USO COMBINADO DE TECNOLOGÍAS SEMÁNTICAS Y ANÁLISIS VISUAL PARA LA ANOTACIÓN AUTOMÁTICA DE IMÁGENES Y SU RECUPERACIÓN

Sergio Rodríguez-Vaamonde, Pilar Ruiz-Ibáñez y Marta González-Rodríguez

Sergio Rodríguez-Vaamonde es ingeniero de telecomunicaciones por la *Universidad del País Vasco* y diploma de estudios avanzados en ingeniería telemática. Actualmente es doctorando, centrando sus investigaciones en la anotación automática de imágenes. Es investigador en el centro tecnológico *Tecnalia*, donde lleva a cabo trabajos de aplicación de las tecnologías de visión artificial a diferentes sectores como las TICs o el sector industrial.

Unidad de Sistemas de Información e Interacción
División ICT-European Software Institute, Tecnalia
Parque Tecnológico de Bizkaia
Ibaizabal Bidea, Edif. 202. 48170 Zamudio, España
sergio.rodriguez@tecnalia.com

Pilar Ruiz-Ibáñez, licenciada en geografía e historia por la *Universidad de Granada*, es investigadora en la *Unidad de Sistemas de información e Interacción* de *Tecnalia*. Antigua responsable del Centro de Documentación de *Robotiker*. Directora del proyecto de fusión de gestión del conocimiento de *Tecnalia R&I*.

Unidad de Sistemas de Información e Interacción
División ICT-European Software Institute, Tecnalia
Parque Tecnológico de Bizkaia
Ibaizabal Bidea, Edif. 202. 48170 Zamudio, España
pilar.ruiz@tecnalia.com

Marta González-Rodríguez, licenciada en informática por la *Universidad del País Vasco* y máster en tecnologías avanzadas de fabricación por la misma universidad, es gerente de *Tecnologías Semánticas* de *Tecnalia*. Participa en proyectos de I+D de ámbito nacional y europeo en el campo de las tecnologías semánticas e investiga su aplicación a diferentes sectores.

Unidad de Sistemas de Información e Interacción
División ICT-European Software Institute, Tecnalia
Parque Tecnológico de Bizkaia
Ibaizabal Bidea, Edif. 202. 48170 Zamudio, España
marta.gonzalez@tecnalia.com

Resumen

Se analizan los actuales sistemas semánticos para la indexación y recuperación de la información no textual en internet. El uso combinado de técnicas de visión artificial y el contenido textual en el que se enmarcan las imágenes, conforman los algoritmos más efectivos para conseguir que los motores de búsqueda en la nube puedan llegar a generar los mejores resultados en la recuperación de la información pertinente. El futuro inmediato de la Web pasa por lograr una contextualización automática de las imágenes que permita establecer similitudes entre los contenidos no textuales, y poder recuperarlos de forma efectiva.

Palabras clave

Anotación automática de imágenes, Visión artificial, Tecnologías semánticas, Procesamiento de lenguaje natural, Contextualización de imágenes, *Content-based image retrieval* (*cbir*).

Title: Joint use of semantics and visual analysis for automatic image annotation and retrieval

Abstract

The present stage of development of semantic systems used in the indexing and retrieval of non-text information on the internet is described. The most effective algorithms for retrieval of non-text information are the combined use of computer vision and content analysis of the text associated with the images. These techniques can lead to the best results in the retrieval of relevant information. The Web's immediate future lies in automatic contextualisation of images to establish similarities between them and to be able to effectively retrieve non-text content.

Artículo recibido el 16-11-11
Aceptación definitiva: 16-12-11

Keywords

Automatic image annotation, Computer vision, Semantic technologies, Natural language processing, Image contextualisation, Content-based image retrieval (cbir).

Rodríguez-Vaamonde, Sergio; **Ruiz-Ibáñez, Pilar**; **González-Rodríguez, Marta**. "Uso combinado de tecnologías semánticas y análisis visual para la anotación automática de imágenes y su recuperación". *El profesional de la información*, 2012, enero-febrero, v. 21, n. 1, pp. 27-33.

http://dx.doi.org/10.3145/epi.2012.ene.04

1. Introducción

La mayoría de las aplicaciones y tecnologías que están surgiendo en internet se centran en permitir a los usuarios utilizar diferentes tipos de contenidos multimedia, como vídeos o imágenes. Por ejemplo, *Facebook* publicaba a principios de 2011 que en el último fin de semana del año anterior se subieron 750 millones de fotografías (*Facebook*, 2011). Esta tendencia crece de forma imparable, ya que *Youtube* reportaba una subida de 8 horas de vídeo por minuto en 2007 mientras que en mayo de 2011 mostraba unas estadísticas de 48 horas por minuto (*Youtube*, 2011).

