

# Realidad Aumentada en HbbTV

## Implementación de una plataforma Hypervideo para la Televisión Digital Conectada

Toni Bibiloni, Miquel Mascaró, Pere Palmer, Antoni Oliver

Universitat de les Illes Balears – Departamento de Matemáticas e Informática

Laboratorio de Tecnologías de la Información Multimedia, LTIM

Palma de Mallorca, Illes Balears, España

{toni.bibiloni, mascport, pere.palmer, antoni.oliver}@uib.es

*Abstract*—En este artículo se presenta un sistema de Realidad Aumentada para la Televisión Digital Conectada a través de la implementación de una plataforma Hypervideo. Se presenta un back-office para la edición de contenido audiovisual con “hot-spots”, la gestión de un repositorio audiovisual y un player de vídeos interactivos en terminales HbbTV. El player permite al usuario obtener información adicional sobre los puntos de interés que aparecen sobre las imágenes. El presente proyecto presenta una plataforma útil para los equipos de producción audiovisual, al permitir crear contenidos audiovisuales que combinan, de forma sencilla, tres conceptos tecnológicos: vídeo interactivo, realidad aumentada y televisión conectada a Internet. Se concluye con diversas posibilidades de mejora y ampliación de la plataforma en varios dispositivos.

*Augmented Reality; HbbTV; Hypervideo; Interactive Digital Storytelling*

### I. INTRODUCCIÓN

El concepto de Hypervideo, o “vídeo interactivo” [1], se define como un flujo de contenido audiovisual que ofrece al usuario una navegación no lineal. El espectador puede interactuar con el contenido mediante enlaces o hipervínculos, que son complementados con otros mecanismos, como búsqueda, información adicional, salto de secuencia o contenido audiovisual, etc. todo ello enfocado a mejorar el acceso a la información y con el propósito de llevar al espectador de un estado pasivo al de un participante activo [2].

Al mezclar conceptos de hypervideo con imágenes reales, nos acercamos a las aplicaciones de realidad aumentada. La RA es el término que se usa para definir una visión directa o indirecta de un entorno físico del mundo real, cuyos elementos se combinan con elementos virtuales para la creación de una realidad mixta.

En este artículo presentamos el desarrollo de una plataforma de vídeos interactivos, la plataforma Hypervideo, capaz de ofrecer información aumentada en receptores de televisión HbbTV, garantizando en consecuencia aspectos como la sincronización, interactividad directa sobre las imágenes reales, accesibilidad universal al utilizar un receptor tan conocido como la televisión y facilidad de uso.

En los siguientes apartados se explican en detalle los módulos desarrollados y las fases necesarias, desde la

filmación, edición de contenido audiovisual, creación de hypervideos, y finalmente la visualización del contenido audiovisual interactivo en el player HbbTV. Finalmente, presentamos las investigaciones actuales y futuras, para terminar con la presentación de las conclusiones extraídas.

### II. SITUACIÓN TECNOLÓGICA

Una de las primeras implementaciones que se llevaron a cabo de un hypervideo fue en [3], donde los enlaces entre escenas del vídeo permiten al espectador elegir cómo se desarrolla el argumento del contenido audiovisual mediante el cambio de escenas. En [4], [5] se propone un cambio en el comportamiento de los enlaces: éstos pueden ser utilizados para obtener información adicional sobre el contenido que se visualiza en ese momento, dando un paso hacia la RA.

En [6] se presenta el diseño y evaluación de “Hvet”, un entorno hypervideo para la enseñanza de cirugía veterinaria, cuyas conclusiones demuestran el potencial de estos entornos en sistemas educativos.

El proyecto LinkedTV [7] propone una plataforma hypervideo basada en las especificaciones web y broadcast de la industria (HTML5, HbbTV), incorporando información adicional de manera automática accediendo a la riqueza de datos disponibles en la WWW a través de LinkedData.

La tendencia actual es la media interactiva, como demuestran servicios de imagen interactiva, como Thinglink [8], o vídeo interactivo, como Wirewax [9]. Aunque se detectan pocas plataformas tecnológicas destinadas a la producción de contenido audiovisual interactivo, y prácticamente son inexistentes las aplicaciones de realidad aumentada en entornos de televisión digital. La mayoría de estas aplicaciones están basadas en soluciones de “segunda pantalla”. Al analizar el uso de la segunda pantalla, se detectan varios problemas: falta de sincronización entre los contenidos audiovisuales mostrados en el televisor y la información aumentada, mínima o nula interacción del usuario con la emisión, limitación de acceso a los usuarios, debido a que es imprescindible disponer de un *Smartphone* e instalar una app, y la complejidad de utilizar dos dispositivos simultáneamente.

