

Tecnologías digitales al servicio de los archivos de imágenes¹

Lorenzo Vilches

Universitat Autònoma de Barcelona

Departament de Periodisme i de Ciències de la Comunicació
08193 Bellaterra (Barcelona). Spain

Resumen

La información televisiva, las series de televisión y la publicidad ocupan la mayor parte del tiempo en las *broadcast*, tanto de sistemas analógicos y digitales como terrestre y de satélite. Estos géneros televisivos constituyen, más allá de la industria y del mercado, también el espejo de una identidad social y cultural. Por eso, una de las cuestiones clave de la cultura del próximo milenio será la constitución, documentación, almacenamiento, reutilización y análisis del patrimonio audiovisual. El archivo, la catalogación, el análisis y la consulta *on line* de este acervo cultural digitalizado pone una serie de problemas técnicos y teóricos que han de resolverse con la alianza entre investigadores de las ciencias humanas y la industria de la información.

Palabras clave: indexación de imágenes, teoría de la imagen, anotación de vídeo, información televisiva, espot publicitario, informática, narrativa, retórica, documentación audiovisual.

Abstract. *Digital technologies at the service of images archives*

Television news, TV series, and spot advertising occupy the greater part of broadcast time. This is as much the case for analogue and digital systems as for cable and satellite TV. These television genres constitute, in addition to the TV industry and market aspects, a mirror of social and cultural identity. Therefore, one of the key issues for culture in the next millennium will be communications, documentation, storing, reusing and analysing the audio-visual heritage. The archive, cataloguing, the analysis and on-line consulting of this digitalised common property poses a series of technical and theoretical problems that will have to be resolved in a joint effort involving researchers in the social sciences and the information industry.

Key words: indexing images, theory of the image, video tagging, television news, advertising spots, computing, narrative, rhetoric, audio-visual documentation.

1. Este trabajo se ha realizado con la ayuda de la Comisión Interministerial de Ciencia y Tecnología, del Ministerio de Ciencia y Tecnología.

Sumario

- | | |
|---|--|
| Introducción | Tareas para una colaboración entre la informática y la teoría de la imagen |
| Estado de la cuestión | Cómo organizar la investigación sobre la indexación semántica |
| Avance de la ciencia computacional y un nuevo problema para la teoría de la imagen: la imagen mosaico | Conclusión |
| | Bibliografía |

Introducción

La información televisiva ocupa una parte importante de la programación en las cadenas públicas y privadas y en los sistemas analógicos y digitales, terrestre y satélite. Las miles de horas producidas y emitidas por las cadenas nacionales suponen una dificultad creciente para productores, periodistas, realizadores, editores y documentalistas, a la hora de acceder y procesar una cantidad de información útil que puede ser reciclada en reportajes y documentos históricos en las televisiones.

El advenimiento de Internet y el desarrollo de las tecnologías de la información han permitido también una aceleración del desarrollo del audiovisual tradicional en su conjunto. En la actual situación de la sociedad de la información, los programas informativos de las televisiones han pasado a ser, junto con la producción de ficción propia (cine y series de televisión), el reflejo de características culturales del país y de la representación de la democracia. Nadie duda hoy de la importancia de la información televisiva, que es, por lo menos, comparable con la información producida por la prensa escrita. Mientras que la prensa escrita se halla en las hemerotecas de todas las ciudades y universidades, la información audiovisual no se halla al alcance de todos debido a las dificultades en la conservación de los archivos y a la falta de un sistema de documentación audiovisual normalizado.

La publicidad impresa y audiovisual, desde que entró en la universidad de mano de R. Barthes, no ha dejado de estudiarse por parte de los teóricos de la imagen. Al mismo tiempo que la publicidad se ha labrado un futuro entre la producción cultural debido a sus cualidades estéticas y su atribución de reflejo de la sociedad de consumo, se ha hecho también imprescindible para la supervivencia de las empresas audiovisuales. El efecto de todo ello es la necesidad de la gestión de millones de horas de producción y emisión televisiva publicitaria a través de sistemas veloces de almacenamiento, documentación y explotación.

La información audiovisual, la publicidad y, en general, toda la creación multimedia constituye una fuente de riqueza y de conocimiento que no ha sido aún sistematizada y puesta al alcance de la sociedad. En algunos países (en España, por ejemplo) no existe una legislación sobre la centralización de los archivos audiovisuales. La conservación selectiva de muestras significativas

de contenidos vehiculados por la televisión no es aún una preocupación prioritaria de la autoridad pública encargada de facilitar la transmisión de la cultura y de la información, que hoy constituyen pilares básicos de la democracia. Aunque esto tampoco parece inquietar demasiado a otros sectores sociales, especialmente a aquéllos que realizan su actividad profesional, investigadora, docente o artística en ámbitos cercanos al universo de la comunicación. Mientras que la literatura y el arte han asegurado felizmente la conservación y el mantenimiento de las obras en bibliotecas y museos, la producción audiovisual no se considera aún con la importancia que sí merece en los mercados internacionales.

La conservación de la cultura audiovisual se encuentra en un estado de degradación semejante al que mostraban en el siglo XX las ruinas históricas de los países mediterráneos. La documentación audiovisual es escasa, los archivos cinematográficos se encuentran dispersos y los criterios de conservación y búsqueda de los contenidos de las televisiones son, cuando existen, en la mayoría de los casos, arbitrarios. Además, mientras que la gestión y transmisión digital de las televisiones comienza a generalizarse en el mundo desarrollado, los sistemas de conservación de la información diaria son aún clasificados según criterios pertenecientes al universo analógico, lo cual dificulta aún más la búsqueda y el procesamiento útil en el tiempo. Y, sin embargo, el tratamiento, la explotación y la difusión de la información constituye una exigencia prioritaria para la industria nacional y la investigación social e histórica de mañana.

