

Tipo de artículo: Artículo original

Temática: Procesamiento de imágenes

Recibido: 13/11/2013 | Aceptado: 17/12/2013 | Publicado: 21/01/2014

Generación de resúmenes escalables de video

Generating scalable video summaries

Abel Díaz Berenguer

Centro de Geoinformática y Señales Digitales. Departamento de Señales Digitales. Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Torrens, Boyeros, La Habana, Cuba. CP.: 19370

*Autor para la correspondencia: aberenguer@uci.cu

Resumen

Los procedimientos para la generación de resúmenes de video, permiten obtener una secuencia sintetizada de la información contenida en el mismo. En este contexto la duración del resumen juega un papel fundamental pues en ocasiones se desean sinopsis con mayor o menor longitud, por lo que los resúmenes escalables permiten dar respuesta a esta necesidad. En el presente trabajo se expone un procedimiento desarrollado con el objetivo de obtener resúmenes escalables de videos según la duración de los mismos. El procedimiento se basa en un proceso de creación del resumen con etapas bien establecidas que garantizan la generación eficiente del mismo. Los resultados muestran la compatibilidad con diversos formatos de codificación, generándose eficazmente los resúmenes para varias longitudes. La eficacia del procedimiento, confirmada con los resultados del trabajo, garantiza sus potencialidades de aplicación en sistemas para la gestión, procesamiento y transmisión de materiales audiovisuales.

Palabras clave: Escalable, procedimiento, resumen, video.

Abstract

Procedures for generating video summaries, allows to obtain a synthesized sequence of the information contained in a video. In this context, the summary duration plays an important role because sometime summaries with varying duration are needed, so the scalable summaries allow to respond to this necessity. In the present work a new

procedure is proposed, which has been developed with the aim of obtaining scalable summaries according to their duration. The procedure developed is based on a summary creation process with very well established stages for an efficient generation of it. The results obtained show compatibility with various encoding formats, what generates summaries for various lengths in an efficient way. The effectiveness of the procedure, confirmed by means of the work results already obtained, guarantees its application potentialities in systems developed for managing, processing and transmission of audiovisual materials.

Keywords: Procedure, scalable, summary, video.

Introducción

En los últimos años se ha evidenciado un notable aumento de la producción y distribución de contenidos audiovisuales. Estudios realizados demuestran que los usuarios de internet prefieren, en su mayoría, consumir los archivos audiovisuales disponibles en la red (Maass, *et al.*, 2005; López, *et al.*, 2011); sitios como YouTube, dedicados a compartir materiales audiovisuales en la red, diariamente reciben más de dos millones de visitas (Gramsci, 2011). Lo anterior, unido al incremento de las capacidades de cómputo y almacenamiento de los sistemas, ha impuesto el reto de perfeccionar los ambientes de producción, distribución y recuperación de contenido audiovisual, pues se predice que continúe aumentando la demanda de estos contenidos (Cisco, 2012).

El acceso al contenido multimedia es una actividad cada vez más difícil, por lo general implica un proceso de pre-visualización del video por parte del usuario, que debe emplear determinado tiempo en esta tarea. En este contexto las técnicas de procesamiento de video se han convertido en una necesidad ante el volumen de materiales audiovisuales existentes. Varias investigaciones se han dedicado a establecer procedimientos para describir, procesar, almacenar y recuperar información del contenido audiovisual. Como parte de estos se pueden observar los que procesan el contenido de un video con la finalidad de generar automáticamente resúmenes del mismo (Truong, *et al.*, 2007). Los procedimientos para generar resúmenes automáticos del video permiten crear una representación visual compactada de la información en el video, es decir, proporcionan una secuencia sintetizada, pero con las representaciones del contenido original (Truong, *et al.*, 2007; Valdés, *et al.*, 2008; Over *et al.*, 2008; Herranz, 2010; Wan, *et al.*, 2010).

Los procedimientos que generan sinopsis de videos están diseñados para facilitar la navegación sobre una base de datos de videos. Además se pueden utilizar los resúmenes obtenidos como un producto final, que garantice al usuario acceder rápidamente a posiciones semánticas relevantes en la secuencia (Berkhin, 2006; Truong, *et al.*, 2007). La mayoría de los algoritmos desarrollados en este ámbito generan la salida de un único resumen. Esto puede ser

insuficiente, ya que en ocasiones es deseable una representación personalizada para cada usuario y los resúmenes escalables resultan útiles para hacer frente a la diversidad de preferencias y lograr mayor usabilidad.

