en cada píxel. Existen distintos formatos de archivos de imagen, algunos de ellos comprimen la información con pérdida (por ejemplo, jpg) y otros sin pérdida (por ejemplo, png).
- **Video.** Es una secuencia de *frames*, en donde cada *frame* representa una imagen en un instante de tiempo. Para dar la sensación de continuidad, un video tiene entre 25 y 30 *frames* por segundo.
- **Modelo 3D.** Son versiones digitales de objetos que pueden ser reales (autos, tazas, árboles, barcos, etc.) o ficticios (creados por un diseñador). Por lo general se representan utilizando una malla de triángulos, aunque existen otras representaciones posibles (por ejemplo, nubes de puntos en el espacio 3D o los denominados *surfels*, que son discos que cubren la superficie del modelo).
- **Serie temporal.** Es una secuencia de valores, provenientes de alguna fuente, medidos a lo largo del tiempo. Cualquier dato que pueda obtenerse en intervalos regulares de tiempo puede ser representado usando una serie temporal.
- **Texto.** También se puede considerar al texto plano del documento como información multimedia, por ejemplo cuando un documento no tiene estructura o ésta no se utiliza para describir su contenido.

ÁREAS DE INVESTIGACIÓN

La investigación en Multimedia abarca distintos ámbitos de la Ciencia de la Computación, entre los cuales se cuentan Computación Gráfica, Procesamiento de Señales (imágenes, audio, etc.), Recuperación de Información, Bases de Datos, Interacción Humano Computador, Sistemas Distribuidos, etc. Debido a esto, existen distintas áreas de especialización en Multimedia. Por ejemplo, la conferencia ACM Multimedia este año 2012 solicitó contribuciones en las siguientes áreas [ACM12]:

1. **Media Content Analysis and Processing.** Esta área se enfoca en el estudio de métodos para comprender y procesar información multimedia, por ejemplo algoritmos de extracción de características, segmentación y detección de objetos, algoritmos de aprendizaje de máquina para el análisis de contenido, etc.
2. **Multimedia Activity and Event Understanding.** Se enfoca en detectar eventos o actividades (información semántica de alto nivel) a partir del contenido multimedia.
3. **Multimedia Search and Retrieval.** Se enfoca en la búsqueda y recuperación de información a partir del contenido multimedia, para resolver tareas como búsqueda o detección de objetos, detección de copias, detección de "casi duplicados", etc., en grandes volúmenes de datos.
4. **Mobile and Location-Based Media.** Se enfoca en la investigación multimedia (análisis, búsqueda, interfaces, despliegue de contenido, etc.) para ambientes móviles.
5. **Social Media.** Se enfoca en estudiar cómo combinar la información generada en las redes sociales (tags, enlaces, información asociada a comunidades, etc.) con los datos multimedia asociados a dicha información.
6. **Multimedia Systems and Middleware.** Se enfoca en la investigación de sistemas multimedia en sus diferentes componentes, por ejemplo sistemas operativos, arquitecturas distribuidas, o representación de medios continuos o dependientes del tiempo, tanto a nivel de software como de hardware.
7. **Media Transport and Sharing.** Se enfoca en los métodos de transmisión o formas para compartir datos multimedia, por ejemplo datos producidos por sensores en tiempo real o datos transmitidos en vivo.
8. **Multimedia Security and Forensics.** Comprende temas como la seguridad, privacidad, y autenticación de contenido multimedia.
9. **Multimedia Authoring, Production and Consumption.** Estudia el ciclo de vida del material multimedia (creación, distribución, almacenamiento, acceso, etc.).
10. **Multimedia Interaction and Applications.** Estudia la interacción de los usuarios con datos multimedia en ámbitos como la educación, la salud, las relaciones sociales, etc., y en cómo hacer que dicha interacción humano-multimedia sea lo más natural posible.
11. **Multimedia Art, Entertainment and Culture.** Se enfoca en el estudio del uso de multimedia para la conservación del patrimonio cultural, para la creación artística y para el entretenimiento en general.

En este artículo voy a profundizar en particular en el área de búsqueda y recuperación de información multimedia, específicamente en métodos de búsqueda por similitud basados en el contenido multimedia.

BÚSQUEDA BASADA EN CONTENIDO EN COLECCIONES DE DATOS MULTIMEDIA

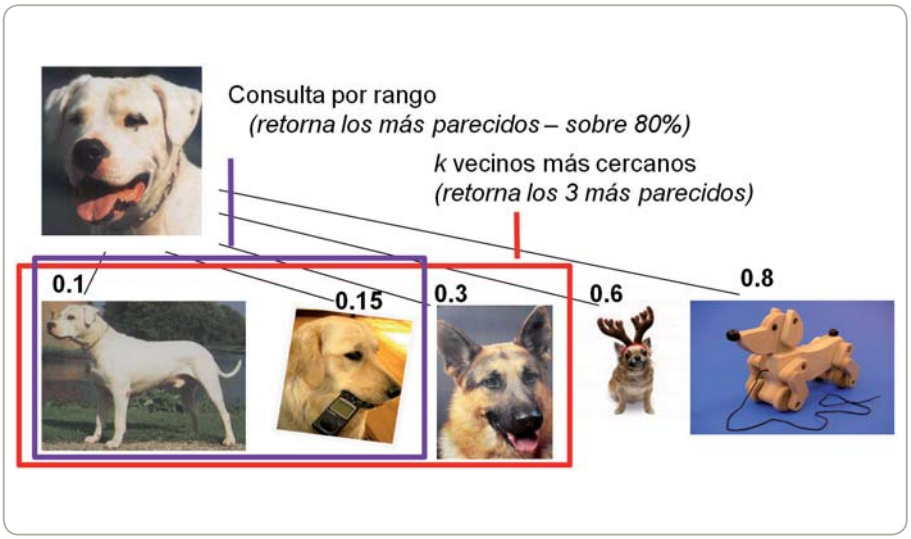
Ya se ha expuesto previamente que la producción de material multimedia ha crecido en forma muy importante en la última década, por la amplia disponibilidad de aparatos de digitalización y que se prevé que el intercambio de datos multimedia en Internet seguirá creciendo en los próximos años. Toda esta vasta cantidad de datos requiere entre otros, de métodos de búsqueda para poder encontrar información relevante. El problema que surge entonces es cómo poder realizar las búsquedas en forma eficiente (rápidamente) y eficaz (obteniendo resultados de buena calidad)

Los buscadores actuales en la Web de información multimedia, por ejemplo Google Images (images.google.com) están basados

en consultas textuales, es decir, el usuario debe ingresar palabras claves con la cuales buscar la información multimedia requerida. Esto ha sido un desarrollo natural a partir de toda la tecnología ya existente utilizada para buscar información textual en la Web. El problema es que los resultados que se pueden obtener a través de estas búsquedas textuales no siempre son satisfactorios, dada la complejidad de los datos multimedia. En efecto, en general los datos multimedia se almacenan directamente como "datos crudos" (la salida directa generada por los algoritmos de digitalización) y es común que no estén asociados a las anotaciones textuales que requerirían los motores tradicionales de búsqueda en la Web. Pregúntese usted mismo, ¿a cuántas de las últimas fotos que ha subido a su red social favorita le ha asociado anotaciones (tags)? Más aún, incluso si se diera el trabajo de anotar todas sus fotos, dichas anotaciones son subjetivas (salvo que sean metadatos que se extraigan automáticamente, como fecha de producción, posición geográfica, autor, etc.) y no necesariamente describen en forma completa lo que aparece en ellas. Por último, es tan vasta la cantidad de información multimedia que se genera constantemente, que sólo para aplicaciones muy acotadas sería posible agregar anotaciones a estos datos en forma manual.

Una forma alternativa para realizar una búsqueda en colecciones multimedia se basa en usar el contenido digital mismo de los datos multimedia para realizarla. Por ejemplo, en el caso de imágenes existen variadas técnicas para analizar en forma automática los colores, bordes, texturas y formas que aparecen en ella, y utilizando directamente esta información se puede realizar la búsqueda. Es decir, en vez de colocar una o varias palabras como términos de consulta (por ejemplo, escribir "torre Eiffel" en el buscador), uno coloca una imagen de consulta (por ejemplo, una foto de la torre Eiffel), la cual es analizada en su contenido y luego se realiza la búsqueda con dicha información. Esta forma de consultar se conoce como "búsqueda por ejemplo" *(query-by-example)* y es la base de los métodos de búsqueda basados en contenido.

Figura 1



Consulta por rango y consulta de los *k* vecinos más cercanos.

Por supuesto, dependiendo del tipo de dato multimedia con el que se desee trabajar y el problema específico en el que se desee desarrollar la aplicación, serán distintas las formas de analizar el contenido digital para poder realizar las búsquedas. No es lo mismo implementar un buscador de fotos en la Web que implementar un método de comparación de radiografías, ya que las características importantes a analizar pueden ser completamente distintas, a pesar de que ambas son imágenes. Una manera de formalizar y generalizar este proceso es utilizando el concepto de espacios métricos, que se detalla a continuación.

BÚSQUEDA POR SIMILITUD EN ESPACIOS MÉTRICOS

En general, no tiene mucho sentido realizar búsquedas exactas en colecciones multimedia, dado que sólo en el caso de copias digitales el contenido de dos objetos multimedia será igual (bit a bit). Es por esto que en la búsqueda en datos multimedia es fundamental el concepto de búsqueda por similitud.

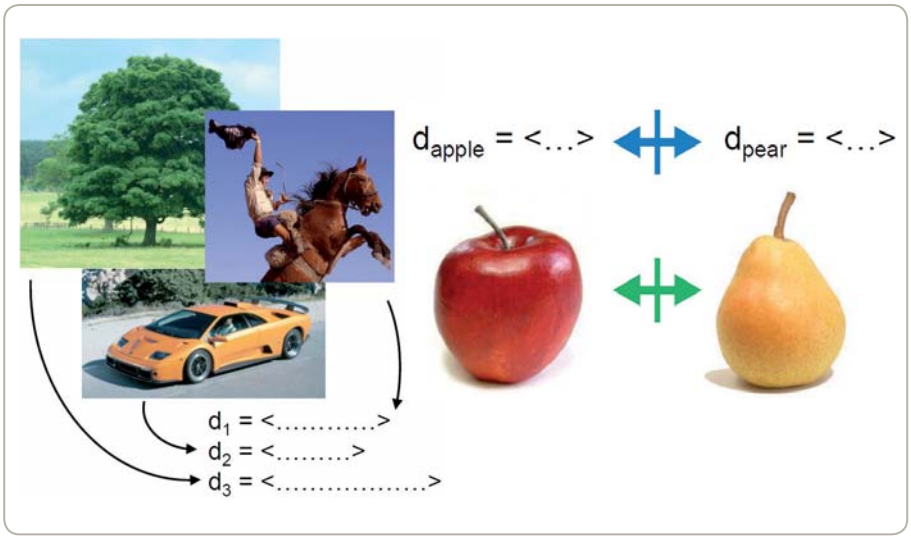
La búsqueda por similitud consiste en encontrar objetos que sean "parecidos" o "relevantes" dentro de una colección de datos. En el caso de definir una consulta multimedia usando "búsqueda por ejemplo", el objetivo consiste en encontrar todos los objetos de una colección multimedia que sean parecidos a la consulta. La similitud entre objetos se define a través de una función que retorna un valor real positivo. Por comodidad, usualmente dicha función representa la disimilitud entre los dos objetos: si la función retorna un valor pequeño es porque los objetos se parecen, si retorna un valor alto es porque son distintos. La Figura 1 muestra un ejemplo de búsqueda por similitud en imágenes y los dos tipos básicos de consulta: por rango y vecinos más cercanos.

El enfoque clásico para realizar búsquedas por similitud basadas en contenido consiste en representar un objetivo multimedia a través de un descriptor. Este descriptor se construye a partir de características o atributos extraídos del objeto multimedia, por ejemplo en el caso de las fotos se puede analizar el color, las formas, las texturas, etc., que aparecen en ellas. Usualmente, las características extraídas son atributos numéricos con el cual se forma un vector característico del objeto multimedia. La Figura 2 muestra un ejemplo del proceso de extracción de características.

La función de extracción de características se define de forma que si dos objetos multimedia son considerados similares por un ser humano (esto es totalmente dependiente del tipo de dato multimedia y del contexto o aplicación en el cual se desea medir la similitud), entonces sus vectores característicos correspondientes estarán cercanos de acuerdo a alguna función de distancia (la función que calcula la disimilitud). Lo usual es definir la función de distancia de tal forma que cumpla con las propiedades de una métrica (positividad estricta, reflexividad, simetría, y la desigualdad triangular), por lo que la colección de datos junto con la función de distancia forman un espacio métrico. Finalmente, el descriptor de cada objeto multimedia se puede almacenar en una estructura de datos ad hoc (un índice) para luego poder realizar las búsquedas en forma eficiente. En resumen, el problema original de búsqueda se transforma desde la comparación directa entre objetos multimedia a la búsqueda de objetos cercanos en un espacio métrico (ver Figura 3).

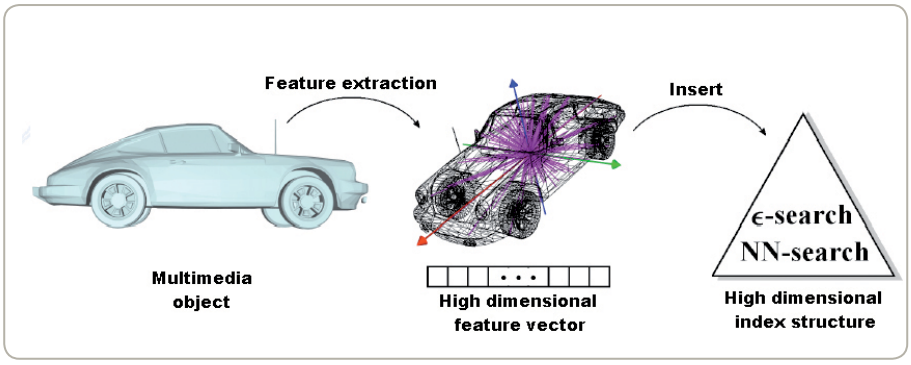
Para realizar una búsqueda, se le debe aplicar al objeto de consulta el mismo proceso de extracción de características, y luego se realiza la consulta especificada (por rango o *k* vecinos más cercanos). Si bien la consulta se puede responder comparando secuencialmente todo objeto de la colección con la consulta, el uso de un índice puede ayudar a mejorar la eficiencia de la búsqueda. Además, se debe considerar qué tan adecuada es la función de distancia utilizada, así como las características que se están extrayendo para realizar la búsqueda, para que ésta tenga sentido y retorne resultados relevantes.

Figura 2



Extracción de características en imágenes. El descriptor obtenido para una manzana debiera ser distinto al descriptor de la pera.

Figura 3



Proceso de extracción de características e indexamiento.

EFICIENCIA Y EFICACIA

La eficiencia de la búsqueda por similitud corresponde al tiempo de procesamiento requerido para obtener la respuesta. Si uno requiere buscar en colecciones de datos masivas, es imprescindible considerar índices que permitan realizar las búsquedas en forma eficiente. Cuando la función de disimilitud que se utiliza es una métrica, es posible indexarla utilizando índices métricos [Zezula06], los cuales hacen uso de la propiedad de la desigualdad triangular para descartar objetos o grupos completos de objetos de la colección durante la búsqueda. Si bien los índices métricos pueden mejorar la eficiencia de la búsqueda, no siempre es fácil determinar cuál es la mejor técnica de indexamiento en cada aplicación particular y, peor aún, la colección de datos puede ser intrínsecamente "difícil de indexar", problema conocido como la "maldición de la dimensión".

Por otra parte, la eficacia de un sistema de búsqueda en colecciones multimedia se define como su capacidad para encontrar objetos relevantes a la consulta especificada y, al mismo tiempo, para descartar aquellos objetos que no sean relevantes. Definir qué es relevante y qué no lo es en forma genérica no tiene mucho sentido, ya que esto depende mucho del tipo de dato multimedia, del contexto en el que se realiza la búsqueda y la aplicación en la cual se utilizarán los resultados. Para datos y aplicaciones específicas se han creado

> Los buscadores actuales en la Web de información multimedia, por ejemplo Google Images están basados en consultas textuales, es decir, el usuario debe ingresar palabras claves con la cuales buscar la información multimedia requerida (...) El problema es que los resultados que se pueden obtener a través de estas búsquedas textuales no siempre son satisfactorios, dada la complejidad de los datos multimedia.

colecciones de referencia, que se componen de un conjunto de objetos multimedia, de un conjunto de consultas y de un conjunto de respuestas a dichas consultas. Es decir, se conoce de antemano qué es lo que el sistema de búsqueda debería retornar para cada consulta definida en la colección. De esta forma se puede medir qué tan parecida es la respuesta del sistema de búsqueda con el ideal definido en la colección de referencia: mientras más parecido sea, más eficaz es el sistema de búsqueda.

Definir una colección de referencia puede ser un trabajo tedioso, ya que idealmente es una persona o un grupo de personas la que define el conjunto de consultas y sus respuestas (lo que implica tener que revisar la colección de datos "manualmente"). También se han utilizado las anotaciones asociadas a los objetos multimedia para clasificarlos y dividirlos en grupos. La ventaja de disponer de colecciones de referencia genéricas y disponibles para toda la comunidad es que permite comparar la eficacia de diferentes sistemas o métodos de búsquedas y así saber cuál retorna los resultados de mejor calidad.

DESAFÍOS A FUTURO

Existen muchos problemas no resueltos en el área de búsqueda en colecciones multimedia, y por supuesto en toda la diversidad de áreas de investigación en Multimedia en general. A continuación detallo algunos de los problemas abiertos en la actualidad:

- **Gap semántico.** Un problema importante del área es que aún es difícil extraer contenido de alto nivel (información semántica) directamente del contenido multimedia (datos de bajo nivel). Por ejemplo, es posible que dos fotos sean muy parecidas si uno analiza exclusivamente la distribución de colores en ellas, pero podrían ser fotos cuyo significado sea completamente distinto. Una de las formas en las que se está abordando este problema es combinar la información de bajo nivel con metadatos o información generada por el usuario (tags, información de bitácoras de búsqueda, etc.), de forma de enriquecer la información multimedia y por ende mejorar la calidad de los resultados obtenidos.
- **Big data.** Las técnicas actuales de indexamiento de datos multimedia pueden funcionar bien en colecciones de millones de objetos, pero hay pocas que tengan escalabilidad a tamaños mayores. Esto es lo que se requeriría para poder buscar en la Web o para buscar en datos que están siendo generados constantemente (datos de sensores, por ejemplo). Esto es tema activo de investigación actual.
- **Heterogeneidad.** ¿Cómo combinar la información digital proveniente de distintas fuentes y que son de distinto tipo?, ¿cómo identificar si pertenecen al mismo contexto?, ¿cómo enriquecer la información digital a partir de sus metadatos? etc.
- **Sistemas distribuidos.** Este tema es de particular importancia, ya que podría ser la forma de resolver algunos de los problemas planteados en Big Data. Es muy interesante por ejemplo, el estudio de algoritmos de búsquedas en redes P2P, donde las técnicas tradicionales de indexamiento no se pueden utilizar directamente.
- **Espacios no-métricos.** Para la búsqueda por similitud, es interesante el estudio de funciones de distancia no-métricas para comparar objetos, ya que en ciertas aplicaciones dichas distancias pueden obtener resultados más eficaces que funciones de distancias métricas más "tradicionales". Sin embargo, si no se cumplen las propiedades métricas, en particular la desigualdad triangular, no se pueden utilizar las técnicas de indexación tradicionales. Queda entonces el problema de cómo indexar en forma eficiente conjuntos de datos multimedia modelados como espacios no-métricos.

AGRADECIMIENTOS

Mis agradecimientos a Tomas Skopal, profesor de la Charles University in Prague, por permitirme usar algunas de sus figuras de ejemplo en este artículo.BITS

REFERENCIAS

[ACM12] ACM Multimedia 2012 Full/Short Paper Area Description. http://www.acmmm12.org/fullshort-paper-area-description/

[Cisco12] Cisco Visual Networking Index: Forecast and Methodology, 2011–2016. Mayo 2012. White paper. Disponible en: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf

[Quora11] http://www.quora.com/How-many-photos-are-uploaded-to-Facebook-each-day

[Zezula05] Zezula, P., Amato, G., Dohnal, V., and Batko, M. 2005. Similarity Search: The Metric Space Approach (Advances in Database Systems). Springer, Berlin, Germany.

El caso de Orand y Chequemático: **investigación y transferencia tecnológica en la industria chilena**

Mauricio Palma
Ingeniero Civil en Computación
Universidad de Chile.
Gerente General Orand S.A.
Founder SafeSigner.com.
mauricio.palma@orand.cl

Juan Manuel Barrios
Director de investigación de Orand S. A.
Doctor (c) en Ciencias mención Computación, Universidad de Chile.
Ingeniero Civil en Computación y Magíster en Ciencias mención Computación Universidad de Chile.
juan.barrios@orand.cl

Paola Cornejo
Periodista Pontificia Universidad Católica de Chile.
Directora de Comunicaciones Orand S.A.
paola.cornejo@orand.cl

Banco Bci, en conjunto con investigadores de la Universidad de Chile y de Orand, llevaron a cabo un proyecto de investigación y transferencia tecnológica que dio origen a Chequemático: el primer autoservicio en el mundo capaz de pagar y depositar cheques automáticamente.

ORIGEN DEL PROYECTO

Chequemático nace de la idea del Banco Bci de contar con un sistema capaz de pagar y depositar cheques en forma automática y disponible a toda hora, con el objetivo de mejorar el servicio al cliente y potenciar al Banco como líder en innovación tecnológica.

El primer desafío de Bci fue diseñar y construir, en conjunto con los fabricantes del hardware, el autoservicio y sus componentes. La funcionalidad de pago de cheques requiere incorporar dispositivos que habitualmente no están presentes en los ATM, como escáner de cheques, dispensador de monedas, lectores de huella dactilar y otros.

Resuelto el diseño del hardware, era necesario emprender el desarrollo del software que opera el autoservicio. Para ello, Bci encargó a Orand S.A. la programación de la aplicación que ejecuta cada máquina y que implementa la interfaz de usuario, el manejo de los dispositivos, y la comunicación con los sistemas transaccionales del Banco.

DESAFÍOS DE INVESTIGACIÓN

En sus inicios, el proyecto Chequemático se presentó como un desafío totalmente ingenieril. Se trataba de una aplicación de usuario que interactuaba con un hardware específico. Sin embargo, el avance del proyecto dio paso a aspectos más complejos ligados a la automatización, verificación del contenido del cheque y

En sus inicios, el proyecto Chequemático se presentó como un desafío totalmente ingenieril. Sin embargo, el avance del proyecto dio paso a aspectos más complejos ligados a la automatización, verificación del contenido del cheque y prevención de fraudes.

prevención de fraudes. Estos problemas no estaban totalmente resueltos por productos disponibles en el mercado mundial, y de hecho siguen siendo materia de investigación científica. Nos dimos cuenta de que para satisfacer los requerimientos específicos de esta aplicación era necesario llegar al estado del arte y crear nuevos algoritmos. Esto nos motivó a acercarnos al DCC de la Universidad de Chile, específicamente al grupo PRISMA liderado por el profesor Benjamin Bustos, en busca de ayuda.

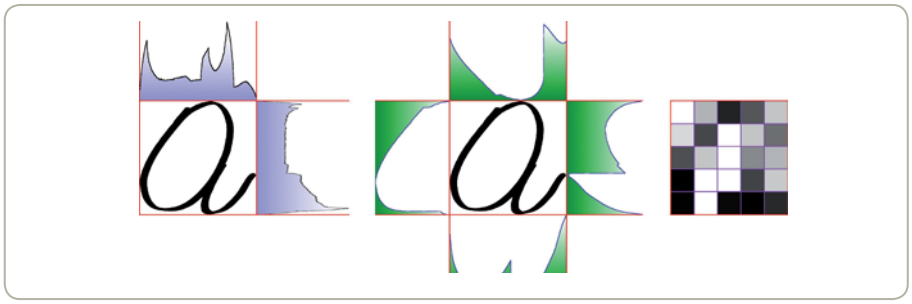
En general, el proceso de investigación realizado por Orand y el grupo PRISMA del DCC tuvo varias etapas, desde la revisión del estado del arte, exploración y descarte de alternativas, hasta la implementación del método más adecuado para enfrentar cada problema.

Entre los problemas a enfrentar, uno particularmente complejo es la automatización del reconocimiento o verificación de textos, especialmente cuando se trata de textos manuscritos. Los contenidos a reconocer incluyen el monto (expresado en dígitos y palabras), el destinatario, la firma y el endoso del cheque. A continuación describiremos el trabajo realizado para implementar la verificación del endoso del cheque.

VERIFICACIÓN DEL ENDOSO DEL CHEQUE

En este caso, el problema de investigación se definió como: dado un texto y una imagen, verificar que la imagen contiene ese texto escrito, ya sea con letra imprenta o manuscrita. Para abordar este problema, inicialmente se consideraron algoritmos como SIFT (Scale-Invariant Feature Transform), métodos basados en proyecciones de la imagen del texto y algoritmos morfológicos. Finalmente, se obtuvieron mejores resultados con búsquedas del vecino más cercano sobre imágenes de prueba generadas dinámicamente, a partir de una base de datos de letras de diferentes tipos de escrituras. Si la similitud con el vecino más cercano superaba un valor umbral, entonces el endoso era aceptado.

El sistema dispone de imágenes de entrenamiento para cada letra que fueron obtenidas de cheques representativos. Dada una imagen de entrada, el problema de segmentación de sus letras constituyentes es un problema difícil y de activa investigación. Para simplificar, se aprovechó que el objetivo es verificación. Usando el texto de entrada, un generador de imágenes combina letras de una base de datos de entrenamiento de diferentes tipos de escrituras, produciendo una gran cantidad de imágenes posibles. Las imágenes generadas son comparadas con la imagen de entrada para determinar un valor de similitud global. El problema por tanto se dividió en dos etapas: generación de candidatos y comparación con la imagen original.



Ejemplos de obtención de características para el cálculo de distancias: histogramas de proyecciones, histogramas de perfiles y densidad por zonas.

Para la generación de candidatos, la cantidad de combinaciones posibles aumenta exponencialmente con el número de imágenes de entrenamiento para cada letra. Por ejemplo, si se tienen d ejemplos de cada letra o dígito, para una palabra de largo n existen d^n combinaciones. Para no probar todas estas posibles combinaciones se usó un método incremental que considera sólo los mejores candidatos. En la imagen original primero se determinan puntos candidatos para segmentación o corte contando píxeles verticalmente y seleccionando los mínimos locales, es decir, donde hay adelgazamiento o separación de trazos. Luego, el generador crea imágenes con textos de largo un carácter, que son comparados contra la imagen dividida en los puntos candidatos de corte. Los mejores m candidatos son extendidos a largo 2, y se comparan con la imagen dividida en los siguientes puntos de corte, y así sucesivamente hasta encontrar un candidato de largo n coincidiendo con el último punto de corte. De esta forma la cantidad de comparaciones entre imágenes generadas e imagen original se reduce a $O(m \cdot d \cdot n)$. Adicionalmente, el número d se redujo restringiendo las combinaciones entre letras al considerar sólo letras compatibles en cuanto a su tipo.

Para la comparación de imágenes generadas con la imagen original se calculó la distancia según diferentes características visuales y se realizó una combinación lineal de éstas. Como la imagen original fue segmentada eligiendo puntos de corte, se pueden calcular características para cada una de sus letras. Entre las características estudiadas, las que entregaron un buen resultado en función de su simplicidad fueron histogramas de proyección horizontal y vertical, histogramas de perfiles horizontal y vertical, y la densidad por zonas. Estas características representan cada carácter en forma aislada y su suma ponderada entrega la distancia total entre imágenes.

Finalmente, el resultado de la validación corresponde a la distancia entregada por la mejor imagen candidata. Esta distancia es comparada con un valor umbral que define la tolerancia a errores que se desea obtener: un umbral muy pequeño no aceptará falsos positivos, pero también rechazará muchos positivos correctos. Por el contrario un valor umbral muy alto aceptará todos los correctos, pero también muchos incorrectos.

Los Gráficos 1 y 2 muestran un ejemplo de análisis realizado sobre el valor umbral en pruebas de laboratorio cuando se verifica con textos imprenta y con textos manuscritos. En el primer ejemplo, el primer falso positivo se encuentra cuando se ha aceptado un 68,2% de correctos y en el segundo con un 17,6%.

Después de seis meses de investigación, refinando las técnicas explicadas anteriormente, el grupo logró diseñar un nuevo método capaz de reconocer tanto letras como números manuscritos con los niveles de certeza requeridos por la aplicación.

Gráfico 1

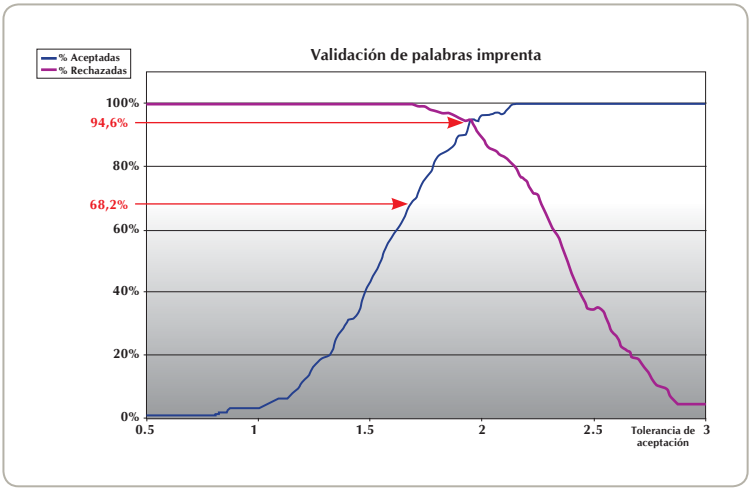
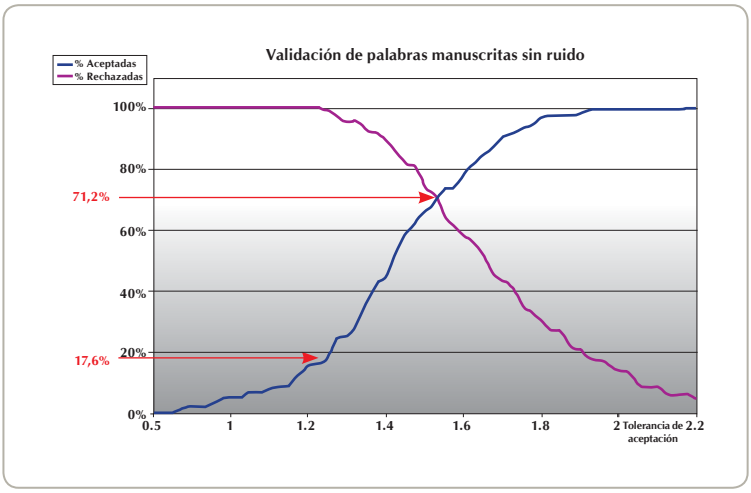


Gráfico 2

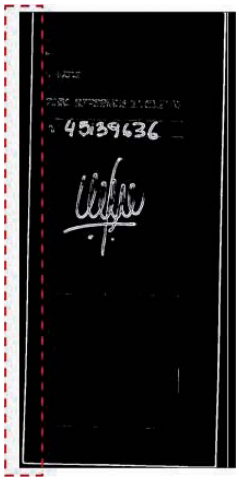


DESAFÍOS DE INGENIERÍA

La conversión de los resultados de la investigación en una solución industrial fue realizada por los ingenieros de Orand. Por una parte, el proceso de verificación estaba incompleto: la digitalización del cheque debe ser alineada, el endoso debe ser encontrado en cualquier parte del reverso del documento, la imagen puede presentar ruido, etc. Además, era necesario implementar la solución de acuerdo a la arquitectura y tecnologías definidas para el resto de la aplicación del Chequemático. Es interesante notar que estos desarrollos significaron esfuerzos y plazos mayores que los de la investigación misma. El preprocesamiento de la imagen del cheque considera los siguientes pasos:

1. Alineamiento de la imagen (Figura 1).
   a. Detección de bordes con método Sobel. Este método es menos sensible al ruido, por lo que elimina gran parte del fondo del cheque.
   b. Obtención del ángulo de alineación. Se analiza el borde izquierdo de la imagen y se detecta la línea usando la transformada de Hough.
   c. Rotación de acuerdo al ángulo de alineación.