Situación actual de la investigación en el ámbito de la imagen

El proceso de digitalización que se lleva a cabo en la sociedad posindustrial, especialmente la *digitalización* de la conservación de la información, de los procesos productivos y de los sistemas de distribución de la señal, hacen posible el establecimiento de sistemas automáticos y de indexación de la información audiovisual en bases de datos que permitan un amplio uso para los profesionales, los actores de la política audiovisual, la educación y la investigación.

La mayoría de proyectos de investigación en el campo de la sociedad de la información audiovisual tienden a trabajar en los aspectos tecnológicos de transporte o conservación, y menos en la indexación semántica de la información. A ello contribuye, entre otras, las dificultades de enfrentarse a tres tipos de soportes de información: textual, sonora y visual. Pero la digitalización reduce todos los documentos a un denominador común, a un mismo soporte físico. Con ello entramos en la era de los hipermedios, donde el audiovisual dicta sus propias leyes. Las bibliotecas, los archivos y los centros de emisión de contenidos audiovisuales se aproximan. Cuando los documentos multimedia producidos por las empresas del ramo sean disponibles en las redes domésticas, y la noción de archivo se una a la edición, el editor será el propietario de bases textuales conservadas en servidores accesibles a una gran demanda.

El acceso y navegación por Internet e Intranet permiten ir hacia sistemas más analíticos y exhaustivos de búsqueda, sistematización y recuperación de la información audiovisual.

La potenciación de sistemas de transporte de imagen y sonido más veloces permitirá una mayor demanda y uso de contenidos audiovisuales por parte de profesionales de la información. Nos encontramos ante nuevas formas tecnológicas y productivas que tomarán los documentos audiovisuales de lengua, sonido e imagen en la era electrónica.

Estado de la cuestión

Estado de la cuestión en el ámbito de la teoría de la imagen

En el ámbito de las nuevas tecnologías se tiene la impresión de que la teoría de la imagen y, en particular, la semiótica de la imagen han dejado paso a conceptos y nociones provenientes de la teoría cognitiva, la informática y la inteligencia artificial. Pero esto es sólo aparente, porque los creadores, artistas y divulgadores de la cultura digital, cibernetica o ciberspacial utilizan una serie de conceptos y categorías provenientes de las teorías de la imagen.

Actualmente la investigación teórica en el campo de la imagen puede dividirse en dos grandes campos: la investigación *de la imagen* y la investigación *por la imagen*.

La teoría *de la imagen* tiene su gran momento de renovación, después del dominio de la iconología (E. Panofsky) y de la percepción visual, en los años sesenta y setenta, con el auge del estructuralismo y de la semiótica de origen norteamericano (Ch.S. Pierce) y europeo (Barthes, Metz, Greimas). La investigación de la imagen, especialmente en Europa, trata de establecer categorías teóricas para analizar las imágenes en soportes artificiales y materiales: el arte (Gombrich), la publicidad (Barthes), el cine (C. Metz), el cómic (Deruelle). Mientras, en América, la semiótica peirciana se aplica especialmente al lenguaje de los gestos, la arquitectura, la proxémica y la semiótica del espacio, en Europa se insiste en la teoría textual y la enunciación.

La investigación *por la imagen* utiliza diversas teorías no semióticas para ilustrar, a través del universo visual, una teoría social de tipo sociológico (Bourdieu), antropológico (Mead) o interaccionista (Goffman). Sus métodos son principalmente cuantitativos o de observación directa.

La historia de la investigación de ambas tradiciones científicas no está exenta de dificultades epistemológicas ni de debilidades empíricas. Las teorías específicamente semióticas han quedado ancladas en un estructuralismo vacío, en la mayoría de los casos debido a un formalismo teórico de inspiración greimasiana bien difundido por la escuela de París. El diálogo con las ciencias sociales es inexistente y la noción de textualidad inmanente se expande de manera tan rápida como repetitiva y superficial ha sido su aplicación, salvo en el campo de la narratología cinematográfica. Una excepción notable representa, sin embargo, la investigación aplicada a la publicidad por parte de J.M. Floch.

Las investigaciones en ciencias sociales, por su parte, disponen del material empírico circunscrito a una fecha y a un espacio, y con él se intenta universalizar principios generales sin tener en cuenta ni el origen, ni la finalidad del material, ni el medio donde se han generado las imágenes (véase M. Emison y P. Smith, 1999). A las teorías de la imagen se les critica su falta de empirismo, es decir, su indiferencia ante el método cuantitativo. A las ciencias sociales se les critica la falta de rigor epistemológico y el desconocimiento de las categorías pertenecientes a la teoría de la imagen. Por ejemplo, la insuficiencia de los análisis de contenido sobre los mensajes audiovisuales si no tiene en cuenta la formalización de los códigos de representación del medio y del género.