> "El futuro de la Web pasa por gestionar toda la información multimedia"

El futuro de la Web pasa por gestionar toda esa información multimedia, y sobre todo por hacer posible su acceso de forma inteligente desde cualquier parte del mundo, teniendo en cuenta que un contenido que no se puede encontrar, *no existe*. Las tecnologías involucradas son numerosas: desde el manejo de grandes cantidades de datos, hasta las redes de telecomunicaciones que soportan el movimiento de estas grandes cantidades de información, pasando por bases de datos distribuidas y de rápida accesibilidad para mejorar los tiempos de consulta. Es fundamental permitir a los usuarios el acceso eficiente a la información multimedia que deseen. El objetivo es favorecer que los diferentes motores de búsqueda sean capaces de encontrar la información multimedia en función de su contenido, lo que se denomina *content based information retrieval (cbir)*.

El caso de las imágenes es un ejemplo claro: si un usuario desea buscar una fotografía del edificio del museo *Guggenheim* de Bilbao, el sistema deberá analizar las existentes en internet y saber que una imagen muestra el *Guggenheim* de Bilbao y no el de New York, Venecia, Berlín o Las Vegas. En la actualidad esto dista de estar resuelto. Los grandes motores, como *Google* o *Bing* en su versión de búsqueda de imágenes, realizan dicha búsqueda en función de las etiquetas html que incluyen los usuarios, del propio nombre de la imagen o del pie de foto de la misma. Esto deriva en búsquedas no siempre adaptadas a las necesidades reales de los usuarios.

Estas necesidades latentes han provocado que tanto los motores de búsqueda como los investigadores de todo el mundo, estén analizando nuevas formas de llevar a cabo búsquedas efectivas sobre grandes conjuntos de contenidos multimedia y, específicamente, sobre imágenes estáticas. De todas ellas, el modelo más utilizado, es la anotación automática de imágenes (**Makadia**; **Pavlovic**; **Kumar**, 2010). En este modelo, dada una imagen, un sistema automático se encarga de analizarla y etiquetarla en función de varios parámetros, de tal forma que al realizar la búsqueda los motores analicen las etiquetas textuales de forma rápida y arrojen los resultados más pertinentes.

El presente artículo se centra en analizar las tendencias actuales a la hora de anotar una imagen de forma eficiente, en base a su contenido visual y a la información que aporta el propio sitio web, lo que se denominaría el "contexto de la imagen". Así mismo, se describirá brevemente el estado actual de los buscadores de imágenes, así como las necesidades planteadas de cara a encontrar lo que los usuarios realmente quieren.

> "Es fundamental permitir a los usuarios el acceso eficiente a la información multimedia que desean"

2. Anotación semántica de imágenes

El primer paso a dar para realizar la búsqueda de imágenes es anotar su contenido. Es primordial analizar la propia imagen mediante tecnologías de visión artificial, con lo que se logrará conocer el contenido de la misma. Estas técnicas permiten obtener su representación matemática de tal forma que posteriormente es necesario interpretar esta información de bajo nivel para encontrar su correspondencia con conceptos semánticos de alto nivel. Este es el principal problema que encuentran los investigadores en la actualidad para poder realizar una anotación de imágenes en función de su contenido y se le denomina *semantic gap*.

Desde el comienzo de la anotación de imágenes no estaba claro si era posible atravesar el *semantic gap*. Diferentes estudios del cerebro humano han investigado cómo se representan los objetos en el córtex visual (**Serre** *et al.*, 2007; **Dicarlo**, 2009), pero no se podía saber cómo se realiza el cambio a una representación semántica de la imagen formada en la retina (**Tousch**; **Herbin**; **Audibert**, 2011). En cuanto a la posibilidad de realizar la interpretación semántica por parte de sistemas automáticos de procesamiento, no ha sido hasta 2011 cuando se ha demostrado que cuanto más similitudes visuales tengan dos imágenes, mayor será su similitud semántica (**Deselaers**; **Ferrari**, 2011).

La barrera del *semantic gap* es crítica a la hora de realizar la anotación y se ha abordado desde diferentes puntos:

– utilizando técnicas estadísticas para mapear características de bajo nivel a vocabularios no estructurados de alto nivel (**Csurka** *et al.*, 2004);
– mediante representaciones intermedias compartidas entre varios objetos (**Torresani**; **Szummer**; **Fitzgibbon**, 2010);
– usando jerarquías semánticas de vocabularios limitados que aportan un mayor grado de riqueza semántica (**Gao**; **Chia**; **Cheng**, 2010);
– mediante la utilización de técnicas complejas de procesamiento de lenguaje natural para definir, no sólo el contenido de las imágenes sino también las relaciones existentes entre el propio contenido (**Lee**; **Trauman**, 2010).