Figura 1



2. Recorte de los bordes del cheque (Figura 2).

   a. Se calcula la proyección vertical de la imagen.

   b. Se analiza la imagen desde el centro hasta los extremos izquierdo y derecho, y se detectan puntos de corte.

Figura 2

3. Selección de zonas de interés (encontrar la ubicación del texto de endoso) (Figura 3).

   a. Binarización con un método adaptivo local.

   b. Cálculo de varianza de bloques de tamaño 4x4.

   c. Unión de bloques y selección del mejor candidato según la forma del bloque.

4. Extracción y binarización de las zonas a verificar (Figura 4).

El ejemplo de verificación de endoso sirve también para entender cómo se abordaron otros proyectos de investigación y transferencia que se requirieron para el autoservicio. Chequemático se convirtió en un exitoso caso de transferencia tecnológica para el Banco Bci, que fue destacado con el premio Innovación Empresarial en Tecnología 2009, como un avance en innovación al servicio de sus clientes.

Figura 3

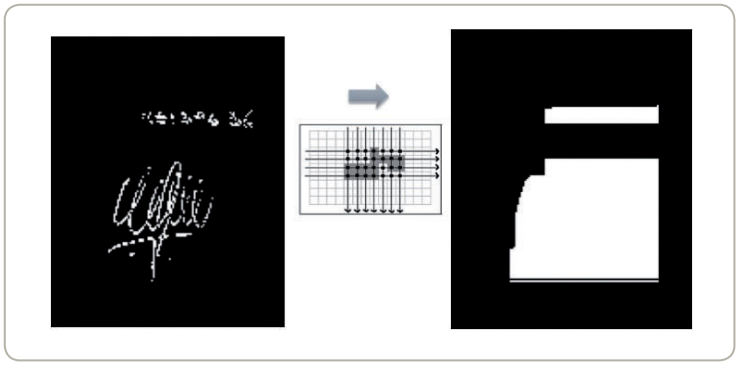
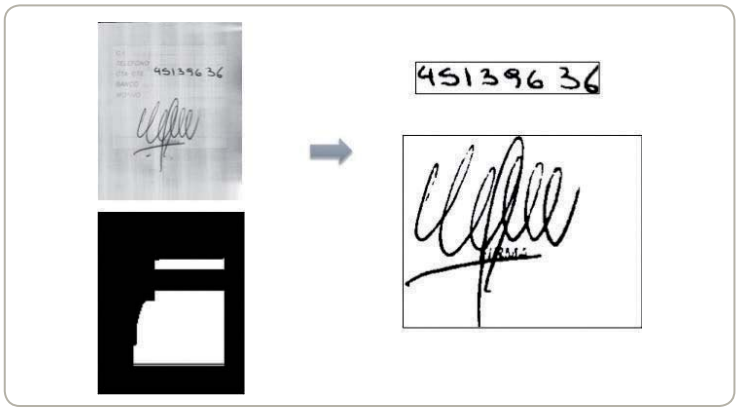


Figura 4



EL MODELO ORAND DE INVESTIGACIÓN Y TRANSFERENCIA TECNOLÓGICA

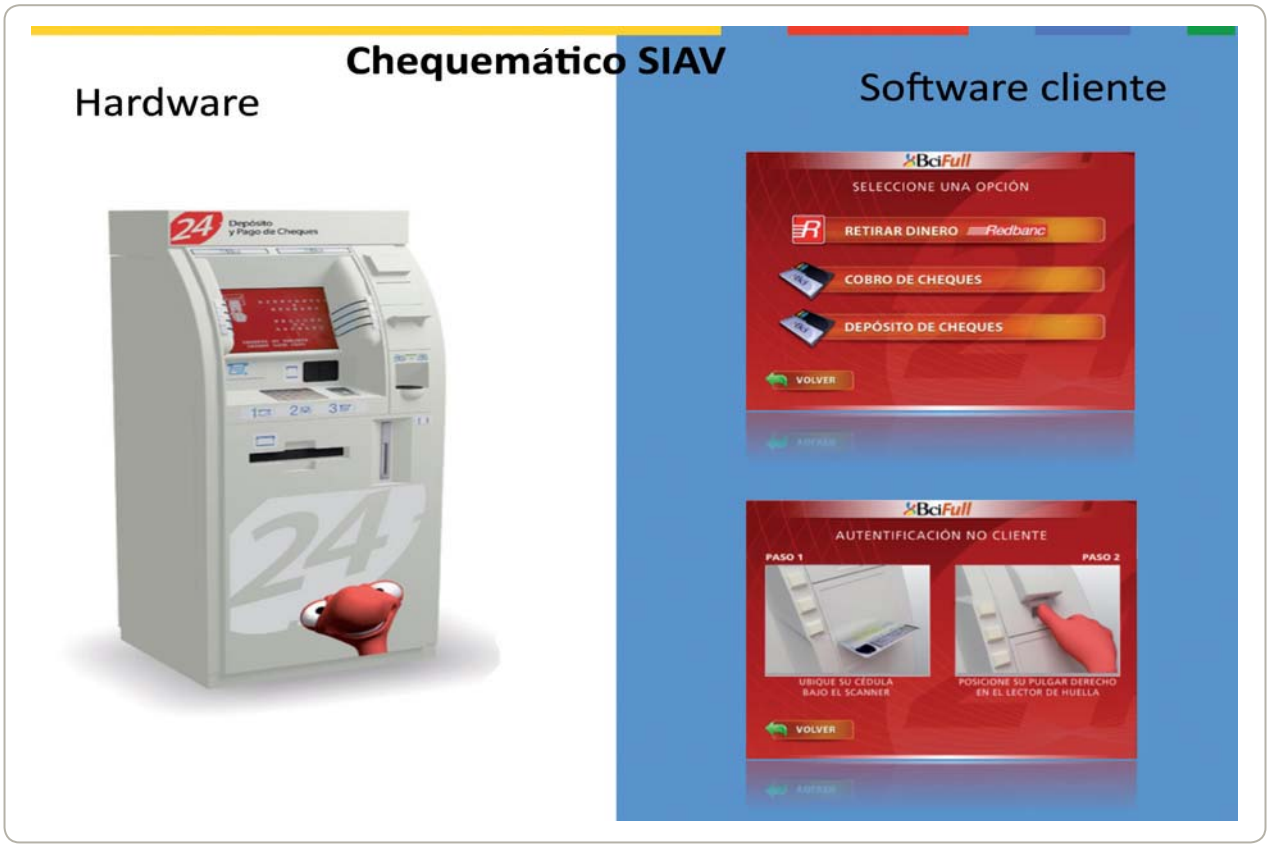
A partir de la experiencia de Chequemático, Orand ha continuado colaborando con la academia y se ha consolidado como la primera empresa privada de nuestro país constituida como Centro de Investigación y Transferencia Tecnológica en Ciencia de la Computación.

En Chile, y también en nuestra Región, existe poca colaboración entre la industria y la academia, lo que se explica por razones no triviales relacionadas con los incentivos, intereses, capacidades y culturas de ambos mundos. Luego del proyecto Chequemático nos dimos cuenta de que podíamos convertirnos en el eslabón necesario para unir ambas partes y emprender proyectos en conjunto. Entendemos las necesidades, motivaciones e idiosincrasias de ambos actores, y somos capaces de completar las piezas faltantes del rompecabezas, aportando la ingeniería, la capacidad de responder a plazos, de controlar de manera efectiva el riesgo, y de identificar los problemas y oportunidades relevantes.

La estrategia elegida por Orand considera incorporar investigadores internos, especializados en temas afines, cuya misión es realizar investigación propia en áreas de interés para la empresa, mantener y enriquecer la red de colaboración con la academia y participar junto a los ingenieros en el diseño de soluciones para la industria. Simultáneamente, se trabaja estrechamente con los clientes para detectar las oportunidades valiosas para su negocio y que se traducen en el desarrollo de soluciones que combinan investigación aplicada e ingeniería de alto nivel.

PERSPECTIVAS FUTURAS

Orand ha continuado el desarrollo del software y ha configurado productos y soluciones que abarcan todo el ciclo de procesamiento digital de cheques.

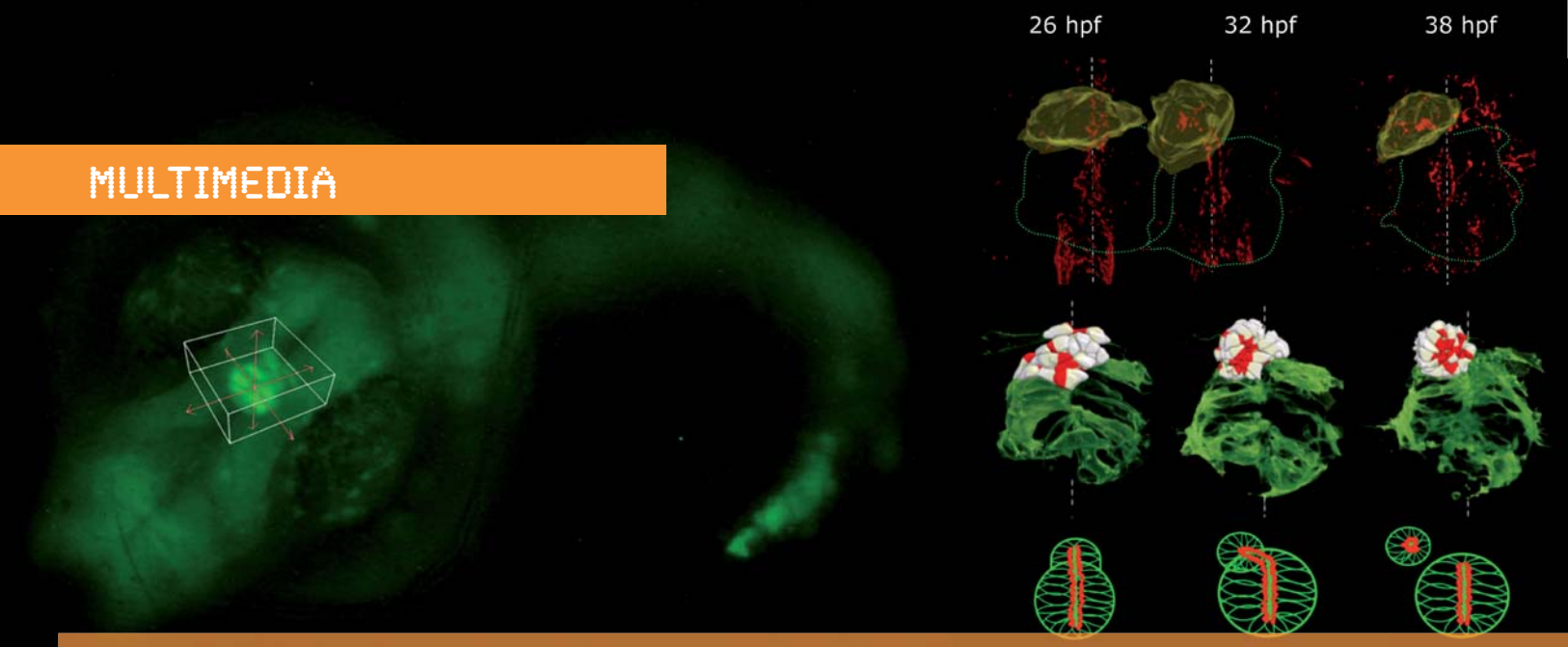


Estos productos están disponibles para autoservicios, cajas de atención al público, smartphones y otros dispositivos. Se complementan con sistemas de prevención de fraudes, workflow de procesamiento y software de intercambio de archivos (estándares X9.37 y relacionados) para implementar canje digital de documentos bancarios. La estrategia de la empresa para estos y otros productos es constituir empresas spin-off que se dediquen a su comercialización global.

De manera similar, se ha proseguido con desarrollos relacionados a reconocimiento de texto manuscrito, tanto para mejorar los resultados obtenidos como para abordar nuevas aplicaciones. Estos trabajos se realizan con investigadores internos y destacados colaboradores extranjeros, y se espera posicionar a Orand dentro de los líderes a nivel mundial de este tipo de soluciones.

> Después de seis meses de investigación, el grupo logró diseñar un nuevo método capaz de reconocer tanto letras como números manuscritos con los niveles de certeza requeridos por la aplicación.

Finalmente, la transferencia tecnológica se ha transformado en el negocio central de la empresa. Actualmente, Orand desarrolla proyectos de investigación y transferencia para bancos, retail, salud y sector público. Se espera a corto plazo trabajar con otras industrias, como telecomunicaciones, minería, astronomía y forestal. Para esto, se ha conformado una red de colaboración que incluye destacados investigadores de las principales universidades del país, universidades extranjeras como la Universidad Federal de Paraná, y centros de investigación como Yahoo! Labs Santiago o Max Planck Institute de Alemania. Los proyectos involucran variadas disciplinas, como seguridad informática, visión computacional, reconocimiento de patrones, detección de tópicos, análisis de sentimientos en textos, marcaje automático, análisis de series de tiempo, machine learning y otras. BITS



Ciencia computacional aplicada a la biomedicina

A la izquierda: imagen de fluorescencia de un embrión de pez cebra. A la derecha: fluorescencia, modelos geométricos y esquemas 3D representan cambios asimétricos entre izquierda y derecha en estructuras del cerebro del embrión, entre las 26 y 38 horas postfertilización. Imagen: C.G. Lemus, J. Jara, M.L. Concha, S. Härtel (LEO, SCIAN-Lab).

Nancy Hitschfeld
Profesora Asociada DCC Universidad de Chile. Obtuvo su doctorado en Technischen Wissenschaften en el ETH-Zurich (Suiza). Su área de estudio es el modelamiento geométrico, en particular la generación de mallas en dos y tres dimensiones, análisis de imágenes y programación orientada a objetos.
nancy@dcc.uchile.cl

Steffen Härtel
Doctor en Ciencias Naturales Universidad de Bremen (Alemania). Profesor Asociado en el Instituto de Ciencias Biomédicas de la Facultad de Medicina. Director de SCIAN-Lab. Su trabajo se orienta al desarrollo de modelos físicos, matemáticos y computacionales para microscopía óptica avanzada y procesamiento de imágenes, con aplicaciones que van desde la investigación básica a la telemedicina.
shartel@med.uchile.cl

Jorge Jara
Alumno del Programa de Doctorado en Ciencias mención Computación, DCC Universidad de Chile. Forma parte del Laboratorio de Análisis de Imágenes Científicas SCIAN-Lab (BNI, Facultad de Medicina U. de Chile). Trabaja en el desarrollo de métodos de segmentación 3D para imágenes de fluorescencia en series de tiempo.
jjara@dcc.uchile.cl

La ciencia e ingeniería computacional es una disciplina que permite entender, predecir y/o resolver problemas complejos que surgen tanto en el diseño y análisis de problemas, no sólo en ingeniería, sino también en el estudio de fenómenos naturales desde la astrofísica a las ciencias de la vida, con creciente frecuencia e impacto. En ciencias biomédicas, el desarrollo de nuevas técnicas microscópicas, nanoscópicas o genéticas generan avalanchas de datos que requieren un enfoque ingenieril en combinación con modelos matemáticos para reconocer patrones, crear nuevos componentes activos a escala nanométrica, o extraer y filtrar información de arreglos de expresión génica. Con la creciente capacidad de adquisición de datos a múltiples escalas, aumenta también la complejidad en el manejo e interpretación de éstos, por lo que se requiere del desarrollo y aplicación de nuevos modelos computacionales tanto de análisis como de simulación con enfoques multidisciplinarios, generalmente asociados a estrategias de computación de alto desempeño.

Actualmente, las aplicaciones de ciencia e ingeniería computacional resultan innumerables. Entre las más conocidas están: simulación de semiconductores y VLSI en ingeniería eléctrica; análisis de elementos finitos en ingeniería civil; simulación de estructuras moleculares y sus propiedades en física/química; simulaciones de Monte-Carlo y eventos discretos, y optimización matemática en ingeniería industrial; modelamiento de reservas de petróleo y yacimientos mineros en geología y minería; simulación de la física del universo y detección automática de cuerpos celestes en astronomía; modelamiento de la física de partículas y cálculo automático de sus interacciones en física; modelos de predicción del tiempo en geofísica; y análisis del genoma de los organismos, simulaciones del comportamiento de órganos, análisis y modelamiento multinivel

de estructuras biológicas en biología y medicina. Dentro de las técnicas usadas hay modelos matemáticos para la solución numérica de ecuaciones diferenciales parciales y ecuaciones lineales, métodos de modelamiento y análisis estadístico, optimización, modelos computacionales representados por algoritmos y estructuras de datos eficientes, tanto para resolver problemas discretos como continuos, algoritmos geométricos, programación paralela, técnicas de visualización y análisis de grandes volúmenes de datos.

Durante las últimas décadas ha habido un enorme avance tanto en la capacidad de cálculo y almacenamiento computacional como en las tecnologías para la adquisición de imágenes, en particular para la medicina y ciencias de la vida. El uso tradicional de las imágenes biomédicas ha sido para visualizar e inspeccionar las estructuras celulares o anatómicas. Hoy en día, las imágenes se han convertido en fuente de información crucial para planear, simular y visualizar cirugías en tiempo real, analizar y modelar el desarrollo de enfermedades, o entender el comportamiento y evolución de organismos vivos y sus estructuras componentes a distintas escalas. En este contexto, el desarrollo de la computación con algoritmos geométricos y análisis de imágenes se potencia con un número creciente de aplicaciones de uso cotidiano o de ciencia básica [1,2]. Uno de los grandes desafíos en este ámbito es mejorar la eficiencia y precisión de los algoritmos actuales para (1) identificar regiones de interés *(regions of interest,* ROIs) en imágenes, a través del proceso de segmentación, (2) construir representaciones geométricas de ROIs, (3) caracterizar y clasificar estas representaciones geométricas a nivel subcelular, celular o supracelular, y (4) analizar y visualizar los resultados en el contexto biomédico correspondiente. Éstas son tareas difíciles debido no sólo a la gran cantidad de datos a manejar, o la complejidad y variabilidad de las estructuras en estudio. Al mismo tiempo aparecen errores e incertidumbres inherentes a los datos y la microscopía, como limitaciones de resolución espacio-temporal, y marcación imperfecta de las estructuras de interés que causa que bordes de ROIs no se vean o no aparezcan cerrados [3].

En este artículo describiremos algunos de los métodos de análisis de imágenes y de representación geométrica que hemos estado desarrollando [4,5] para caracterizar el proceso de generación de órganos en embriones de pez cebra y el uso de procesamiento de imágenes para obtener imágenes de súper resolución en microscopía de fluorescencia.

LA COMPLEJIDAD DE ANALIZAR ORGANOGÉNESIS EN EL PEZ CEBRA

¿Por qué usar el pez cebra? Se trata de un modelo de desarrollo de organismos vertebrados para el cual se conocen diversas técnicas de experimentación genética y de manipulación de embriones. Los embriones de pez cebra son casi transparentes, lo que permite usar técnicas de marcación fluorescente y luz visible para observar al mismo tiempo distintas estructuras subcelulares o celulares con uno o más "colores" (longitudes de onda). La Figura 1 muestra el enfoque general que utilizamos para el procesamiento de imágenes de fluorescencia en embriones de pez cebra. Luego, en la Figura 2 se muestran modelos geométricos para representar y cuantificar la morfología en células del complejo pineal-parapineal en el cerebro, además de los cambios que ocurren durante su desarrollo. Utilizando estos acercamientos estudiamos el desarrollo de la asimetría izquierda/derecha en el sistema nervioso durante la organogénesis del cerebro en el pez cebra. Análisis de este tipo se realizan en condiciones normales y alteradas, con múltiples aplicaciones en estudios de enfermedades o de terapias [6].

La proliferación, migración y cambios en la organización de células durante la formación de cualquier órgano involucran procesos en intervalos de tiempo que van de segundos a días. Para la observación *in vivo* de estos fenómenos se recurre a la microscopía confocal de fluorescencia, que permite observar embriones marcados con moléculas exógenas o proteínas fluorescentes (fluoróforos) que no alteran al sistema vivo y sin necesidad de fijarlos a una placa, como en el caso de la microscopía electrónica. Mediante un sistema de espejos y filtros es posible obtener imágenes de varios canales de fluorescencia en un mismo espécimen (por ejemplo, longitudes de onda en rojo para núcleos y verde para membranas celulares) y realizar capturas de imágenes tanto en dos como tres dimensiones a intervalos de tiempo controlados.

Actualmente existen en Chile más de una decena de microscopios confocales "convencionales" en distintas universidades y centros de investigación, y unos pocos orientados a alta resolución espacio/temporal y la adquisición automatizada para experimentos simultáneos con múltiples muestras: (1) el *spinning disk microscope* (SDM, Facultad de Medicina U. de Chile) es capaz de adquirir imágenes 2D a tasas de ~10 imágenes/seg. (de unos 1.024x1.344 píxeles), permitiendo observar procesos de transporte y reorganización de estructuras intracelulares como el retículo endoplasmático. En 3D, el SDM permite capturar un volumen de datos del orden de 1.024x1.344x70 vóxeles ("píxeles 3D") cada cinco minutos, permitiendo observar la dinámica de conglomerados de células como en el caso del complejo pineal y la habénula en el cerebro (Figura 1 y 2). (2) el *Large scale imaging* (LSI) o *macro-zoom* (U. de Chile), permite observar estructuras en escala de milímetros a nanómetros en placas con hasta 96 muestras, mediante fluorescencia o sin ninguna marcación (contraste de fase), y cuenta con software programable para realizar capturas automatizadas durante horas o días. Un experimento típico en estos microscopios con 1-2 canales de fluorescencia genera 20-80GB de imágenes "crudas" *(raw)*, los que pueden aumentar 5-10 veces según la información extra que se requiera generar, como imágenes con reducción de ruido, mejoras en la resolución y/o modelos geométricos asociados a las estructuras de interés. La

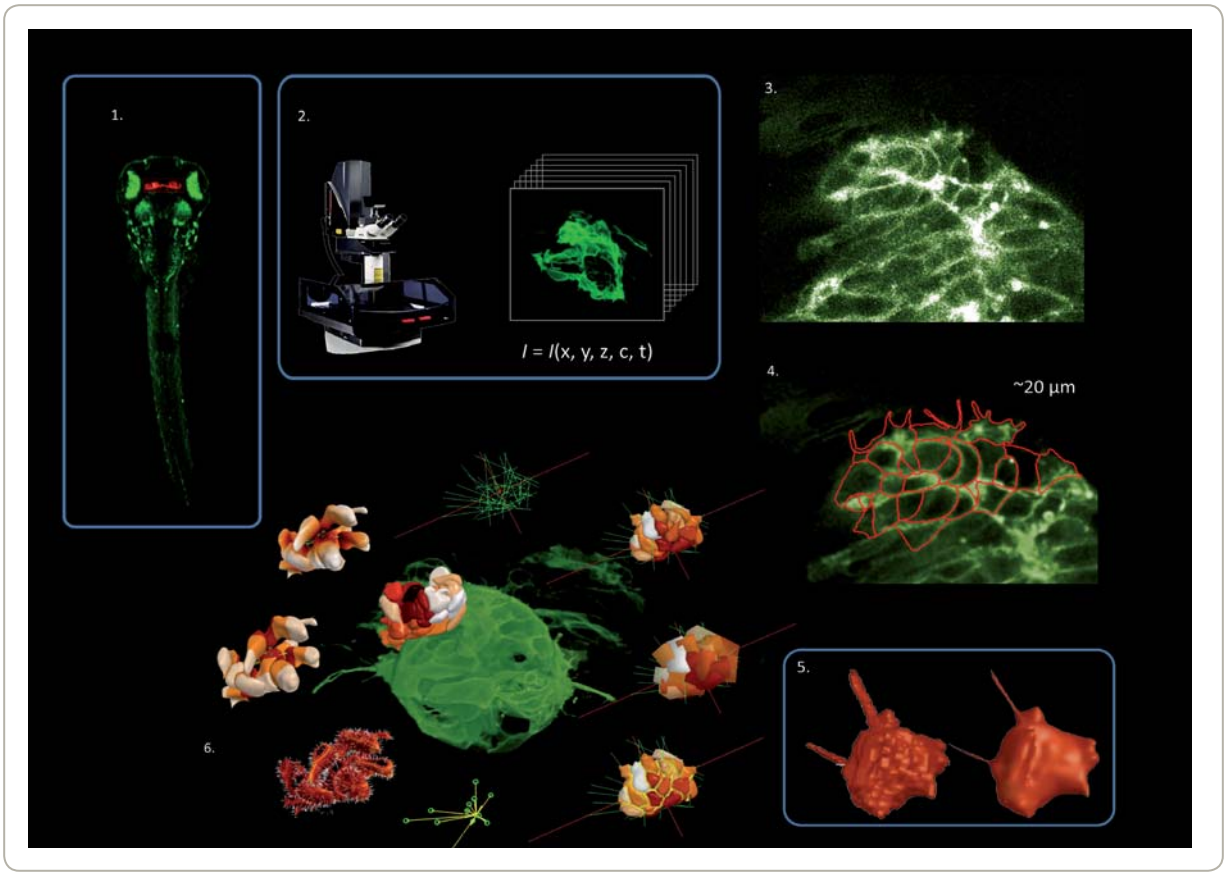
Figura 1 esquematiza una secuencia de procesamiento de imágenes típicas para el estudio del desarrollo asimétrico del órgano parapineal, que ocurre entre las 24 y las 36 horas postfertilización (hpf) de un embrión de pez cebra. Una vez adquiridas, las imágenes de fluorescencia se someten primero a una etapa de *tratamiento* para minimizar el ruido y las distorsiones del microscopio (deconvolución); posteriormente, diversas técnicas de *análisis* se enfocan en extraer información tanto sobre las imágenes como sobre las estructuras de interés que se puedan detectar; finalmente, las tareas de *comprensión* o *interpretación* apuntan a consolidar la información obtenida en las etapas anteriores para responder a las preguntas de alto nivel que motivan el procesamiento. Nos centraremos ahora en la parte de análisis, particularmente en la segmentación, modelamiento y optimización de características de forma y organización de ROIs a nivel subcelular, celular y supracelular.

Segmentación de estructuras y caracterización de morfología, topología y organización

El análisis de imágenes, a nivel de células o de sus organelos constituyentes, complementa al uso de técnicas de biología a nivel molecular, como la genética o proteómica, ya que permite observar fenómenos a una escala mayor que la sola expresión de genes y proteínas (cientos de veces más pequeños). Por ejemplo, la organización de múltiples células en forma de rosetas o intercaladas en capas obedece no sólo a interacciones moleculares a escala de nanómetros, sino también a interacciones en el orden de micrómetros, y características como posición relativa, tamaño y forma de las células. Para llevar a cabo caracterizaciones de este tipo resulta clave contar con algoritmos precisos para:

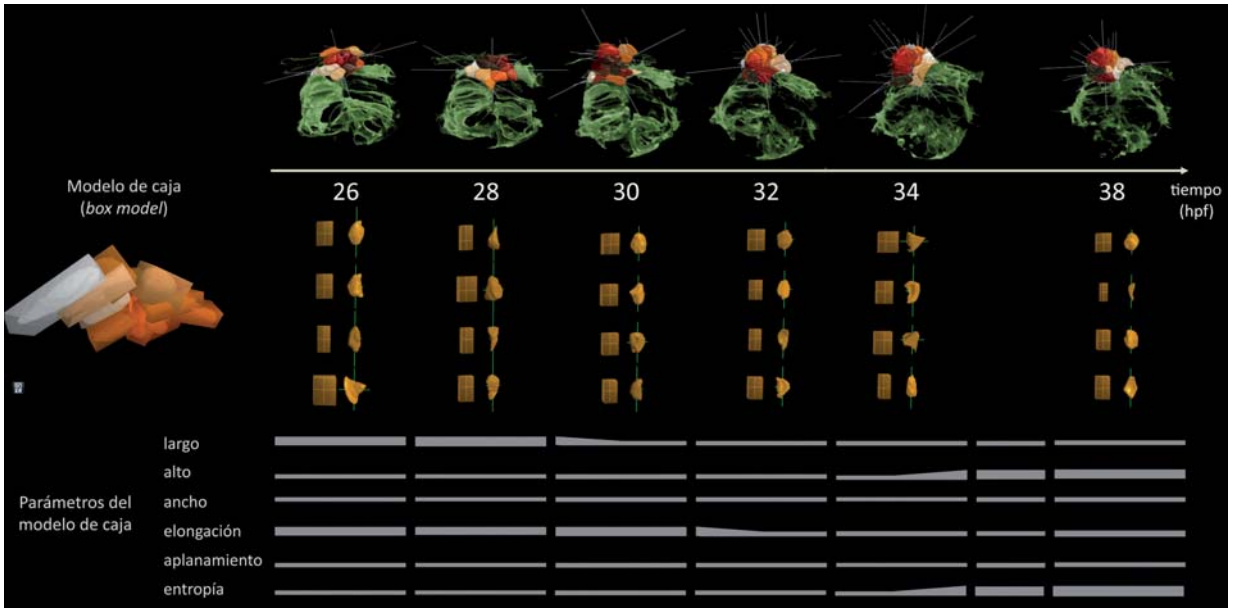
- Identificar ROIs en imágenes 2D/3D, en el llamado proceso de segmentación.
- Cuantificar características morfológicas (e.g. tamaño, orientación, curvatura del borde), topológicas (conectividad de estructuras, ramificaciones) y de organización o patrones (compactación, alineamiento, adyacencia o proximidad entre ROIs).
- Realizar análisis en series de tiempo, incluyendo estimación de movimiento y seguimiento de ROIs *(tracking)*, además de los cambios de características de cada una.

Figura 1



Procesamiento de imágenes en un embrión de pez cebra. Un microscopio láser confocal permite observar embriones marcados con proteínas fluorescentes (1), generando imágenes 3D en el tiempo de estructuras cerebrales (2,3). Sobre estas imágenes se aplican algoritmos de mejora y detección de regiones de interés como membranas celulares (4), luego se construyen modelos geométricos que son optimizados (4) para obtener descriptores de forma y organización (5). Imagen: Carmen Gloria Lemus, Karina Palma, Lorena Armijo, Néstor Guerrero, Miguel Concha (LEO, Facultad de Medicina), Jorge Jara (DCC, SCIAN-Lab) y Steffen Härtel (SCIAN-Lab, Facultad de Medicina).

Figura 2



Cambios morfológicos en las células del órgano parapineal del embrión de pez cebra. Mediante una representación simplificada de modelos de caja, se observan cambios significativos en la morfología de las células del complejo parapineal, en el período de 26 a 38 horas postfertilización (hpf). Imagen: Carmen Gloria Lemus, Miguel Concha (LEO, Facultad de Medicina) y Steffen Härtel (SCIAN-Lab, Facultad de Medicina).

Para segmentar imágenes usamos modelos de contorno activo [7], que representan al borde de cada ROI mediante polígonos simples en 2D, o mallas de superficie en 3D compuestas por caras poligonales (poliedros). La idea general es definir una serie de características deseables en forma de fuerzas que deforman a cada contorno hasta alcanzar un estado de equilibrio que representa una segmentación optimizada. La acción de las fuerzas es modelada por ecuaciones diferenciales parciales discretizadas, que se resuelven en forma iterativa. Típicamente se definen *fuerzas internas* intrínsecas a la morfología del contorno, tales como su continuidad o curvatura, y *fuerzas externas* características de la imagen tales como las transiciones de color o intensidad.