Estado de la cuestión en el ámbito de la computación

Los sistemas de anotación computerizada tradicionales se basan en palabras claves y pequeños extractos descriptivos en sistema textual. Se han propuesto diversos sistemas en entornos de investigación basados en imágenes de vídeo. El más importante es el Photobook del Media Lab (MIT), el Navigator del IIS (consorcio de centros y universidades, Viena) y el QBIC de IBM. Algunos de los sistemas más conocidos son Excalibur y Virage, que ya se comercializan y han sido adoptados, entre otros, por la CNN y Tele 5 (España).

En un principio, los estudios sobre el vídeo digital se centraron sobre la compresión de la información. Esta tarea ha sido tradicionalmente asumida por los compresores de archivo de sonido y de imagen, MPG y JPEG, a través de una formulación de estándares de uso universal. Pero la compresión del material audiovisual es sólo el primer peldaño de la escalera que lleva a su almacenamiento y envío. Esa tarea no puede realizarse sin dar respuesta a la búsqueda de información audiovisual por parte de los usuarios. El MPEG-7 da satisfacción a esta necesidad: en 1996, el Moving Pictures Expert Group (MPEG) constituyó una solución que se afianzó con un nuevo miembro de la familia, el MPEG 7 (Multimedia Description Interface, cuyo objetivo es la estandarización de los descriptores que permitan una búsqueda eficiente de la información multimedia (véase F. Nack y A.T. Lindsay, 1999).

Pero estos estándares no contemplan ni el análisis de la imagen ni el sistema de conocimiento para la búsqueda. Se requiere, por tanto, abordar la tarea de segmentación del material audiovisual, la obtención de *key-frames* y la extracción de contenido semántico. El primer paso en el análisis del vídeo digital consiste en su segmentación temporal (inicio y final de cada *shot*) para poder construir índices que permitan localizar las imágenes del plano de un vídeo.

En el terreno de la investigación de la información televisiva y de la imagen publicitaria, las necesidades algorítmicas se pueden integrar en cuatro apartados:

1. Segmentación del vídeo digital

Mientras que a partir del vídeo comprimido se intenta detectar los cortes de edición como discontinuidades de los vectores de movimiento, en el vídeo sin comprimir, la semejanza entre las imágenes de un plano es la base de la mayoría de los algoritmos. La variación entre los histogramas en color de las imágenes es la técnica más utilizada, tanto por su bajo costo computacional como por sus buenos resultados.

2. Generación de key-frames

La extracción de imágenes clave de una escena constituye un gran ahorro en la búsqueda automática, puesto que esas imágenes actúan como resumen (indicios) de un tamaño mayor de secuencias o planos. El sistema de extracción utiliza los criterios de similitud de imágenes.

3. Indexación y extracción automática de contenidos en vídeo digital

Los métodos de segmentación temporal de vídeo generan información que permite indexar los planos del vídeo. Los problemas de fondo son la representación de la información (extracción de descriptores) y la métrica de similitud entre la representación del plano o imagen mediante la que se interroga a la base de datos y las representaciones almacenadas.

4. Reconocimiento de personas

Dado que en el ámbito computacional no se tratan temas de significado, la búsqueda de elementos distintivos referentes a personas o animales adquiere una complejidad especial. No obstante, se ha estudiado con frecuencia la detección de color de piel en las imágenes con el fin de lograr el reconocimiento automático de rostros (o de piel, como se suele decir). Así, se han utilizado modelos deformables junto con proyección de componentes principales en un espacio de caras aprendidas para determinar localmente la dirección de ajuste del modelo. De este modo se pueden determinar las miradas a la cámara sin interferencias de iluminación que debiliten la búsqueda automática.

Avance de la ciencia computacional y un nuevo problema para la teoría de la imagen: la imagen mosaico

La tradición de la teoría de la imagen y de la semiótica de la imagen se ha centrado preferentemente en la imagen como representación del mundo. Normalmente se suele empezar por definir la imagen no narrativa o pictórica definiéndola dentro de la perspectiva renacentista. A partir de este momento se teoriza un *punto de vista fijo* del observador hacia la escena mirada. Mediante este dispositivo, el observador puede obtener una visión de la imagen en forma instantánea. Por ello, la perspectiva renacentista se manifiesta como una acción de bloqueo de toda posibilidad de movimiento de la escena. Ello significa que también la imagen se bloquea en el tiempo, constituyendo un *punto de vista*

único. Cuando se habla de movimiento en la imagen fija, ésta viene inferida por el espectador a través del descentramiento de las figuras en un cuadro o a través del efecto de *trompe l'oeil*. De este modo se pasa desde un punto de vista durativo estático hasta un punto de vista durativo móvil.

Pero con la invención del cine, la imagen en movimiento cambia la visión del observador instantáneo en un lector secuencial. Con el advenimiento de la imagen secuencial, incluyendo a la televisión, el acento de la teoría de la imagen recae sobre el aspecto narrativo, lo que se remonta a una tradición de análisis que va desde la identificación de códigos (como en el caso de Metz) hasta el análisis de las figuras actanciales del observador en el filme, de la cámara y el narrador como elementos sincréticos del punto de vista. El denominador común en los estudios sobre la imagen en los últimos treinta años es la determinación de la acción en la imagen y los métodos para analizar los cambios lógicos y cronológicos de un estado de cosas en un texto audiovisual. Se trata, en síntesis, de definir el cambio espacio-temporal e identificar las unidades sintácticas y significativas que producen el cambio.