Esta situación, junto a la falta de producciones hypervideo con fines educativos, promocionales, o divulgativos fue la motivación de este proyecto: especificar y desarrollar una plataforma tecnológica destinada a la producción, distribución

y visualización de contenido audiovisual interactivo sobre el estándar de TV interactiva HbbTV.

### III. PLATAFORMA HYPERVIDEO

En esta sección se presenta la arquitectura de la plataforma de creación, publicación y visualización de contenidos audiovisuales Hypervideo. La estructura del vídeo interactivo está formada por tres niveles de datos: contenido audiovisual, información de los puntos de interés (PoI, *Point of Interest*) y los marcadores o “hotsposts” que representan la ubicación de los PoIs sobre las imágenes del vídeo.

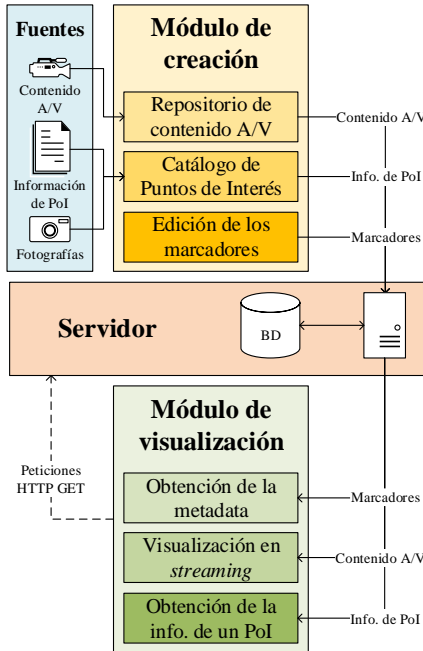


Figura 1. Arquitectura completa de la plataforma Hypervideo

La Figura 1 muestra la arquitectura propuesta de la plataforma, compuesta por el módulo de creación, el módulo de visualización y la comunicación de ambos con el servidor hypervideo.

#### A. Módulo de creación

El módulo de creación comprende los componentes básicos para la gestión del repositorio audiovisual y de la información aumentada, ambas descritas en detalle en las secciones “Repositorio de contenido audiovisual y “Catálogo de Puntos de Interés”.

Estas dos fases, aunque totalmente independientes entre sí, generan conjuntamente los metadatos espaciotemporales necesarios para relacionar los puntos de interés identificados en las imágenes reales con la posición visual de los marcadores en la reproducción del contenido audiovisual.

#### B. Módulo de visualización

El módulo de visualización representa el player hypervideo, una aplicación dedicada a reproducir el contenido del repositorio audiovisual en un entorno hypervideo, ofreciendo: filtrado de capas de información, selección de *hot-spots*,

visualización de RA, acceso y búsqueda de contenidos AV a partir de PoIs.

#### C. Modelo de datos

La información utilizada por la plataforma –contenido A/V, puntos de interés y marcadores– se almacena en una base de datos MySQL; el modelo de datos se muestra en la Figura 2 y se detalla a continuación:

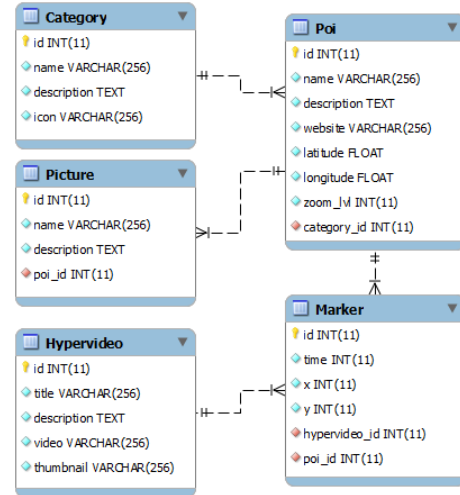


Figura 2. Modelo de datos de la plataforma Hypervideo

La clase Hypervideo contiene la información básica para la reproducción del contenido audiovisual, indicando el *path* del fichero, además de los datos descriptivos del proyecto: título, descripción e icono.