En nuestro caso, el proceso de segmentación consiste en obtener primero una estimación inicial de las ROIs en cada imagen, mediante operaciones (filtros) a nivel de píxeles, tales como clasificación semiautomática [8], o bien mediante dibujos de biólogos expertos para los casos más difíciles. Las imágenes filtradas terminan en blanco y negro, con los píxeles en blanco representando a la(s) ROI(s) y en negro al fondo. Para imágenes 3D, esto se hace de modo similar con varias imágenes 2D (plano xy) que se "apilan" en el eje *z* para formar un volumen compuesto por vóxeles (cubos cuyo lado es un píxel). Un algoritmo de tipo *marching cubes* [9] genera una malla de triángulos que representa la superficie de cada ROI reconocida. Las triangulaciones generadas están compuestas de hasta cientos de miles de vértices, arcos y triángulos. Debido a la complejidad de las estructuras celulares, las triangulaciones generalmente no son conformes, es decir que sus elementos constituyentes no están bien conectados, generando casos como superficies que no encierran un volumen, o cuyas caras son polígonos que se intersectan entre sí. Además, los triángulos pueden tener ángulos muy pequeños, muy grandes, o un área muy pequeña. Dado que estas triangulaciones las usamos para calcular propiedades morfo-topológicas de las estructuras, y los algoritmos que las calculan requieren que la malla sea conforme y de buena calidad, realizamos un proceso de reparación y mejoramiento de calidad. Para reparar la malla usamos implementaciones de algoritmos en librerías de software abierto [10], que detectan y resuelven las inconsistencias. Para mejorar los ángulos mínimos manteniendo las características geométricas usamos el algoritmo de Delaunay, y para simplificar la malla eliminando triángulos de área pequeña usamos una implementación local de un algoritmo basado en el colapso de aristas [11].

Cuantificación de características morfológicas mediante skeletons

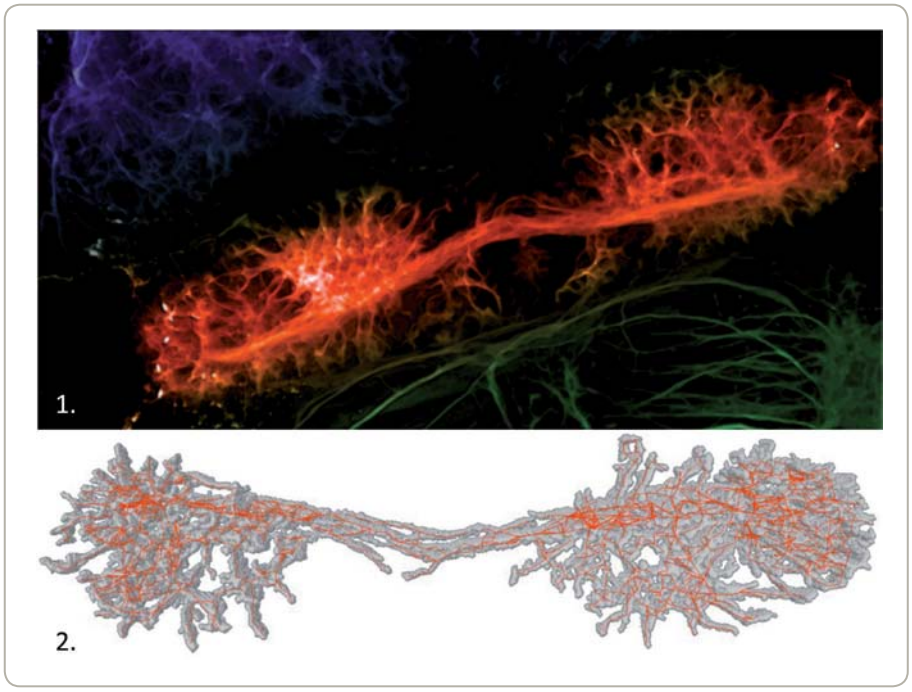
Usaremos el concepto de eje medial o esqueleto *(skeleton)* para un objeto o ROI 2D (polígono) o 3D (poliedro), como una estructura geométrica de tipo grafo formando líneas (1D) ubicadas en el interior del objeto original, y conservando varias de sus propiedades. En el caso de los sistemas biológicos, esto nos permite cuantificar, por ejemplo, el número de nodos (puntos terminales y de unión), la cantidad de bifurcaciones en cada nodo, largo de arcos y ángulo entre los nodos, entre otras propiedades, que dan cuenta de patrones arquitectónicos a nivel celular y subcelular,

tales como ramificaciones y conexiones entre neuronas, o redes de transporte de proteínas (retículo endoplasmático) y fibras estructurales (citoesqueleto) al interior de células. La Figura 3 muestra ejemplos de *skeletons* de la habénula en pez cebra, desde imágenes 3D de fluorescencia.

Existen varios algoritmos para calcular *skeletons*, pero no todos conservan las mismas propiedades del objeto original. Para representar las conexiones neuronales, nos interesa un algoritmo que mantenga el número de componentes conectados, túneles y cavidades, que sea invariante bajo transformaciones isométricas (como rotaciones o escalamiento), que esté aproximadamente centrado con respecto al objeto que representa y que sea poco sensible al ruido del borde de éste. Debido a esto elegimos trabajar con un algoritmo que recibe como entrada la triangulación que describe la superficie del objeto y le aplica (1) un proceso de contracción, (2) la transformación en líneas para generar un *skeleton*, y (3) un postproceso a las partes del *skeleton* que queden fuera o cerca del borde del objeto original, o que sean poco representativas de la forma original [12,13]. Para esto se utiliza una representación de caras (triángulos) compuestas por aristas (segmentos de recta) y vértices (puntos).

La etapa de contracción consiste en desplazar los vértices de la triangulación hacia el interior del objeto, en una dirección que se calcula en función de los vértices vecinos. Este proceso se realiza hasta que el área de la superficie contraída alcanza el 15% del área de la triangulación original aproximadamente. Si se visualiza esta malla, se asemeja a un *skeleton* (Figura 3). La etapa de transformación a un *skeleton* elimina todos los triángulos a través de un *colapso de aristas* dejando sólo aristas conectadas entre sí. Se usa un algoritmo "avaro" o *greedy*, que iterativamente calcula el costo de colapso para todas las aristas de la malla, y luego remueve la de menor costo. La función de costo incluye un término de forma y un término de muestreo. El costo de forma cuantifica la distorsión producida al colapsar una arista particular y el costo de muestreo cuantifica el largo de las aristas (o espaciamiento entre los puntos que las unen). Finalmente, en la etapa de postprocesamiento se evalúa si existen partes del *skeleton* que necesitan centrarse y, si es así, se usa la información de los vértices originales que fueron colapsados a cada parte para centrarla.

Figura 3

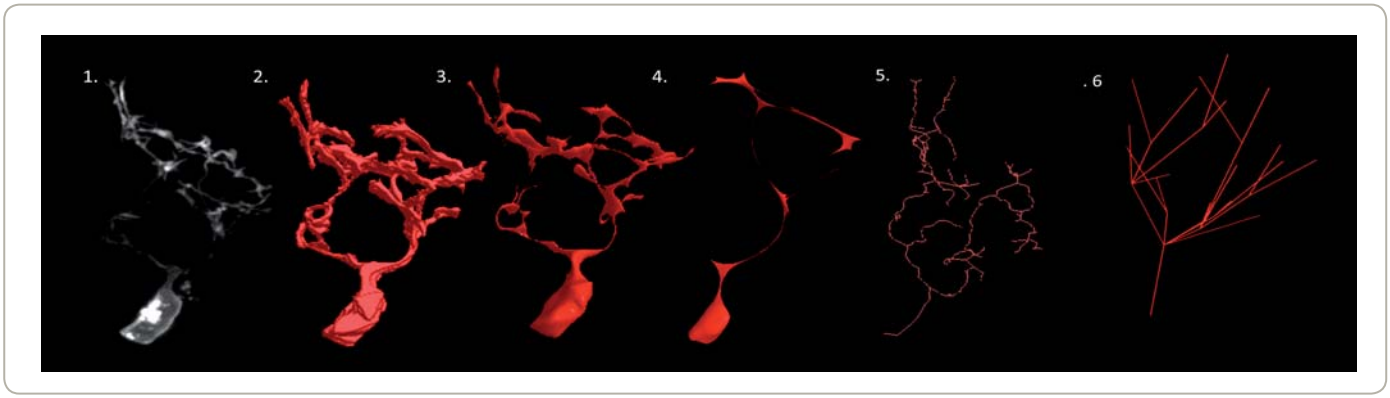


***Skeletons* de proyecciones neuronales en la habénula de un embrión de pez cebra. A partir de la imagen de fluorescencia 3D (1), se identifican las proyecciones neuronales y se representan con modelos de superficie 3D. Un algoritmo de esqueletonización genera un modelo de líneas que representa el patrón de conectividad de las proyecciones (2). Imagen: Karina Palma (LEO, Facultad de Medicina) y Pablo Aguilar (DCC). Distinción en concurso Nikon Small World 2009.**

MICROSCOPÍA DE SÚPER RESOLUCIÓN SOFI

La resolución de la microscopía óptica clásica está limitada por la difracción de la luz (ley de Abbe). En la última década esta restricción ha sido superada mediante técnicas avanzadas de óptica combinadas con análisis matemático y procesamiento de imágenes. Una de ellas, SOFI *(Superresolution Optical Fluctuation Imaging)* se basa en las fluctuaciones temporales estocásticas e independientes de los emisores de fluorescencia (asociados a las estructuras de interés) y en el procesamiento estadístico postadquisición de imágenes que registran estas fluctuaciones [14]. Mediante el cálculo de correlaciones, mediante cumulantes de orden superior y análisis de Fourier, es posible "filtrar" la fluorescencia cuya fluctuación se ajusta al patrón de los marcadores, localizándola con precisión subpíxel (o subvóxel) y eliminando la fluorescencia espuria como ruido fotónico o autofluorescencia (Figura 5.1). La generación de una imagen de súper resolución requiere de miles de imágenes adquiridas a muy alta velocidad (~200-500 GBytes en 2D), y para incrementar la resolución en factor *n se* requieren $O(n^2)$ de memoria temporal y cantidad de cálculos. La ventaja de SOFI sobre otras técnicas similares recae en que no se necesita equipamiento electrónico o sistemas de adquisición sofisticados, lo que repercute significativamente en su costo. En colaboración con el laboratorio del Dr. Jörg Enderlein en la Universidad de Göttingen (Alemania), inventor de la técnica, hemos desarrollado algoritmos de SOFI, protocolos para tinción simultánea de dos proteínas en longitudes de onda distintas (rojo y verde), adquisición de imágenes y análisis de co-localización de receptores de neurotransmisores en neuronas de

Figura 4



Construcción del modelo de *skeleton* para una neurona de un embrión de pez cebra: imagen de fluorescencia de una neurona (1), mallas de superficie 3D detectadas (3) y optimizadas (4), y las etapas del algoritmo de *esqueletonización* (4-6). Imagen: Karina Palma (LEO/SCIAN-Lab, Facultad de Medicina), Liliana Alcayaga (DCC), Mauricio Cerda, Jorge Jara (DCC, SCIAN-Lab).

hipocampo. Utilizando un microscopio de epifluorescencia obtenemos series de diez mil imágenes por cada longitud de onda, para generar un incremento de resolución de $\sqrt{2}$ (SOFI de orden 2, Figura 5.2).

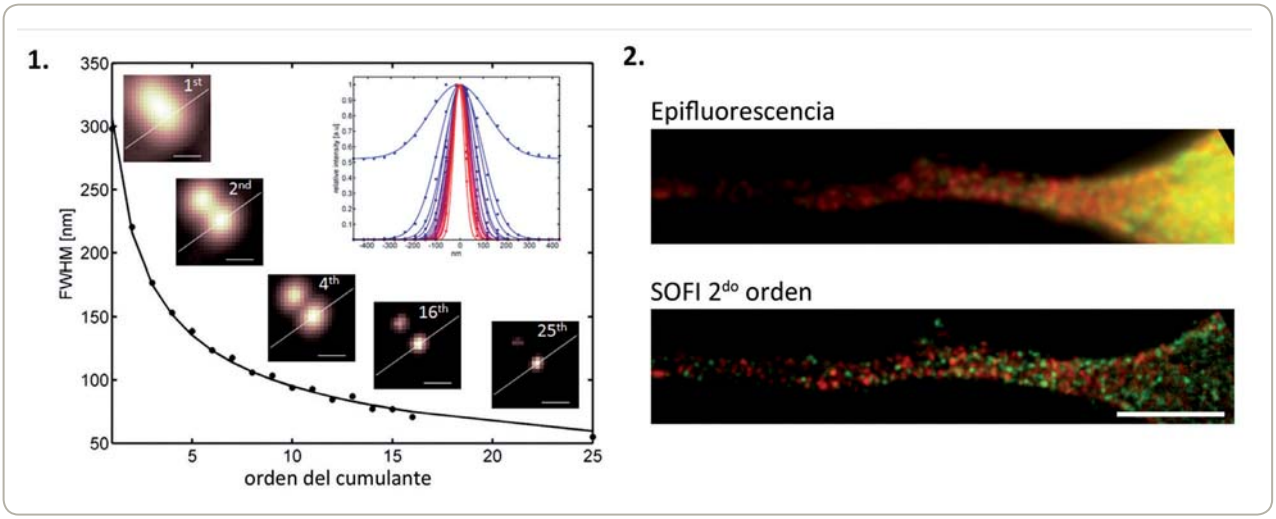
TRABAJO ACTUAL Y FUTURO

Actualmente nos encontramos desarrollando métodos para combinar la segmentación con técnicas de estimación de movimiento, para conocer, además de la posición de cada ROIs en el tiempo, qué tipo de movimiento realiza y los cambios que experimenta en relación a su entorno. La dinámica de estructuras biológicas observadas *in vivo* incluye movimientos con deformaciones de diversos tipos, como la formación y retracción de membranas con formas características, cambios de orientación en conjuntos de células y formación de adhesiones célula-célula, por nombrar algunas. En el caso del desarrollo del órgano parapineal, sus células se desacoplan del resto del complejo, para luego migrar en una dirección preferente, a la vez que se forma una especie de roseta con un punto de convergencia. Hasta ahora se conocen algunos genes y proteínas que participarían del proceso y, aunque sus patrones de expresión se han visto sincronizados con la dinámica descrita en el párrafo anterior, los mecanismos a nivel celular y supra-celular se desconocen. Un entendimiento detallado de este fenómeno permitiría acercarse a las causas de enfermedades del sistema nervioso, además de sugerir blancos y estrategias terapéuticas. Una problemática similar encontramos en otros órganos del pez cebra, además de modelos de células en cultivo sujetas a migración y deformaciones. Es por esto que nos encontramos trabajando en métodos variacionales de flujo óptico para estimación de movimiento, buscando una herramienta flexible que podamos aplicar en diversos escenarios [5,15]. En este contexto, la participación como plataforma biomatemática en el Instituto Milenio de Neurociencia Biomédica (BNI) diversifica las tareas y desafíos de nuestras colaboraciones. Aplicando una estrategia integrada y multidisciplinaria, el BNI explora la organización estructural y funcional del cerebro en condiciones normales y patológicas, tanto a nivel de organismos completos como a nivel celular; capacita una nueva generación de investigadores y clínicos; produce investigación clínica de alto nivel, y transfiere sus resultados a la sociedad mediante nuevos enfoques diagnósticos y terapéuticos para mejorar la calidad de vida de los pacientes neurológicos o con trastornos psiquiátricos.

En nuestra experiencia, una vez ideado o encontrado un algoritmo descrito para abordar un problema de imágenes, existe una brecha importante entre la teoría y su implementación en imágenes biomédicas. Esto se debe tanto a consideraciones numéricas y de calibración de parámetros como al tipo de imágenes, que suelen ser más complejas y requieren márgenes de error muy acotados. Los algoritmos probados con éxito en *benchmarks* sintéticos resultan insatisfactorios en este contexto, y hemos debido desarrollar *benchmarks* sintéticos con estructuras biológicas representativas, para poder escoger los métodos numéricos a usar, los parámetros óptimos de cada algoritmo, entender sus capacidades y limitaciones en cada escenario y proponer mejoras [13,15,16]. Esto mismo nos ha servido para dar un marco de trabajo a los biólogos, que pueden ajustar las condiciones de adquisición de los microscopios de modo de optimizar la segmentación y la estimación de movimiento, con miras a automatizar procesos que aún son realizados en forma manual y/o que pueden ser propensos a error.

También consideramos importante la reproducibilidad de los resultados obtenidos, así como la disponibilidad de los algoritmos implementados. Es por esto que publicamos algoritmos en repositorios de domino público como ImageJ (NIH), y hemos comenzado a participar de un proyecto para publicación arbitrada de algoritmos en imágenes, acompañados de implementaciones documentadas y ejemplos de uso, y de acceso abierto al público general. Se trata de la revista electrónica Image Processing Online (IPOL), iniciada en Francia. Mediante una colaboración con el Grupo de Tratamiento de Imágenes de la Universidad de la República en Uruguay, tenemos un primer algoritmo en revisión y proyectamos seguir nuestros

Figura 5



Imágenes de súper resolución con SOFI. Se muestra el aumento de resolución obtenido para: imágenes de prueba con *quantum dots* sobre un cubreobjetos (1) a distintos niveles de súper-resolución (órdenes de cumulante), y subunidades de receptores para neurotransmisores en neuronas de hipocampo (2). Barra de escala: 5 μm. Imagen: Omar Ramírez, Felipe Santibáñez (SCIAN-Lab, Facultad de Medicina).

desarrollos futuros no sólo publicando nuestros métodos en revistas de teoría y aplicación, sino también poniéndolos a disposición de la comunidad científica y general, tanto para su evaluación y crítica como para su uso y difusión.

El volumen de datos y tiempo de procesamiento de imágenes con el equipamiento actual ya sobrepasa las capacidades de un computador personal. Los algoritmos de *skeleton* para estructuras complejas pueden tardar de horas a días en calcularse. Esto nos lleva a trabajar en estrategias para cómputo de alto rendimiento, aprovechando el nuevo clúster de computación en el National Laboratory for High Performance Computing (NL-HPC) del Centro de Modelamiento Matemático (CMM). Finalmente, apuntamos a generar una infraestructura de red de alta velocidad que conecte a las instalaciones de imágenes en la Facultad de Medicina de la Universidad de Chile (campus norte, en Independencia) con el clúster, permitiendo el desarrollo futuro de aplicaciones en telemedicina (a través de centros de análisis remotos como el recién formado Centro de Espermiogramas Digitales Asistidos por Internet (CEDAI) o el Centro de Patología Digital (CPD), biología a nivel celular y bioinformática a gran escala, contribuyendo al desarrollo de la ciencia y la salud en el país.

OTROS AUTORES DE ESTE ARTÍCULO

Carmen Gloria Lemus

Licenciada en Ciencias Biológicas de la Pontificia Universidad Católica de Chile. Actualmente realiza su tesis de Doctorado en Ciencias Biomédicas en la Facultad de Medicina Universidad de Chile, estudiando la morfogénesis del órgano parapineal en pez cebra (LEO, SCIAN-Lab, BNI).

Felipe Santibáñez

Ingeniero Civil Electrónico, Universidad de Concepción. Actualmente trabaja como asistente de investigación en SCIAN-Lab (BNI, Facultad de Medicina Universidad de Chile) en procesamiento de imágenes para SOFI y modelamiento de sistemas celulares mediante propiedades tensiles y de adhesión.

Omar Ramírez

Doctor en Ciencias Biomédicas Universidad de Chile. Actualmente realiza su postdoctorado en SCIAN-Lab (BNI, Facultad de Medicina U. de Chile). Estudia la relación entre estructura y procesos de transporte intracelular que se realizan en el retículo endoplasmático, en el contexto de función sináptica en neuronas.

Miguel Concha

Doctor en Ciencias Biomédicas Universidad de Chile. Profesor Titular del Instituto de Ciencias Biomédicas de la Facultad de Medicina, y director del Laboratorio de Estudios Ontogénicos LEO (BNI, Facultad de Medicina U. de Chile). Su trabajo se enfoca en el estudio de la genética del desarrollo, morfogénesis celular y desarrollo neuronal, con peces teléosteos (pez cebra, anual, medaka, entre otros) como modelo de desarrollo.

ACERCA DE SCIAN-LAB

SCIAN-Lab reúne a un equipo en la interfaz de computación, matemáticas e investigación biomédica, y colabora estrechamente con laboratorios del Centro de Modelamiento Matemático (CMM) y del DCC, representado por la profesora Nancy Hitschfeld y tesistas de pre y postgrado, formando la Advanced Imaging and Bioinformatics Initiative (AI·BI). En años recientes esta red se ha constituido en una iniciativa de colaboración única en Chile y América del Sur, combinando experticias en biología molecular, neurociencia, biología del desarrollo, neuropatología, reproducción humana, computación, telemedicina y análisis de imágenes in vivo para abordar problemas de ciencia básica y salud, con investigación y desarrollo de nivel internacional.BITS

REFERENCIAS

[1] C. Vonesch and F. Aguet and J-L Vonesch and M. Unser. The Colored Revolution of Bioimaging. IEEE Signal Processing Magazine, 23(3): 20-31, 2006.

[2] Editorial: Microscopic Marvels. Nature 7247(459): 615, 2009.

[3] Tim McInerney and Dimitri Terzopoulos. Deformable models in Medical Image Analysis: a survey. Medical Image Analysis, 1(2):91-108, 1996.

[4] Proyecto FONDECYT 1090246, Partial differential equations for 3D photon denoising, optical flow and adjacent active surface models for high throughput in vivo spinning disk microscopy. 2009-2012. Investigador responsable: Steffen Härtel, co-investigador: Nancy Hitschfeld Kahler.

[5] Proyecto FONDECYT 1120579, Fast Computational Schemes for the Analysis of Morpho-Topological Data from High Throughput Microscopy. 2012-2015. Investigador responsable: Steffen Härtel, co-investigador: Nancy Hitschfeld Kahler.

[6] Pascal Haffter, Michael Granato, Michael Brand, Mary C. Mullins, Matthias Hammerschmidt, Donald A. Kane, Jörg Odenthal, Fredericus J. M. van Eeden, Yun-Jin Jiang, Carl-Philipp Heisenberg, Robert N. Kelsh, Makoto Furutani-Seiki, Elisabeth Vogelsang, Dirk Beuchle, Ursula Schach, Cosima Fabian, Christiane Nüsslein-Volhard. The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. Development 123, 1–36, 1996.

[7] Jorge Jara. Contornos activos para segmentación en imágenes digitales. Revista Bits de Ciencia 5:68-73, 2011.

[8] Luis Felipe Olmos. Segmentación de Objetos en Imágenes de Microscopía mediante Contornos Activos con Adaptabilidad Topológica. Memoria para optar al título de Ingeniero Civil en Computación e Ingeniero Civil Matemático, F-Med/FCFM, Universidad de Chile, 2009.

[9] William E. Lorensen, Harvey E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. SIGGRAPH Computer Graphics, 21(4): 163-169, 1987.

[10] Marco Attene. A lightweight approach to repairing digitized polygon meshes. The Visual Computer, 26(11): 1393-1406, 2010.

[11] Nicolás Silva. Modelamiento del crecimiento de árboles usando mallas de superficie. Memoria para optar al título de Ingeniero Civil en Computación. Universidad de Chile, 2007.

[12] Oscar Kin-Chung Au, Chiew-Lan Tai, Hung-Kuo Chu, Daniel Cohen-Or, Tong-Yee Lee. Skeleton extraction by mesh contraction. ACM SIGGRAPH 2008 papers: 44:1-44:10, 2008.

[13] Liliana Alcayaga. Generación de skeletons a partir de mallas de superficie. Memoria para optar al título de Ingeniero Civil en Computación, F-Med/FCFM, Universidad de Chile, 2012.

[14] Thomas Dertinger, Ryan Colyer, G. Iyer, Shimon Weiss, Jörg Enderlein. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). Proceedings of the National Academy of Sciences of the United States of America, 106(52):22287-22292, 2009.

[15] José Delpiano, Jorge Jara, Jan Scheer, Omar Ramírez, Javier Ruiz del Solar, Steffen Härtel. Performance of optical flow techniques for motion analysis of fluorescent point signals in confocal microscopy. Machine Vision and Applications, 23(4): 675-689, 2012.

[16] María Laura Fanani, Steffen Härtel, Luisina De Tullio, Jorge Jara, Felipe Olmos, Rafael Oliveira, Bruno Maggio. The action of sphingomyelinase in lipid monolayers as revealed by microscopic image analysis. Biochimica et Biophysica Acta (BBA) - Biomembranes 1798(7): 1309-1323, 2010.

ENLACES

Advanced Imaging and Bio-Informatics Initiative (AI·BI) www.aibi.cl

Centro de Espermiogramas Digitales Asistidos por Internet (CEDAI SpA) www.cedai.cl

Centro de Modelamiento Matemático (CMM) www.cmm.uchile.cl

Centro de Patología Digital (CPD) www.microscopiavirtual.cl

Image Processing Online (IPOL) www.ipol.im

Instituto Milenio de Neurociencia Biomédica (BNI, Iniciativa Científica Milenio) www.bni.cl

Laboratorio de Análisis de Imágenes Científicas (SCIAN-Lab) www.scian.cl

Laboratorio de Estudios Ontogénicos (LEO) www.ontogenesis.cl

National Laboratory for High Performance Computing (NL-HPC) www.nlhpc.cl

Procesando más allá de lo visible

Domingo Mery
Grupo de Inteligencia de Máquina (GRIMA) Departamento de Ciencia de la Computación Pontificia Universidad Católica de Chile. Profesor Titular DCC, Pontificia Universidad Católica de Chile. Ph.D. Electrical Engineering, Technical University of Berlin (2001), MSc Electrical Engineering, Karlsruhe Institute of Technology (1992), Ingeniero Electrónico, Universidad Nacional de Ingeniería, Perú (1989). Líneas de especialización: Visión por Computador, Reconocimiento de Patrones, Procesamiento de Imágenes, Aplicaciones Industriales.
http://dmery.ing.puc.cl

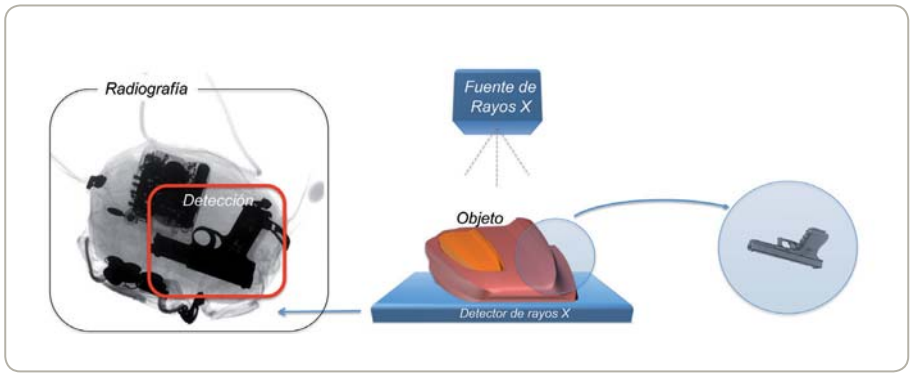
En algunas aplicaciones de visión por computador –relacionadas por ejemplo con la seguridad de personas o con la calidad de productos–, el análisis de las imágenes no puede llevarse a cabo usando fotografías digitales convencionales. En estos casos se requiere del uso de imágenes obtenidas fuera del espectro visible, como los rayos X, que son capaces de mostrar lo que ocurre dentro de un objeto. En el Grupo de Inteligencia de Máquina (GRIMA) hemos desarrollado diversas aplicaciones de visión por computador con rayos X en la identificación de objetos peligrosos en equipajes, detección de objetos no deseados en alimentos y reconocimiento de fallas internas en materiales. En este artículo revisamos los principales logros en investigaciones y desarrollos que hemos llevado a cabo en la última década.

INTRODUCCIÓN

Es sabido por todos nosotros que cuando viajamos en avión nuestro equipaje debe ser revisado exhaustivamente, con el fin de controlar si es que llevamos algún objeto prohibido. Estamos acostumbrados, por ejemplo, a que cada bolso de mano u otro objeto personal debe pasar por una cabina de rayos X en la que se obtiene una radiografía para que un operador humano inspeccione su contenido y detecte, si es que los hay, objetos que puedan atentar contra la seguridad del vuelo, como se ilustra en la Figura 1.

Como una fotografía a color de la mochila de la Figura 1 no nos proporcionaría información alguna para inspeccionar su contenido, es necesario contar con imágenes

Figura 1



Esquema del proceso de inspección de un objeto (mochila) usando rayos X.

que nos muestren lo que está más allá del espectro visible. Necesitamos entonces sistemas de rayos X que sean capaces de penetrar los objetos y que nos proporcionen una imagen de lo que está en su interior. Un sistema de visión por computador de rayos X consta de un equipo encargado de la adquisición de imágenes (fuente y detector de rayos X), muchas veces un manipulador robótico que posicione el objeto de análisis (o el equipo mismo de adquisición) en la posición adecuada, y un computador donde corran algoritmos de visión por computador encargados de realizar la detección o reconocimiento de manera automática. Estos sistemas son utilizados en un sinnúmero de aplicaciones relacionadas con la seguridad de personas y con la calidad de productos o materiales.

Los algoritmos se basan en técnicas de procesamiento de imágenes (relacionadas principalmente con el mejoramiento de la calidad de las imágenes obtenidas y con la segmentación que separe en la imagen los objetos de interés); reconocimiento de patrones (extracción de características que nos proporcionen información discriminativa de los objetos a detectar, y clasificadores); visión por computador (uso de múltiples vistas, visión activa y análisis moderno de gran cantidad de imágenes) y aprendizaje de máquina (estrategias de entrenamiento y selección de datos). En cada uno de estos problemas es necesario contar con una base de datos de imágenes con casos representativos que pueden ser empleados para el aprendizaje y las pruebas de nuestros algoritmos. Con este fin hemos creado una base de datos pública de imágenes de rayos X en la que hemos recopilado las principales radiografías que hemos utilizado en nuestras investigaciones disponibles en nuestra página web (grima.ing.puc.cl).

En este artículo mostraremos investigaciones y desarrollos llevados a cabo en la última década por el Grupo de Inteligencia de Máquina (GRIMA) del Departamento de Ciencia de la Computación de la Universidad Católica, que hoy cuenta con un laboratorio único en Chile, dotado de un poderoso sistema de visión por computador de rayos X, en el que han trabajado estudiantes de postgrado que realizan investigación en el área.

APLICACIONES

Dentro de los problemas que hemos investigado se encuentran la revisión de equipaje (detección de algunos objetos prohibidos en bolsos), análisis de alimentos (detección de espinas en salmones y truchas, así como control de calidad de kiwis y arándanos) e inspección de materiales (detección de fallas en soldaduras y en piezas de automóviles). A continuación haremos una breve descripción de ellos.

Revisión de equipajes

La revisión de equipajes se realiza hoy en día de manera manual. Nuestra investigación se ha concentrado en tratar de ofrecer algún grado de automatización a este proceso. La dificultad de este problema radica principalmente en el inmenso número de variantes en la apariencia que pueda tener una categoría de objetos prohibidos como muestra la Figura 2. En los dos últimos años hemos obtenido algunos resultados preliminares en la detección de hojas de afeitar y en la detección de pistolas, que a continuación describiremos con mayor detalle.

En el primer problema, creamos una base de datos de descriptores SIFT de radiografías de hojas de afeitar en distintas posiciones. La idea principal de la estrategia de detección, se basa en buscar en la radiografía de un objeto a revisar (por ejemplo una mochila) descriptores SIFT que sean similares a los de la base de datos y que estuvieran agrupados en una zona compacta de la imagen. A partir de la pose de la hoja de afeitar de la base de datos cuyos descriptores SIFT sean los más similares a los obtenidos de la imagen

Figura 2

Alta variabilidad en la apariencia de pistolas y cuchillos (imágenes obtenidas de Google Images al buscar estos ítems).