Si una de las funciones sociales de la tecnología en cada época es acelerar el conocimiento de sus beneficiarios, en el caso del vídeo esto ocurre con generosidad, pues permite un notable avance para la comprensión de las estructuras del audiovisual como generador de información recuperable. La invención del reproductor del vídeo primero, la tentativa frustrante del videodisco después y la compresión del vídeo más tarde, abren la puerta a la explotación y gestión de la información visual. La emergencia del soporte vídeo en los años ochenta y noventa se presenta como una fuente y base poderosa de una base de datos icónicos que puede procesarse a través de la informática, abriendo la puerta a nuevas formas de acceso, de visión y de manipulación de la imagen.

La forma secuencial de almacenamiento de la imagen de vídeo provoca que la búsqueda de imágenes sea onerosa, debido al alto nivel de redundancia de las imágenes y planos en una secuencia. La única forma de examinar la imagen de vídeo es escaneando una secuencia en movimiento. De modo que para saber la información que hay en un vídeo se debe emplear tanto tiempo como dure la visión del vídeo, es decir, una operación de lectura que avanza linealmente, *frame a frame*. Si el analista debe escoger un plano de una secuencia para conservar una síntesis de ella, sabe que perderá parte de la información contenida. Pero si no se selecciona una imagen, la representación de las imágenes para la indexación se convierte así en una unidad de análisis redundante. ¿Cómo obtener la información necesaria pero económica? Una respuesta a este problema lo constituye la navegación basada en la representación mosaical de una secuencia de escenas para su indexación.

La representación mosaical o imagen mosaico comprime las imágenes de las escenas que componen una secuencia en una sola escena compacta que representa a toda la secuencia. Esta imagen mosaico muestra, sin embargo, en forma explícita, todos los movimientos y las acciones que se han realizado dentro de la secuencia. La colección de imágenes secuenciales concentradas en una sola escena constituye un verdadero «sumario» o resumen de una serie de datos del

vídeo. Con esta operación se ha logrado pasar de un solo *frame* a un mosaico de escenas que contiene todos los planos de una secuencia. Así se logra recuperar un sentido de *coherencia* de la secuencia, que en el caso de la selección de un *frame* por secuencia se perdía. Como se sabe, el principio de coherencia es un concepto que pertenece al análisis del discurso, una categoría que en el campo de la imagen permite analizar el recorrido de la lectura por parte del espectador a través de una tematización (cfr. Vilches, 1983).

¿Cómo se crea la información fundamental que obtiene el observador en una imagen mosaico?

La representación de la escena base obtenida se divide en tres partes:

1. Una imagen mosaico panorámica: la imagen mosaico se construye alineando todos los planos producidos por el movimiento panorámico sobre una escena fija.
2. Las representaciones geométricas que contienen la información necesaria para diseñar el lugar de cada escena anterior y posterior entre la imagen mosaico panorámica y el plano aislado.
3. La información dinámica que se obtiene acerca de los objetos en movimiento que son capturados por la imagen mosaico panorámica. De este modo, la información del movimiento es capturada completamente a través de la representación de las trayectorias temporales realizadas por el objeto, además de su apariencia.

La imagen mosaico suscita una serie de reflexiones relativas al concepto de la visión, el lugar que ocupa en nuestra cultura la categoría del ojo como ventana de la mente y, por tanto, dotado de la capacidad de conocer a través de las estructuras geométricas representadas en una escena. Si la teoría del punto de vista había insistido en la localización y temporalización del *sujeto observador* que es capaz de ver sólo desde una perspectiva siempre parcial, la investigación sobre la imagen mosaico plantea la posibilidad de una subordinación del concepto de la observación del sujeto en beneficio de la *escena del objeto*. Pero, por otro lado, la imagen mosaico tiende a comprimir la experiencia temporal en una visión dinámica que permite seguir la evolución de un objeto en el espacio. Aunque lo que tiene delante de sí el espectador es una compresión de la experiencia durativa en una visión instantánea, semejante a la fotografía donde sólo podíamos intuir la experiencia del paso del tiempo, aquí, por el contrario, se muestran explícitamente las huellas del paso del tiempo entre escena y escena.

Con la imagen mosaico se abre un gran campo para la manipulación de la imagen. Si en la secuencia de vídeo siempre era posible insertar un nuevo punto de vista, y el observador se encontraba en un régimen de visión abierto y subjetivo, en la imagen mosaico ya están incluidos todos los posibles puntos de vista. En ese sentido es una imagen cerrada, es decir, una imagen máquina. El punto de vista pasa a ser así histórico, contingente y local, una categoría de totalidad sinóptica.

Tareas para una colaboración entre la informática y la teoría de la imagen

La creación de sistemas de indexación semántica de vídeo para géneros de información y publicidad

Se hace imprescindible la investigación en un marco de convergencia de disciplinas con el fin de progresar en la construcción de bases de datos inteligentes en el campo del audiovisual.

El objetivo general de esta colaboración ha de ser construir un modelo de uso interactivo sobre la masa de imágenes vídeo aprovechando los recursos de las tecnologías digitales y basado en el análisis semántico de imágenes, las estructuras narrativas y el procesamiento inteligente.

Su campo de intervención han de ser todos los géneros a los que aplicar modelos de indexación semántica, esto es, los géneros cinematográficos, las series televisivas, los informativos diarios y magacines, los espots publicitarios y el emplazamiento del producto.

Para ello, se ha de partir de las aportaciones teóricas y empíricas, científicas o experimentales existentes en ambos campos de trabajo, con el fin de crear un sistema de indexación semántica estructurada según la lógica del género y del soporte o tecnología de los medios de comunicación.