La clase PoI representa, juntamente con las clases Picture y Category, la información relacionada con un Punto de Interés. Esta última agrupa los PoI que tienen aspectos en común, indicando el atributo *icon*, que indica la imagen del marcador que representa los PoI de esa categoría en el player. La clase PoI contiene los atributos descriptivos del punto de interés, junto con la tripleta de localización: latitud, longitud y nivel de zoom, necesaria para representar el mapa en el módulo de visualización. La clase Picture almacena la información gráfica relativa a las fotografías enlazadas a un punto de interés.

Finalmente, las apariciones de los puntos de interés en un hypervideo se plasman en la clase Marker, que indica el tracking de la posición (x,y) a lo largo del tiempo de vida del marcador de un PoI en concreto. La estructura de esta clase permite la aparición de un mismo punto de interés en varios hypervideos de manera independiente.

#### D. Comunicación con el servidor

El servidor hypervideo actúa de agente intermedio entre los módulos de creación y de visualización con la base de datos. El servidor Apache recibe las llamadas HTTP de ambos módulos: el primero introduce los datos mediante peticiones HTTP POST, mientras que el de visualización los obtiene, en formato JSON, mediante llamadas HTTP GET.

En la Figura 3 se muestra la respuesta a la petición de los metadatos por parte del módulo de visualización, que se divide en cuatro secciones: información del hypervideo, listado de categorías, listado de PoIs con su categoría y tracking de marcadores.

```

{
  "hypervideo": {
    "title": "This is an example",
    "video": "http://ltim.uib.es/example.mp4",
    "categories": {
      "1": {
        "name": "First category",
        "icon": "http://ltim.uib.es/cat1.png"
      },
      "2": {
        "name": "Second category",
        "icon": "http://ltim.uib.es/cat2.png"
      }
    },
    "pois": {
      "4": 1,
      "5": 1,
      "6": 2,
      "7": 2
    },
    "markers": {
      "21": {
        "4": {
          "x": 29.27,
          "y": 18.12
        },
        "5": {
          "x": 62.55,
          "y": 18.12
        }
      },
      "22": {
        "4": {
          "x": 28.34,
          "y": 18.47
        },
        "5": {
          "x": 63.64,
          "y": 18.33
        },
        "6": {
          "x": 13.82,
          "y": 36.89
        }
      }
    }
  }
}

```

Figura 3. Metadatos de un Hypervideo en formato JSON

En la Figura 4 se muestra la información adicional relacionada con un punto de interés, codificada en JSON, tal como la entrega el servidor hypervideo al recibir la llamada GET de un PoI en concreto.

```

{
  "poi": {
    "name": "This is the name",
    "description": "Why it is important",
    "location": {
      "latitude": 39.637,
      "longitude": 2.644,
      "zoom_level": 10
    },
    "pictures": [
      "http://ltim.uib.es/pic1.jpg",
      "http://ltim.uib.es/pic2.jpg",
      "http://ltim.uib.es/pic3.jpg"
    ]
  }
}

```

Figura 4. Información de un PoI en formato JSON

#### IV. MÓDULO DE CREACIÓN

El objetivo de este módulo es facilitar la creación de hypervideos, ofreciendo interfaces para las tareas de gestión del repositorio audiovisual, la creación del catálogo de puntos de interés y la especificación espaciotemporal de los marcadores. La Figura 5 muestra la secuencia del proceso de creación, cuyas fases se explican en los siguientes apartados:

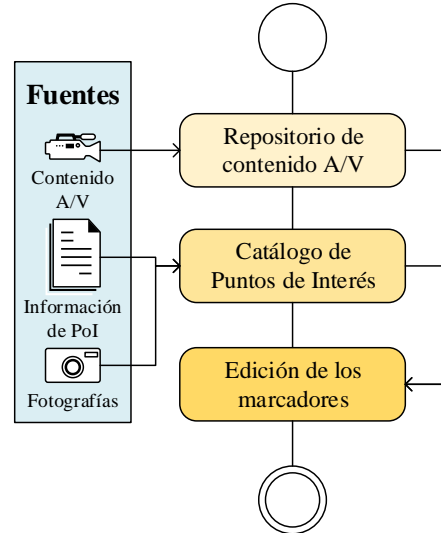


Figura 5. Detalle del módulo de creación de Hypervideos

##### A. Repositorio de contenido audiovisual

Con el propósito de garantizar la compatibilidad con el estándar HbbTV 1.1 [10] los contenidos audiovisuales deben cumplir los siguientes parámetros: códec vídeo H264/AVC, códec audio HE-AAC y contenedor MP4.