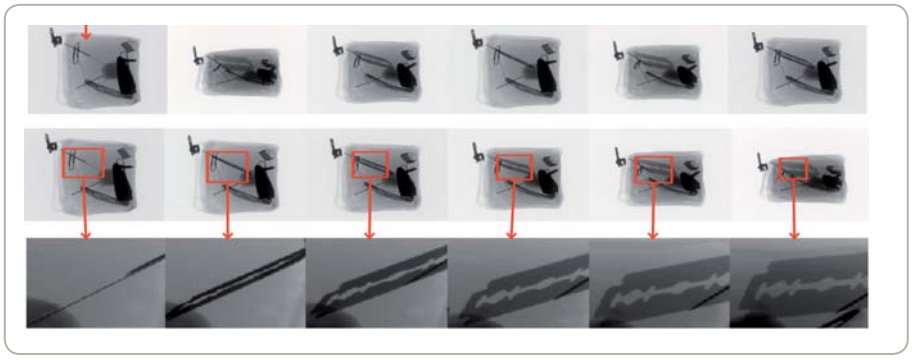
a revisar, se utiliza un manipulador robótico que gire el objeto con el fin de obtener una vista frontal de la hoja de afeitar detectada. De esta manera, basándonos en estrategias de visión activa en los que dirigimos al manipulador para obtener una mejor vista, es posible obtener una segunda, tercera y hasta una cuarta vista del objeto a revisar para mejorar y corroborar la detección (Riffo & Mery, 2012).

En el segundo problema, para poder disminuir la complejidad del problema de la alta variabilidad en la apariencia de las pistolas, se desarrolló una estrategia que buscara solamente los gatillos, ya que todas las pistolas cuentan con uno de ellos. Con este fin se entrenó un detector de gatillos, basándose en características geométricas obtenidas de la parte *(bounding-box)* de la imagen donde están los gatillos de diversas fotografías de pistolas obtenidas de Internet. Utilizando la estrategia de ventana deslizante *(sliding-windows)* se barre entonces la radiografía del objeto a inspeccionar (por ejemplo la mochila) detectando posibles gatillos. Luego, se usa una estrategia de múltiples vistas en la que se obtienen diversas vistas del objeto y mediante algoritmos de correspondencia en las vistas se corroboran las detecciones correctas y se eliminan las falsas alarmas, donde las detecciones no estén presentes en las siguientes imágenes en las posiciones donde según el modelo geométrico del movimiento deberían estar (Mondragón, 2012).

También se ha trabajado en diseñar una estrategia general para la detección de objetos dentro de equipajes, con algoritmos de segmentación que separen objetos de interés y que puedan ser seguidos a lo largo de una secuencia de imágenes (Mery, 2011a). Una vez obtenida la secuencia es posible analizar las múltiples vistas, desarrollando un algoritmo de clasificación que tome como entrada las apariencias desde distintos puntos de vista del objeto, y que luego realice una tomografía computada del objeto a seguir. Algunos resultados preliminares se muestran en las Figuras 3, 4 y 5.

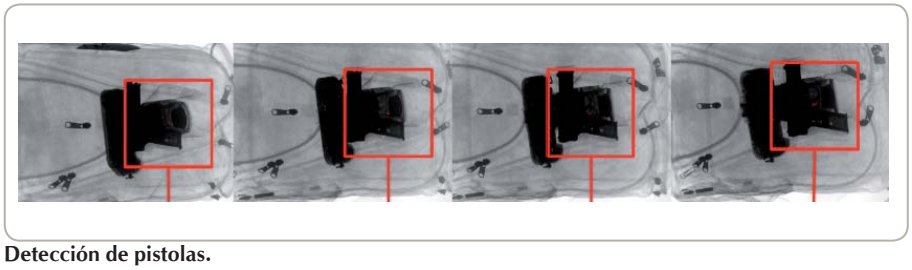
En esta área de investigación se han obtenido resultados preliminares satisfactorios: en las categorías "hojas de afeitar" y "pistolas" se alcanzaron índices de desempeño de un 90% aproximadamente (en 130 y 27 experimentos respectivamente), sin embargo, resulta evidente que se necesita una validación experimental mayor, y ampliar considerablemente el número de categorías a detectar. Nuestra aspiración es poder contar con un equipo semiautomático que asista y alivie la tediosa inspección humana indicándole a los operadores cuáles objetos pueden ser sospechosos, y en algunos casos mostrar una reconstrucción 3D como en la Figura 5, que le ayude al operador a tomar una decisión y así aumentar los índices de desempeño en la detección de objetos prohibidos.

Figura 3



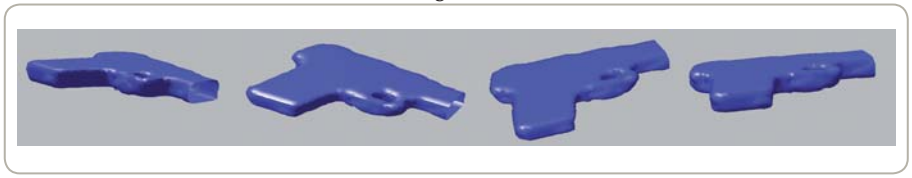
Detección de hojas de afeitar.

Figura 4



Detección de pistolas.

Figura 5



Reconstrucción 3D de pistola detectada.

Análisis de alimentos

En el análisis de alimentos hemos desarrollado una metodología general basada en reconocimiento de patrones en la que se extrae –en una etapa de entrenamiento– un número inmenso de características, por ejemplo del orden de tres mil, y que luego de manera automática se seleccionan las características y el clasificador que obtengan el mejor desempeño en una estrategia supervisada (Mery *et al.*, 2012). Basándonos en esta metodología, que se ilustra en la Figura 6, hemos podido desarrollar un proyecto Fondef[1] con la industria salmonera, cuyo fin fue poder detectar de manera automática las espinas presentes en los filetes de trucha y salmón. Los resultados entregaron índices de desempeño superiores al 95% (Mery *et al.*, 2011). Asimismo, hemos trabajado en el control de calidad de kiwis (Mondragón *et al.*, 2011) y arándanos (Leiva *et al.*, 2011) usando rayos X con resultados similares.

1 SalmonX: Sistema de Rayos X para inspección automática de filetes de salmón. Fondef Nº D07I-1080 (2009-2010).

Figura 6

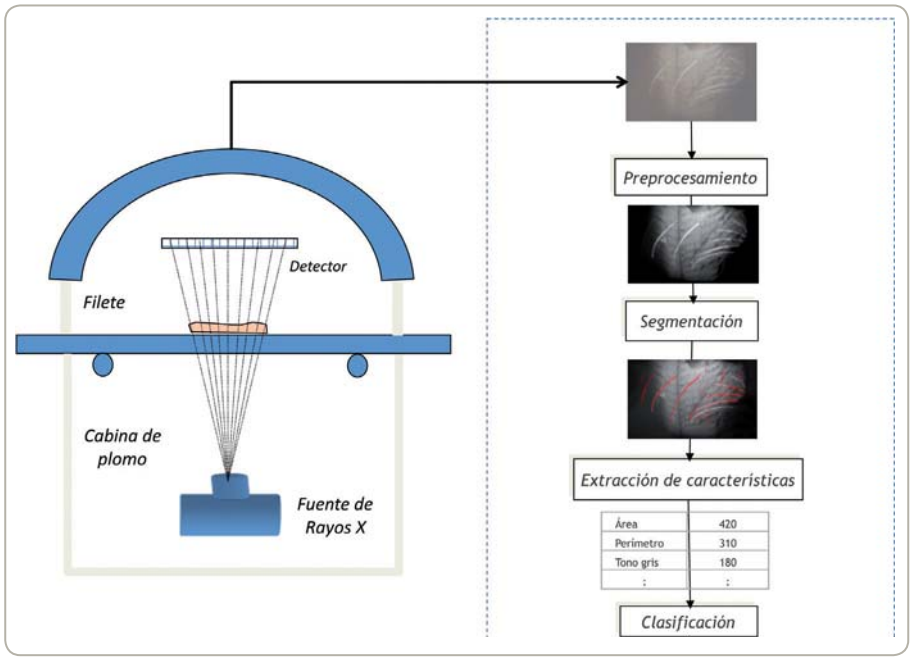


Diagrama de bloques de la detección automática de espinas en filetes de salmón.

Inspección de materiales

Dos aplicaciones en las que el análisis de imágenes radiográficas juega un rol primordial son inspección de soldaduras (Silva & Mery, 2007) e inspección de partes metálicas de automóviles (Mery, 2006). En estos ejemplos industriales, en que las imágenes de rayos X son utilizadas para determinar si estos objetos presentan fallas internas, la inspección debe observar criterios de seguridad sumamente rigurosos. En este caso, un control de calidad errado que no detecte una fisura o una burbuja interna en una soldadura de una tubería que suministre gas a una ciudad, o en una llanta de automóvil, puede ocasionar serios accidentes como se ilustra en la Figura 7. Cabe mencionar, que este tipo de inspección no es por muestreo, ya que es necesario examinar el 100% de los objetos.

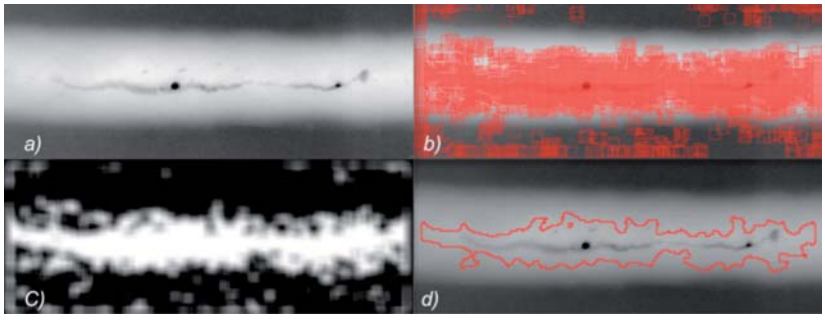
En la inspección de soldaduras hemos desarrollado estrategias clásicas como las explicadas en la sección anterior basándonos en características de textura (Berty & Mery, 2003) y también técnicas más modernas en las que hemos aplicado metodologías de ventanas deslizantes ampliamente usadas por la comunidad de visión por computador en la detección de caras, por ejemplo. En este caso, como se logra apreciar en la Figura 8, hemos obtenido desempeños del orden del 92% (Mery, 2011b).

En la inspección de piezas de automóviles, se trabaja con múltiples imágenes tomadas desde distintos puntos de vista gracias al posicionamiento adecuado de la pieza a través de un manipulador robótico (Mery & Filbert, 2002). En estos algoritmos la modelación geométrica de la proyección 3D → 2D juega un rol muy importante debido a que es necesario encontrar correspondencias en las múltiples vistas (Mery, 2003). Para validar los algoritmos de detección en variados casos, se cuenta con herramientas de simulación de fallas (Mery 2001; Mery et al., 2005) que gracias a la modelación de superficies poligonales de fallas, se logra sobreponer zonas más claras, que simulan las fallas, en radiografías reales. En los últimos años, los algoritmos de detección han logrado índices de desempeño suficientemente altos (Pieringer & Mery, 2010; Carrasco & Mery, 2011), lo que ha llevado a que hoy en día este tipo de industria esté completamente automatizada.

Figura 7

Accidente en una llanta producto de una inspección deficiente

Figura 8



Detección de fallas en una soldadura: a) radiografía original, b) ventanas detectadas, c) ventanas sobrepuestas, d) detección final.

CONCLUSIONES

El Grupo de Inteligencia de Máquina (GRIMA) ha desarrollado diversas aplicaciones de visión por computador con rayos X en la identificación de objetos peligrosos en equipajes, detección de objetos no deseados en alimentos y reconocimiento de fallas internas en materiales. En este artículo revisamos los principales logros en cada una de estas aplicaciones. Observamos que algunas de ellas (partes de automóviles, detección de espinas en salmones) están bastante desarrolladas, alcanzando niveles de automatización aceptables; otras aplicaciones (soldaduras) han obtenido desempeños altos pero aún no es posible automatizar el proceso completamente debido a que los índices de desempeño deben bordear el 100%, ya que está en juego la seguridad de las personas; y otras aplicaciones (revisión de equipaje) son aún un problema sin resolver debido principalmente a la alta variabilidad de categorías a detectar y a que las imágenes presentan un grado de complejidad elevado con bajo contraste y con oclusión.

Gran parte de los adelantos en esta comunidad se deben a poder compartir tanto el código de los algoritmos, así como las imágenes utilizadas. Por esta razón, una contribución de nuestro grupo ha sido crear y actualizar un Toolbox en Matlab, con los algoritmos implementados y probados, disponible en nuestra página web para fines no comerciales. Asimismo, las imágenes empleadas también se pueden descargar de nuestro sitio.

AGRADECIMIENTOS

El autor desea agradecer por su apoyo a los proyectos Fondecyt 1100830 y 1040210, Fondef D07I-1080, LACCIR R0308LAC003, Anillos ACT32 y CopecUC; y también a los miembros de GRIMA, quienes han colaborado en el desarrollo de esta línea de investigación, especialmente a Miguel Carrasco, Gabriel Leiva, Germán Mondragón, Christian Pieringer, Vladimir Riffo, Álvaro Soto e Irene Zuccar.BITS

BIBLIOGRAFÍA

Carrasco, M.; Mery, D. (2011): Automatic Multiple View Inspection using Geometrical Tracking and Feature Analysis in Aluminum Wheels. Machine Vision and Applications, 22:157-170.

Leiva, G.; Mondragón, G.; Mery, D.; Aguilera, J.M (2011): The automatic sorting using image processing improves postharvest blueberries storage quality. Proceedings of International Congress on Engineering and Food.

Mery, D. (2001): Flaw simulation in castings inspection by radioscopy, Insight, 43(10):664-668.

Mery, D.; Filbert, D. (2002): Automated Flaw Detection in Aluminum Castings Based on The Tracking of Potential Defects in a Radioscopic Image Sequence. IEEE Transactions on Robotics and Automation, 18(6):890-901.

Mery, D.; Berti, .M.A. (2003): Automatic Detection of Welding Defects using Texture Features. Insight, 45(10):676-681.

Mery, D. (2003): Explicit Geometric Model of a Radioscopic Imaging System. NDT & E International, 36(8):587-599.

Mery, D.; Hahn, D.; Hitschfeld N. (2005): Simulation of Defects in Aluminium Castings Using CAD Models of Flaws and Real X-ray Images. Insight, 47(10):618-624. (Ron Halmshaw Award 2005).

Mery, D. (2006): Automated Radioscopic Testing of Aluminium die Castings. Materials Evaluation, 64(2):135-143.

Mery, D. (2011a): Automated Detection in Complex Objects using a Tracking Algorithm in Multiple X-ray Views. Proceedings of the 8th IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS 2011), in Conjunction with Computer Vision and Pattern Recognition (CVPR 2011). Colorado Springs, USA.

Mery, D. (2011b): Automated Detection of Welding Discontinuities without Segmentation. Materials Evaluation, June-2011:657-663.

Mery, D.; Lillo, I.; Loebel, H.; Riffo, V.; Soto, A.; Cipriano, A.; Aguilera, J.M. (2011): Automated Fish Bone Detection using X-ray Testing. Journal of Food Engineering, 105(2011):485-492.

Mery, D.; Pedreschi, F.; Soto, A. (2012): Automated Design of a Computer Vision System for Visual Food Quality Evaluation" International Journal of Food and Bioprocess Technology (in Press doi 10.1007/s11947-012-0934-2).

Mondragón, G.; Leiva, G.; Aguilera, J.M.; Mery, D. (2011): Automated detection of softening and hard columella in kiwifruits during postharvest using X-ray testing. Proceedings of International Congress on Engineering and Food.

Mondragón, G. (2012): Metodología para el desarrollo de un clasificador de múltiples vistas de radiografías. Aplicación para la búsqueda de armas pequeñas dentro de equipajes. Pontificia Universidad Católica de Chile.

Pieringer, C.; Mery, D. (2010): Flaw Detection in Aluminum Die Castings Using Simultaneous Combination of Multiple Views. Insight, 52(10):548-552.

Riffo, V.; Mery, D. (2012): Active X-ray testing of Complex Objects. Insight, 54(1):28-35.

Silva, R.; Mery, D. (2007): State-of-the-Art of Weld Seam Inspection by Radiographic Testing: Part I – Image Processing. Materials Evaluation, 65(6):643-647.

Silva, R.; Mery, D. (2007): State-of-the-Art of Weld Seam Inspection by Radiographic Testing: Part II – Pattern Recognition. Materials Evaluation, 65(9):833-838.

Reconocimiento visual con técnicas de aprendizaje de máquina

Pablo Espinace
Doctor en Ciencias de la Ingeniería y Magíster en Ciencias de la Ingeniería de la Pontificia Universidad Católica de Chile (2011). Actualmente es investigador en el Departamento de Ciencia de la Computación de la Escuela de Ingeniería de la Pontificia Universidad Católica de Chile. Sus intereses de investigación incluyen Robótica Móvil, Inteligencia de Máquina y Visión por Computador.
pespinac@ing.puc.cl

Billy Peralta
Candidato a Doctor en Ciencias de la Ingeniería de la Pontificia Universidad Católica de Chile. Magíster en Ciencias de la Ingeniería (2007). Sus intereses de investigación principales son Inteligencia de Máquina y Visión por Computador.
bmperalt@uc.cl

Álvaro Soto
Profesor Asociado del Departamento de Ciencia de la Computación de la Escuela de Ingeniería de la Pontificia Universidad Católica de Chile. Ph.D. en Ciencias de la Computación Carnegie Mellon University (2002); Magíster en Ingeniería Eléctrica y Computación en Louisiana State University (1997). Sus intereses principales en investigación son: Aprendizaje de Máquina, Robótica Cognitiva y Reconocimiento Visual.
asoto@ing.puc.cl

La gran versatilidad de nuestro propio sistema de percepción visual, es un claro ejemplo de la gran relevancia de poder contar con sistemas artificiales de reconocimiento visual que faciliten la operación de robots y máquinas embebidas en nuestros ambientes cotidianos. Sin embargo, complejidades del mundo visual como cambios de pose, escala o condiciones de iluminación, así como situaciones de oclusión o cambios en la configuración de objetos deformables, han resultados desafíos de envergadura mayor, que han complicado el desarrollo de posibles soluciones.

Afortunadamente, la exitosa reciente aplicación de técnicas de aprendizaje de máquina al área de reconocimiento visual ha abierto nuevas puertas que han llenado de optimismo a esta área [1,2,3]. La Figura 1 ejemplifica el aporte medular que ha ofrecido el aprendizaje de máquina al reconocimiento visual. En la Figura, pese a la mala calidad de la imagen, nuestro sistema visual es capaz de distinguir que se trata de una mesa con sillas a su alrededor, sobre la cual hay un objeto decorativo. En forma notable, mediante la integración de conocimiento aprendido en nuestro largo interactuar con el mundo natural, nuestro sistema visual es capaz de realizar una inferencia que va mucho más allá de la información contenida en los ruidosos píxeles de esta imagen. Esta capacidad, de ir más allá de la información en una imagen determinada, es el ingrediente clave que ha aportado la inteligencia de máquina al reconocimiento visual. De esta manera, los algoritmos de aprendizaje de máquina están permitiendo extraer desde miles o millones de imágenes, patrones de apariencia visual y relaciones contextuales relevantes, las cuales luego pueden ser usadas para reconocer elementos en una imagen particular. Este tipo de aprendizaje

visual está permitiendo por primera vez crear sistemas de reconocimiento visual capaces de operar exitosamente en ambientes naturales [1,4].

En este artículo, mediante dos casos ilustrativos, queremos mostrar algunos de los aportes realizados por nuestro Grupo de Investigación en Inteligencia de Máquina, GRIMA [5], en el área de reconocimiento visual. Nuestro primer ejemplo corresponde a un sistema de reconocimiento de escenas de ambientes interiores que desarrollamos para nuestro robot móvil. Este sistema utiliza relaciones contextuales entre objetos y escenas, de manera que si nuestro robot encuentra una sala con refrigerador y microondas pensará que se encuentra en una cocina. Interesantemente, nunca entregamos esta información directamente a nuestro modelo, sino que nuestro sistema es capaz de inferir (aprender) este tipo de relaciones por sí mismo, analizando la información textual contenida en millones de imágenes del popular sitio web Flickr. En nuestro segundo ejemplo, exploramos la relación inversa entre objetos y escenas, es decir, cómo conocimiento holístico sobre la escena condiciona las relaciones contextuales entre objetos. De esta manera, podemos construir un modelo capaz de distinguir que en una escena de parque es altamente probable ver personas caminando al lado de un perro, pero que esa misma relación es altamente improbable para el caso de un edificio de oficinas.

RECONOCIMIENTO DE ESCENAS A TRAVÉS DE DETECCIÓN DE OBJETOS

En el ámbito de reconocimiento de escenas, los primeros métodos buscaban extraer características y categorizar la escena en la cual se tomó una imagen mediante representaciones "holísticas", esto es, analizando la información contenida dentro de la imagen como un todo [6,7]. Luego, algunos trabajos buscaron utilizar representaciones intermedias [8,9,10], en algunos casos incluyendo información espacial [11]. Todos estos métodos tuvieron relativo éxito, especialmente en categorización de imágenes de exterior, sin embargo, su desempeño en escenas de interior resulta ser muy pobre. Las razones para este bajo rendimiento saltan a la vista: una escena de interior (o habitación) puede ser muy similar, o incluso igual a otra, excepto por los objetos que contiene, los cuales son muy diversos en apariencia y posibles poses.

A modo de ejemplo, consideremos una habitación de 3x3 metros inicialmente vacía. Sin objetos en su interior, muy difícilmente se podrá distinguir de qué tipo de habitación se trata. Ahora, si en esta habitación ponemos una silla y un escritorio, pasa de inmediato a tomar la forma de una oficina. Luego, si sacamos la silla y el escritorio, y en su lugar ponemos una cama y una lámpara, la habitación tomará la forma de un dormitorio. Como se ve, en términos globales la habitación es la misma, pero son los objetos dentro de ella los que hacen la diferencia.

Considerando lo anterior, diversos trabajos han comenzado a utilizar la detección de objetos como parte central de la detección de escenas [12,13,14]. La idea principal es utilizar los objetos como partes componentes de una escena, además de aprovechar la configuración espacial de estos para mejorar el rendimiento y los resultados obtenidos. A continuación presentamos el trabajo realizado al respecto en una investigación conjunta entre GRIMA y el Laboratorio de Ciencia de la Computación e Inteligencia Artificial del Instituto Tecnológico de Massachusetts (CSAIL-MIT).

Figura 1

A pesar de lo borroso de la imagen, nuestro sistema visual es capaz de inferir información a partir de conocimiento aprendido previamente.

Modelamiento matemático del problema

El objetivo de nuestro método es encontrar la distribución de probabilidad de los valores que puede tomar una variable ξ, que representa las distintas etiquetas que se pueden asignar a una escena. En el caso de escenas de exterior estas etiquetas podrían ser bosque, playa, ciudad, etc., mientras que en escenas de interior podrían ser oficina, dormitorio, cocina, etc. Como punto de

partida, se tienen las características visuales que se pueden extraer directamente desde los píxeles de la imagen correspondiente, bajo un método de ventana deslizante. Así, si asumimos que se tienen w_L ventanas, a cada una de las cuales se le extrae un conjunto $\vec{f}$ de características, tendríamos un conjunto total de $\vec{f}_{1:w_L}$ características extraídas a todas las ventanas. De esta forma, lo que buscamos es la distribución de probabilidad de ξ dada la información $\vec{f}_{1:w_L}$, esto es, $p(\xi \mid \vec{f}_{1:w_L})$.

El hecho de ocupar un método de ventana deslizante nos permite obtener características locales en diversas partes de la imagen, en lugar de características globales. Dado que nuestro objetivo es reconocer una escena a través de los objetos presentes en ella, nuestro método implementa un clasificador que determina la probabilidad de que cada uno de S objetos distintos esté presente en cada una de las ventanas, a partir del conjunto de características $\vec{f}$ extraído a cada ventana. Se representa la salida de este clasificador para el conjunto de w_L ventanas como $c_{1:w_L}$. De esta forma, las características locales extraídas en diversas partes de la imagen (ventanas), se transforman en probabilidades de que distintos objetos estén presentes en la escena donde fue tomada la imagen.

Teniendo información probabilística sobre los objetos que están presentes en la escena donde fue tomada la imagen, se debe establecer una relación entre el conjunto de objetos presentes y la etiqueta a asignar a esta escena. De esta manera se podrá saber cuál es la escena más probable dados los objetos presentes en ella. Así, nuestro método implementa una relación contextual, la cual determina a partir de información de frecuencia la probabilidad de que una escena tome cada uno de los valores posibles para sus etiquetas. Se representa la presencia o ausencia de cada uno de los objetos disponibles a través de la variable conjunta $o_{1:s}$, la cual tiene un valor por cada categoría de objeto.

Las variables asociadas a las salidas de los clasificadores de objetos, $c_{1:w_L}$, y la combinación de objetos presente en la escena, $o_{1:s}$, son incorporadas al cálculo de $p(\xi \mid \vec{f}_{1:w_L})$ a través de la Ley de Probabilidad Total:

$$p(\xi \mid \vec{f}_{1:w_L}) = \sum_{o_{1:s}} \sum_{c_{1:w_L}} p(\xi \mid o_{1:s}, c_{1:w_L}, \vec{f}_{1:w_L}) p(o_{1:s}, c_{1:w_L} \mid \vec{f}_{1:w_L})$$

Simplificando y aplicando la Ley de Probabilidad Conjunta al segundo término, obtenemos:

$$p(\xi \mid \vec{f}_{1:w_L}) = \sum_{o_{1:s}} \sum_{c_{1:w_L}} p(\xi \mid o_{1:s}) p(o_{1:s} \mid c_{1:w_L}) p(c_{1:w_L} \mid \vec{f}_{1:w_L})$$

De esta manera, se construye nuestro método de reconocimiento de escenas a través de detección de objetos en base a tres términos principales:

- $p(c_{1:w_L} \mid \vec{f}_{1:w_L})$, que transforma las características extraídas en las ventanas a salidas de un clasificador de objetos.
- $p(o_{1:s} \mid c_{1:w_L})$, que representa la confianza que se tiene en el clasificador de objetos utilizado.
- $p(\xi \mid o_{1:s})$, que relaciona una combinación de objetos encontrados a través de la clasificación a una escena.

Podemos ver en estos términos que las variables incorporadas funcionan como intermediarios entre las características extraídas a partir de la información de los píxeles de la imagen, de bajo nivel semántico, y la escena donde fue tomada la imagen, de alto nivel semántico. Como se verá en la siguiente sección, el cálculo de los términos expuestos anteriormente requiere de información externa a la imagen misma desde donde se extrae la información de los píxeles, requiriéndose ir a buscar información más allá de la propia imagen.

Adicionalmente, nuestro modelo agrega características asociadas a información tridimensional extraída a través de un sensor de profundidad. Matemáticamente, esta información tridimensional se representa a través de un conjunto de características tridimensionales $\vec{d}$, extraídas en cada ventana las que se agregan a las características visuales ya disponibles. Así, se incorpora el conjunto de características tridimensionales extraídas en todas las ventanas $\vec{d}_{1:w_L}$, al término asociado al clasificador de objetos:

$$p(c_{1:w_L} \mid \vec{f}_{1:w_L}) \rightarrow p(c_{1:w_L} \mid \vec{f}_{1:w_L}, \vec{d}_{1:w_L})$$

Como se verá a continuación, la incorporación de información tridimensional no sólo permite tener información más rica de la imagen en cuestión, sino que también permite implementar nuestro modelo de manera eficiente.

Implementación

Como se vio en la sección anterior, nuestro modelo depende de tres términos principales. A continuación, veremos cómo se implementa cada uno de estos tres términos para lograr el reconocimiento de una escena a través de la detección de los objetos presentes en ella.

El término $(c_{1:w_L} \mid \vec{f}_{1:w_L}, \vec{d}_{1:w_L})$ relaciona las características extraídas de la información visual y 3D a la salida de un clasificador de objetos. En nuestro caso, y de modo de aprovechar la información tridimensional de la mejor forma posible, se usa esta información como un primer filtro, construyendo un foco de atención que define sectores de la imagen donde luego se utilizará un clasificador con las características visuales. La Figura 2 muestra los resultados de este foco de atención aplicado al caso de la detección de un monitor de computador.

Para el caso de la información tridimensional, se construye un modelo en base a ejemplos de cada objeto, el que define distintas características geométricas sobre una categoría de objetos, por ejemplo tamaños típicos, los cuales se utilizan para calcular la probabilidad de que un objeto esté presente en una ventana en base sólo a información tridimensional. Para el caso de la información visual, se construye por cada objeto un clasificador bajo el método de AdaBoost, similar a lo hecho en [1], usando como datos de entrenamiento

Figura 2

Ejemplo de ejecución del foco de atención con información 3D. (a) Imagen donde dos monitores son detectados. (b) Resultados del foco de atención, ejecutado como paso previo a la clasificación visual y que acota el espacio donde se buscan monitores.

imágenes etiquetadas que contienen a los objetos, extraídas de distintos sets de datos disponibles en la Web.

El término $p(o_{1:s}|c_{1:w_L})$ representa la confianza que se tiene en las salidas de la clasificación de objetos, esto es, qué tan probable es que una cierta configuración de objetos esté presente en la escena dado que la clasificación dijo que esa configuración estaba presente. Esta confianza se mide a través de probar los clasificadores con conjuntos de imágenes que contienen los objetos respectivos, distintas a las utilizadas para el entrenamiento, las que una vez más son extraídas desde distintos sets de datos disponibles en la Web.

El término $p(\xi\,|o_{1:s})$ relaciona una combinación de objetos encontrados a través de la clasificación a una escena. Esto se puede ver como una relación de contexto objeto-escena, donde se busca saber qué tan probable es la aparición conjunta de un objeto, o un conjunto de estos, con cada escena particular. Para calcular este término nuevamente se utiliza un gran conjunto de imágenes extraídas de la Web, en este caso de Flickr, con objetos y escenas etiquetadas. Luego, se calcula utilizando frecuencias la probabilidad de que una combinación de objetos esté en una escena dada, a través de analizar cuántas veces aparecen etiquetas asociadas a todos estos objetos en imágenes etiquetadas como la escena en cuestión.

Como se puede ver, para calcular cada uno de estos tres términos debemos recurrir a fuentes de datos externas a la imagen misma, lo que muestra lo indispensable que es recurrir a información que va más allá de la imagen analizada. Las técnicas de aprendizaje de máquina han resultado efectivas en esta tarea, renovando los aires en el ámbito del reconocimiento visual.

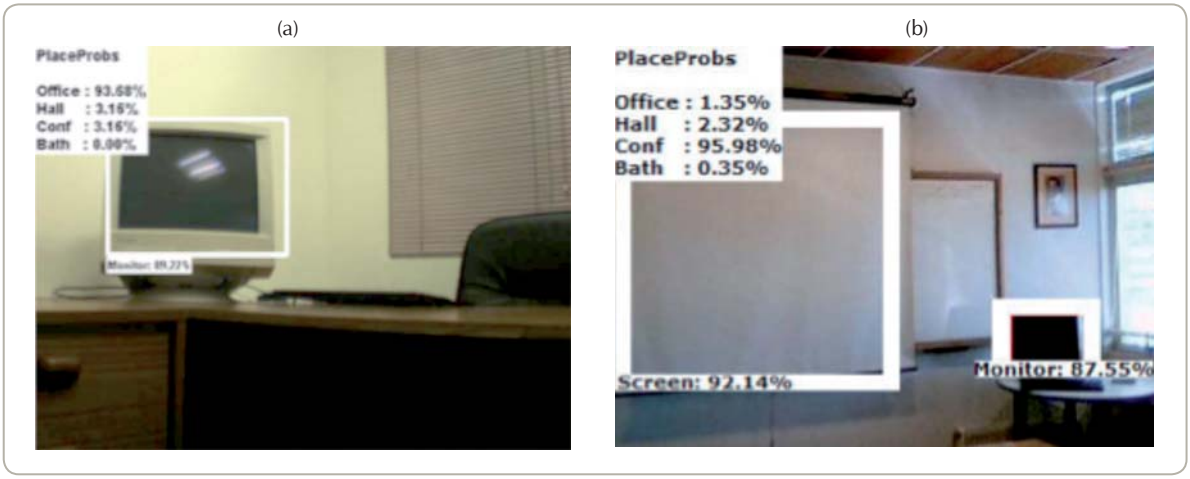
Resultados

A continuación presentamos algunos resultados obtenidos con nuestro método, en pruebas realizadas en un ambiente de oficinas, utilizando siete objetos entre los que se encontraban monitor y pantalla de proyector, y cuatro posibles escenas entre las que se encontraban oficina y sala de conferencias.