Dado que ya existen ciertas utilidades informáticas que permiten la segmentación del vídeo y la búsqueda de imágenes, la teoría de la imagen puede colaborar activamente en mejorar esa tarea a través de estructuras que permitan la búsqueda por núcleos significativos y no sólo a través de imágenes aisladas.

Con ese objeto, se propone trabajar en dos niveles o planos que permitan una lectura y recuperación de la información homogénea: planos plástico-figurativos que pertenecen al orden de soportes visuales y a las unidades elementales de la imagen (color, iluminación, volumen, planos, etc.), y planos semánticos de significados que permitan el reconocimiento de personajes, escenarios y acciones codificadas. La novedad de este modelo metodológico consiste en la posibilidad de deducir del mismo plano expresivo (el plano de las formas puras y de las técnicas audiovisuales de la realización) los tópicos semánticos y las inferencias temáticas para evitar la aleatoriedad y el subjetivismo de la anotación archivística actual.

La teoría de la imagen, especialmente la semiótica, la retórica visual y la teoría narrativa, pueden aportar a las ciencias de la información computerizada utilidades semánticas. Ellas pueden consistir, por ejemplo, en el caso de la información televisiva: en identificación de los actores, el tipo de rutina productiva, el contexto comunicativo y persuasivo de la noticia (escenarios, jerarquías, grado de importancia). En el caso de la producción audiovisual publicitaria, el estudio del espot y de otros géneros (como el *product placement*): marcos visuales, puntos de vista del producto, tematización de la marca y catalogación de valores del consumo con el fin de relacionarlos con los sistemas de recepción o interpretación de los espectadores.

Tareas del objetivo común

Teoría de la imagen e informática tienen ante sí una nueva tarea, cuyos objetivos concretos son proceder a la creación y experimentación de un sistema inteligente de recuperación y sistematización de información televisiva, a través de la creación de un sistema de indexación de grandes cantidades de información audiovisual proveniente de la información televisiva diaria y periódica para su aprovechamiento en bases de datos. El sistema de indexación tiene una aplicación inmediata en la recuperación de archivos que contengan unidades y secuencias de imágenes que permitan a los profesionales de la información el análisis y el reaprovechamiento documental. En este nuevo campo de intervención de la teoría de la imagen se hacen necesarias otras aportaciones clave para la información, como es el caso de los métodos y las técnicas de la documentación audiovisual en un sistema que permita el análisis de la imagen y sus estructuras informativas y, al mismo tiempo, la identificación concreta del comportamiento de la audiencia en relación con las secuencias de información.

El sistema de indexación semántica es una herramienta de búsqueda textual e icónica que permite una localización veloz y sistemática de contenidos y expresiones estéticas de la información.

Esta herramienta se basa en:

- Segmentación de vídeo en planos de forma automática y generación de una representación icónica con capacidad para resumir el contenido de cada plano y secuencia.
- La conservación sistematizada de información de vídeo, escaletas o sinopsis documentales de los informativos e imagen sumario de un spot publicitario.
- Extracción y rastreo de descriptores significativos a partir del análisis automático de secuencias de vídeo.
- Integración de anotaciones textuales externas e internas a cada programa de información (datos técnicos de producción, identificación de personajes, escenas geográficas, información textual al espectador, tales como tiempo y modalidad del programa que se emite) o emisión de publicidad (comercial e institucional).

Los objetivos de la investigación en el marco de la producción del audiovisual y de la sociedad de la información

Se consideran como objetivos de la investigación los siguientes tópicos:

- Sistema de indexación basado en la duración del plano visual de la escena como unidad de segmentación.
- Establecimiento de un sistema de identificación automática de elementos espaciales, temporales y actoriales de una secuencia informativa audiovisual.
- Establecimiento de códigos automáticos de indexación en soportes digitales en el momento de la emisión de la información y de la post-producción publicitaria.

- Creación de un sistema de reconocimiento facial para identificación de profesionales de la información televisiva y de personajes de la noticia, y en general, de todos los actantes de un espot publicitario.
- Creación de un sistema de identificación de los espacios de la información y escenarios de la publicidad.
- Creación de un sistema de reconocimiento de los diversos parámetros temporales utilizados en la producción y la emisión.

La investigación sobre la indexación automática de bases de datos audiovisuales se enmarca en los grandes objetivos de la sociedad de la información. Se trata de construir un modelo de conocimiento interactivo sobre la masa de información audiovisual aprovechando los recursos de las tecnologías de la televisión y de las tecnologías avanzadas de la información para la creación de *software multimedia*.

Un efecto coherente con lo anterior es la mejora de los sistemas de búsqueda, navegación, identificación y recuperación de imágenes de la industria televisiva. De este modo se puede contribuir a la creación de una base de datos sobre la información audiovisual para uso social más amplio, destinada a los profesionales de la información, a los responsables de políticas audiovisuales, así como también a investigadores, expertos y educadores. En este sentido, el aprovechamiento de un sistema de análisis de los archivos de la información puede permitir una ayuda preciosa en la investigación para periodistas, historiadores, sociólogos, polítólogos, psicólogos, antropólogos.

Cómo organizar la investigación sobre la indexación semántica

Indexación de la información televisiva

Dado que la información es la construcción de un espacio semántico discursivo, la metodología para crear un sistema de indexación semántica está determinada por los factores espaciales, espacio-temporales y actoriales que intervienen en la producción de un programa de informativos en televisión.