Se muestra la interfaz de gestión del repositorio en la Figura 6.

Figura 6. Interfaz de gestión del repositorio A/V

##### B. Catálogo de Puntos de Interés

El catálogo de PoIs está organizado en categorías. Para cada PoI se especifica a qué categoría pertenece y la información descriptiva: texto, imágenes, URL y posición geográfica. La interfaz de gestión del catálogo de PoIs se muestra en la Figura 7.

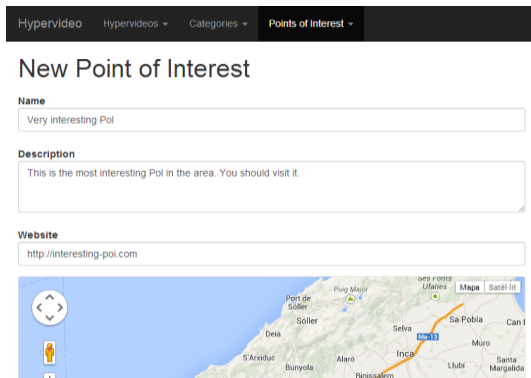


Figura 7. Interfaz de gestión del catálogo de Puntos de Interés

Con el propósito de garantizar un rendimiento de usabilidad aceptable en el módulo de visualización, se han tenido que definir las siguientes restricciones para las imágenes asociadas a un PoI: JPEG, 450x338px y alrededor de 200KB.

### C. Edición de los marcadores

El propósito de los marcadores es el de posicionar los puntos de interés sobre las imágenes del vídeo. Para facilitar la edición de marcadores, se ha desarrollado una herramienta en el módulo de creación, permitiendo al productor navegar sobre el contenido audiovisual, seleccionar los intervalos temporales donde aparece un punto de interés, y especificar la posición del marcador, indicando tantas posiciones clave como sean necesarias. En la Figura 8 se muestra la interfaz de edición de marcadores.

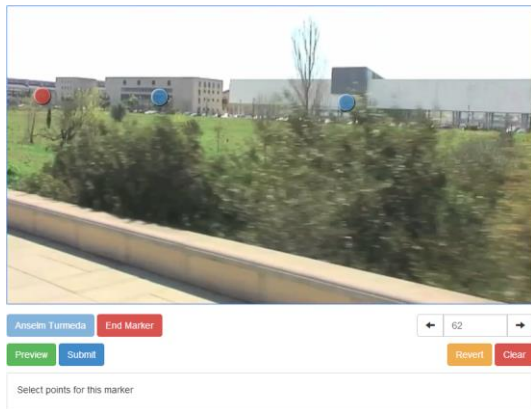


Figura 8. Interfaz de edición de los marcadores de un Hypervideo

Una vez definidas las posiciones clave para un marcador, el sistema genera mediante interpolación lineal las posiciones intermedias para cada segundo de visualización. A posteriori, si es necesario, el editor puede corregir las posiciones generadas automáticamente.

El resultado de especificar los marcadores se plasma en la base de datos como el tracking temporal de todos los marcadores asociados como una tupla, como se describe en (1): en función del hypervideo  $h$ , el segundo  $t$ , y el PoI  $p$ , qué posición  $(x,y)$  relativa al tamaño del vídeo adopta un marcador.

$$f(h,t,p) = (x,y) \quad (1)$$

## V. MÓDULO DE VISUALIZACIÓN

El módulo de visualización representa el “player hypervideo”, desarrollado en HbbTV 1.1, siguiendo el estándar CE-HTML [11], haciendo uso de CSS TV Profile 1.0 [12] para el diseño y Javascript para la interactividad.