La Figura 3 muestra cómo el encontrar distintos objetos, o combinaciones de estos, define la detección de distintas escenas. En la Figura 3(a) se puede ver que al detectarse sólo un monitor en la escena, la detección más probable es una oficina, ya que de acuerdo a los datos utilizados es en las oficinas donde en mayor cantidad se encuentran monitores, en relación a las otras escenas utilizadas. Aún así, se mantiene cierta probabilidad de que sean otras las escenas reales en las que se tomó la imagen, debido a que pueden haber también monitores en ellas, además de existir cierta incerteza en la detección del monitor. Resultados similares se pueden ver en la Figura 2(a).

En la Figura 3(b) se puede ver un segundo caso en el que también se ha detectado un monitor, sin embargo en este caso, al estar acompañado de una pantalla de proyector, la escena que se detecta es una sala de conferencias, ya que la combinación de objetos hace que ésta sea la escena más probable. Nuevamente, se mantiene cierta probabilidad de que sea otra la escena, dada la incerteza en las detecciones.

Figura 3



Detecciones en ambiente de oficina. (a) Sólo un monitor es detectado, llevando al reconocimiento de una oficina. (b) Un monitor junto a una pantalla de proyector son detectados, llevando a la detección de una sala de conferencias.

La Tabla 1 muestra una matriz de confusión de las detecciones encontradas en imágenes tomadas en las distintas escenas utilizadas, haciendo una comparación de nuestro método (OM) con tres métodos alternativos que no utilizan objetos para la detección de escenas: OT-G [7], LA-SP [11] y pLSA [10]

Se puede ver que los resultados de nuestro método muestran una clara inclinación hacia la diagonal, mejorando bastante los resultados de los métodos alternativos. Aún así, existen ciertos errores, derivados principalmente de falsos positivos o falsos negativos en las detecciones de objetos, además del hecho de que no todos los objetos en una escena se pueden observar en una imagen tomada sólo en un área parcial de la misma.

RECONOCIMIENTO DE OBJETOS A TRAVÉS DE CONTEXTOS JERÁRQUICOS ADAPTIVOS

En el ámbito del reconocimiento de objetos, existen diversos métodos que han tenido éxito a través del uso de distintas características visuales y algoritmos asociados. Hasta hace algunos años, prácticamente la totalidad de estos métodos se basaba en extraer características a un objeto como un todo [1,15], mientras que recientemente se han comenzado a utilizar representaciones basadas en partes de los objetos y las relaciones espaciales entre éstas, las que han tenido gran éxito [3].

Tabla 1

	OM				OT-G			
Scene	Off.	Hall	Conf.	Bath.	Off.	Hall	Conf.	Bath.
Office	**91%**	7%	2%	0%	**83%**	4%	13%	0%
Hall	7%	**89%**	4%	0%	7%	**86%**	5%	2%
Conference	7%	7%	**86%**	0%	17%	3%	**79%**	1%
Bathroom	0%	6%	0%	**94%**	3%	5%	3%	**89%**
	LA-SP				**pLSA**			
Scene	Off.	Hall	Conf.	Bath.	Off.	Hall	Conf.	Bath.
Office	**72%**	6%	22%	0%	**88%**	4%	7%	1%
Hall	19%	**71%**	9%	1%	4%	**87%**	5%	4%
Conference	24%	6%	**67%**	3%	11%	2%	**85%**	2%
Bathroom	2%	4%	3%	**91%**	1%	4%	3%	**92%**

Comparación de nuestro método con métodos alternativos.

Una idea atractiva es la de incorporar información de contexto, esto es, información de las relaciones entre distintos objetos, o entre estos y la escena donde se encuentran, con lo cual se ha logrado mejorar el rendimiento de diversos clasificadores de objetos que no incorporan estas relaciones [16,17]. En particular, la información de contexto local, donde se cuantifica la relación entre partes de la imagen o entre ubicaciones particulares de objetos en ella, ha tenido varios casos de éxito [18,19].

Dentro de los trabajos mencionados anteriormente, uno que resulta muy relevante para modelar eficientemente la relación entre objetos es el de Choi et al. [16]. En este caso, se usa una red bayesiana en forma de árbol para representar las probabilidades condicionales entre objetos. Una gran ventaja del uso de una red bayesiana es que permite hacer inferencia de forma eficiente aprovechando la factorización del modelo gráfico. Sin embargo, una restricción de este modelo es que las interrelaciones entre objetos están fijas, es decir, no dependen de la escena donde éstas hipotéticamente ocurren. Por ejemplo, consideremos el caso de los objetos "perro" y "persona". Si consideramos la escena "oficina", la co-ocurrencia de ambos objetos tiende a ser bastante baja. Por otro lado, al analizar la misma relación en la escena "parque", la co-ocurrencia de ambos objetos tiende a ser mucho más alta.

A continuación, presentamos un trabajo realizado en GRIMA que incorpora conocimiento holístico sobre la escena en que fue tomada una imagen, condicionando las relaciones contextuales entre objetos, de modo de poder construir un modelo adaptivo que mejore las debilidades del algoritmo de Choi et al.

Modelamiento matemático del problema

Con el objetivo de incorporar información adaptiva acerca de los distintos contextos en los que podría haber sido tomada una imagen, nuestro método modela el problema de reconocimiento de objetos en el marco de una mezcla de expertos [20]. En particular, el método aprende relaciones adaptivas condicionales entre objetos de acuerdo a distintos árboles, los que son ponderados según la información de la escena.

Cada árbol sigue la estructura dada en [16], donde se acoplan los modelos a priori y de verosimilitud. El modelo a priori representa el conocimiento previo acerca de la ocurrencia y ubicación de los objetos. Se denomina b_i a la variable que representa la presencia de un objeto de categoría i y L_i a la variable que representa la ubicación y escala de los objetos de esta categoría. El modelo de verosimilitud predice la presencia de un objeto de categoría i en una ventana w, utilizando la información de la imagen misma. Aquí se utiliza un clasificador de objetos para cada categoría, el cual corresponde a una implementación del clasificador presentado en [3], además de información holística de la escena calculada a través del método Gist [7].

Debido a que posiblemente hay múltiples detecciones de un objeto, el modelo se simplifica al considerar las K mejores detecciones para cada objeto, ordenadas por score. Para cada detección $k = 1...K$, de cada objeto i, se incorpora en el modelo la validez o confianza de la detección C_{ik}, el score de la clasificación S_{ik}, la ubicación y escala de cada detección W_{ik} y la información global de la imagen g_L.

Es importante notar que la información global de la escena g_L se considera en forma individual para la detección de cada objeto y no considera la co-ocurrencia de los mismos, de modo de alcanzar una detección adecuada que incorpore la información del conjunto de objetos como un todo.

Dado que lo que buscamos es incluir adaptivamente la información de varios contextos, nuestro trabajo se centra en el modelo a priori de la ocurrencia de los objetos. En el trabajo de Choi et al. esta variable se representa por medio de una red bayesiana en forma de árbol. Nosotros planteamos una mezcla de expertos de redes bayesianas, donde el peso de cada una es función de las características de la imagen de entrada x_G.

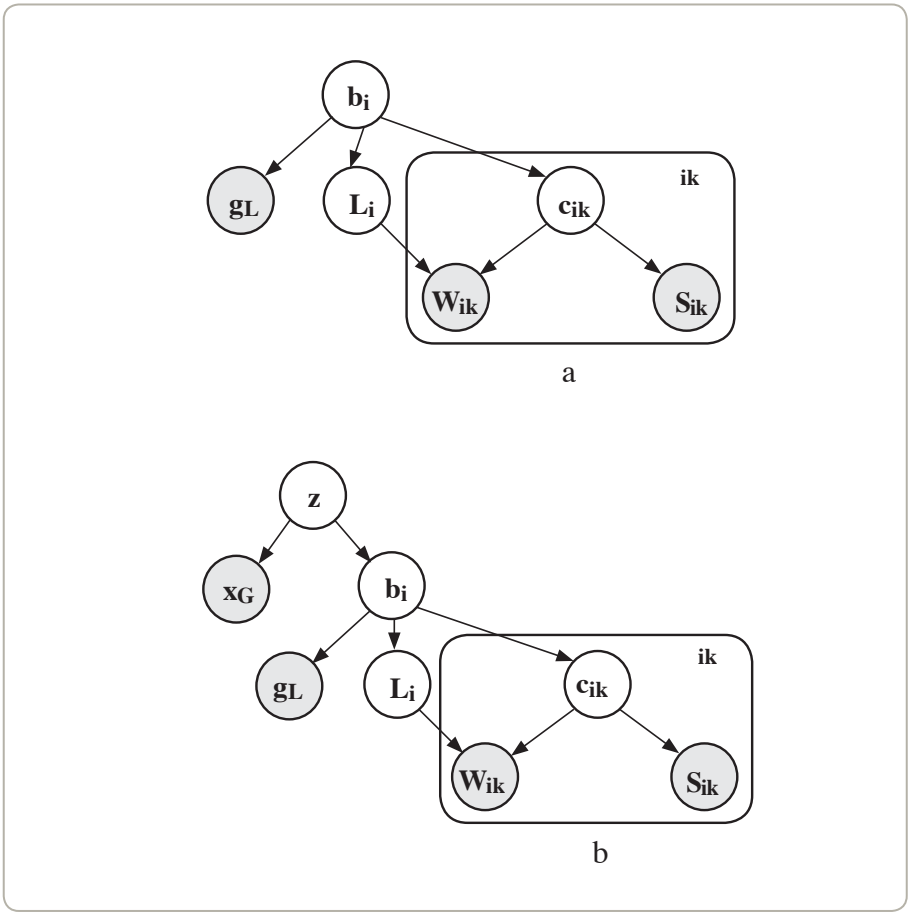
Asumiendo un conjunto de N imágenes de entrenamiento y D categorías, podemos construir N pares (x_G, b) donde $b \in 2^D$. Luego, relacionamos ambas variables usando una variable latente z que representa la escena a la cual pertenece la imagen (Figura 4.b). Asumiremos que hay K valores para z . En este punto sólo consideramos las variables x_G, b y z en forma conjunta por medio de la mezcla de expertos, con lo cual tenemos:

$$p(b|x_G) = \sum_{i=1}^{K} p(b, z_i|x_G) \quad =$$

$$\sum_{i=1}^{K} p(b|z_i, x_G)p(z_i|x_G) \quad =$$

$$\sum_{i=1}^{K} p(b|z_i)p(z_i|x_G)$$

Figura 4



Modelos de árbol para reconocimiento de objetos. (a) Sólo un árbol, equivalente a la implementación en [16]. (b) Mezcla de árboles propuesta en nuestro trabajo.

Aquí podemos distinguir dos componentes correspondientes al modelo de la mezcla de expertos. La función de compuerta, que corresponde al término $p(z_i|x_C)$, indica la influencia de cada escena de acuerdo a la información global y está representada por una función de kernel Gaussiano [21]. La función de experto, que corresponde al término $p(b|z_i)$, representa la probabilidad de ocurrencia de los objetos de acuerdo a la escena. En nuestro caso, esto es representado por una red bayesiana. Este modelo es similar al modelo de mezcla de árboles de Meila y Jordan [22], sin embargo la diferencia es que su modelo usa un conjunto fijo de pesos, mientras que en nuestro caso los pesos se adaptan a la información de entrada.

Implementación

Para resolver los parámetros del modelo aplicamos el algoritmo EM [23]. Asumiendo conocida la responsabilidad de cada componente en los datos podemos maximizar (paso M) los parámetros de la compuerta, es decir, los pesos, medias y varianzas de los kernel Gaussianos, además de los parámetros de los expertos, es decir, la estructura y las probabilidades condicionales de cada red bayesiana. Para el caso de las redes bayesianas, se utiliza una versión modificada del algoritmo de Chow-Liu, que permite encontrar los parámetros relevantes de forma óptima para árboles [22]. En el caso del valor esperado de variables latentes (paso E) se obtiene:

$$p(z_i|x_n, b_n) =$$

$$\frac{p(z_i|x_n)p(b_n|z_i, x_n)}{p(b_n|x_n)} = \frac{p(z_i|x_n)p(b_n|z_i)}{\sum_{j=1}^{K} p(z_j|x_n)p(b_n|z_j)}$$

Para realizar la inferencia, seguimos el mismo proceso alternado de [16] para cada árbol y luego combinamos los scores usando los pesos de las compuertas.

En relación a los términos base de nuestro modelo, podemos ver que en general ellos dependen de datos de entrenamiento externos a la imagen de la detección actual, los que permiten aprender sobre distintos componentes que necesitamos para la aplicación de nuestro método. Por ejemplo, el término S_{ik} corresponde a la salida de un clasificador de objetos, entrenado con un conjunto de imágenes etiquetadas. Lo mismo sucede con el caso de los contextos a través de información global, el término que representa las confianzas de las clasificaciones C_{ik}, etc. Con esto podemos ver una vez más que la información de la imagen misma que está siendo analizada debe ser complementada con información externa, correspondiente a datos que entregan conocimiento acumulado sobre los objetos y escenas utilizadas, los que resultan esenciales para un buen rendimiento del método implementado.

> En este artículo hemos presentado dos trabajos realizados por nuestro Grupo de Investigación en Inteligencia de Máquina, GRIMA, en el área de reconocimiento visual. Estos muestran cómo el uso de técnicas de aprendizaje de máquina permite incorporar en el análisis de una imagen información que va más allá de ésta.

RESULTADOS

Para nuestro trabajo consideramos dos base de datos reales de objetos, las que denominamos OUTDOR y SUN09. OUTDOOR contiene 2.600 imágenes de ocho escenas distintas tales como costa, montaña, bosques, etc. En este set de datos consideramos 21 categorías de objetos. SUN09 contiene aproximadamente 8.500 imágenes. En este set de datos consideramos 111 categorías de objetos. Ambos sets de datos fueron divididos en partes iguales para el entrenamiento y el testeo. Como clasificador de objetos individuales utilizamos el detector de objetos de Felzenswalb et al.[3]. Se usa el promedio de la curva Precision-Recall (APR) [24] como métrica para la detección de objetos.

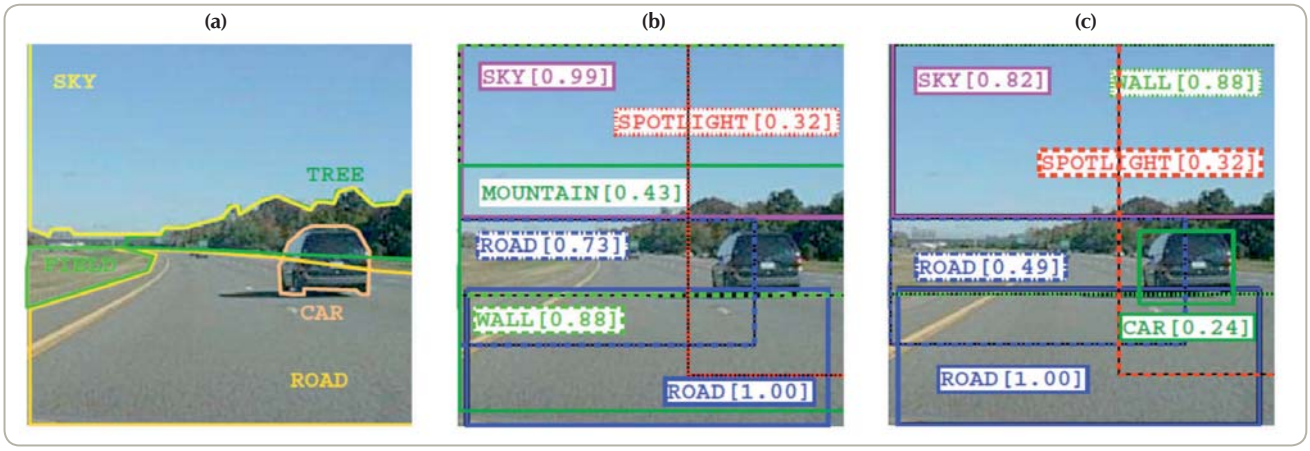
La Tabla 2 muestra una comparación entre los resultados utilizando el detector de objetos en [3] por sí sólo, el detector de un árbol único en [16] y nuestro método utilizando diversos números de árboles en la mezcla de expertos. Se puede apreciar que el mejor modelo resulta ser el de seis árboles para ambos sets de datos, con una mejora relativa de 5.5% y 5.7% respecto a un solo árbol para OUTDOOR y SUN09, respectivamente.

Tabla 2

Método	OUTDOOR	SUN09
Detector de Objetos en [3]	14.02 (6.5%)	6.82 (13.2%)
Método de un árbol en [16]	15.00 (0.0%)	7.87 (0.0%)
Mezcla de 2 árboles	15.07 (0.5%)	7.98 (1.5%)
Mezcla de 3 árboles	14.87 (-0.9%)	8.09 (2.9%)
Mezcla de 4 árboles	15.12 (0.8%)	8.06 (2.5%)
Mezcla de 5 árboles	15.25 (1.7%)	8.03 (2.2%)
Mezcla de 6 árboles	**15.83 (5.5%)**	**8.31 (5.7%)**
Mezcla de 7 árboles	14.84 (-1.1%)	7.88 (0.3%)

Resultados de APR sobre set de pruebas.

Figura 5



Comparación de resultados en una imagen de prueba. (a) Detecciones reales (ground-truth). (b) Método de un árbol en [16]. (c) Mezcla de árboles propuesta en nuestro trabajo.

La Figura 5 muestra un ejemplo de ejecución de nuestro método con seis árboles, en comparación con las detecciones reales (ground-truth) y con el método de un árbol en [16]. Se observa cómo el algoritmo es capaz de detectar un auto además de remover la falsa detección de montaña, mejorando lo entregado por el método de un árbol.

Como podemos ver, nuestro método presenta importantes ventajas en los resultados, las que de acuerdo a nuestra hipótesis, se deben al hecho de que incorporar contextos adaptivos ayuda a realizar una mejor detección al considerar la información más relevante en cada caso particular.

CONCLUSIONES

En este artículo hemos presentado dos trabajos realizados por nuestro Grupo de Investigación en Inteligencia de Máquina, GRIMA, en el área de reconocimiento visual. Éstos muestran cómo el uso de técnicas de aprendizaje de máquina permite incorporar en el análisis de una imagen información que va más allá de ésta, lo que se logra a través de aprender información relevante sobre objetos, escenas y las relaciones entre ellos desde miles de ejemplos de escenas cotidianas.

El método de reconocimiento de escenas a través de detección de objetos presentado muestra buenos resultados en el ambiente de oficinas probado, mostrando claras ventajas respecto a métodos que no utilizan objetos para el reconocimiento. Más aún, hemos probado este método en otro ambiente de interior, una casa, con distintos objetos y escenas, obteniendo resultados similares. Además, para aliviar la limitante de que una imagen particular de una escena no contiene todos los objetos presentes en ella, el método implementado fue probado sobre nuestro robot móvil utilizando una implementación secuencial que iba detectando objetos a medida que el robot se movía. Así, en el caso de la sala de conferencias, al entrar el robot detectó sólo el monitor y creyó que estaba en una oficina, sin embargo, gracias a la implementación probabilística utilizada, cuando posteriormente encontró la pantalla de proyector pudo actualizar sus creencias y definir que realmente estaba en una sala de conferencias.

Cabe destacar, que tal como se mencionó anteriormente, el método requiere de una gran capacidad computacional para ejecutar, lo que lo hace difícil de implementar para una operación en el mundo real. Dado esto, se implementó una aproximación a través de muestreo, utilizando el método de Monte Carlo, para poder realizar un eficiente cómputo del reconocimiento de escena.

En el caso del método de reconocimiento de objetos utilizando contextos jerárquicos adaptivos, se logró superar las limitaciones relevantes de un modelo de árbol de contexto fijo, a través de un modelo que utiliza una mezcla condicional de árboles para lograr resultados que se adapten a las distintas condiciones en las que pudo haber sido tomada la imagen. Nuestros experimentos usando distintos conjuntos de datos indican que el modelo propuesto mejora el rendimiento del reconocimiento de objetos con respecto a los métodos comparados, al considerar información de la escena subyacente que influye en las relaciones de objeto a objeto.

Cabe destacar que este método puede ser utilizado con distintos clasificadores de objetos, los cuales deberían mejorar su rendimiento debido a la inclusión de información adaptiva sobre el contexto. Así, por ejemplo, para alcanzar una buena escalabilidad al ejecutar el método con un gran conjunto de categorías de objetos, se puede incluir políticas adaptativas para

controlar la ejecución de los clasificadores de objetos, en forma similar al método propuesto en [25].

Más detalles de la implementación de cada uno de los métodos presentados pueden ser encontrados en [14] y [26].

AGRADECIMIENTOS

Este trabajo fue financiado en parte por Fondecyt, proyectos 1120720 y 1095140.BITS

REFERENCIAS

[1] P. Viola y M. Jones. "Rapid object detection using a boosted cascade of simple features". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001.

[2] J. Sivic y A. Zisserman. "Video Google: A Text Retrieval Approach to Object Matching in Videos". International Conference on Computer Vision (ICCV), 2003.

[3] P. Felzenszwalb, D. McAllester, y D. Ramanan. "A discriminatively trained, multiscale, deformable part model". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[4] Microsoft Corp. Redmond WA. Kinect for Xbox 360.

[5] Grupo de Inteligencia de Máquinas DCC PUC, GRIMA, http://grima.ing.puc.cl.

[6] C. Siagian y L. Itti. "Rapid biologically-inspired scene classification using features shared with visual attention". IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 29 (2), 300–312 , 2007.

[7] A. Oliva y A. Torralba. "Modeling the shape of the scene: a holistic representation of the spatial envelope". International Journal of Computer Vision (IJCV), 42, 145–175, 2001.

[8] L. Fei-Fei y P. Perona. "A bayesian hierarchical model for learning natural scene categories". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[9] A. Bosch, A. Zisserman y X. Muñoz. "Scene classification via plsa". European Conference on Computer Vision (ECCV), 2006.

[10] J. Sivic, B. Russell, A. Efros, A. Zisserman, y W. Freeman. "Discovering objects and their localization in images". International Conference in Computer Vision (ICCV), 2005.

[11] S.Lazebnik, C. Schmid y J. Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.

[12] A. Quattoni y A. Torralba. "Recognizing indoor scenes". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[13] L. Li, H. Su, E. Xing y L. Fei-Fei. "Object Bank: A High-Level Image Representation for Scene Classification and Semantic Feature Sparsification". Neural Information Processing Systems (NIPS), 2010.

[14] P. Espinace, T. Kollar, A. Soto y N. Roy. "Indoor Scene Recognition Through Object Detection". IEEE International Conference on Robotics and Automation (ICRA), 2010.

[15] N. Dalal y B. Triggs. "Histograms of oriented gradients for human detection". European Conference on Computer Vision (ECCV), 2005.

[16] M. Choi, J. Lim, A. Torralba, y A.Willsky. "Exploiting hierarchical context on a large database of object categories". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[17] C. Desai, D. Ramanan, and C. Fowlkes. "Discriminative models for multi-class object layout". International Journal of Computer Vision (IJCV), 95(1), 2011.

[18] A. Torralba, K. Murphy y W. Freeman. "Contextual models for object detection using boosted random fields". Advances in Neural Information Processing Systems (NIPS), 2005.

[19] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora y S. Belongie. "Objects in context". International Conference on Computer Vision (ICCV), 2007.

[20] R. Jacobs, M. Jordan, S. Nowlan y G. Hinton. "Adaptive mixtures of local experts". Neural Computation, 3, 79-87, 1991.

[21] L. Xu, M. Jordan y G. Hinton. "An alternative model for mixtures of experts". Advances in Neural Information Processing Systems (NIPS), 1994.

[22] M. Meila y M. Jordan. "Learning with mixtures of tres". Journal of Machine Learning., 1:1–48, 2001.

[23] A. Dempster, N. Laird y D. Rubin. "Maximum likelihood from incomplete data via the EM algorithm (with discussion)". Journal of the Royal Statistical Society, Series B, 39:1–38, 1977.

[24] J. Davis y M. Goadrich. "The relationship between precision-recall and roc curves". International Conference on Machine Learning, pages 233–240, ACM Press, 2006.

[25] P. Espinace, T. Kollar, A. Soto y N. Roy. "Indoor scene recognition through object detection using adaptive objects search". European Conference on Computer Vision (ECCV), Workshop on Robotics for Cognitive Tasks, 2010.

[26] B. Peralta, P. Espinace y A. Soto. "Adaptive hierarchical contexts for object recognition with conditional mixture of trees". British Machine Vision Conference (BMVC), 2012.

I've seen the **FUTURE**

It's in my **BROWSER**

HTML5: ¿nuevas perspectivas o déjà vu?

Imagen: CC-BY W3C (http://www.w3.org).

Nelson Baloian
Profesor Asociado DCC, U. de Chile. Doktor rer. nat, Universität Duisburg, Alemania (1997); Ingeniero Civil en Computación, Universidad de Chile (1988). Líneas de especialización: Instrucción Asistida por Computador, Trabajo Colaborativo, Sistemas Distribuidos.
nbaloian@dcc.uchile.cl

LA NECESIDAD DE ESTÁNDARES EN LA COMPUTACIÓN

Los estándares son de gran ayuda en cualquier disciplina, pues permiten que distintas entidades puedan desarrollar diversos artefactos que pueden interactuar o complementarse sin necesidad que se comuniquen explícitamente, ya que el estándar cumple dicha función. En computación esto es tremendamente cierto, ya que tenemos millones de entidades de los más variados tamaños y formas de organización desarrollando software en todo el mundo. No pocas veces hay empresas que se han sentido con la fortaleza suficiente para lanzar al mercado productos que difieren de las normas, como lo hizo IBM con el sistema de codificación EBCDIC.

Es difícil imaginar que la computación sería lo que es hoy, sin estándares como TCP/IP, HTTP y HTML. Si bien hasta hace algunos años parecía que la computación se estabilizaba alrededor de tres plataformas principales (Windows, Unix/Linux y Mac) la irrupción masiva de la computación móvil ha sumado una serie de diversas plataformas al paisaje computacional, complicando así el panorama de los desarrolladores. No sólo hay diferencias grandes en el desarrollo de aplicaciones para equipos fijos y móviles, sino también entre los mismos equipos móviles hay variedades irreconciliables. Y como si esto fuera poco, tenemos una variedad de tamaños de pantalla, dificultando el diseño de una interfaz de aplicación que sea compatible con todos los equipos.

Es por todo esto que algunos desarrolladores han puesto sus ojos en HTML5 como una herramienta que les permite desarrollar aplicaciones altamente interactivas que sean capaces de correr en varias plataformas. Para ponerlo en pocas palabras, podríamos

decir que HTML5 se diferencia de su predecesor HTML4 en cuanto a que este último incorpora una serie de elementos nuevos que pueden ser instanciados y manipulados desde pedazos de código JavaScript insertos en la página web.

NUEVOS ELEMENTOS PRESENTES EN HTML5

En efecto, HTML5 provee nuevas funcionalidades que permiten enriquecer significativamente sitios desarrollados con HTML a través de nuevos elementos que permiten usar nuevas técnicas de presentación, comunicación y manejo de datos. A continuación nombramos algunos de estos elementos:

Canvas: a través del objeto CANVAS HTML5 provee la posibilidad de incorporar gráficos 2D en forma fácil y flexible, además de capturar y procesar los eventos de interacción del usuario. Su función se parece mucho al canvas de Java, permitiendo dibujar líneas simples, elipses, rectángulos, etc. Además se pueden colocar imágenes sobre un CANVAS que pueden ser movidas, rotadas y escaladas (cambio de tamaño). También se pueden manejar algunas características gráficas como transparencia. Todas estas funciones están a disposición del desarrollador como simples comandos en JavaScript. Un ejemplo básico del uso de éste se ve en la Figura 1. Este código define un elemento CANVAS de tamaño 500x250 píxeles y de color verde, como se muestra en la Figura al hacer clic sobre él.

WebSockets: dado el protocolo HTTP, un sitio web sólo permite que el servidor reaccione a un requerimiento de un cliente del tipo request-replay, permitiendo un modo de comunicación esencialmente unidireccional. La única forma de que el cliente recibiera datos del servidor cuando el servidor los tuviera disponible era mediante métodos de "pull" (por ej., AJAX) que básicamente consisten en que el cliente pregunta a intervalos constantes si existe nueva información. Esto cambia con la implementación de WebSockets, ya que este enfoque provee una comunicación bidireccional, tal como en los sockets tradicionales de TCP/IP pero sobre protocolo HTTP. Para recibir datos desde un WebSocket, la página debe implementar los siguientes métodos en JavaScript, los cuales serán invocados por el navegador cuando sucedan los eventos correspondientes:

- **onopen:** se llama cuando se abre un WebSocket, es decir, cuando la comunicación se establece.
- **onmessage:** se llama cuando un mensaje desde el servidor es recibido, el mensaje es parámetro de este método.
- **onclose:** se llama cuando la conexión al servidor es cerrada.

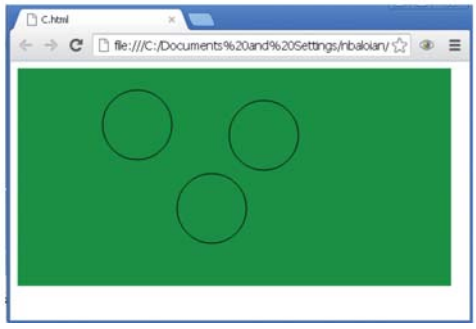
Web Database: HTML5 provee una implementación de SQLite que permite asociar aplicaciones Web con una base de datos local, permitiendo guardar datos recolectados en forma remota en una base de datos local.

LocalStorage: consiste en una tabla de hashing que permite guardar hasta 10Mb de pares (nombre, valor); estos datos están relacionados con la aplicación web en modo online y offline.

FileSystem, Manifest y modo offline: la clave para el desarrollo de aplicaciones web es el uso combinado del sistema de archivos FileSystem y el archivo de Manifest. FileSystem es una API que proporciona un sistema de archivos independiente de otras aplicaciones que estén corriendo y de los archivos de usuario. Manifest es una declaración de cuáles son los archivos que componen la aplicación proveyendo referencias locales a ellos. El Manifest, es utilizado por los navegadores para descargar estos archivos y cargarlos a un

Figura 1

```
<canvas style="top:0;left:0; background-color:#180" id="can" width="500" height="250" ></canvas>
<script>
// get the canvas element and its context
var canvas = document.getElementById('can');
var context = canvas.getContext('2d');
function draw(event){
   context.beginPath();
   context.arc(event.pageX,event.pageY,40,0,2*Math.PI);
   context.stroke();
}
// attach click event listener
canvas.addEventListener('click',draw, false);
</script>
```



El código en HTML y JavaScript, y su resultado.

sistema de archivos local, lo que permite el funcionamiento de la aplicación sin conexión.

Geolocation: dadas las capacidades de los dispositivos móviles actuales, hay muchas aplicaciones para ellos que utilizan la API para hacer uso de esta información. HTML5 también provee acceso a esta API, que permite conocer las coordenadas donde se encuentra el dispositivo en el momento, y así integrar esta información a la aplicación. Los datos disponibles son: latitud, longitud, altitud, exactitud de la medida de altitud provista y velocidad de desplazamiento.

WebWorker: permite al desarrollador trabajar con hilos de ejecución (threads) en un entorno web. Normalmente, el navegador es quien controla todos los threads activos corriendo dentro de su ambiente y estos no son referenciables desde el código JavaScript inserto en la página que se está interpretando. Sin embargo, un WebWorker permite operaciones síncronas y asíncronas, sobre threads independientemente de las que maneje el navegador.

WebGL: adicionalmente a las capacidades de dibujo 2D que provee el CANVAS de HTML5 también se provee el manejo de dibujos 3D basado en OpenGL, llamado WebGL. Al hacer uso de esta tecnología es posible enriquecer la presentación de las interfaces gráficas basadas en HTML5.