> Es primordial analizar la propia imagen mediante tecnologías de visión artificial para conocer su contenido

El contenido de la imagen es muy importante a la hora de anotarla para su posterior recuperación. Pero hay situaciones en las que la información visual no aporta todo el contenido de la propia imagen y es evidente que la combinación de la información visual con otra información textual, es beneficiosa para los sistemas de anotación (**García-Serrano**, 2011). Para lograr una mayor riqueza de anotación, destacamos una línea de investigación a la que cada vez se están dedicando más esfuerzos y será muy relevante en el futuro: la "contextualización de la imagen". Generalmente, una imagen estará localizada dentro de una o varias webs que tratarán sobre temas relacionados con ella. De esta forma, la contextualización de la imagen analiza las propias webs, en vez de la imagen, y busca obtener un mayor conocimiento del contenido de la imagen además del que se pueda extraer de la misma. Por ejemplo, si se presenta en una web una imagen del cuadro de *Las meninas*, el sistema de procesamiento visual sabrá que hay un perro, un espejo, varias personas, incluso podría llegar a saber quiénes son las personas. Pero, del texto que rodea la imagen se puede extraer información sobre quién fue el autor, cuándo se pintó, pudiendo reconocer que se trata del cuadro original pintado por Velázquez o la reinterpretación de Picasso.

De todo ello se hablará más en profundidad en los siguientes apartados. Primero, se describirán las líneas de investigación actuales en el campo del análisis de imagen para la detección de objetos y anotación de imágenes a gran escala, mientras que en segundo lugar se mostrará el funcionamiento de la contextualización de las imágenes con el objetivo de disminuir el *semantic gap* enriqueciendo la anotación automática de las imágenes.

2.1. Procesamiento automático de imágenes para anotación del contenido visual

Su objetivo es utilizar tecnologías de visión artificial para detectar qué "elementos" componen la imagen. Como es de suponer, el concepto "elemento" es muy amplio y por tanto, las anotaciones de las imágenes también pueden ser extremadamente extensas, desde la detección de las personas que están presentes hasta la del tipo de árboles existentes o las acciones que se están realizando.

> Desde el comienzo de la anotación de imágenes no estaba claro si era posible atravesar el *semantic gap*

Detección de objetos a pequeña escala

Una de las líneas de investigación que tiene más fuerza en el campo de la visión artificial es la categorización de objetos en imágenes, es decir, identificar qué clases de objetos existen en una imagen (persona, coche o edificio).

Esta línea de investigación no es nueva y los primeros trabajos se remontan a los años 70 (**Fischler**; **Elschlager**, 1973), con resultados modestos. Hoy en día, la evolución del campo de detección y categorización de objetos está aumentando de forma exponencial. En los últimos trabajos el objetivo es crear y entrenar un algoritmo específico para detectar un tipo de objeto, existiendo varias aproximaciones que se pueden ver en diversos artículos de referencia (**Dicarlo**, 2009).

Para lograr avanzar en este campo, numerosas comunidades de investigadores están creando bases de datos de imágenes comunes y públicas sobre las que cualquiera pueda realizar pruebas, de tal forma que sea posible establecer

Figura 1. *Las meninas* de Velázquez

Figura 2. *Las meninas* de Picasso

comparaciones entre los diferentes métodos que se proponen. Un ejemplo es el denominado *Visual object classification challenge* (**Everingham** *et al.*, 2010). Esta competición se celebra una vez al año y reúne decenas de algoritmos que tratan de obtener la mejor tasa de aciertos sobre un conjunto definido de 20 objetos, entre los que se encuentran por ejemplo "persona", "botella", "avión" o "planta"; y en la cual se puede ver el avance del estado de la técnica. En la tabla 1 se comparan los algoritmos ganadores de la competición entre los años 2007-2011. Se muestra para cada año, la mejor puntuación obtenida por el algoritmo ganador, la peor y la puntuación media de los 20 objetos. Además, muestra qué algoritmo ha sido el ganador, así como los objetos que han conseguido la mejor y la peor puntuación.