### A. Estados de la aplicación

En la Figura 9 se muestra la secuencia de los estados del player hypervideo y las acciones que se toman en cada momento. El comportamiento en cada estado se describe a continuación:

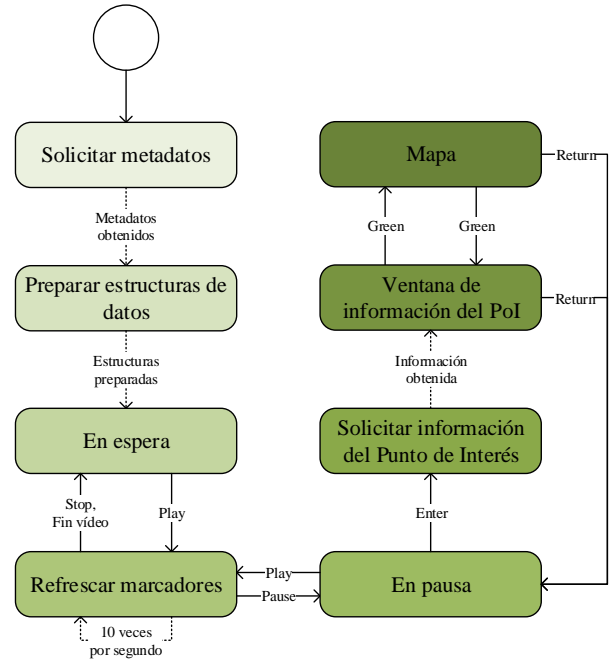


Figura 9. Diagrama de secuencia del módulo de visualización

Al iniciar el módulo de visualización, éste realiza una llamada HTTP GET al servidor mediante el objeto XMLHttpRequest [13]. Los metadatos de respuesta (véase la Figura 3), en formato JSON, se decodifican y se utilizan para preparar las estructuras de datos necesarias, que se enlazan con el fichero de vídeo a reproducir bajo protocolo *streaming*. Entonces, el player se pone en espera.

Al activarse el *play* empieza la reproducción *streaming* del contenido audiovisual como *background* del entorno hypervideo.

Durante la reproducción audiovisual, el player se encarga de sincronizar el *streaming* vídeo con el posicionamiento de los marcadores de forma eficiente, refrescando su posición diez veces por segundo para garantizar una visualización suave de acorde con las imágenes manteniendo un bajo coste computacional.

Si bien en la fase de creación se ha calculado la posición de los marcadores una vez por segundo, para obtener mayor fluidez en el movimiento de los *hot-spots*, se realiza un cálculo en el mismo terminal HbbTV por interpolación lineal a partir de las posiciones calculadas para el segundo actual y el

siguiente, junto con el valor de tiempo actual, de acorde con el pseudocódigo detallado en la Figura 10:

```

t = segundo_actual;
d = segundo_actual % 1;
// para cada PoI que tiene que aparecer
p0 = posicion_en(t);
p1 = posicion_en(t + 1);
p = {'x': (p1.x - p0.x) * d + p0.x,
     'y': (p1.y - p0.y) * d + p0.y};

```

Figura 10. Pseudocódigo de interpolación para el suavizado de la posición de los marcadores

Durante la reproducción, el usuario puede filtrar los puntos de interés según la categoría a la que pertenecen. Esta actividad se realiza de forma síncrona a la reproducción a través del menú de categorías.

Al activar *pause*, el player permite al espectador seleccionar los marcadores que actualmente se visualizan en pantalla y seguir el hiperenlace al PoI que representan. Al hacerlo, el módulo de visualización solicita al servidor, mediante una petición asíncrona HTTP GET al servidor, la información aumentada relacionada con el PoI seleccionado, en formato JSON (véase la Figura 4), y se utiliza para construir la ventana de realidad aumentada, permitiendo al usuario ver la descripción textual, navegar entre las fotografías asociadas y visualizar la ubicación del PoI en un mapa del servicio Google Static Maps [14].

### B. Interfaz de usuario

La interfaz del player se describe a continuación y consta de tres partes, como se aprecia en la Figura 11: el *streaming* de vídeo ocupa la esquina superior izquierda de la pantalla, con los marcadores que representan los puntos de interés sobre él; el listado seleccionable de categorías, siempre visible para permitir el filtrado de PoIs, se ubica en la esquina superior derecha; y en la parte inferior se detalla una referencia a los botones que el módulo permite pulsar y qué función desempeñan.



Figura 11. Interfaz del módulo de visualización

Una vez se muestra la ventana de información aumentada, como vemos en la Figura 12, ésta solapa el vídeo, mostrando la información textual del punto de interés: nombre, descripción y enlace web, así como la galería de fotografías.