La lista de elementos adicionales que tiene HTML5 nos hace pensar que algunos de ellos fueron diseñados con la idea de apoyar (también) la computación móvil. En efecto, el uso de la geolocalización tiene mayor sentido cuando el usuario está en movimiento, o cuando la posición del usuario varía más o menos frecuentemente. También los elementos que permiten el almacenamiento local con conexión y sin conexión ayudan al desarrollo de este tipo de aplicaciones, pues en el escenario móvil es común que se pierda la conexión a ratos y, por lo tanto, conveniente que se trabaje con datos localmente hasta que se vuelva a tener conexión. Suponemos que esto no es casualidad, pues es justamente en el ámbito de la computación móvil donde un estándar se hace hoy en día urgentemente necesario, dada la heterogeneidad de los dispositivos disponibles en el mercado.

DESAFÍOS EN LOS ESCENARIOS MÓVILES DE HOY

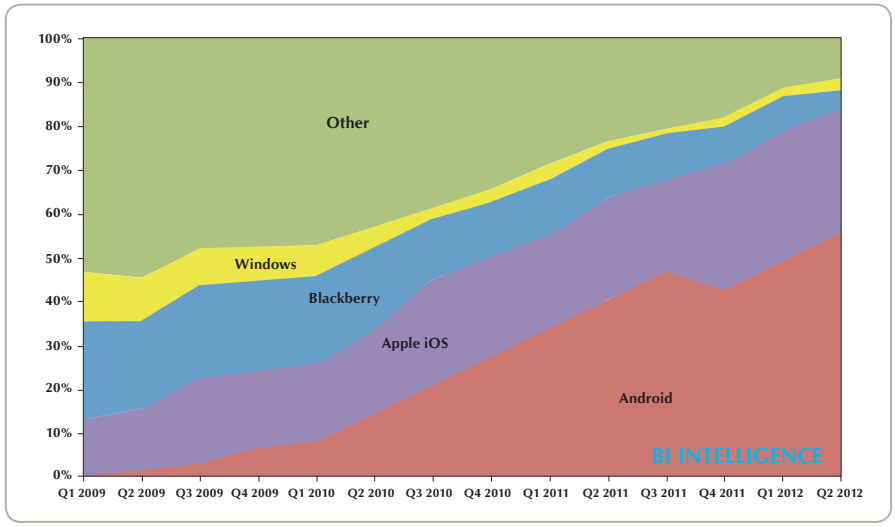
Hoy nos enfrentamos a varios desafíos al tratar de desarrollar aplicaciones móviles. Como describe G.Avellis et al: 2003, algunos de los desafíos podrían ser resueltos por un enfoque arquitectónico que permite la adaptación flexible de las diferentes representaciones de los mismos datos. Con este planteamiento varios problemas, como las pantallas de diferentes tamaños, diferentes mecanismos de interacción y el poder computacional, pueden ser abordados de manera satisfactoria.

Otro desafío importante desde el punto de vista técnico en escenarios móviles de hoy es la heterogeneidad de las plataformas de sistemas operativos móviles. Por lo menos hay cinco grandes jugadores actualmente en el mercado: iOS, Android, Symbian, Blackberry y WindowsMobile. Todos ellos proporcionan interesantes plataformas de desarrollo integrado que permiten el desarrollo de aplicaciones específicas para la plataforma.

De acuerdo con Álex Cocotas (http://www.businessinsider.com/mobile-platform-market-share-2012-8), la distribución de las ventas de dispositivos móviles por proveedor de los diferentes sistemas operativos móviles, en el tercer cuarto de 2012 fue la siguiente: el sistema operativo más extendido es el móvil Android (58%). Segundo está iOS (26%), sistema de los teléfonos móviles de Apple. Le sigue Blackberry con cerca del 5%, Windows Mobile apenas un 3% y Otros cerca del 11% (ver Figura 2). Si bien se podría pensar en que Android domina el mercado, se ha visto que éste es bastante dinámico y aún no se puede decir quién tendrá la última palabra.

Por lo tanto, deben considerarse alternativas distintas de la implementación de las aplicaciones dependientes de la plataforma para dispositivos móviles. En la actualidad, el enfoque más prometedor parece ser la creación de aplicaciones basadas en HTML5 con JavaScript.

Figura 2



Evolución de las ventas de los sistemas operativos para dispositivos móviles. Fuente: http://www.businessinsider.com/mobile-platform-market-share-2012-8.

ENFOQUES PARA EL DESARROLLO DE APLICACIONES HTML5 INDEPENDIENTES DE LA PLATAFORMA

Podemos decir que existen principalmente dos enfoques diferentes para desplegar (deploy) y correr aplicaciones basadas en la Web con HTML5 dependiendo de las condiciones del ambiente. La primera, la más obvia, es usar un navegador Web que entienda HTML5 y simplemente ingresar la URL del sitio (página) que contenga el código y JavaScript, el cual será interpretado. Este enfoque bien general implica que la aplicación no tendrá acceso a los accesorios específicos del dispositivo móvil, pues el código JavaScript no tiene cómo acceder a ellos, ya que son específicos de cada máquina (excepción a esto es el acceso a los datos de geolocalización, como se vio más arriba). Esto es así dado el carácter general y portable que debe tener el código escrito en JavaScript.

Un segundo enfoque para el despliegue de aplicaciones basadas en HTML5, es usar una aplicación nativa especialmente desarrollada para la plataforma que se desea usar, que contenga un motor HTML5, es decir, una funcionalidad capaz de bajar una página de un servidor e interpretarla, y que tenga acceso a los accesorios del dispositivo. Obviamente, esta aplicación no será totalmente portable a otra plataforma. Un enfoque que se perfila como el más conveniente cuando se necesita que una aplicación converse con los accesorios del dispositivo, es usar el segundo enfoque pero programar lo más que se pueda en HTML, de modo que al momento de portar la aplicación se tenga que reimplementar lo menos posible.

BIBLIOTECAS ABIERTAS EN HTML5

Durante los últimos meses han aparecido algunas bibliotecas de JavaScript que facilitan el desarrollo de aplicaciones más complejas en HTML, algunas de ellas especialmente orientadas a trabajar con los elementos de HTML5. A continuación describimos algunas de ellas:

JQuery: es quizá la herramienta más popular para trabajar con JavaScript. Si bien no es exclusiva de HTML5, es ampliamente usada para trabajar en la Web. El núcleo de JQuery consiste en la redefinición del símbolo $ del lenguaje para que éste contenga una serie de funciones que se enfocan en hacer fácil lo que con JavaScript era complejo de hacer. Por ejemplo, la selección de elementos de HTML (DOM) se realiza de una manera práctica e intuitiva, incorporando una estructura de consultas sobre los elementos que permite retornar una o más coincidencias. En particular, podemos seleccionar todos los links de un documento y manipularlos, o sólo los links que se encuentren dentro de una lista. JQuery además provee de utilitarios que simplifican las operaciones que son más repetitivas, como llamadas AJAX, conversión del formato de retorno (texto, XML, JSON) a objetos manipulables y otras. Gracias a JQuery actualmente vemos avanzadas interfaces gráficas en la Web, que nos permiten incluso editar imágenes en tiempo real sin tener que subir datos al servidor.

> Podríamos decir que HTML5 se diferencia de su predecesor HTML4 en cuanto a que este último incorpora una serie de elementos nuevos que pueden ser instanciados y manipulados desde pedazos de código JavaScript insertos en la página web.

Modernizr: en la actualidad tenemos un sinfín de navegadores con formas diferentes de interpretar HTML/HTML5. Muchos de estos soportan operaciones de manera parcial lo que hace complejo diseñar una aplicación sin saber dónde se ejecutará. Una alternativa es generar múltiples versiones de nuestra aplicación y entregar la que corresponde al navegador. La otra es usar Modernizr, una librería que se encarga de detectar la disponibilidad de HTML5 en el explorador y según esto se encarga de incorporar código en demanda de manera dinámica y a la medida. Esto permite que un mismo documento se pueda comportar de manera diferente según las capacidades del navegador.

Kendo / Ext js: al diseñar un sistema, comúnmente se definen qué componentes se utilizarán, cómo se manejará si existen múltiples idiomas, se definen las interfaces gráficas y, en el caso de utilizar un modelo vista controlador, corresponde definir cada uno también. Kendo y Ext js son recopilaciones de todo lo anterior para JavaScript. Es decir, de manera fácil se puede especificar el soporte para múltiples idiomas, componentes gráficos, el "tema" de la "interface" los modelos, vistas y

controladores de la aplicación. La ventaja es que, dado que se ejecuta en el navegador, no existe sobrecarga de los servidores para la gestión de la aplicación, lo cual permite tener plataformas escalables con el mismo o menor esfuerzo que antes.

PhoneGap: debido a que el manejo de los accesorios de multimedia y posicionamiento de los dispositivos móviles es muy dependiente de su tipo, se han desarrollado bibliotecas que tienden a uniformizar este manejo para hacer así más portable el código. La API más utilizada para este propósito externo es probablemente PhoneGap, compatible con al menos cuatro de los cinco principales sistemas operativos para dispositivos móviles. El único que falta aquí es Windows Mobile, que los desarrolladores de PhoneGap se comprometieron a integrar en el futuro también. Por desgracia, la API PhoneGap no permite el acceso a todos y cada dispositivo de hardware depende del sistema operativo móvil (la Tabla 1 muestra qué tipo de hardware está soportado bajo qué sistema operativo móvil). Sin embargo, PhoneGap proporciona suficiente apoyo de manejo de accesorios específicos a una plataforma para los sistemas operativos móviles habituales, de manera de apoyar el desarrollo de aplicaciones independientes de la plataforma para variados escenarios de la computación móvil.

PERSPECTIVAS

Al revisar con detención las características de HTML5 y las posibilidades que ofrece para el desarrollo de aplicaciones altamente interactivas y complejas corriendo dentro de un navegador Web, el desarrollador que ha visto la evolución de las aplicaciones basadas en la Web se preguntará qué hay de diametralmente nuevo en este enfoque que no se haya visto aún. En efecto, si el lector ya desarrollaba aplicaciones basadas en la Web hacia finales de los noventa recordará que HTML5 tiene muchos elementos que son característicos de los Applets de Java. Sin embargo, hoy vemos que son pocas las páginas que integran esta tecnología a pesar de los grandes augurios con la que fue recibida en su momento.

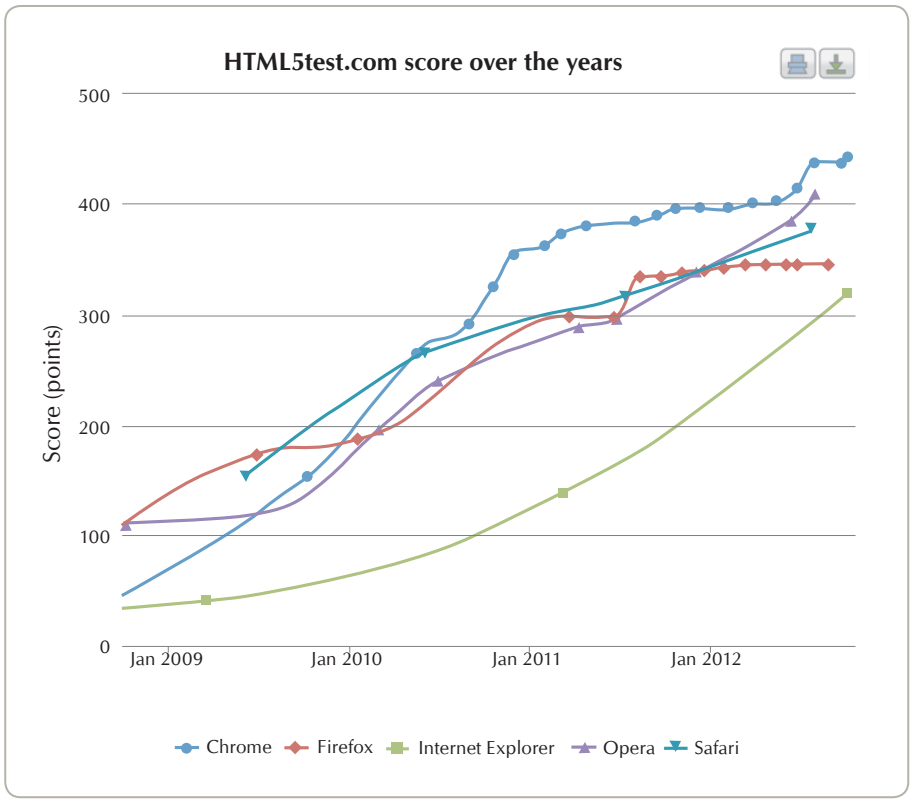
Quizás para preguntarse qué perspectivas tiene HTML5 sea necesario también analizar cuáles fueron las razones por las cuales los Applets de Java no llegaron a ser el estándar que se pensaba llegarían a ser. Ciertamente en todo establecimiento de un estándar los intereses económicos de las compañías juegan un papel importante. Java fue desarrollado y mantenido por una compañía específica (Sun Microsystems) y si bien fue liberado sin costo, podríamos pensar que la intención detrás era que los productos de otras compañías no se impusieran (por lo tanto, habían intereses económicos de por medio entre proveedores del mercado). En este sentido HTML5 comienza con mejor perspectiva ya que no fue desarrollado ni "apadrinado" por ninguna compañía: las versiones de HTML son sancionadas y liberadas por la organización W3C, que se encarga de establecer los estándares para la Web, por lo que podríamos pensar que su establecimiento como estándar estará relativamente libre de intrigas. Sin embargo, también hubo factores técnicos que quizás pesaron más aún que los económicos que jugaron en contra de los Applets y que hasta cierto punto se da también en la adopción de HTML5 como estándar. Esto tiene que ver con la velocidad con que los navegadores incorporan las capacidades para interpretar completamente el código. Hacia mediados de los noventa se libera la primera versión de Java y muchos de los navegadores de la época incorporan la máquina virtual de Java (JVM) capaz de interpretar los Applets, que son desarrollados con la primera versión de Java. Sin embargo, Java se desarrolla muy rápido y los Applets desarrollados con las versiones posteriores no pueden ser interpretados por la máquina virtual de versiones anteriores. En esos tiempos la autoactualización del software, tal como ocurre hoy, no estaba implementada en los navegadores. Esto causó entonces

Tabla 1

	Android	iOS	Symbian	Blackberry		Android	iOS	Symbian	Blackberry
Acelerómetro	x	x	x	x	Notificación	x	x	x	x
Cámara	x	x	x	x	Geolocalización	x	x	x	x
Brújula	x	x	-	-	Multimedia	x	x	-	-
Lista de contactos	x	x	x	x	Red	x	x	-	x
Información del dispositivo	x	x	-	x	Sistema de archivo	x	x	-	x
Eventos	x	x	-	x	Almacenamiento	x	x	-	x

Soporte para el manejo de accesorios específicos del hardware por sistema operativo que brinda PhoneGap.

Figura 3



Curva de implementación de HTML5 para los principales navegadores. La medición (score) supone un máximo de 500 puntos para un 100% de implementación de las funcionalidades definidas por el estándar.

que muchos sitios que integraban Applets desarrollados con las últimas versiones no pudiesen verse en un porcentaje importante de los navegadores.

Si bien el problema de los navegadores con los Applets fue el no contar con la versión correcta de la máquina virtual de Java, hoy en día vemos que la implementación de todas las funcionalidades que define el estándar por parte de los navegadores ha sido más bien lenta. En el sitio http://html5test.com, es posible medir en qué medida el browser que contacta el servidor del sitio tiene implementadas las funcionalidades de HTML5. En esta misma página se puede ver también un gráfico que muestra cómo ha sido la evolución en la adopción de las características de HTML5 por parte de los navegadores más comunes de hoy, esto es, en qué medida los navegadores son capaces de entender el protocolo HTML5 e implementar las funcionalidades que se requieren para interpretarlo (ver Figura 3). Si bien la tendencia ha sido positiva, se han requerido tres años para llegar a una situación donde la mayoría de los navegadores ha implementado más del 60% de las funciones del estándar. Sólo dos superan el 80% y ninguno el 90%.

Otro problema que también atenta fuertemente contra el establecimiento de HTML5 como un estándar, es el hecho de que los resultados que se obtienen en cuanto a la interpretación del código varían de navegador a navegador, e incluso entre distintas versiones de un mismo navegador, especialmente entre las versiones para dispositivos móviles y para desktop. Por ejemplo, Chrome utiliza un webkit que provee algunas funcionalidades, tales como ver imágenes captadas por la cámara, pero no permite grabar de la misma. Lo mismo ocurre con el sonido. Las llamadas para controlar estos accesorios de multimedia son diferentes en Android, Mozilla y Opera. Por ahora, Opera es el único que respeta el estándar, tanto la versión para dispositivos móviles como en la de dispositivos de escritorio (desktop). Los navegadores Firefox e Internet Explorer implementan sólo las funcionalidades más populares: WebDatabase, LocalStorage, FileSystem y Canvas. Websockets funcionan en la versión desktop de Chrome pero no en móvil. Lo mismo sucede con Firefox. Chrome incorporó el sonido hace poco en la versión 2.2, en las anteriores no estaba disponible. Por esto es que existen librerías que autodetectan todas las características implementadas y autorrenombran las llamadas para que se ajusten al estándar.

Del análisis anterior podemos concluir que HTML5 aún no está en condiciones de ser nombrado "el" estándar para estos tiempos. En el caso de los Applets, cuando se quiso presentarlos como una solución estándar para varios problemas que había en ese entonces, las variadas dificultades e inconsistencias que mostraban los navegadores al momento de interpretarlos terminaron por hacer que los desarrolladores de páginas web evitaran incorporarlos en sus páginas. Lo curioso es que hoy en día es probable que los Applets puedan funcionar sin problemas y sin diferencias en casi cualquier navegador. Esto se debe a que Java está mucho más maduro y estable, y por lo tanto ya no evoluciona de manera tan dinámica, y además es normal que los navegadores se autoactualicen bajando nuevas versiones y parches casi sin que los usuarios se den cuenta. Es probable que de aquí a un tiempo no muy lejano la interpretación que los navegadores hagan de HTML5 también sea más completa y estandarizada incluso para las versiones móviles y desktop. Cabe entonces hacerse la pregunta de si para ese entonces los desarrolladores de sitios web habrán decidido ya evitar esta tecnología y pase algo parecido a lo sucedido con los Applets o si para entonces ya habrá disponible otra tecnología que prometa (¿de nuevo?) ser el estándar del futuro de la Web.BITS

La televisión en la Universidad de Chile

Nicolás Beltrán
Profesor Asociado DIE, Universidad de Chile, Dr. in App. Sc. (1985), M.Elect. Eng. (1981), ambos de la Katholieke Universiteit Leuven, Bélgica, Ingeniero Civil Electricista Universidad de Chile (1974). Trabajó en el Departamento de Ingeniería Eléctrica de la U. de Chile desde 1985 hasta el 2012, acogiéndose recientemente a jubilación. Fue Director de Departamento en tres períodos, Jefe Docente, Consejero de Departamento, miembro de la Comisión Asesora de TV Digital del Ministro de Transporte y Telecomunicaciones (2007-2010) y de la Comisión para la TV Digital nombrada por el Consejo Universitario (2009-2012). Sus líneas principales de interés académico fueron la Electrónica y Sistemas de Instrumentación, y de interés profesional Comunicaciones en Sistemas móviles y en Sistemas broadcasting. Actualmente es consultor en comunicaciones electrónicas.
nicolasbelt@gmail.com

La Universidad de Chile ha estado involucrada en el desarrollo de la televisión abierta en nuestro país desde sus orígenes. La revista de la Facultad de Ciencias Físicas y Matemáticas en su número 42 de otoño de 2008, incluye un artículo titulado "La televisión chilena nació en Beauchef". El artículo muestra cómo un desafío tecnológico relevante fue enfrentado con éxito y talento por ingenieros formados en esta Facultad. Creo necesario destacar en este aspecto que junto al desarrollo de esta solución tecnológica, se sumó un interés genuino por hacer de esta tecnología un vehículo de cultura y entretención que perduró por un poco más de una década.

A partir de mediados de los años sesenta la televisión abierta en Chile se masifica, las generaciones a partir de esos años crecen con esta ventana que se introdujo en nuestro círculo más cercano con un poder no bien evaluado, que en cierta forma nos manipula y con contenidos que en buena parte han ido degradándose hoy día hacia una morbosidad que persigue retener audiencia para mejorar la venta de publicidad que asegure el financiamiento, situación que parte de la sociedad chilena intenta contrarrestar al interior de la familia.

¿Qué sucedió con este medio de comunicación en nuestra Universidad? No es simple encontrar una causa para una trayectoria que una vez demostrada su factibilidad, la misma Universidad no logró estructurar un esquema organizacional a su interior que le permitiera autofinanciarse generando contenidos atractivos. Hoy la Universidad prácticamente no participa en estas decisiones y su único activo es una concesión de frecuencia para una red nacional que, en el nuevo paradigma de la televisión digital, no está claro a esta fecha que lo vaya a mantener.

En este artículo se revisa superficialmente lo que mucho de los lectores con todo derecho podrían llamar su propia prehistoria televisiva, siguiendo con una explicación técnico-conceptual de lo que es hasta hoy día la TV análoga y las oportunidades que surgen con la mal denominada TV digital. Digo mal denominada porque, en estricto rigor, el futuro es mucho más que TV entendido como imágenes en movimiento que permiten entretener, educar o informar. La implantación de esta tecnología de comunicación permitirá incorporar otros esquemas de presentación de información multimedial, por ello prefiero denominarla Plataforma de Servicios Digitales. En Chile su implantación ha sido demorada por la discusión en el Congreso del proyecto de ley que ya lleva cuatro años y que permite el uso de este medio. Si nuestra Universidad decide participar en este nuevo paradigma, respetando nuestra tradición, nuestra misión y fundamentalmente nuestra vocación de servicio a un país que requiere con cierta urgencia redefinir sus esquemas económicos y sociales, las oportunidades que se abren sólo tienen como límite la creatividad de su comunidad.

HISTORIA [1]

En 1959 el Departamento de Electricidad de la Facultad de Ciencias Físicas y Matemáticas, hoy también conocida al interior de nuestra Universidad como la "República de Beauchef", le informa al Rector de la época, profesor Juan Gómez Millas, que se ha logrado alcanzar el nivel de *"expertise"* suficiente con los equipos de televisión desarrollados y adaptados en el laboratorio, que permitían asegurar una continuidad de transmisiones. El Rector encarga entonces al Secretario General de la Universidad, profesor Álvaro Bunster, que organice la actividad de manera que en la frecuencia concesionada por la Secretaría de Servicios Eléctricos y Telecomunicaciones, surja el Canal de TV de la Universidad de Chile, el Canal 9.

Esta iniciativa del Rector tiene, como cualquier acción innovadora, sus detractores y también sus impulsores al interior del Consejo Universitario. Así varios decanos consideran esta actividad un despilfarro de recursos habiendo necesidades de tipo docente que era necesario y urgente cubrir.

Por el lado de los decanos que impulsan la iniciativa se destaca el profesor Eugenio González, Decano de la Facultad de Filosofía y Educación, quien es un convencido de que la TV universitaria es sinónimo de medio para la difusión de la cultura.

Tal vez aquí vale la pena hacer una digresión aprovechando el convencimiento del Decano González. La TV universitaria como medio para masificar la cultura NO fue producto de una demanda de la sociedad chilena, como es una creencia común, sino que muy por el contrario, se debió al convencimiento de personalidades fuertes que coyunturalmente se encontraban en posiciones de poder tanto en el gobierno como en las universidades. Es así como el Presidente Jorge Alessandri, conocido por su austeridad y cierta tozudez, pensaba que una variable relevante e impulsadora del desarrollo de un país era contar con una buena educación pública. Él se opuso tenazmente a la instalación de la TV comercial en Chile porque consideraba que su instauración significaría un mayor gasto de divisas para el país, por tener que importar equipos caros y además era un gasto superfluo para la población que debían adquirir los receptores.

Así el mito de que parte de la sociedad chilena, entre ellos intelectuales y academia habían influido para que en nuestro país se implantara una TV cultural y de entretenimiento, es y seguirá siendo eso, un mito.

Lo real es que las autoridades de Gobierno de fines de los años cincuenta entregaron a tres universidades la posibilidad de experimentar este modelo de TV cultural, con la confianza de que lo desarrollarían exitosamente. Sin embargo, como sucede hasta hoy, cuando los compromisos de financiamiento especialmente del Estado son débiles y las expectativas de resultados esperados son altas, quienes se comprometieron a lograrlo, o cambian el modelo con el ingenio y viveza de nuestra nación, o sucumben en el esfuerzo.

Así, la TV universitaria en la Universidad de Chile nace en espacios de ingeniería con parte de sus equipos desarrollados en la propia Universidad, a los que se les suma posteriormente los contenidos generados por las capacidades internas de ella. En la organización universitaria para soportar el Canal, surge el Departamento Audiovisual que incluye al cine y la fotografía que permiten dar un soporte a las emisiones de Canal 9.

Un hito en la década de los sesenta lo constituye el mundial de fútbol realizado en Chile en 1962, que produce un quiebre en este modelo. Hasta ese año la TV universitaria presentaba las dificultades propias en la generación de contenidos o programas "en vivo" debido a lo exiguo de los presupuestos.

Con motivo del evento de fútbol, el Gobierno del Presidente Alessandri disminuye las barreras arancelarias para la importación de receptores y de transmisores de TV, aumentando fuertemente el parque de receptores.

También en esa época en nuestra Universidad se experimentaba un crecimiento en apoyo a la cultura del país, con la realización de docencia e investigación en las artes de la representación y de expresión musical, junto con el desarrollo de la Escuela de Periodismo.

Así el escenario cambia y se tiene, por una parte, una demanda de la ciudadanía por más programas de televisión con contenidos diversos, que abarquen desde la información hasta la entretención y, por otra, un conjunto de profesionales formados en la Universidad y que se encuentran en condiciones de mejorar cualitativamente los contenidos transmitidos.

Al interior de la Universidad de Chile, el canal de TV comenzó a verse como un competidor adicional a la distribución del presupuesto universitario, perdiendo el débil carácter corporativo que había tenido en su primera fase. Así comenzamos a tener una TV al estilo *"chilean way"*. Sin el financiamiento adecuado de la Universidad, con un débil financiamiento del Estado ya que no había legislación específica, esto es, no había una

ley sobre TV, se complementó entonces el financiamiento con publicidad disfrazada.

La historia en la década siguiente sigue siendo de altos y bajos. En junio de 1973 y como una manifestación más de la polarización que vivía Chile en ese momento, los conflictos internos de Canal 9 llevan la situación a tal extremo que origina la salida al aire de un nuevo canal de TV por la frecuencia 6 y que correspondía al sector antagónico al Gobierno del Presidente Allende en la Universidad de Chile. El conflicto se resuelve cuando una comisión negociadora acuerda con el sindicato de Canal 9 que devuelva los estudios de transmisión, a cambio de que la Universidad de Chile venda los equipos a la Universidad Técnica del Estado, quienes transmitirían en esa frecuencia reservando el Canal 6 para la Universidad de Chile. Este acuerdo se trunca dos días después al producirse el golpe de estado del 11 de septiembre de 1973.

Finalmente la Universidad de Chile decide recuperar la frecuencia 9, sin embargo las dificultades económicas obligan a disminuir la producción propia e incrementar la envasada, programándose en su mayoría series extranjeras.

En 1978 comienza a transmitir en color junto a las otras operadoras de TV. Un par de años más tarde una inyección de recursos a través de un crédito permite renovar los equipos y el Canal cambia de frecuencia desde la 9 a la 11.

El endeudamiento del canal crece a pesar de las distintas estrategias para conseguir más publicidad y el deseado autofinanciamiento. Se suceden diferentes directores, hay más intervención de la dictadura y su audiencia no mejora. En 1989 la deuda era cercana a cinco veces el patrimonio de Canal 11 y en este escenario una modificación a la normativa permite el ingreso de más actores al mercado televisivo, Megavisión Canal 9 y Red Televisión Canal 4.

El inicio de la década de los noventa no presenta una situación distinta en relación al financiamiento y nuevamente las estrategias y planes de revertir la situación fracasan. La deuda hacia fines del año 1993 de Teleonce se acercaba a los dos millones de dólares y la Universidad decide formar una Sociedad Anónima: RTU de la Universidad de Chile. Esta sociedad enfocaría el desarrollo de Canal 11 hacia el deporte, la cultura y la familia, esperándose como resultado ocupar el lugar cuarto en la venta de publicidad. La Sociedad se formó con un 49% de las acciones en manos del Grupo Cisneros de Venezuela que pagó ocho millones de dólares por ellas y el 51% restante quedaba en propiedad de la Universidad.

Al cabo del primer año Chilevisión, que fue la nueva denominación de la señal televisiva, mostraba pérdidas por mil millones de pesos, pero la red abarcó una buena parte del país y la infraestructura se mejoró significativamente. Sin embargo, la sociedad no funcionó y el Rector Lavados decide vender el 51% restante de la sociedad, manteniendo la titularidad de la concesión de la frecuencia que fue entregada a nuestra Universidad por decreto. En el contrato se especifica que se entrega el usufructo de la frecuencia hasta el año 2018. El año 2004, Chilevisión es comprada por Bancard, controlada por la familia Piñera, y en el año 2011 es traspasada al grupo Times Warner, actuales propietarios.

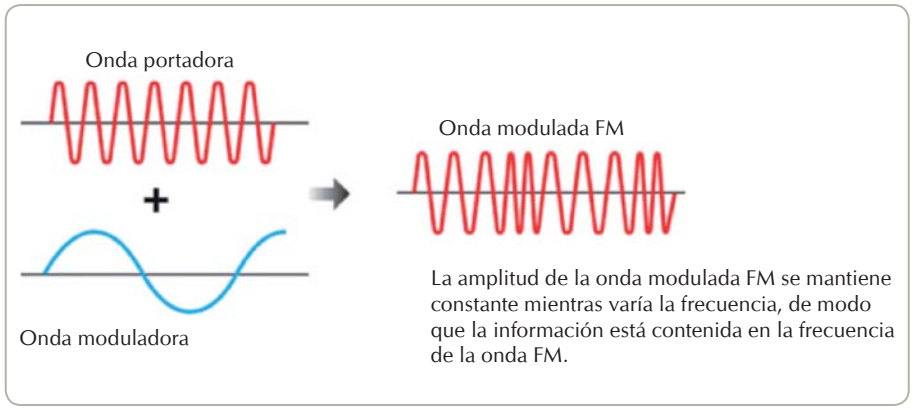
TELEVISIÓN ANALÓGICA

La televisión en su concepto más básico es el proceso de convertir imágenes estacionarias o en movimiento a señales eléctricas, transmitirlas vía ondas electromagnéticas y reconvertir esas señales a las imágenes capturadas en la fuente.

La televisión abierta, como cualquier otra fuente que haga uso del espectro electromagnético, está regulada en el uso de este espectro. Así la autoridad reguladora estableció que en Chile cada canal de televisión podía ser transmitido en un intervalo del espectro de 6 MHz. Para ello, la Subsecretaría de Telecomunicaciones (Subtel) destinó dos intervalos del espectro o bandas. La banda que va desde los 54 MHz hasta los 216 MHz, denominada VHF (Very High Frequency) por sus siglas en inglés y desde los 512 MHz hasta los 698 MHz, denominada UHF (Ultra High Frequency), para otorgar concesiones a lo largo del país a los solicitantes cumpliendo los requisitos técnicos.

La forma cómo se transmite la señal de audio y video hasta ahora, es modulando una portadora en frecuencia o en amplitud. En la Figura 1 se muestra una señal sinusoidal pura o portadora que, al ser modulada por otra señal de frecuencia más baja, produce una variación de la frecuencia donde va contenida la información que se está transmitiendo. Como la variación de frecuencia (también puede ser la amplitud) de la portadora es proporcional a la frecuencia de la señal que modula, la resultante es una señal que varía de forma análoga a la que modula. En términos de ingeniería de comunicaciones, este proceso se conoce como modulación analógica o análoga y como ésta es la forma de transmisión del audio y del video en la televisión, se le denomina televisión análoga o analógica.