La determinación de las modalidades espaciales de la información constituye uno de los terrenos privilegiados de la manipulación de la imagen

Con el fin de determinar los escenarios de la noticia, se procederá a definir un *espacio vídeo*, que se compone de *espacio televisivo* (en el cual se diferenciarán los espacios de los estudios centrales de las televisiones respecto de los centros regionales, corresponsalías y extranjero) y de *espacio extratelevisivo* (en el cual se diferenciarán las localizaciones habituales de las noticias de las localizaciones inhabituales).

En relación con las formas de producción de los escenarios, se procederá a distinguir las siguientes formas de *producción de la información*: el material real grabado tanto en plató como en exteriores; el material de la posproducción y el relativo a la sonorización de la imagen, digitalización y editaje, y el material cuya existencia se deba exclusivamente a tratamiento digital.

Dado que toda la producción de la información se visualiza en la pantalla del televisor, se analizará en un nivel sintáctico como una *unidad visual* de lectura (códigos arbitrarios e icónico-cinéticos) realizada durante una *unidad de tiempo* de lectura. Estas unidades de lectura permitirán la identificación de una *continuidad y homogeneidad* sintagmática del flujo del vídeo (según códigos narrativos de secuencialidad) y al mismo tiempo la individualización de estructuras de empaquetamiento mediante *agrupaciones heterogéneas* de secuencias paradigmáticas (según códigos de géneros y de temas). Estas secuencias paradigmáticas podrán señalar, en el momento de la anotación de las características visuales de la información, las imágenes provenientes de operaciones de *reconocimiento*, tales como escenarios, viñetas, indicios y huellas materiales; las imágenes provenientes de operaciones de *ostensión*, tales como presentadores y actores de las noticias; las imágenes provenientes de operaciones de *réplica*, tales como vectores de movimientos de cámara, movimientos dentro del plano, iconogramas y pictogramas; las imágenes provenientes de operaciones de *invención* o imágenes virtuales, tales como proyecciones y efectos especiales.

Determinación de las modalidades temporales

La construcción de unidades de medición temporal de la información será determinada a partir de la diferenciación de tres orígenes del material en vídeo:

- La temporalidad tecnológica: formas temporales producidas por el conjunto de tecnologías que intervienen en la información (satélite, terrestre, cable).
- La temporalidad del género informativo: se construirán unidades de segmentación que permitan diferenciar:
 - a) El tiempo de la noticia: el tiempo del acontecimiento real en relación con el tiempo de la narración, el tiempo de los protagonistas de la información respecto al tiempo de los observadores.
 - b) El tiempo de la estructura de la información: las temporalidades de la simultaneidad y los tiempos del diferido, las duraciones del programa, de las secciones, de las secuencias, presentaciones visuales y caretas, ritmo de las secuencias informativas en tiempo real, tiempo y ritmo de la edición en vídeo.
 - c) El tiempo de referencia del informativo respecto al espectador (tiempo de la emisión), respecto a los informadores (tiempo de procesamiento de la información) y respecto del acontecimiento acaecido (tiempo de las acciones). Estos tiempos de referencia serán discriminados a través de la comprobación de concomitancia/no concomitancia entre la duración de la información y su efectiva recepción por parte del espectador.
- La temporalidad narrativa: temporalidades de las formas de presentación, de los presentadores y periodistas, los tiempos de la información en relación con los géneros (tiempo de la política, del deporte, de sucesos, etc.).

identificación de figuras narrativas del tiempo, tales como la duración, la elipsis y *raccord*, la dilatación, las pausas, las interrupciones publicitarias, etc.

La temporalidad narrativa permitirá también identificar uno de los asuntos centrales de la teoría de la información periodística: la tematización. Ésta se establecerá a través de la frecuencia de aparición de ítems de contenido que permitirá diferenciar entre unidades de tematización aislada o serializada (durante un ítem que se continúa en una semana de información, por ejemplo). *La identificación de procesos de tematización permitirá a los usuarios la búsqueda por palabras claves de temas concernientes a las agendas informativas objeto de la consulta.*

Determinación de los actores y objetos de la información

Las categorías de los actores y objetos de la información serán definidas a través del concepto de **actantes**, lo cual permite identificar y diferenciar entre actantes personas, animales, naturaleza y objetos que aparecen en la información.

La identificación de los actantes de la noticia se hará a través del establecimiento de las estructuras textuales de un programa informativo. Esto supone determinar tanto:

a) Los niveles de expresión y contenido audiovisual:

Plano de la expresión: — Imágenes y sonidos de la información.
— Textos escritos y verbales de la información.

Plano del contenido: — Identificación de imágenes y sonidos.
— Identificación de textos escritos y verbales.

b) Las formas de visualización de los actores de la información:

Plano de expresión: — Actantes visualizados
— Actantes verbalizados } Formas de presentación icónico-perceptivas

Plano de contenido: — Actantes visualizados
— Actantes verbalizados } Identificación de nombres y roles

Con el fin de determinar con exactitud las formas de presentación de los actantes o actores de la información, se crearán sistemas analíticos de visualización que permitan su identificación en los escenarios, planos y efectos mediante mediciones de intensidad perceptiva que irán desde la máxima visualización hasta la mínima visualización.

Con el objetivo de integrar variables de movimientos de cámara y composición de las escenas informativas, se establecerán tablas analíticas de control de la actividad de las telecámaras y sus modalidades perceptivas (zonas de interés visual), comunicativas (mostrar y dar a conocer) y actanciales (puntos de vista).