La referencia de botones de la parte inferior se mantiene visible en todo momento para permitir al usuario navegar en la galería de fotografías, visualizar la ubicación del PoI en el

mapa, cerrar la ventana para seleccionar otro punto de interés que aparezca en escena o proseguir con la visualización.



Figura 12. Módulo de visualización: ventana de información adicional

Finalmente, como se muestra en la Figura 13, a la hora de mostrar el mapa con la ubicación del punto de interés, éste reemplaza toda la otra información adicional en la ventana, obteniendo la atención del usuario.



Figura 13. Módulo de visualización: mapa del PoI

## VI. TRABAJO FUTURO Y CONCLUSIONES

### A. Trabajo futuro

La creación del catálogo de PoIs es un proceso costoso. Actualmente estamos integrando el repositorio de puntos de interés con la base de datos de Open Street Map, realizando llamadas a su API pública [15], para reducir el coste de trabajo del equipo de producción.

Se han identificado dos técnicas para automatizar el proceso de posicionamiento de los marcadores, según la naturaleza del contenido audiovisual:

Si se trabaja con contenidos audiovisuales filmados en el exterior, al disponer de PoIs relativamente grandes y separados entre sí, como por ejemplo iglesias, estadios u otro tipo de edificios, se puede implementar el posicionamiento automático mediante GPS. Para poder utilizar esta técnica se debe introducir el registro de la posición GPS y la orientación de la cámara durante la grabación del contenido audiovisual. Esta información se utiliza, juntamente con los datos de ubicación de los PoI, para calcular el volumen de visión, como se muestra en la Figura 14, que permite obtener la posición (x,y) en el plano de visión de los objetos contenidos en el volumen.



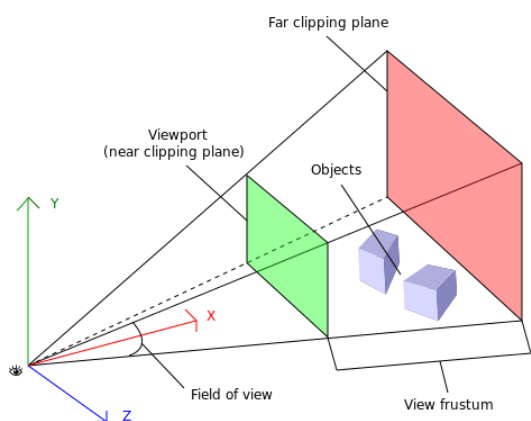


Figura 14. Volumen de visión para el posicionamiento automático de marcadores

En cambio, al utilizar video grabado en interiores, con condiciones de filmación muy estables, como por ejemplo, las obras de arte de un museo, estamos investigando en aplicar una técnica alternativa: la detección del contorno de los objetos y el seguimiento de su forma sobre la secuencia de imágenes originales. El resultado es la detección de posibles PoIs y el posicionamiento de sus marcadores de forma automática, pero se introduce una nueva problemática: discernir los PoIs interesantes de los no deseados.

Al aplicar cualquiera de estas técnicas, aunque se facilita el proceso de creación, se debe realizar el proceso de revisión de la ubicación de los marcadores para no influir negativamente en la experiencia del espectador durante la visualización.

Actualmente se está realizando un segundo módulo de visualización, complementario al módulo HbbTV, implementado en la plataforma Android TV, con el propósito de comparar y validar la aplicación en ambas plataformas.

Finalmente, se está estudiando la posibilidad de ampliar la plataforma hypervideo para soportar vídeos 360°, aumentando en gran medida la interactividad del usuario con el contenido.

## B. Conclusiones

De la fase de validación, realizada por productores audiovisuales y estudiantes universitarios se identificaron los siguientes aspectos: 1) la importancia de las fuentes AV, por lo que se recomienda encarecidamente realizar la planificación de qué puntos de interés y categorías se desean representar, para poder llevar a cabo el guión de filmación (travellings, zooms, panos...) de acorde a su visualización posterior en un entorno hypervideo, y 2) es imprescindible tener presente la percepción del usuario en la visualización del hypervideo: el intervalo temporal de visualización de los marcadores no debe ser demasiado corto, no deben aparecer agrupados y no es útil representar puntos de interés muy lejanos de la cámara. Tener presente estos consejos garantiza la usabilidad del sistema al facilitar al usuario centrar su atención y distinguirlos de forma intuitiva.