Figura 1

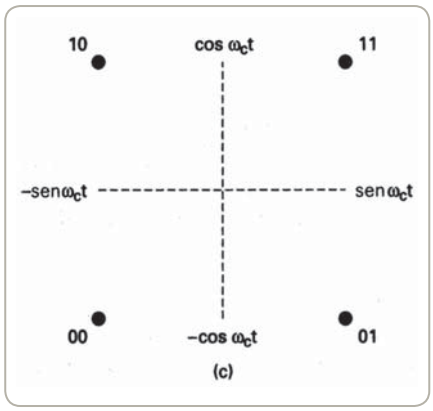


Señal sinusoidal pura o portadora, que al ser modulada por otra señal de frecuencia más baja, produce una variación de la frecuencia donde va contenida la información que se está transmitiendo.

TELEVISIÓN DIGITAL

Cualquier proceso de comunicación tiene por objetivo transmitir información entre dos puntos no vecinos en el espacio físico. Una forma alternativa de transmitir la información análoga producida por una fuente, es tomar muestras de ellas, codificarlas en un sistema binario, montar estos códigos en una portadora (la misma señal sinusoidal portadora del caso analógico), transmitirla y recuperarla en el receptor, decodificándola. En la Figura 2 se grafica la correspondencia entre las fases de la onda y la codificación transmitida. Esta codificación se denomina QPSK.

Figura 2



Correspondencia entre las fases de la onda y la codificación transmitida. Esta codificación se denomina QPSK.

La ventaja de este método de transmitir la información se basa en que los códigos pueden llegar muy deteriorados al receptor, debido a interferencias electromagnéticas encontradas en la trayectoria recorrida por la portadora. Sin embargo, en el receptor, ellos aún pueden ser identificados y reconstruidos, permitiendo recuperar la información tal como salió de la fuente. Esta forma de transmitir información se denomina comunicaciones digitales, porque la información es codificada digitalmente para ser transmitida.

La televisión digital hace uso de estos mismos conceptos. Es más, como las posibilidades de introducir códigos en las portadoras de información es muy alta hoy día, aumenta la eficiencia de uso del espectro electromagnético ya que es posible enviar más información por unidad de frecuencia que en la transmisión analógica. Ésta es una ventaja clara de esta forma de transmisión, puesto que tanto la imagen como el audio pueden ser reconstruidos en el receptor con una mejor calidad que en el caso análogo.

Por otra parte, la televisión digital puede usar codificaciones diferentes para distintas fuentes de información y utilizar la misma banda de frecuencias para transmitirlas. En este caso una información no afecta a la otra, puesto que están codificadas en forma distinta y pueden ser reconstruidas independientemente en el receptor.

Como es posible concentrar mayor cantidad de información por unidad de frecuencia, el canal de 6 MHz puede ser dividido en tres o cuatro canales distintos de definición estándar, para transmitir tres o cuatro programas distintos en la misma porción de espectro electromagnético, en que la televisión análoga puede transmitir sólo uno. Otra ventaja la constituye el hecho de que en la misma banda se pueda transmitir una señal en alta definición que utiliza el espectro de los cuatro canales de definición estándar y un canal para equipos móviles.

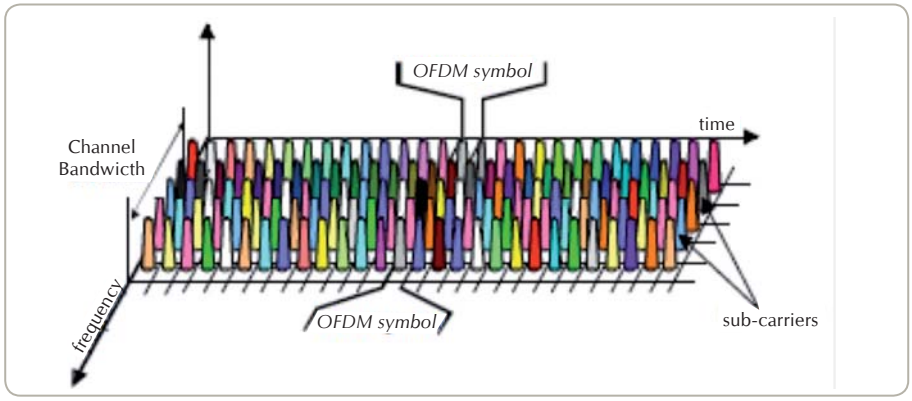
Finalmente, en esta revisión conceptual daré algunos comentarios sobre la norma seleccionada por Chile para la televisión digital abierta (ISDB-Tb). Esta norma utiliza para el transporte de la información codificada un número grande de portadoras (entre dos mil y ocho mil), que se caracterizan por ser ortogonales entre ellas (esto es, la fase de las ondas electromagnéticas sinusoidales están 90° adelantadas unas con otras), tal como se muestra en la Figura 3.

Como cada una de ellas lleva una fracción de la información codificada, cualquier pérdida de información en la trayectoria de propagación puede ser recuperada en el receptor aplicando algoritmos que determinan la información faltante. Esta característica hace que la recuperación de las imágenes y el audio en el receptor sea muy robusta en condiciones de interferencia en la trayectoria, por reflexiones múltiples en accidentes geográficos o vegetación muy densa. Por otra parte la norma considera un protocolo de transmisión de datos que permite que se transfieran datos en las mismas condiciones de propagación del audio o video de televisión. Así es posible transferir video, archivos de datos de imágenes médicas para aplicaciones docentes, presentaciones audiovisuales aplicadas a cursos de programas académicos, etc., lo que constituye en realidad una Plataforma de Servicios Digitales que puede ofrecer muchos más servicios que la televisión. En Chile sólo falta que el Congreso apruebe la legislación para el uso de esta tecnología en sus señales abiertas y permita que la mayor oferta de contenidos mejore la situación actual, dependiente casi exclusivamente de los recursos financieros destinados a la publicidad.BITS

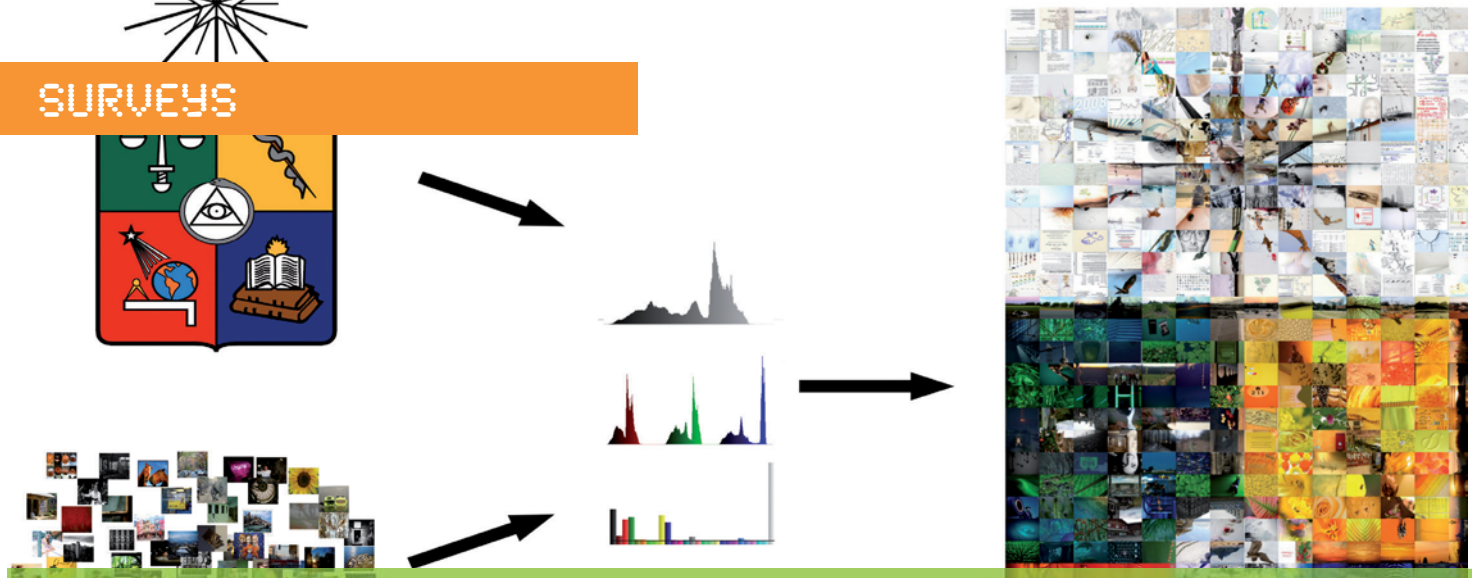
REFERENCIAS

[1] María de la Luz Hurtado. "Historia de la Televisión Chilena entre 1959 y 1973". CENECA. Octubre 1989.

Figura 3



Estándar ISDB-Tb: Múltiples Portadoras desfasadas en 90° cada una. Esta norma utiliza para el transporte de la información codificada, un número grande de portadoras (entre dos mil y ocho mil) que se caracterizan por ser ortogonales entre ellas.

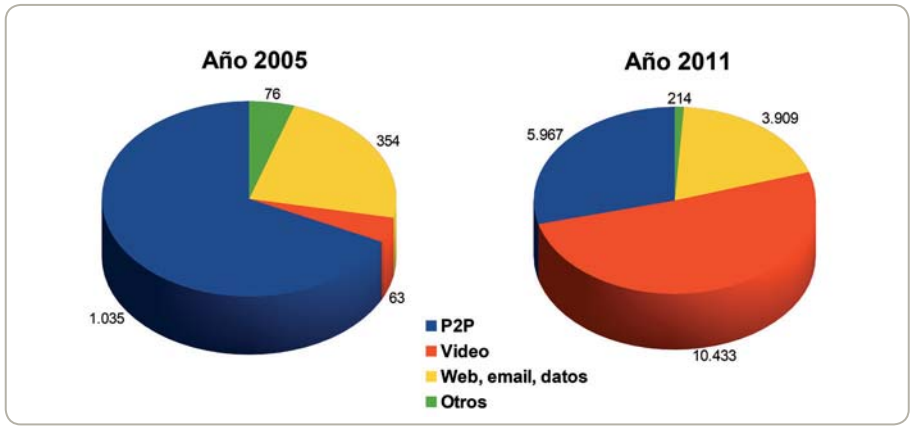


Búsquedas por contenido en imágenes y videos

Juan Manuel Barrios
Doctor (c) en Ciencias mención Computación, Universidad de Chile. Ingeniero Civil en Computación y Magíster en Ciencias mención Computación Universidad de Chile. Director de Investigación Orand S.A.
jbarrios@dcc.uchile.cl.

El tráfico de datos en Internet ha aumentado exponencialmente en la última década. En 2005 las redes P2P dominaban las estadísticas del tráfico de datos de usuarios de Internet, mientras que la visualización de videos en línea recién comenzaba a masificarse con el lanzamiento de YouTube. Seis años después, el tráfico global de datos de usuarios ha aumentado más de doce veces, y específicamente la visualización de videos (esto es, reproducción de cortos de videos, visualización de películas, streaming de canales televisión, webcams públicas, etc.) corresponde a más de la mitad (ver Figura 1). Este crecimiento muestra un sorprendente aumento en el tráfico de datos, y en particular del impacto de los videos en el uso de ancho de banda.

Figura 1



Tráfico de datos de usuarios de Internet para 2005 y 2011 (fuente: [1] y [2]).

Por otra parte, la investigación académica en tópicos de video se ha desarrollado desde hace varias décadas. Por ejemplo, la industria del software ha provisto de aplicaciones para edición de videos desde previo a este siglo. Así también, algunas de las técnicas que veremos más adelante provienen de las decadas de 1970 y 1980. Sin embargo, la actual masificación y universalidad en el uso de videos crea nuevas necesidades y presiona el desarrollo de nuevos y mejores algoritmos. Actualmente, muchos usuarios, investigadores y empresas estarán interesados en explorar posibilidades como administración, análisis, y gestión de grandes colecciones de videos, monitoreo y automatización de procesos en tiempo real, etc.

Multimedia Information Retrieval (MIR) es el área de investigación que tiene por objetivo la búsqueda y recuperación de información semántica desde documentos multimedia [3]. Un documento multimedia se puede entender como cualquier repositorio de información, ya sea estructurado o no. En particular, en este artículo nos enfocaremos en documentos audiovisuales, esto es imágenes, audio y videos, aunque un documento multimedia también comprende fuentes de información más genéricas como grafos, series de tiempo, páginas web, documentos XML, secuencias de ADN, etc. En general, se pueden destacar dos procesos fundamentales en cada sistema MIR: un proceso de *descripción de contenido* que calcula uno o más descriptores para cada documento, y un proceso de *búsqueda por similitud* que analiza la distribución de descriptores para encontrar descriptores parecidos de forma efectiva y eficiente.

Computer Vision (CV) es un área de investigación muy vinculada con MIR. Ambas áreas comprenden la adquisición y procesamiento de imágenes y videos, y la toma de decisiones según el análisis de descriptores. La principal diferencia proviene del contexto de uso: CV se enfoca mayormente en procesos en tiempo real, por ejemplo, detectar cuando un objeto específico aparece frente a la cámara de un robot; mientras que MIR se enfoca mayormente en técnicas de búsqueda y análisis de grandes colecciones, por ejemplo, buscar un objeto específico en imágenes de Internet.

PROCESAMIENTO DE IMÁGENES

Llamaremos procesamiento de imágenes a las técnicas que dada una imagen de entrada producen una imagen de mejor calidad de salida. El objetivo básico de estas técnicas es facilitar o mejorar la extracción de características.

Una imagen es una señal bidimensional discreta $I(x,y)=c$, donde cada celda (x,y) se denomina píxel ("picture element"). En imágenes grises c corresponde a un valor unidimensional con la intensidad del píxel, mientras que en imágenes en color c es un valor de (al menos) tres dimensiones. Cada dimensión de c se denomina canal, y la cantidad de valores posibles para representar cada canal se denomina profundidad. Comúnmente se usa profundidad 8 bits/canal, lo que permite representar hasta 2^8 tonos de gris o hasta $(2^8)^3 \approx 16$ millones de colores.

Operadores

Los operadores puntuales modifican el valor de cada píxel en forma independiente: dada una imagen de entrada I un operador puntual f produce una imagen de salida I' definida como $I'(i,j)=f(I(i,j))$. Algunos ejemplos de operadores puntuales son el filtro de binarización ($f(x)=0$ si $x<t$, o 1 si $x \geq t$), corrección gamma ($f(x)=x^\gamma$) y ajuste de brillo y contraste ($f(x)=ax+b$).

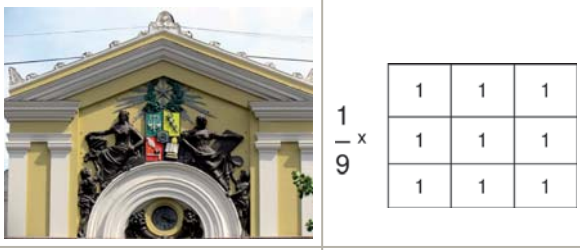
La convolución es un operador lineal donde el nuevo valor de un píxel corresponde a una combinación lineal de los píxeles de su vecindad. A la ventana de ponderación se le conoce como filtro, máscara, o kernel. Existen diferentes filtros usados con diferentes objetivos. Por ejemplo, el efecto de desenfoque o *blur* se logra con la convolución de una imagen con un filtro promedio (ver Figura 2) o con un filtro gaussiano (en general, un filtro pasa bajo). Estos filtros descartan información de detalles y son usados para disminuir ruido o para reducir el tamaño de la imagen. Otros filtros comúnmente usados son Laplaciano (para resaltar detalles) y filtros de Roberts, Prewit y Sobel (para detección de bordes).

Algunos operadores no lineales comunes son el filtro de mediana, donde el nuevo valor es la mediana de los valores de la vecindad de cada píxel; filtro bilateral, el que es un filtro gaussiano con ponderaciones dinámicas por píxel; y filtros morfológicos, que marcan formas relevantes en la imagen según cómo se comportan al expandirse o contraerse al usar un elemento estructurante.

Detección de bordes

Los bordes corresponden a las zonas de la imagen donde existe un cambio abrupto en la tonalidad (ya sea de blanco a negro o de negro a blanco). Una técnica común para detectar bordes es el método del gradiente. En una función bidimensional el gradiente corresponde a un vector conteniendo las derivadas parciales según ambos ejes. La orientación del gradiente entrega la dirección de máximo cambio, mientras que su magnitud representa la intensidad de ese cambio. La convolución con filtros de Sobel permite aproximar las derivadas parciales, y estimar el valor del gradiente

Figura 2



a | b | c

a) Imagen original. b) Filtro promedio 3x3. c) Convolución con un filtro promedio.

Figura 3

-1	0	1	-1	-2	-1	0	1	0
-2	0	2	0	0	0	1	-4	1
-1	0	1	1	2	1	0	1	0

a **b** **c**

a) Filtros de Sobel ejes *x* e *y* . b) Filtro Laplaciano. c) Resultado de detección de Canny.

para cada píxel. Los píxeles de borde se definen como los puntos donde la magnitud del gradiente supera cierto valor umbral. Esta técnica tiende a dar bordes gruesos, los que pueden ser adelgazados usando un criterio basado en la segunda derivada. El Laplaciano se define como la suma de las segundas derivadas parciales de una función, y puede ser estimado por medio de una convolución con el filtro del mismo nombre. Luego, se pueden obtener bordes delgados si seleccionan los píxeles donde la magnitud del gradiente supera cierto umbral *t* y además el valor del Laplaciano es cero. Esta técnica fue formalizada por Canny [4], quien además incluye dos mejoras: realiza una selección incremental de puntos recorriendo en dirección perpendicular al gradiente (es decir, "caminando" por sobre el borde), e incluye un umbral incerteza t', donde un píxel cuya magnitud de gradiente está entre *t'* y *t* puede ser seleccionado como borde si es contiguo a un píxel que ya fue seleccionado como borde (ver Figura 3).

Una segunda técnica para detectar bordes es la Diferencia de Gaussianas (DoG). Ésta consiste en aplicar un filtro gaussiano y restar la imagen desenfocada con la original. Como el desenfoque afecta mayormente las zonas donde hay gran variación, al comparar el desenfoque con la imagen original las mayores diferencias se producen en los píxeles sobre el borde. Para reducir el nivel de ruido, se usa una imagen base I_1 con filtro gaussiano de desviaciones estándar σ, luego se calcula una imagen I_2 con un filtro gaussiano $k\sigma$ (para cierto paso k), y finalmente se aplica una binarización sobre la imagen de diferencia $I'(x,y) = |I_2(x,y)-I_1(x,y)|$ (ver Figura 4).

Figura 4

Resultado de bordes usando DoG.

En el caso de imágenes de objetos, el Análisis de Componentes Principales (PCA) permite obtener una imagen invariante a rotaciones. PCA busca un sistema de coordenadas donde los datos no tengan correlación lineal por medio de calcular valores y vectores propios de la matriz de covarianzas. En este caso, aplicando PCA sobre las coordenadas de los puntos de borde, con una rotación inversa de acuerdo al eje principal encontrado (esto es, el vector propio asociado al mayor valor propio) se puede normalizar la orientación de la imagen en función de su contenido (ver Figura 5). Cabe señalar que esta técnica no funciona correctamente con formas mayormente simétricas.

Figura 5

a) Imagen original. b) Uso de PCA para determinar los ejes principales.

Una vez seleccionados los píxeles de bordes, una tarea común es la de buscar diferentes estructuras que ellos forman, como rectas, circunferencias u otras estructuras paramétricas*. Dado un tipo de estructura a buscar (por ejemplo, rectas), existen dos técnicas comúnmente usadas para determinar la existencia de esa estructura y sus parámetros. Random Sample Consensus (RANSAC) es un algoritmo aleatorio que iterativamente crea y evalúa distintos candidatos para seleccionar el que obtuvo un mayor apoyo. Por ejemplo, en el caso de buscar rectas, se elige un par de puntos al azar, se calcula la ecuación de recta y se cuentan los puntos de borde que están sobre esa recta. Con suficientes intentos, se encontrará la recta que pasa sobre más puntos de borde. Sin embargo, cuando la probabilidad de encontrar un modelo correcto por selección aleatoria es muy baja (es decir cuando el modelo buscado es cumplido por muy pocos puntos) es recomendable preferir una técnica alternativa. La Transformada de Hough es un algoritmo exhaustivo y determinístico que reduce los parámetros posibles de la estructura a un conjunto finito. La ecuación de la recta se determina por dos parámetros, por tanto el espacio bidimensional de los parámetros se discretiza a una matriz de cierta dimensión fija. Para evitar problemas con los rangos de parámetros se debe usar un sistema de coordenadas polares. Cada punto de borde agrega un voto a los parámetros de todas las rectas que pasan por ese punto. La estructura buscada está definida por los parámetros que obtienen mayor votación.

DESCRIPCIÓN DEL CONTENIDO

La descripción del contenido consiste en analizar y resumir el contenido de cada documento creando uno o más descriptores o vectores característicos. Los descriptores de alto nivel representan características semánticas, como metadatos, tags o anotaciones, mientras que los de bajo nivel corresponden a estadísticas o patrones del contenido que pueden ser creados automáticamente, como promedios,

* Un caso particular es el de segmentación por contornos activos, tema que fue revisado en el número 5 de la Revista Bits de Ciencia.

varianzas e histogramas. Un tópico amplio de investigación es la generación automática de descriptores de alto nivel a partir de información de bajo nivel, es decir, predecir los conceptos que una persona asignaría a un documento analizando su contenido. Este problema está ligado a las áreas de machine learning, data mining, e inteligencia artificial. En el contexto de búsquedas por contenido, la diferencia que se produce entre los conceptos que un usuario asigna a un documento y los que un computador puede asignar automáticamente es conocida como "brecha semántica" (semantic gap) [31].

Descriptores globales de imágenes

Un descriptor global es un valor o vector que representa el contenido de toda la imagen. El nivel de parecido entre dos descriptores se mide por medio de una función de distancia o de disimilitud. Cuando la distancia entre dos descriptores es cercana a cero, es decir, cuando son espacialmente cercanos, se espera que las dos imágenes que produjeron esos descriptores sean parecidas entre sí. Una familia de distancias comúnmente usadas para comparar vectores son las distancias de Minkowski:

$$L_p(\vec{x},\vec{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}} \text{ para } \quad p \geq 1$$

En particular, p=2 corresponde a distancia euclidiana.

Un descriptor global muy simple consiste convertir la imagen a grises, escalarla a un tamaño fijo de *WxH* píxeles, y crear un vector de *W·H* dimensiones, donde cada valor corresponde a la intensidad gris de cada píxel escalado. Los descriptores pueden ser comparados con distancia euclidiana. Este descriptor permite comparar directamente pares de imágenes, sin embargo, el valor de la distancia es altamente afectado por ajustes simples de brillo o contraste.

Una variante común es usar la medición ordinal (ordinal measurement) [6], que consiste en reemplazar los valores gris de cada píxel por su posición relativa con respecto a los demás píxeles al ordenarlos de menor a mayor. Por ejemplo, si la imagen se reduce a 5x5 la medida ordinal es una permutación de los números 1 a 25 donde 1 corresponde al píxel más oscuro y 25 al más claro. Este descriptor es invariante a cambios en brillo y contraste, por lo que es usado para buscar imágenes duplicadas, sin embargo es altamente afectado por modificaciones parciales como inserción de subtítulos o logos.

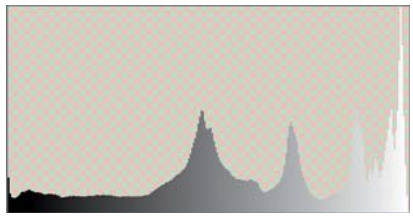
Otra variante es reducir la dimensionalidad de los descriptores usando PCA. Esto consiste en determinar los ejes principales, rotar los descriptores según estos ejes y luego descartar los ejes de menor varianza. Esta alternativa es especialmente útil cuando las imágenes de la colección tienen características comunes, por ejemplo una colección de rostros de personas. En este caso usando PCA se obtienen vectores propios que al ser combinados linealmente permiten obtener las imágenes originales de la colección. Esta es la idea base de la técnica conocida como *eigenobjects* para reconocer objetos, y que es aplicada en la autenticación de personas bajo el nombre de *eigenfaces* (Ver Figura 6).

Figura 6

Ejemplos de vectores propios (*eigenfaces*) de un conjunto de imágenes de rostros (fuente: prtools.org).

En una imagen en escala de grises, el histograma normalizado contiene la fracción de píxeles que tiene cada nivel de gris descartando su ubicación espacial. Dadas n observaciones (pixeles) y un conjunto de *k* categorías (rangos de intensidades) un histograma cuenta el número de observaciones que caen dentro de cada categoría o bin (ver Figura 7). El histograma puede ser visto como una distribución de probabilidad de que un píxel al azar tenga cierto valor de gris. Esta información es útil tanto para el procesamiento de imágenes como para representar el contenido. Un histograma es un vector de *k* dimensiones que puede ser comparado por distancias vectoriales o por test estadísticos como χ^2 o divergencia Kullback-Leibler.

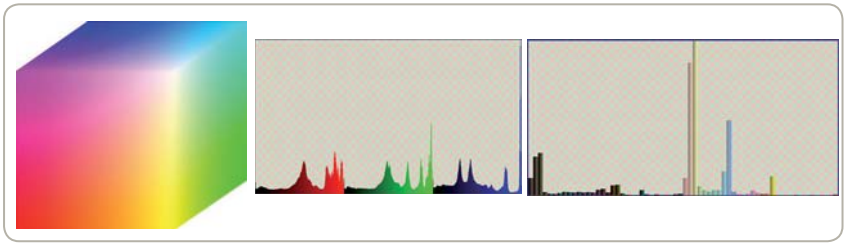
Figura 7



Histograma de grises para la imagen 2-a.

En el caso de histogramas de color existen diferentes enfoques. Un primer enfoque es calcular un histograma para cada canal en forma independiente, de esta forma el histograma de color es la concatenación de tres histograma de grises. Un segundo enfoque (y el más utilizado) es crear una partición regular fija del espacio de colores y calcular un histograma tridimensional. Por ejemplo, el espacio de colores RGB forma un cubo regular, dividiendo cada canal en cuatro rangos se obtendrán histogramas de $4^3=64$ bins (ver Figura 8). En este caso, la comparación de dos histogramas se puede realizar con distancia euclidiana, sin embargo, esta distancia no considera similitudes entre bins. La distancia Forma Cuadrática permite calcular distancias entre vectores usando una matriz de similitud entre dimensiones. Esta función permite ser más precisos en el cálculo de similitud entre histogramas, a un costo de más operaciones. La matriz de similitud se puede definir por medio de distancias de colores en el cubo RGB, sin embargo

Figura 8



Cubo RGB, histograma de color por canal, histograma de 27 colores de la imagen 2-A.

esto implica que la similitud entre blanco y verde es igual que entre blanco y azul, aún cuando perceptualmente la primera es más parecida que la segunda. Los espacios de color HSV, HLS y variantes corresponden a transformaciones geométricas de RGB. Se puede calcular histogramas directamente sobre estos espacios de colores, sin embargo en la práctica siguen estando fuertemente ligados a RGB. Una alternativa diferente es basarse en representaciones perceptuales del color, como las definidas por la Commission Internationale de L'Eclairage (CIE). Esta comisión desde los años veinte estudia la naturaleza física y perceptual del color, definiendo modelos basados en experimentos con observadores humanos. Durante los años setenta se definieron los espacios L*a*b* y L*u*v* que permiten calcular distancia entre colores asemejando la similitud percibida por un observador. Estos espacios de colores pueden ser usados para el cálculo de matrices de similitud entre colores.

Un tercer enfoque para crear histogramas de color consiste en utilizar una división dinámica del espacio de colores dependiendo de los colores en la imagen a representar. Esto es, determinar la mejor división y número de bins para representar fielmente el contenido de cada imagen. Esta división se logra por medio del análisis de la distribución de colores y/o por medio de clustering en el espacio de colores. Una vez creados los histogramas, su comparación se puede realizar con las funciones Earth Mover's Distance [7], que resuelve un problema de transporte entre bins según una función de costo, y Signature Quadratic Form [8], que extiende la forma cuadrática para incluir matrices de similitud intrahistogramas e interhistogramas.

Los bordes de la imagen entregan información valiosa sobre su contenido. Una alternativa consiste en representar los gradientes en la imagen por medio de un histograma de orientaciones [9]. Otra alternativa consiste en dividir la imagen en grupos de 2x2 píxeles y usar filtros de orientación para testear el tipo de borde [10] [11] (ver Figura 9). Una tercera alternativa es determinar bordes según el método de Canny y luego representar las ubicaciones dominantes de los bordes [9].

La información de texturas también puede ser usada para representar el contenido. Para esto, un análisis de frecuencias de la imagen revela patrones de textura en la imagen. Un descriptor consiste en obtener la transformada de Fourier de la energía y representar la energía de diferentes zonas del espacio de frecuencias de la imagen por medio de la energía obtenida por filtros Gabor [10]. Otra alternativa consiste en describir la imagen según el valor de los primeros coeficientes de la transformada discreta coseno [12]. Estos coeficientes tienen la propiedad que permiten reconstruir la imagen original a una aproximación ajustable según el número de coeficientes guardados.

Figura 9

1	-1	1	1	√2	0	0	√2	2	-2
1	-1	-1	-1	0	-√2	-√2	0	-2	2

Filtros de orientaciones de bordes.

Descriptores locales de imágenes

Un descriptor local representa sólo una pequeña zona de la imagen, y por tanto la imagen se representa por una cantidad variable de descriptores (del orden de cientos o miles para una sola imagen). Existen diferentes enfoques para decidir la ubicación y el tamaño de las zonas de interés en una imagen. Un enfoque es la detección de esquinas, que se basa en seleccionar zonas que mantienen una alta diferencia consigo misma bajo cualquier desplazamiento pequeño. Algunos algoritmos que usan esta idea son los detectores de Harris-Stephens [5], Shi-Tomasi y Harris-Laplace. Un segundo enfoque es la detección de manchas, que se basa en localizar zonas oscuras dentro de zonas claras (o zonas claras rodeadas de zonas oscuras). Estas zonas se pueden localizar por medio de detectores puntuales basados en DoG, o de formas arbitrarias llamadas Maximally Stable Extremal Regions (MSER). Otro enfoque es el llamado *dense sampling*, que consiste en utilizar una grilla densa de zonas de interés. Esta grilla permite obtener una gran cantidad de regiones, incluso en zonas donde los detectores anteriores detectan muy pocas.

Para cada zona de interés se calcula un descriptor del contenido en la zona. Los descriptores locales más usados son SIFT [13], SURF [14], y alguna de sus variaciones y extensiones. El descriptor SIFT divide la zona de interés en 4x4 regiones y calcula histogramas de orientaciones del gradiente, produciendo un vector de 128 dimensiones (ver Figura 10). Una variación común es PCA-SIFT [15] que reduce el largo de los vectores. SURF es un vector de 64 dimensiones que también representa las orientaciones del gradiente, pero usando la imagen integral para un procesamiento más rápido. SURF muestra casi los mismos resultados de efectividad que SIFT a un menor costo de procesamiento.

Figura 10

Descriptores SIFT sobre dos imágenes.

Para comparar dos imágenes I_1 e I_2, cada descriptor local de I_1 es comparado con cada descriptor local de I_2. En el caso de SIFT esta comparación se realiza con distancia euclidiana. Un descriptor *p* de I_1 se asocia con un descriptor *q* de I_2 cuando se cumplen dos condiciones: 1) *q* es el más cercano a *p* entre todos los descriptores de I_2 de acuerdo a la distancia entre descriptores; 2) la razón entre la distancia de *p* a *q* con la distancia de *p* al segundo más parecido en I_2 es menor a un parámetro *s* (usualmente 0.8). El primer criterio exige que se asocien zonas parecidas entre sí, mientras que el segundo criterio descarta calces que no sean suficientemente seguros o discriminativos (ver Figura 11).

Figura 11

Calces entre imágenes representadas por descriptores locales.