Año	Mejor objeto detectado	Mejor puntuación (%)	Peor objeto detectado	Peor puntuación (%)	Puntuación media (%)
2007	persona	85,9	botella	33,1	59,4
2008	persona	86,9	planta	29,2	54,9
2009	avión	88,0	planta	36,6	66,5
2010	avión	93,0	planta	48,6	73,8
2011	avión	95,0	planta	56,5	78,6

Tabla 1. Algoritmos ganadores de la competición entre 2007-2011

Desde 2007 hasta 2011 se ha producido una evolución en las tasas de acierto, de tal forma que todas (incluida la menor) han aumentado con el paso de los años.

El salto cuantitativo ha sido muy grande, por lo que cada vez los sistemas de detección de objetos son más perfectos en tareas como detectar un objeto concreto en un determinado tipo de imagen.

> Uno de los sistemas de contextualización que más se está trabajando es la relación de una imagen con sus etiquetas asociadas en la página web

Detección de objetos a gran escala

A pesar del avance en la detección de objetos, estos trabajos se están centrando en el análisis de imágenes a pequeña escala (se trabaja con unos pocos cientos de imágenes en escenas más o menos controladas). Cuando se pasa a una escala del tamaño de internet, donde las imágenes pueden contener todo tipo de objetos en todo tipo de situaciones, se deben tener en cuenta otros aspectos. Este ámbito comienza a tener una gran relevancia y se están generando bases de datos comunes para la comparativa, como puede ser *Saiapr-TC 12* (**Escalante** *et al.*, 2010), con unas 20.000 imágenes, *MIR Flickr* (**Huiskes**; **Lew**, 2008), con 1 millón de imágenes en su edición de 2010, o las competiciones *The imagenet large scale visual recognition challenge* o *ImageCLEF*.

Existen dos grandes aproximaciones para solventar el problema de la anotación automática: modelizar todo el universo de elementos que existen o reducir el mundo a la hora de realizar la anotación centrándose en un conjunto limitado de imágenes.

Una aproximación muy empleada, debido a su similitud con los algoritmos clásicos mencionados anteriormente, es modelizar de forma individual todo el universo de objetos a anotar (**Hao** *et al.* 2009). Otra que está teniendo muy buenos resultados es la de combinar información estadística de todos los tipos de conceptos a anotar en una imagen y realizar una modelización conjunta (**Weston**; **Bengio**; **Usunier**, 2011). En este caso sigue siendo necesario modelizar todo el universo, pero la combinación de información relativa a relaciones semántica y visual entre conceptos posibilita mejorar la anotación.

> Para lograr una mayor riqueza de anotación, destaca una línea de investigación que será muy relevante en el futuro, la "contextualización de la imagen"

Existe una segunda aproximación: reducir el problema de anotación antes de modelizar los objetos. La ventaja clara de este tipo de sistemas es que no es necesario modelizar todos y cada uno de los elementos existentes en el mundo. La propuesta, cuyo máximo exponente se encuentra en el artículo de **Makadia**, **Pavlovic** y **Kumar** (2010), se basa en que a la hora de anotar una imagen, ésta se compare visualmente con un conjunto de entrenamiento ya anotado. Una vez se conozcan las imágenes más similares, se analizan sus anotaciones y se trata de inferir, en base a ellas, las anotaciones de la imagen que deseamos anotar. Esta aproximación es mucho más sencilla que otras propuestas anteriores, pero según sus autores es la que mejores resultados obtiene, con una tasa de aciertos del 35% mientras que con otros métodos de modelización de objetos más complejos se obtienen valores inferiores, como el 32% (**Llorente**; **Manmatha**; **Rüger**, 2010). Debido a su simplicidad, han comenzado a aparecer evoluciones de este método que permiten alcanzar una tasa de acierto hasta del 52% (**Verbeek** *et al.*, 2010).

2.2. Contextualización de las imágenes en la web

A pesar de las mejoras en el procesamiento de imágenes, hay más información adicional que puede extraerse conociendo el entorno web que rodea a una imagen y que se denomina "contexto de la imagen". Este contexto web es capaz de mejorar la anotación automática de imágenes gracias a que permite eliminar la ambigüedad de los conceptos a anotar. En este apartado se mostrarán dos de las tecnologías más relevantes de contextualización web, la primera de ellas se basa en las propias "etiquetas" relacionadas con las imágenes, mientras que la segunda trata de un nuevo modelo de contextualización fundamentado en las redes sociales.