A nivel de rendimiento, la visualización hypervideo sobre un receptor HbbTV únicamente es posible si se realiza un cálculo previo de las posiciones de los *hot-spots*, debido a que

su capacidad de cálculo es insuficiente para calcular posiciones de varios PoIs y reproducción de streaming MPEG-4 de forma paralela, por lo cual, es imprescindible generar unos metadatos optimizados para cada fichero audiovisual.

El resultado de la fase de pruebas realizada entre consumidores (perfil universitario) ha sido muy positivo, han resaltado el valor añadido del hypervideo, identificando dos factores: 1) el aumento del interés por consumir contenido audiovisual de forma interactiva, y 2) la valoración de los enlaces entre hypervideos, demostrando que el concepto de *storytelling* es aceptado y entendido por el usuario.

## AGRADECIMIENTOS

Este trabajo se enmarca dentro del proyecto de investigación ConTVLab IPT-2012-0871-430000, apoyado por la Secretaría de Estado de Investigación, Desarrollo e Innovación – Ministerio de Economía y Competitividad.

## REFERENCIAS

- [1] N. Sawhney, D. Balcom, and I. Smith, "Hypercafe: Narrative and Aesthetic Properties of Hypervideo", in *Proceedings of the seventh ACM conference on Hypertext*, 1996. p. 1–10.
- [2] G. Landow and P. Kahn, "Where's the Hypertext? The Dickens Web as a System-Independent Hypertext", in *Proceedings of the ACM conference on Hypertext*, 1992. p. 149–160.
- [3] N. Sawhney, D. Balcom, and I. Smith, "Authoring and navigating video in space and time", *IEEE Multimedia*, no. 4, pp. 30–39, October-December 1997.
- [4] J. Doherty et al., "Detail-on-demand hypervideo", in *Proceedings of the 11th ACM international conference on Multimedia*, 2003, pp. 600–601.
- [5] F. Shipman, A. Girgensohn, and L. Wilcox, "Combining spatial and navigational structure in the hyper-hitchcock hypervideo editor", in *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, 2003, pp. 124–125.
- [6] C. Tiellet et al., "Design and evaluation of a hypervideo environment to support veterinary surgery learning", in *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, 2010, pp. 213–222.
- [7] *Proyecto LinkedTV* [sitio web]. 2014. [Consulta: 1 febrero 2014]. Disponible en: <http://linkedtv.eu>.
- [8] *Thinglink* [sitio web]. 2014. [Consulta: 1 febrero 2014]. Disponible en: <https://thinglink.com/>
- [9] *Wirewax* [sitio web]. 2014. [Consulta: 1 febrero 2014]. Disponible en: <https://wirewax.com/>
- [10] ETSI, "HbbTV specification Version 1.1", 2010. [Consulta: 1 febrero 2014]. Disponible en: [http://www.hbbtv.org/pages/about\\_hbbtv/specification.php](http://www.hbbtv.org/pages/about_hbbtv/specification.php).
- [11] W. Dees, P. Shrubsole, "Web4CE: accessing web-based applications on consumer devices", in *proceedings of the 16th ACM international conference on World Wide Web*, 2007, pp. 1303–1304.
- [12] W3C Consortium, "CSS TV profile 1.0", *W3C Candidate Recommendation*, 2003. [Consulta: 1 febrero 2014]. Disponible en: <http://www.w3.org/TR/css-tv>.
- [13] WHATWG, "XMLHttpRequest", *Living Standard*, 2014. [Consulta: 1 febrero 2014]. Disponible en: <http://xhr.spec.whatwg.org/>.
- [14] Google, "Google Static Maps API V2", 2014. [Consulta: 1 febrero 2014]. Disponible en: <https://developers.google.com/maps/documentation/staticmaps/>.
- [15] OpenStreetMap Foundation, "OpenStreetMap API v0.6", 2009. [Consulta: 1 febrero 2014]. Disponible en: <http://wiki.openstreetmap.org/wiki/API>.
- [16] Google, "GoogleTV", 2014. [Consulta: 1 febrero 2014]. Disponible en: <http://www.google.com/tv/>