Una vez asociados descriptores locales similares, un proceso de correspondencia geométrica determina el mayor subconjunto de asociaciones que cumplen con una misma función de transformación espacial. Para esto, primero se debe definir un modelo de transformación geométrica a buscar. Los más comunes son escala+traslación, rotación+escala+traslación, transformación afín, y transformación de perspectiva u homográfica. Cada modelo requiere de una cantidad mínima de asociaciones para estar definido (por ejemplo, escala+traslación se define por dos asociaciones, mientras que una de perspectiva por cuatro). Luego se debe determinar el modelo de transformación geométrica que es cumplido por la mayor cantidad de asociaciones. En este caso los algoritmos RANSAC y Transformada de Hough ya descritos son aplicables (ver Figura 12).

Figura 12

Subconjunto de calces con coincidencia espacial usando RANSAC.

La correspondencia geométrica entrega la transformación que se debe llevar a cabo en la I_1 para que la mayor cantidad de zonas coincida con I_2. Esta información de postura puede ser aplicada en la composición de imágenes (llamado *image stitching*) para producir vistas panorámicas (ver Figura 13). También puede ser aplicada en realidad aumentada, donde objetos virtuales son transformados para ser incluidos en una escena real manteniendo coherencia con su entorno.

Descriptores de videos

Un video es la composición de una secuencia de imágenes del mismo tamaño más una pista de audio. La imágenes, que en este contexto se les llama cuadros o frames, se deben mostrar a cierta velocidad mínima (del orden de 24 frames por segundo) para percibir un movimiento fluido en vez de imágenes aisladas. Un shot se define como una secuencia de frames consecutivos de una misma cámara representando una acción continua en tiempo y espacio [16]. Los detectores de límites de shots más comunes comparan frames consecutivos, y reportan un cambio de shot cuando la diferencia entre dos frames supera un umbral. Algunas variaciones de esta técnica utiliza diferencias por zonas, diferencias de histogramas, ventanas temporales de comparación y uso de umbrales adaptativos [17].

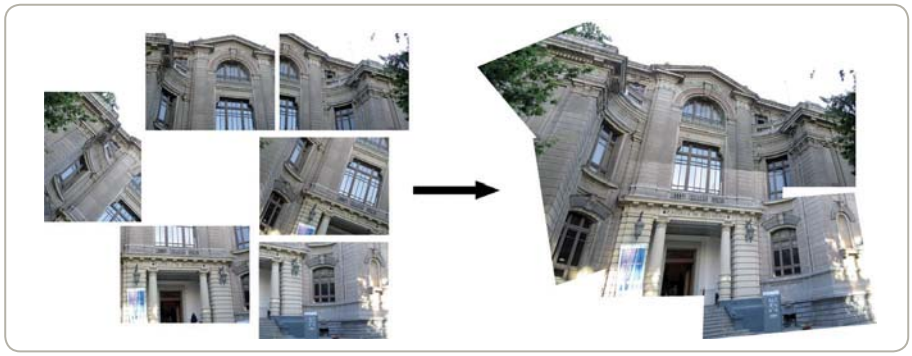
El flujo óptico consiste en calcular un vector de movimiento aparente entre frames consecutivos para cada píxel y/o grupos de píxeles. Es relativamente costoso de calcular, ya que requiere medir diferencias de frames consecutivos a distintas escalas y desplazamientos. El flujo óptico es usado para el tracking de objetos, codificación de videos y selección de keyframes.

Un keyframe es un frame que representa el contenido en una secuencia de video. En búsquedas por similitud, los keyframes permiten descartar frames redundantes y por tanto reducir la cantidad de información a describir. Los frames con mucho movimiento tienden a ser poco definidos, por tanto un criterio de selección es elegir frames mayormente estáticos. Para esto, una alternativa consiste en seleccionar los frames con menor diferencia con sus frames previos. Otra alternativa es usar el flujo óptico para seleccionar los frames cuyo largo total de vectores de movimiento sea mínimo. En el caso de querer seleccionar keyframes representativos para secuencias largas (por ejemplo, películas), se puede utilizar un algoritmo de clustering sobre frames para seleccionar los frames más cercanos a los centroides.

Para el cálculo de descriptores de videos, un primer enfoque conocido como hashing visual o video signature consiste en calcular un único descriptor para la secuencia completa. Este enfoque es útil para detectar videos duplicados con invariancia a pequeñas modificaciones visuales debidas a re-encoding o reducción de calidad [18]. Sin embargo, para poder detectar videos con segmentos comunes, se requiere calcular descriptores a un nivel más fino como keyframes o segmentos cortos.

La mayoría de los descriptores globales de imágenes pueden ser usados directamente sobre keyframes de videos. Una técnica común es calcular descriptores espacio-temporales para segmentos de videos por medio de agregación de descriptores de sus frames [19]. Otra técnica es calcular distancias temporales entre conjuntos de descriptores [20] [21].

Figura 13



Composición de imágenes al combinarlas según el modelo de transformación de perspectiva.

En el caso de descriptores locales para videos, una técnica consiste en calcular descriptores sobre keyframes y mantener los descriptores de las zonas de interés permanentes entre frames consecutivos [22]. Además, se pueden clasificar los tipos de zonas en estáticas o en movimiento [23]. También se ha realizado investigación para extender las técnicas de detección [24] y descripción [25] de zonas de interés para zonas espacio-temporales.

Con el objetivo de reducir la cantidad de descriptores locales producidos por un video y poder realizar búsquedas eficientes se desarrolló la técnica conocida como Bag-of-Visual-Words o Bag-of-Features [26]. Esta técnica se basa en usar un algoritmo de clustering sobre una gran cantidad de descriptores locales para determinar un conjunto discreto de N símbolos representativos, denominados vocabulario visual o codebook. Luego, los descriptores locales en una misma imagen o frame se resumen en un vector de dimensión N con las apariciones de cada símbolo. Las imágenes conteniendo cada símbolo se organizan en un índice invertido de N entradas. Luego, dada una imagen de consulta, el índice invertido permite determinar eficientemente todas las imágenes o frames que comparten un mismo símbolo. El índice invertido puede también contener información espacial de la aparición de cada símbolo, o en caso contrario, es necesario realizar un proceso de consistencia espacial entre imágenes o frames candidatos. Esta técnica ha sido foco de investigación durante los últimos años, abordando diferentes problemas como detección de videos duplicados [27], búsqueda de objetos en videos [28], clasificación de imágenes [29] y búsqueda de imágenes similares en la Web [30].

Una pista de audio es una señal unidimensional discreta $f(t)=a$, donde cada unidad de tiempo t se conoce como sample y el valor a corresponde a la amplitud de la onda de sonido. Una resolución de al menos ocho mil samples por segundo permite escuchar un sonido de calidad media mientras que 44 mil samples por segundo o más se considera de alta calidad. El análisis de audio se realiza por medio de ventanas pequeñas a intervalos regulares en la señal acústica. En vez de describir la señal directamente, los descriptores se basan en llevar la señal al dominio de las frecuencias (por medio de la Transformada de Fourier) y describir la energía de las frecuencias audibles. El descriptor más usado para audio es el Mel-frequency cepstral coefficients (MFCC) [32]. Otros descriptores son el de Philips [33], chroma features [34] y un descriptor enfocado en la duplicidad de pistas de audio [35].

Una vez descrito cada video por medio de descriptores globales, locales y/o acústicos ya sea por keyframes o shots, la comparación entre videos debe considerar todas estas variables. Las técnicas conocidas como *late fusion* [36] dividen la comparación en subsistemas independientes para cada tipo de descriptor y el resultado de cada uno se une ya sea por unión, intersección o agregación de scores. Por otra parte, las técnicas conocidas como early fusion intentan realizar una comparación combinada entre frames usando todos los descriptores en una única función de distancia [35].

Finalmente, una comparación de videos basada en keyframes requiere de un proceso de correlación espacio-temporal entre videos candidatos. La comparación entre dos videos puede ser modelada por un grafo bipartito donde los keyframes son nodos y la similitud entre keyframes son pesos de aristas. La similitud global de dos videos se puede determinar mediante un algoritmo de flujo máximo [37]. Otra alternativa es extender los modelos de coherencia espacial a modelos de coherencia espacio-temporal [38].

APLICACIONES

Existen diferentes aplicaciones para las técnicas de análisis y búsqueda de imágenes y videos. Por ejemplo, mediante análisis en tiempo real se pueden lograr usos industriales como inspección automática de control de calidad, autenticación biométrica, monitoreo y detección de eventos en cámaras de seguridad, o interfaces humano-computador a través de cámaras, etc. Mediante análisis de grandes volúmenes de imágenes y videos se pueden resolver problemas como organización automática de colecciones multimedia, detección de imágenes y videos similares, descubrimiento de imágenes o videos relacionados, detección de falsificación de imágenes, asignación automática de etiquetas semánticas, reconocimiento de eventos multimedia, etc. Además, existen subáreas especializadas para distintos dominios como el análisis de imágenes médicas, satelitales y astronómicas.

En el caso específico de videos, TRECVID es una conferencia organizada anualmente por el National Institute of Standards and Technology (NIST) de EE.UU que fomenta la investigación en la recuperación de información en videos [39]. Esta conferencia presenta problemas o desafíos relacionados con el análisis de videos, publica colecciones de referencia, y evalúa la participación de equipos que se inscriben libremente. Se han realizado evaluaciones para la detección de límites de shots, detección de copias de videos, búsquedas de objetos conocidos, monitoreo de cámaras de seguridad, asignación de etiquetas y detección de eventos multimedia. En estas evaluaciones es común ver equipos de universidades y empresas de diferentes partes del mundo, lo que prueba que el análisis de documentos multimedia es un tema desafiante que requiere de la participación de equipos multidisciplinarios, y la unión de la industria con la academia prueba la relevancia académica del área así como su proyección económica.

Un equipo del DCC de la Universidad de Chile participó en TRECVID en la evaluación de detección de copias de videos durante 2010 y 2011. Los resultados fueron satisfactorios al ser comparados con otros equipos y sistemas del estado del arte [20] [35]. El software desarrollado durante esa participación, llamado P-VCD, ha sido liberado como Open Source con licencia GPL. Actualmente, la empresa Orand ha apoyado el desarrollo de este software, con el que se ha participado en la evaluación de búsqueda de objetos (Instance Search) en TRECVID 2012 [40].BITS

REFERENCIAS

[1] Cisco Systems Inc. Global IP Traffic Forecast and Methodology, 2006–2011. White paper, 2007.

[2] Cisco Systems Inc. Cisco Visual Networking Index: Forecast and Methodology, 2011-2016. White paper, 2012.

[3] M. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. ACM Transactions on Multimedia Computing, Communications and Applications, 2(1):1-19, 2006.

[4] Canny, J., A Computational Approach To Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.

[5] C. Harris and M. Stephens. A combined corner and edge detector. In Proc. of the Alvey Vision Conference, 147-151. The Plessey Company, 1988.

[6] C. Kim. Content-based image copy detection. Signal Processing: Image Communication, 18(3):169-184, 2003.

[7] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision, 40(2):99-121, 2000.

[8] C. Beecks, M. S. Uysal, and T. Seidl. Signature Quadratic Form Distance. In Proc. of ACM Int. Conf. on Image and Video Retrieval (CIVR), 438-445, 2010.

[9] A. Hampapur and R. Bolle. Comparison of distance measures for video copy detection. In Proc. of the IEEE int. conf. on Multimedia and Expo (ICME), 737-740. IEEE, 2001.

[10] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. IEEE Transactions on Circuits and Systems for Video Technology, 11(6):703-715, 2001.

[11] K. Iwamoto, E. Kasutani, and A. Yamada. Image signature robust to caption superimposition for video sequence identification. In Proc. of the int. conf. on Image Processing (ICIP), 3185-3188. IEEE, 2006.

[12] X. Naturel and P. Gros. A fast shot matching strategy for detecting duplicate sequences in a television stream. In Proc. of the int. workshop on Computer Vision meets Databases (CVDB), 21-27. ACM, 2005.

[13] D. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91-110, 2004.

[14] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (SURF). Computer Vision and Image Understanding, 110(3):346-359, 2008.

[15] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In Proc. of the intl. conf. on Computer Vision and Pattern Recognition (CVPR), II-506-513. IEEE, 2004.

[16] A. Hanjalic. Shot-boundary detection: unraveled and resolved? IEEE Transactions on Circuits and Systems for Video Technology, 12(2):90-105, 2002.

[17] J. S. Boreczky and L. A. Rowe. Comparison of video shot boundary detection techniques. Journal of Electronic Imaging, 5(2):122-128, 1996.

[18] B. Coskun, B. Sankur, and N. Memon. Spatio-temporal transform based video hashing. IEEE Transactions on Multimedia, 8(6):1190-1208, 2006.

[19] X. Wu, A. G. Hauptmann, and C.-W. Ngo. Practical elimination of near-duplicates from web video search. In Proc. of the int. conf. on Multimedia (ACMMM), 218-227. ACM, 2007.

[20] J. M. Barrios and B. Bustos. Competitive content-based video copy detection using global descriptors. Multimedia Tools and Applications, Springer, 2012.

[21] C. Kim and B. Vasudev. Spatiotemporal sequence matching for efficient video copy detection. IEEE Transactions on Circuits and Systems for Video Technology, 15(1):127-132, 2005.

[22] X. Anguera, T. Adamek, D. Xu, and J. M. Barrios. Telefonica research at trecvid 2011 content-based copy detection. In Proc. of TRECVID. NIST, 2011.

[23] J. Law-To, O. Buisson, V. Gouet-Brunet, and N. Boujemaa. Robust voting algorithm based on labels of behavior for video copy detection. In Proc. of the int. conf. on Multimedia (ACMMM), 835-844. ACM, 2006.

[24] G. Willems, T. Tuytelaars, and L. V. Gool. An efficient dense and scale-invariant spatio-temporal interest point detector. In Proc. of the european conf. on Computer Vision (ECCV), 650-663. Springer, 2008.

[25] G. Willems, T. Tuytelaars, and L. V. Gool. Spatio-temporal features for robust content-based video copy detection. In Proc. of the int. conf. on Multimedia Information Retrieval (MIR), 283-290. ACM, 2008.

[26] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In Proc. of the IEEE int. conf. on Computer Vision (ICCV), 1470-1477. IEEE, 2003.

[27] M. Douze, A. Gaidon, H. Jegou, M. Marsza lek, and C. Schmid. Inria lear's video copy detection system. In Proc. of TRECVID. NIST, 2008.

[28] D.-D. Le, C.-Z. Zhu, S. Poullot, V. Q. Lam, D. A. Duong, and S. Satoh. National institute of informatics, japan at trecvid 2011. In Proc. of TRECVID. NIST, 2011.

[29] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluation of color descriptors for object and scene recognition. In Proc. of the intl. conf. on Computer Vision and Pattern Recognition (CVPR), 1-8. IEEE, 2008.

[30] H. Jegou, M. Douze, and C. Schmid. Packing bag-of-features. In Proc. of the IEEE int. conf. on Computer Vision (ICCV), 2357-2364. IEEE, 2009.

[31] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12):1349-1380, 2000.

[32] F. Zheng, G. Zhang, and Z. Song. Comparison of Different Implementations of MFCC. Journal of Computer Science & Technology, 16(6): 582–589, 2001.

[33] J. Haitsma and T. Kalker. A highly robust audio fingerprinting system. In Proc. of the int. symp. on Music Information Retrieval (ISMIR), 2002.

[34] D. P. W. Ellis. Classifying music audio with timbral and chroma features. In Proc. of the int. symp. on Music Information Retrieval (ISMIR), 2007.

[35] J. M. Barrios, B. Bustos, and X. Anguera. Combining features at search time: Prisma at video copy detection task. In Proc. of TRECVID. NIST, 2011.

[36] C. G. Snoek, M. Worring, and A. W. Smeulders. Early versus late fusion in semantic video analysis. In Proc. of the int. conf. on Multimedia (ACMMM), 399-402. ACM, 2005.

[37] H.-K. Tan, C.-W. Ngo, R. Hong, and T.-S. Chua. Scalable detection of partial near-duplicate videos by visual-temporal consistency. In Proc. of the int. conf. on Multimedia (ACMMM), 145-154. ACM, 2009.

[38] A. Joly, O. Buisson, and C. Frelicot. Content-based copy retrieval using distortion-based probabilistic similarity search. IEEE Transactions on Multimedia, 9(2):293-306, 2007.

[39] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVid. In Proc. of the int. workshop on Multimedia Information Retrieval (MIR), 321-330. ACM, 2006.

[40] J. M. Barrios and B. Bustos. PRISMA-ORAND: Instance Search Based on Parallel Approximate Searches. In Proc. of TRECVID. NIST, 2012

Entrevista
Daniel Gatica

Por Benjamin Bustos

Daniel Gatica está a cargo del Grupo de Computación Social de la École Polytechnique Fédérale de Lausanne (EPFL), donde es profesor. Tiene más de cien publicaciones, 3.800 citas según Google Scholar y un índice h de 33. Sus principales aéreas de investigación son Social Computing, Mobile Computing, Social Media, Multimedia y Ubiquitous Computing.

¿Cuál es el impacto que están teniendo los avances en computación móvil con respecto a la investigación que se realiza en Multimedia? ¿Cómo influye en este aspecto la heterogeneidad de las fuentes de información?

No es aventurado decir que todo en multimedia será móvil (en cierto sentido ya lo es). El contexto móvil introduce nuevas condiciones de uso para dispositivos multimedia tradicionales y añade nuevos usos para tecnologías emergentes. Por ejemplo, el consumo de video en casa usando múltiples dispositivos en paralelo (TV como monitor, iPad para búsqueda de información sobre el programa que muestra la tele) es un ejemplo (de entre muchos posibles) de cómo los problemas de investigación en multimedia tienen que adaptarse a estos nuevos contextos, desde modelar usuarios que simultáneamente accesan distintos dispositivos con diferentes intenciones, hasta analizar automáticamente dominios múltiples.

¿Cuál es la importancia del contexto (temporal, geográfico, etc.) y cómo se puede considerar en el análisis de datos multimedia?

El asunto del contexto es interesante si lo analizamos históricamente. En los años ochenta y noventa, se analizaban videos, imágenes y audio por separado, no porque estos medios existiesen necesariamente aislados unos de otros (la TV es audiovisual por definición), sino porque, como objetos de estudio, se facilitaba estudiarlos así. El dominio de multimedia viene exactamente del reconocer que los medios múltiples se

generan conjuntamente de forma natural y que, por tanto, su estudio por separado es en cierto punto arbitrario. Además, muchas tareas se facilitan si las fuentes de información se integran.

Lo mismo puede decirse del contexto temporal o geográfico hoy en día. Los eventos que son capturados por los medios ocurren en el espacio y el tiempo. No integrar estas variables en los problemas de investigación que se plantean equivale a estudiar, hace veinticinco años, el audio o el video por separado.

¿Qué avances ha habido con respecto al problema del "gap semántico"?

El "gap semántico" sigue ahí, como el dinosaurio de Tito Monterroso, gozando de buena salud. En realidad se ha desarrollado, porque lo que entendemos hoy por "semántico", comparado con lo que denotaba cuando el término apareció, se ha expandido, y no sólo significa ponerle nombres a las objetos que aparecen en una imagen sino explicar su uso, su contexto descrito en la pregunta anterior.

Lo mejor que pudo pasar con respecto al "gap semántico" es la aparición, por una parte, de los medios sociales y, por otra, parte del crowdsourcing, que están dando pauta para poder plantear soluciones usando datos y participación social a gran escala. Así que hay esperanza.

¿Cuál es un buen enfoque para realizar emprendimiento e innovación con los resultados de la investigación en Multimedia?

Mi opinión es que un factor clave es el entendimiento de las necesidades reales (presentes o futuras) de quienes usan o usarán multimedia. Como ejemplo, empresas como Pinterest se están concentrando en comunicación social "especializada", respondiendo a la necesidad de los usuarios de comunicarse a un nivel más específico que lo que se puede hacer, por ejemplo, en Facebook. Otro punto importante está resumido en una frase de Ramesh Jain (University of California, Irvine), que traducida del inglés más o menos dice: *"No te preguntes lo que los medios sociales pueden hacer por multimedia, sino lo que multimedia puede hacer por los medios sociales"*. En otras palabras, el enfoque ya no es solamente hacer investigación sobre los medios como tales, sino pensarlos en la nueva realidad establecida por la ubicuidad de los medios sociales, que están generando una nueva serie de problemas de investigación.BITS

> El contexto móvil introduce nuevas condiciones de uso para dispositivos multimedia tradicionales y añade nuevos usos para tecnologías emergentes.

Computación y colaboración: el grupo CARL (Collaborative Applications Research Laboratory)

Nelson Baloian
Profesor Asociado DCC Universidad de Chile. Doktor rer. nat, Universität Duisburg, Alemania (1997); Ingeniero Civil en Computación, Universidad de Chile (1988). Líneas de especialización: Instrucción Asistida por Computador, Sistemas Distribuidos.
nbaloian@dcc.uchile.cl

Sergio Ochoa
Profesor Asociado DCC Universidad de Chile. Dr. en Ciencias de la Computación, Pontificia Universidad Católica de Chile (2002); Ingeniero de Sistemas, UNICEN, Argentina (1996). Líneas de especialización: Sistemas Colaborativos, Ingeniería de Software -arquitectura de software, mejora de procesos en micro y pequeñas empresas de software, ingeniería de requisitos, educación apoyada con tecnología.
sochoa@dcc.uchile.cl

José A. Pino
Profesor Titular DCC Universidad de Chile. Máster y calificado a Ph.D. in Computer Science (1977); Master of Science in Engineering, University of Michigan (1972); Ingeniero en Matemáticas, Universidad de Chile (1970). Líneas de especialización: Sistemas Colaborativos, Sistemas de Apoyo a Decisiones, Interacción Humano-Computador, Educación apoyada con tecnología.
jpino@dcc.uchile.cl

Los profesores de jornada completa del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile, Nelson Baloian, Sergio F. Ochoa y José A. Pino forman un grupo de investigación en Sistemas Colaborativos. Es el área que internacionalmente se denomina CSCW (Computer Supported Cooperative Work) o también Socio-technical Systems.

El foco de esta área está en posibilitar y facilitar, mediante el uso de tecnología, el trabajo de grupos o equipos de personas. Esto representa una diferencia con métodos tradicionales de computación, los cuales consideran al computador como el instrumento que debe resolver problemas, y a las personas como las responsables de alimentar con datos a estas máquinas. En CSCW, quienes resuelven problemas son las personas (colectivamente) apoyadas por tecnología apropiada.

Con esta definición tan amplia, caben muchas especializaciones y el grupo ha explorado varias de ellas, como se describe en las secciones siguientes. El desarrollo de sistemas CSCW puede considerarse como una rama específica de desarrollo de software que incluye una o más de estas características: desafíos de diseño asociados a objetivos organizacionales, dinámicas de grupo, comunicación, coordinación y colaboración entre personas, resolución de conflictos y toma de decisiones, contexto social de las actividades, y efectos positivos y negativos de la tecnología en tareas, grupos y organizaciones.

EDUCACIÓN COLABORATIVA

En esta subárea, el grupo ha graduado tres estudiantes de Doctorado (César Collazos, Olivier Motelet y Jens Hardings) y varios estudiantes de Magíster e Ingeniería. También ha tenido colaboración de otros investigadores, como el profesor Ulrich Hoppe (Universidad de Duisburg-Essen,

Alemania), Flavia Santoro (Universidad Federal del Estado de Río de Janeiro, Brasil), Mitsuji Matsumoto (Universidad de Waseda, Japón) y Luis Guerrero (Universidad de Costa Rica).

Entre los proyectos interesantes podemos mencionar COSOFT, que fue uno de los pioneros en proponer el uso de la computación en forma colaborativa dentro de la sala de clases (1993-1997). Este proyecto contemplaba el uso de una pizarra electrónica desde la cual el profesor podía crear o recopilar material, repartirlo a los estudiantes y recopilar material creado por los alumnos, los cuales trabajaban en sus computadores personales desde su pupitre. Versiones más recientes de esta idea contemplan el uso de dispositivos móviles por parte de los alumnos.

Otro proyecto realizado incluyó el desarrollo de un sistema interactivo que permite a grupos de estudiantes de enseñanza media, aprender sobre terremotos usando datos reales de la red nacional de sismógrafos, para calcular los epicentros de ciertos sismos. El uso de esta herramienta debiera llevar a los estudiantes a aprender a convivir con movimientos sísmicos.

Varios de los últimos proyectos de esta subárea se han basado en la teoría del aprendizaje situado que, entre otras, cosas, postula que el aprendizaje ocurre más efectivamente en el lugar donde el conocimiento necesita ser aplicado. La proliferación de dispositivos móviles inteligentes ha facilitado el desarrollo de sistemas para ser usados fuera de la sala de clases misma, dando apoyo al alumno durante visitas a museos, zoológicos y salidas a terreno. Estos sistemas permiten contar con la información y/o apoyo necesario durante la salida, así como también facilitan la recopilación de datos para ser usados posteriormente en la sala de clases o laboratorio.

Los resultados de las investigaciones en esta subárea se han publicado en journals tales como: Interactive Learning Environments, Knowledge and Information Systems, Educational Technology and Society, J. of Web Engineering, Computer Applications in Engineering Education.

PROCESOS DE NEGOCIO, REUNIONES DE DECISIÓN, MEMORIA ORGANIZACIONAL

En esta subárea, el grupo ha hecho investigación conjunta con los profesores Pedro Antunes (actualmente en la Universidad Victoria de Wellington, Nueva Zelandia) y Marcos R.S. Borges (Universidad Federal de Río de Janeiro, Brasil).

Uno de los temas de investigación es la especificación de procesos de negocio por grupos de personas, con la aplicación de conceptos tales como *group storytelling* y *storyboards*. Otro tema es la generación de herramientas para apoyar a los participantes durante el ciclo de reuniones de decisión. Esto incluye herramientas específicas para la preparación grupal de las reuniones (prereunión), para las reuniones mismas con objeto de aumentar su efectividad, y para la implementación posterior de las decisiones (postreunión).

Uno de los trabajos más citados del grupo en estos temas es un artículo en que se estudia el tipo de computadores más apropiados para distintos escenarios, dependiendo de la movilidad de los participantes, el tipo de trabajo que ellos están realizando y las características del ambiente.

La investigación resultante ha sido publicada en journals que incluyen Group Decision and Negotiation, Group Support Systems, Advanced Engineering Informatics, European J. of Operational Research, Computing & Informatics, Information Sciences, e International Journal of Information Technology & Decision Making.

MOVILIDAD Y COLABORACIÓN

Los dispositivos móviles presentan un interesante desafío y oportunidad para colaborar, y el grupo CARL recientemente ha dedicado mucha atención a esta subárea. Los estudiantes de Doctorado Andrés Neyem y Valeria Herskovic (hoy profesores de la Pontificia Universidad Católica de Chile), quienes se graduaron con tesis vinculadas a este ámbito, continúan trabajando con investigadores del grupo CARL en este dominio de aplicación. Además de estas tesis de Doctorado se han guiado varias tesis de Magíster y memorias de Ingeniería.

En esta subárea se han realizado diversas propuestas que van desde la infraestructura básica para soportar las interacción entre los colaboradores, hasta aplicaciones concretas para apoyar la colaboración móvil en diversos escenarios, por ejemplo en asistencia a emergencias urbanas o en inspecciones de obras de civiles.

En la infraestructura básica se crearon patrones de colaboración móvil y se desarrolló un protocolo de alto nivel para redes MANET (Mobile Ad hoc Network). También se creó un modelo para desarrollar aplicaciones colaborativas móviles.

> El foco de esta área está en posibilitar y facilitar, mediante el uso de tecnología, el trabajo de grupos o equipos de personas. Esto representa una diferencia con métodos tradicionales de computación, los cuales consideran al computador como el instrumento que debe resolver problemas, y a las personas como las responsables de alimentar con datos a estas máquinas.

Nelson Baloian, Sergio Ochoa y José A. Pino.

Por otro lado, se desarrolló un modelo acerca de las diferentes situaciones de conexión y disponibilidad en que puede encontrarse un grupo de personas, y cómo esta situación va cambiando en el tiempo debido a la movilidad de los participantes con sus equipos. Este modelo es últil para verificar la adecuación del software cuando se están desarrollando aplicaciones colaborativas móviles.

La interacción gestual también ha sido objeto de estudio por parte del grupo CARL. En efecto, la manera más frecuente de interactuar con los dispositivos móviles es mediante gestos con el dedo o un lápiz sobre una pantalla táctil. Esta es una forma relativamente nueva de interactuar con los dispositivos, por lo tanto hay poca literatura con recomendaciones para el diseño de las interfaces de aplicaciones basadas en gestos, tal como existe hoy para el caso de las interfaces de aplicaciones para computadores de escritorio. El problema es más interesante aún de investigar si se toma en cuenta que las dimensiones de la pantalla de los dispositivos móviles es muy variada, por lo que cabe preguntarse si las recomendaciones que se puedan derivar de los experimentos realizados en un tipo de dispositivo son extrapolables a los demás.

Respecto a las aplicaciones colaborativas móviles, se desarrolló un software apropiado para inspectores de obras de la construcción que utilizan los dispositivos móviles. Sin embargo, es el área de aplicación de la computación móvil colaborativa al manejo de emergencias lo que ha atraído más la atención del grupo.

El manejo de emergencias por parte de Bomberos presenta una situación ideal para la colaboración móvil. El medio típico de comunicación por parte de los bomberos son los sistemas de radio VHF. Sin embargo, el uso de esos dispositivos tiene muchas limitaciones, por ejemplo: no permite el envío de información digital, los canales de comunicación asignados a Bomberos son insuficientes, se pierden mensajes porque la transmisión es borrada por la transmisión a través de otro dispositivo más potente, etc. El uso de dispositivos con comunicación digital es entonces una alternativa o complemento apropiado para apoyar el proceso de respuesta a una emergencia. El grupo ha hecho contribuciones relevantes en el modelamiento de las situaciones, desarrollo de herramientas y aplicaciones concretas.

Una de las herramientas desarrolladas más exitosas para Bomberos ha sido MapaMóvil, una aplicación que posibilita a los Bomberos prepararse en viaje al lugar de la emergencia, guiar la conducción a través de calles y avenidas al co-piloto de los carros, ubicar puntos de interés (tales como grifos u hospitales), conocer detalles de los edificios siniestrados. Actualmente, MapaMóvil está en su cuarta versión, y funciona sobre notebooks, tablet-PCs y teléfonos inteligentes.

La investigación realizada en esta subárea ha sido publicada en journals que incluyen a Journal of Network and Computer Applications, Journal of Systems and Software, Expert Systems with Applications, Future Generation Computer Systems, y Group Decision and Negotiation.

El trabajo en este ámbito ha contado con la participación de otros investigadores extranjeros, como por ejemplo los profesores Marcos R.S. Borges (Universidad Federal de Rio de Janeiro, Brasil), Roc Meseguer (Universidad Politécnica de Cataluña, España) y Rodrigo Santos (Universidad Nacional del Sur, Argentina).

EVALUACIÓN DE SISTEMAS COLABORATIVOS

La evaluación de sistemas colaborativos presenta múltiples desafíos, comenzando por aclarar si se va a evaluar calidad, eficiencia o efectividad, ¿qué variables?, ¿para la organización, el equipo de trabajo, cada usuario? El grupo CARL ha tratado de dar respuestas en este tema, para lo que ha contado con la cooperación internacional del profesor Pedro Antunes, mencionado previamente.

Los primeras propuestas son técnicas de evaluación. Se desarrolló una técnica formal de evaluación y otra basada en modelos de procesamiento humano de información.

El trabajo más maduro es una propuesta global de cómo enfrentar el tema de la evaluación de los sistemas colaborativos. Esta propuesta pone en perspectiva la investigación previa en el tema, incluyendo marcos de referencia, variables de medida, y técnicas.

Los resultados se han publicado en los journals Information Research y ACM Computing Surveys.BITS

Crea y desafía al mundo a través de la Computación

www.dcc.uchile.cl

Departamento de Ciencias de la Computación
UNIVERSIDAD DE CHILE

REVISTA
BITS de Ciencia
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

UNIVERSIDAD DE CHILE

fcfm
Ciencias de la Computación
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

www.dcc.uchile.cl/revista

revista@dcc.uchile.cl