Contextualización en base a etiquetas

Uno de los sistemas de contextualización que más se está trabajando en la actualidad es la relación de una imagen con sus etiquetas asociadas en su página del sitio web. Un ejemplo claro es *Flickr*, donde cada usuario puede subir sus imágenes y etiquetarlas en función de sus intereses personales,

y donde gracias a su importancia están surgiendo numerosas iniciativas (**Ulges**; **Worring**; **Breuel**, 2011). Una de las más innovadoras es el hecho de utilizar la estructura del contenido de la web que proporcionan los usuarios. Por ejemplo, la comunidad *Flickr* organiza sus fotos en base a los grupos. Existen 200.000 grupos que han sido definidos y relacionados con todo tipo de temas como "fotos de fiestas" o "fotografía natural". La utilización de los grupos se centra en la fase de aprendizaje del sistema, donde para cada grupo se genera un modelo. Durante la anotación, la información del grupo se asume como proporcionada por el usuario, es decir utilizan los grupos como fuente adicional de información para eliminar la ambigüedad (**Ulges**; **Worring**; **Breuel**, 2011).

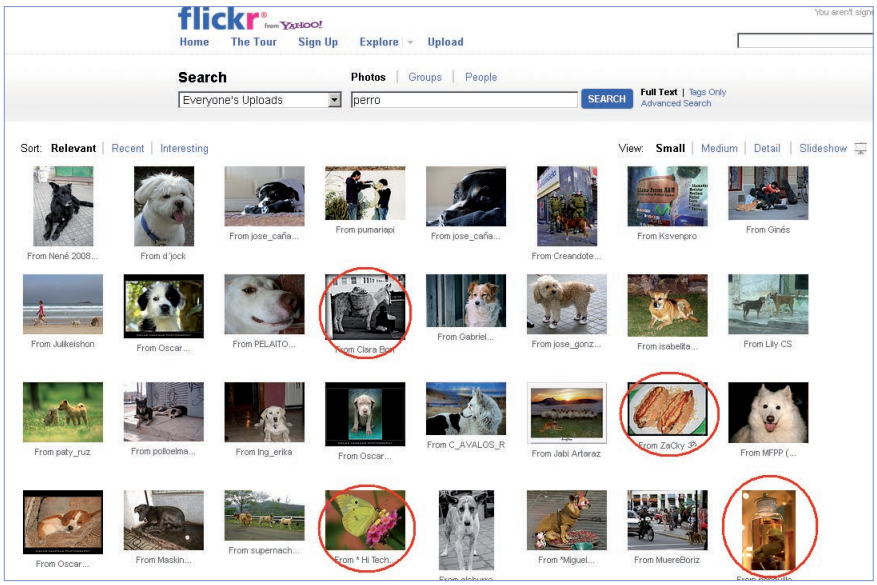

Figura 3. Resultados en *Flickr* para búsqueda con la palabra "perro"

Otros autores consideran que los resultados obtenidos de *Flickr* son muy "ruidosos" es decir, se obtienen muchas imágenes que tienen poco o nada que ver con la imagen anotada. Este fenómeno se puede observar en la figura 3, donde se ilustran los resultados de la búsqueda con la palabra "perro": entre los primeros 36 resultados obtenemos 4 que no corresponden (11%).

Por ello, cuando se utiliza *Flickr* para construir una base de datos de entrenamiento para la anotación de imágenes, es necesario hacer un filtrado previo, aplicando la máxima de que una imagen es relevante para una cierta etiqueta cuando la etiqueta describa el contenido de una o más regiones de la imagen (**Tang** *et al.*, 2011).

Contextualización en base a redes sociales

Otras iniciativas utilizan las redes sociales (**Elhai**; **Karlsen**; **Akselsen**, 2009) para adquirir los metadatos contextuales sociales. Las redes sociales son una de las plataformas más utilizadas por las comunidades online para compartir texto, imágenes y vídeos. Se basan en perfiles de usuarios que ofrecen una descripción de cada miembro. Además de las imágenes subidas por un usuario, su perfil contiene comentarios y opiniones positivas/negativas sobre las mismas.

Un trabajo de análisis del contexto más complejo es el llevado a cabo por **Elhai**, **Karlsen** y **Younas** (2011). En él se propone un sistema que genera de forma semi-automática anotaciones de imágenes sobre la ontología *OntoCAIM*, considerando el contexto de una red social y haciendo uso de las anotaciones manuales de imágenes proporcionadas por el usuario más activo. Para ello, infieren a este usuario mediante *social network analysis* y se toman sus anotaciones como base para las imágenes objetivo.