Dado que la aparición de personajes públicos (sobre todo políticos) forma parte del material «sensible» de la información, se construirán plantillas analíticas que permitan la identificación del declarante, las formas de presentación, duración, verbalización, explicitación del contexto y calidad de la imagen y su relación con los temas y repertorios informativos.

Los resultados previsibles a los que la teoría de la imagen y la indexación informática pueden llegar en el campo de la información televisiva se pueden sintetizar en la obtención de un sistema de tratamiento semántico de la información que permita una aplicación de parámetros de análisis en un nivel formal y en un nivel de contenidos. Estos parámetros se hallan:

a) En un nivel formal:

- Identificación de los actores de la noticia y de los productores de un programa de información.
- Identificación de los espacios de la noticia, tanto geográficos como de producción.
- Contexto perceptivo y comunicativo de los actores de la noticia (perspectiva, lateralidad, angulación, planos y marcos, escenarios, mensajes lingüísticos, focalizaciones, movimientos de cámara y de objeto, etc.).

b) En un nivel de contenidos:

- Identificación de las agendas informativas a través de la representación icónica, tematizaciones e importancia de la información y de sus actores a través de los parámetros de jerarquía, orden, duración, repetición, serialización.
- Identificación de estructuras de compaginación de la información audiovisual, segmentación de secciones de escaletas y emisiones de programas, discriminación de géneros y tópicos de la información.

La indexación de la imagen publicitaria: análisis semántico del espot comercial

Las categorías de la teoría de la imagen que se pueden aplicar a la información televisiva son también las que pueden aplicarse en el caso de la imagen publicitaria. Existen grandes diferencias entre ambos géneros, no obstante, que han de tenerse en cuenta en la construcción de una metodología para la indexación automática. Entre las diferencias más importantes se hallan los tamaños de los productos (de treinta a sesenta minutos para la información, menos de un minuto en el espot publicitario); la ausencia de narraciones secuenciales propiamente dichas; la estética predominante en cada uno de los productos; las restricciones de género (información versus seducción); la profusión de efectos de posproducción y de imágenes no reales en la publicidad, que obliga a

un análisis pormenorizado de cada imagen, frente al predominio de la imagen real y los pocos efectos de posproducción de la información televisiva.

La investigación de los espots comerciales de televisión se estructura en torno a las estrategias comunicativas del espot a través del estudio de su presentación visual, de modo que permita una indexación rápida de grandes cantidades de imágenes.

Se trata de establecer un sistema de identificación e indexación del espot comercial de televisión basado en el análisis visual y semántico del programa.

La posibilidad de construir una parrilla de análisis de grandes cantidades de espots basados en su aspecto visual y en sus funciones semánticas está facilitada por los sistemas de anotación de vídeo, que ya hemos expuesto en la primera parte de este trabajo. Por tanto, se puede incluso ir más allá del reconocimiento de parámetros espaciales, temporales y geométricos para situar la imagen dentro del universo económico y social del producto.

Esto es, se pueden encontrar elementos estándares de color, espacio, ritmo, personajes y objetos que pueden asociarse a valores semánticos del consumo que los espectadores relacionan dentro de su cultura. No se trata sólo de productos o de marcas que puedan ser reconocidos por los espectadores, sino de veredaderos códigos icónicos que tienen un funcionamiento muy estandarizado en la cultura de masas. Entre estos, los valores del consumo.

Se entiende por *valores del consumo* las estrategias discursivo-visuales de las imágenes de los espots comerciales en relación con posibles tópicos de interpretación por parte de los espectadores. Estos tópicos están representados por valores visuales que, elaborados según la estética del género publicitario, pueden representar significaciones precisas. Éstas pueden ser *prácticas*, si el espot busca sólo la presentación del producto; de fantasía *utópica*, si el espot se permite plantear situaciones oníricas y de fantasía al espectador; de evaluación *crítica* del producto, o de una relación entre producto o marca y espectador basada en una estética de tipo *lúdico*.

Si estos estándares visuales se pueden relacionar con valores del consumo, se puede construir una parrilla de espots que pueden ser identificados, anotados y procesados en una base de datos para su posterior gestión interactiva por un productor o un analista de marketing.

La definición operativa de las variables de análisis se realiza a través de un estudio del concepto de *aspecto visual*, utilizando la teoría de la imagen, y un estudio del concepto *valor del consumo*, utilizando la teoría textual del análisis del marketing (J.M. Floch).

Esta definición de las categorías teóricas se lleva a cabo a través del análisis textual y el análisis semántico. Diversas teorías son necesarias para la preparación de las categorías que han de servir para la investigación: la psicología de la percepción visual, especialmente la Gestalt, así como la teoría narrativa, la teoría de la imagen, en particular la lectura de unidades de la expresión y lo relativo a la construcción de personajes o actantes visuales.

La retórica visual tiene también una gran importancia a la hora de definir los sistemas de manipulación de las formas, los colores y los marcos visuales

de la representación publicitaria. Se trata de los *procedimientos de manipulación visual*.

Para ello se procede a establecer dos niveles de análisis, correspondientes a los códigos de la expresión visual y a los códigos del contenido visual.

Esto es, buscar los procedimientos de producción y posproducción correspondientes a:

- a) Manipulación en soportes y superficies del vídeo: supresiones, adjunciones, sustituciones y conmutaciones operadas a nivel de los soportes en vídeo durante la fase de posproducción: colores, textos, figuras, fondos, etc.
- b) Manipulación de contenidos del vídeo: supresiones, adjunciones, sustituciones y conmutaciones operadas a nivel del contenido semántico y sintáctico.