La ontología *OntoCAIM* reutiliza las ontologías *FOAF* (representa el perfil del usuario de la red social), *SIOC* (representa comunidades online), *EXIF* (representa metadatos básicos de fotografías) y *WordNet* (ayuda en la desambiguación del lenguaje natural), y por lo tanto describe la red social y las imágenes. Como se ha comentado, este sistema gira entorno al contexto social de una imagen que estaría formado entre otros por la geo-referencia, fecha y hora y la granularidad de las relaciones entre actores de la red social:

- para obtener el actor central o correcto utilizan técnicas de SNA (*social network analysis*);
- la fecha y hora es utilizada porque se parte de la hipótesis de que fotografías que hayan sido sacadas en un corto espacio de tiempo, seguramente sean muy similares porque la cámara no puede haber estado en localizaciones muy distantes en el umbral de 5 minutos que plantean inicialmente;
- la geo-referencia se utiliza para agrupar fotografías tomadas dentro de un radio concreto, siempre y cuando contengan en sus metadatos las coordenadas GPS.

Otra opción de utilización de las redes sociales como ayuda a la anotación es el reconocimiento facial para la anotación.

> Otra opción de utilización de las redes sociales como ayuda a la anotación es el reconocimiento facial

En el trabajo de **Stone**, **Zickler** y **Darrel** (2008) se propone mejorar los algoritmos de reconocimiento utilizando las redes sociales. Así, una imagen facial puede no ser reconocida por diversos motivos (iluminación, maquillaje). En este caso, se utilizan imágenes parecidas anotadas de forma manual en las redes sociales y se analizan las personas que aparecen en la misma imagen, identificando grupos de amigos. De esta forma, la tasa de anotaciones automáticas de personas aumenta considerablemente gracias al uso del contexto social.

3. Sistemas de búsqueda de imágenes actuales

Se ha visto cómo es posible anotar de forma automática el contenido de una imagen. El objetivo final de esta anotación es permitir a los motores de búsqueda la recuperación de imágenes de forma inteligente. Existen numerosos intentos por crear buenos buscadores de imágenes basados en su

Buscador de imágenes	Referencia web
TinEye	*http://www.tineye.com*
Cydral	*http://www.cydral.com*
Quintura (y versión Kids)	*http://www.quintura.com*
Google similar images	*http://images.google.com*
Ithaki	*http://www.ithaki.net*
LTU	*http://www.ltutech.com*
Pixlogic	*http://www.pixlogic.com*

Tabla 2. Buscadores de imágenes

contenido. Algunos de los más avanzados se pueden encontrar en la tabla 2.

Hay que destacar el buscador *Google images* en su versión de búsqueda de imágenes similares (opción a la derecha de una imagen hallada). Tras realizar numerosas pruebas, se ha comprobado que es capaz de realizar búsquedas de muy distintos tipos dependiendo de la imagen subida. Por ejemplo, si se le indica la imagen frontal de un coche, puede reconocer la marca y modelo, así como su color. En cambio, si se le muestra el mismo coche en una vista lateral, sólo intuye que es un vehículo y devuelve imágenes genéricas de vehículos. Esto demuestra que los buscadores comienzan a anotar los contenidos para su búsqueda, pero el resultado no es del todo correcto, ya que existen casos en los que se buscan imágenes de un determinado elemento y aparecen imágenes de otros que no esperamos. Este fenómeno está relacionado con la experiencia de uso, que no es satisfactoria porque el usuario espera otra información.

Por ello, como reto futuro en el campo de la recuperación inteligente de imágenes se puede destacar la necesidad de avanzar en la elaboración de herramientas de medición de la satisfacción de la experiencia de usuario, considerando sus expectativas y estado emocional. El resultado de estas mediciones debería influir en el proceso de búsqueda y presentación de los resultados. Así mismo, la anotación de imágenes posee un factor de medición subjetivo, es decir, esperamos que la anotación sea tan buena como aquella que realizaría un ser humano. Por ello, esta subjetividad influye de forma determinante en la recuperación de información: ésta será mejor cuanto mejor sea la anotación. Pero si además se une a la evolución de la capacidad de medición de la calidad de la experiencia -aprendiendo qué es lo que cada usuario espera en cada ocasión- la recuperación inteligente de información daría un paso de gigante.

> Existen numerosos intentos por crear buenos buscadores de imágenes basados en su contenido

4. Conclusiones

La búsqueda de imágenes es uno de los retos del futuro de internet. En la actualidad se están proponiendo modelos de anotación automática de imágenes en base a su contenido.

Se ha querido destacar que aunque el campo del análisis de imagen está aún en evolución, va alcanzando unas tasas de anotación efectiva muy altas. En un futuro las anotaciones según el contenido serán mucho más eficientes.

Por otra parte, se ha descrito cómo las imágenes en la Web están rodeadas de información de contexto, útil para ayudar en el proceso de anotación de las imágenes. Las redes sociales, los sitios para compartir fotografías, etc., son fuentes de dicho contexto y sirven, por un lado para construir nuevas y mejores bases de entrenamiento, y por otro para mejorar la precisión, problemas clave en todo proceso de anotación de imágenes.

> Como reto futuro es necesario avanzar en la medición de la satisfacción de la experiencia de usuario, considerando sus expectativas y estado emocional

5. Agradecimientos

Este trabajo ha sido elaborado dentro del proyecto *Buscamedia* de programa *Cenit-E*, y ha sido subvencionado parcialmente por el *CDTI* (*Centro para el Desarrollo Tecnológico e Industrial*), dependiente del *Ministerio de Ciencia e Innovación*. Los autores también agradecen el trabajo a todos los socios del proyecto.
http://www.cenitbuscamedia.es

6. Bibliografía

Csurka, Gabriella; **Dance, Christopher**; **Fan, Lixin**; **Willamowski, Jutta**; **Bray, Cédric**. "Visual categorization with bags of keypoints". En: *Workshop on statistical learning in computer vision*, *ECCV*, 2004.
http://www.cs.cmu.edu/~efros/courses/AP06/Papers/csurka-eccv-04.pdf

Deselaers, Thomas; **Ferrari, Vittorio**. "Visual and semantic similarity in imageNet". En: *IEEE Computer vision and pattern recognition (CVPR) conf.*, 2011.
http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/es//pubs/archive/37065.pdf

Dicarlo, James. "A strategy for understanding how the brain accomplishes object recognition". En: Dickinson, Sven J.; Leonardis, Ales; Schiele, Bernt; Tarr, Michael J. *Object categorization*. Cambridge University Press, 2009. ISBN 9780511635465
http://dx.doi.org/10.1017/CBO9780511635465

Elai, Najeeb; **Karlsen, Randi**; **Akselsen, Sigmund**. "A context centric approach for semantic image annotation and retrieval". *Future computing, service computation, cognitive, adaptive, content, patterns*, 2009. *Computation world'09.* Computation World: IEEE, pp. 665-668.
http://dx.doi.org/10.1109/ComputationWorld.2009.30

Elahi, Najeeb; **Karlsen, Randi**; **Younas, Wagas**. "Image annotation by leveraging the social context". En: *ACM Icuimc 2011*, *The 5th Intl conf on ubiquitous information management and communication*, 2011. Seoul, South Korea, 21-23 Febr. ISBN 978 1450305716
http://dx.doi.org/10.1145/1968613.1968662

Escalante, Hugo-Jair; **Hernández, Carlos A.**; **González, Jesús A.**; **López-López, Aurelio**; **Montes, Manuel**; **Morales, Eduardo**; **Sucar, L. Enrique**; **Villaseñor, Luis**; **Grubinger, Michael**. "The segmented and annotated IAPR TC-12 benchmark". *Computer vision and image understanding*, 2010, n. 4, pp. 419-428.
http://ccc.inaoep.mx/~emorales/Papers/2010/hugo.pdf
http://dx.doi.org/10.1016/j.cviu.2009.03.008

Everingham, Mark; **Van-Gool, Luc**; **Williams, Christopher K.**; **Winn, John**; **Zisserman, Andrew**. "The Pascal visual object classes (VOC) challenge". *Intl journal of computer vision*, 2010, v. 88, n. 2, pp. 303-338.
http://research.microsoft.com/pubs/102944/PascalVOC_IJCV2009.pdf
http://dx.doi.org/10.1007/s11263-009-0275-

Facebook. *Facebook Tweet*, 2011.
http://twitter.com/#!/facebook/status/22372857292005376

Fischler, Martin A.; **Elschlager, Robert A.** "The representation and matching of pictorical structures". *IEEE transactions on computers*, 1973, v. 22, n. 1, pp. 67-92.
http://dx.doi.org/10.1109/T-C.1973.223602

Gao, Shenghua; **Chia, Liang-Tien**; **Cheng, Xiangang**. "Web image concept annotation with better understanding of tags and visual features". *Journal of visual communication and image representation*, 2010, v. 21, n. 8, pp. 806-814.
http://202.114.89.42/resource/pdf/5546.pdf
http://dx.doi.org/10.1016/j.jvcir.2010.08.005

García-Serrano, Ana. "UNED-UV experiments using multimodal information approaches". *Image CLEF*, 2011, Amsterdam.