Las estructuras de la expresión temporal del espot se constituyen a través de los conceptos de duración y ritmo de cada plano o escena del producto audiovisual.

El espot es una creación sintética en el cual convergen complejas operaciones de simulación temporal, así como manipulaciones posproducción que son la base de este género comercial. Por ello, se han de aplicar los conceptos de unidades mínimas que serán detectadas en la segmentación automática del vídeo, tales como el número de cortes (*shot*), disolvencias; la duración de estas unidades a través de la identificación de los planos, y los movimientos de las figuras dentro de los planos y movimientos de cámara.

Una de las tareas más difíciles de objetivar son los análisis cualitativos realizados en el campo de la publicidad sobre comprensión y recuerdo de los espectadores. La teoría de la imagen puede mejorar los estándares analíticos de la investigación cualitativa a través de la explicitación de las *actividades cognitivas* que realiza el espectador frente a un texto visual.

Las actividades cognitivas se refieren al aspecto narrativo del espot publicitario, a propósito del cual se pueden determinar el grado de información que posee el espectador, así como la actividad que desarrolla durante su tarea de comprensión. Esto está relacionado tanto con el conocimiento del género publicitario como con la capacidad de establecer un *frame* donde sea posible relacionar las diversas escenas del espot. Esta actividad permite analizar y detectar el grado de importancia o focalización entre escena y producto, así como establecer posibilidades del recuerdo basado en las relaciones jerárquicas y contextuales en relación con la marca y sus posibles valores sociales y simbólicos reconocidos.

La semántica de los colores y sus atribuciones en la industria publicitaria y en los análisis cualitativos permitirá al analista trabajar con elementos típicamente valorativos, porque existe una hipótesis bastante generalizada entre los productores de espots según la cual determinados colores y líneas geométricas que estructuran las escenas pueden permitir la inferencia de un valor afectivo, de un estado de ánimo y de efectos comunicativos entre los destinatarios o receptores.

La espacialidad semántica de las escenas se halla en la producción de los planos a través de las líneas horizontales, verticales y de las perspectivas de los espots. Todos los planos tienen reconocida una atribución de significados tales como calma, relax, dinamismo, agresividad, etc. El estudio de estas características con vistas a ser utilizadas en el análisis de la segmentación automática deberá relacionarse con las categorías del movimiento y de la temporalidad de cada plano, de tal manera que se podrán obtener catálogos de espots de quietud o de euforia a través de la aplicación de criterios perfectamente objetivables y aplicables a una gran cantidad de material publicitario.

Conclusión

Este trabajo presenta los objetivos de una investigación de anotación computacional semántica de la imagen que combina la información de bajo nivel y alto nivel.

La información de bajo nivel se refiere a los estándares de color y espacio, que pueden ser determinados a través del sistema informático de análisis de la imagen. La información de alto nivel se refiere a elementos significativos que normalmente se hallan sistematizados y categorizados por la teoría de la imagen. En este trabajo se ha tratado de mostrar las posibilidades de una convergencia entre dos ámbitos científicos que se ocupan de la imagen, la visión por computador y las teorías de la imagen, a través de las tareas de segmentación, anotación, indexación y navegación sobre una determinada base de datos visuales recogidos a partir de los géneros de la información y la publicidad televisiva. Esta nueva vía de investigación presupone una definición más operativa (y empírica, si cabe) de algunas categorías de la teoría de la imagen, con la ayuda de la percepción visual, la retórica visual y la teoría narrativa, con el fin de facilitar un cuadro de ítems que puedan ser utilizados para la búsqueda semántica de grandes cantidades de imágenes procesadas. Por su parte, los constantes descubrimientos de nuevas herramientas informáticas para el análisis y la conservación de las imágenes suponen, en algunos casos, enfrentarse con nuevos conceptos, como el de la imagen mosaico, que propone nuevas perspectivas teóricas sobre la representación de la imagen.

Bibliografía

- BINEFA, X.; SÁNCHEZ, J.M. (2001). *AudiCom: a Vídeo Analysis System For Auditing Commercial Broadcast*. Centro de Visión por Computador. Departament de Informàtica, Facultat de Ciències, Universitat Autònoma de Barcelona.
- CALIANI, M.; COLOMBO, C.; DEL BIMBO, A.; PALA, P. (1997). *Commercial Video Retrieval by Induced Semantics*. IEE Multimedia.
- EMMISON, M. (2000). *Researching the Visual*. Londres: Sage.
- GOFFMAN, E. (1979). *Gender Advertisements*. Londres y Basingstoke: Macmillan.
- HUIJSMANS, D.P.; SMEULDERS, A.W.M (eds.) (1999). *Visual Information and Information Systems*. Springer.

- IRANI, M.; ANANDAN, P. (1999). *Video Indexing Based on Mosaic Representations*.
- NACK, F.; LINDSAY, Adam T. (1999). *Everything You Wanted to Know About MPEG-7*. IEE Multimedia.
- VILCHES, L. (1999). *La lectura de la imagen*. Barcelona: Paidós, 10^a edición.

Lorenzo Vilches es licenciado en Filosofía, doctor en Ciencias de la Información y profesor de la Universitat Autònoma de Barcelona. Entre sus obras se cuentan *La lectura de la imagen* y *Teoría de la imagen periodística*.