Hao, Su; **Sun, Min**; **Fei-Fei, Li**; **Savarese, Silvio**. "Learning a dense multi-view representation for detection, viewpoint classification and synthesis of object categories". En: *IEEE 12th Intl conf on computer vision*, 2009, pp. 213-220.
http://www.eecs.umich.edu/vision/papers/SuSunLiSavarese_ICCV2009.pdf

Huiskes, Mark J.; **Lew, Michael S.** "The MIR *Flickr* retrieval evaluation". En: *Proceedings of the 1st ACM Intl conf on multimedia information retrieval* (*MIR*), 2008, pp. 39-43. New York: ACM.
http://press.liacs.nl/mirflickr/mirflickr.pdf
http://doi.acm.org/10.1145/1460096.1460104

Lee, Yong-Jae; **Grauman, Kristen**. "Object-graphs for context-aware category discovery". En: *IEEE Conf on computer vision and pattern recognition* (*CVPR*), 2010, pp. 1-8.
ftp://ftp.cs.utexas.edu/pub/AI-Lab/tech-reports/AI-14.pdf

Llorente, Ainhoa; **Manmatha, Raghavan**; **Rüger, Stefan M.** "Image retrieval using Markov random fields and global image features". *Proc. of the ACM Intl conf on image and video retrieval*, 2010, pp. 243-250.
http://oro.open.ac.uk/23503/1/p243-llorente.pdf
http://doi.acm.org/10.1145/1816041.1816078

Makadia, Ameesh; **Pavlovic, Vladimir**; **Kumar, Sanjiv**. "Baselines for image annotation". *Intl journal of computer vision*, 2010, v. 90, pp. 88-105.
http://www.sanjivk.com/ImageAnnotation_IJCV10.pdf
http://doi.acm.org/10.1007/s11263-010-0338-6

Serre, Thomas; **Wolf, Lior**; **Bileschi, Stanley**; **Riesenhuber, Maximilian**; **Poggio, Tomaso**. "Robust object recognition with cortex-like mechanisms". *IEEE transactions on pattern analysis and machine intelligence*, 2007, v. 29, n. 3, pp. 411-426.
http://cbcl.mit.edu/publications/ps/serre-wolf-poggio-PAMI-07.pdf
http://doi.acm.org/10.1109/TPAMI.2007.56

Stone, Zak; **Zickler, Todd**; **Darrel, Trevor**. "Autotagging *Facebook*: social network context improves photo annotation", *IEEE Computer Society Conf on computer vision and pattern recognition workshops*, 2008.
http://www.eecs.harvard.edu/~zickler/papers/Autotag_IVW2008.pdf
http://dx.doi.org/10.1109/CVPRW.2008.4562956

Tang, Jinhui; **Yang, Shuicheng**; **Chua, Tat-Seng**; **Jain**, **Ramesh**. "Label-specific training set construction from web resource for image annotation". *CoRR*, 2011.
http://arxiv.org/pdf/1107.2859v1

Torresani, Lorenzo; **Szummer, Martin**; **Fitzgibbon, Andrew**. "Efficient object category recognition using classemes". En: *Proc of the 11th European conf on computer vision: Part I*, 2010, pp. 776-789. ISBN 978 3642155482
http://dl.acm.org/citation.cfm?id=1886063.1886122

Tousch, Anne-Marie; **Herbin, Stéphane**; **Audibert, Jean-Yves**. "Semantic hierarchies for image annotation: a survey". *Pattern recognition*, 2011, v. 45, pp. 333-345.
http://certis.enpc.fr/~audibert/Mes%20articles/PR11.pdf
http://dx.doi.org/10.1016/j.patcog.2011.05.01

Ulges, Adrian; **Worring, Marcel**; **Breuel, Thomas**. "Learning visual contexts for image annotation from *Flickr* groups". *IEEE transactions on multimedia*, 2011, v. 13.
http://www.science.uva.nl/research/publications/2011/UlgesITM2011/tmm_final.pdf
http://dx.doi.org/10.1109/TMM.2010.2101051

Verbeek, Jakob; **Guillaumin, Matthieu**; **Mensink, Thomas**; **Schmid, Cordelia**. "Image annotation with TagProp on the MIR *Flickr* set". En: *11th ACM Multimedia information retrieval*, 2010, pp. 537-546.
http://hal.inria.fr/docs/00/60/63/92/PDF/verbeek10mir.pdf
http://dx.doi.org/10.1145/1743384.174347

Weston, Jason; **Bengio, Samy**; **Usunier, Nicolas**. "Wsabie: scaling up to large vocabulary image annotation". En: *Proc of the Intl joint conf on artificial intelligence*, *IJCAI*, 2011, pp. 2764-2770.
http://www.thespermwhale.com/jaseweston/papers/wsabie-ijcai.pdf

Youtube Inc. "Thanks, *YouTube* community, for two big gifts on our sixth birthday!" *Broadcasting ourselves. The official Youtube blog*, 25 mayo 2011.
http://youtube-global.blogspot.com/2011/05/thanks-youtube-community-for-two-big.html