La creación de algoritmos que permitan obtener resúmenes de video escalables, despierta cierto interés en investigadores de la temática (Benini, *et al.*, 2006a; Benini, *et al.*, 2006b; Zhu, 2004; Herranz, *et al.*, 2010; Herranz, 2010). Se asume por resumen escalable el caso en que es posible adaptar la sinopsis obtenida a determinadas preferencias o condiciones; por ejemplo, la resolución, la tasa de bits por segundo y la longitud o la duración. Para lograr este objetivo los procedimientos que generan resúmenes escalables deben garantizar la posibilidad de generar varias secuencias sintetizadas dependiendo de las condiciones o preferencias establecidas. En el presente trabajo, se asume la escalabilidad como la capacidad del procedimiento para devolver como salida varios resúmenes con distintas duraciones para una misma secuencia original. Como antecedentes de esta investigación se pueden observar las aproximaciones encaminadas a lograr resúmenes de varias escalas (Zhu, 2004; Benini *et al.*, 2006a; Benini *et al.*, 2006b; Herranz *et al.*, 2010; Herranz, 2010).

Las aproximaciones mencionadas anteriormente poseen buenos resultados aunque presentan dificultades, Zhu y sus colaboradores (Zhu, 2004) asumen la escalabilidad para cada estructura jerárquica con la limitante de que los resúmenes no son escalables dentro de cada nivel. Además a consideración del autor de este artículo un resumen escalable debería lograr mayor número de escalas para adaptarse a situaciones en las que la duración del mismo puede limitarse a un valor específico.

Por su parte la principal desventaja de las investigaciones presentadas por Benini, (Benini, *et al.*, 2006a; Benini *et al.*, 2006b), radica en que se requiere una etapa de procesamiento a priori antes de la anotación de la secuencia, para establecer el umbral a través del que se establece la prioridad de las tomas a representar.

En el caso de los resultados presentados en las investigaciones de Herranz (Herranz, *et al.*, 2010; Herranz Arribas, 2010) se evidencia un procedimiento ajustado a un dominio restringido de codificación de video que es basado en estándares MPEG (Mitchell, *et al.*, 1996). Este procedimiento se centra en las características propias del flujo de datos de las codificaciones MPEG, lo que limita su utilización en otros dominios de codificación. A pesar de las dificultades de los trabajos analizados, según las consideraciones del autor de este artículo, los conceptos de escalabilidad y resumen embebido planteados en esta aproximación (Herranz, *et al.*, 2010; Herranz, 2010) constituyen los más aceptados para la creación de resúmenes escalables de video.

El desarrollo de la presente investigación tiene sus orígenes en la construcción por parte de la Universidad de las Ciencias Informáticas de aplicaciones informáticas para la gestión, procesamiento y transmisión de contenido audiovisual. Estas aplicaciones se caracterizan por el constante aumento de la cantidad y diversidad de materiales audiovisuales gestionados y posteriormente almacenados para su utilización. Los materiales audiovisuales gestionados no necesariamente se codifican en un formato establecido, pues esto depende de los requisitos establecidos por los clientes potenciales.

Por otra parte en las aplicaciones de gestión, procesamiento y transmisión de contenido audiovisual se debe garantizar el proceso de catalogación de los materiales audiovisuales gestionados para facilitar su búsqueda y correcta utilización. Se han desarrollado algoritmos para lograr la catalogación semiautomática, pero muchas veces continúa siendo necesaria la anotación manual de determinados datos. En este proceso, los usuarios obtienen las perspectivas del contenido audiovisual navegando, en ocasiones aleatoriamente, sobre las secuencias de video. Adicionalmente en dichas aplicaciones se persigue el objetivo de proveer a los usuarios de materiales audiovisuales disponibles en la web para consumir bajo demanda. Es común notar en los usuarios que consumen videos, generalmente de larga duración, que antes de proceder a la visualización completa del material, realizan una pre-visualización del mismo para obtener una idea de su contenido y posteriormente visualizarlo completamente si lo consideran de su agrado.

Lo expuesto anteriormente revela la necesidad de un mecanismo para obtener varias secuencias sintetizadas con la información significativa de un mismo video, favoreciendo el proceso de catalogación y pre-visualización de materiales audiovisuales en sistemas de gestión, procesamiento y transmisión de archivos audiovisuales. En el presente trabajo se expone un procedimiento desarrollado para generar resúmenes escalables de videos.

Metodología computacional

Creación del resumen automático de video

La creación de resúmenes de video involucra diversos niveles de complejidad, en el presente trabajo se asumen los resultados del estudio realizado por (Valdés, *et al.*, 2008), donde se evidencia la secuencia de etapas necesarias para lograr un resumen de video. En el siguiente esquema se muestra la secuencia de pasos computacionales que permiten transformar una secuencia de video de entrada en un resumen de la misma.

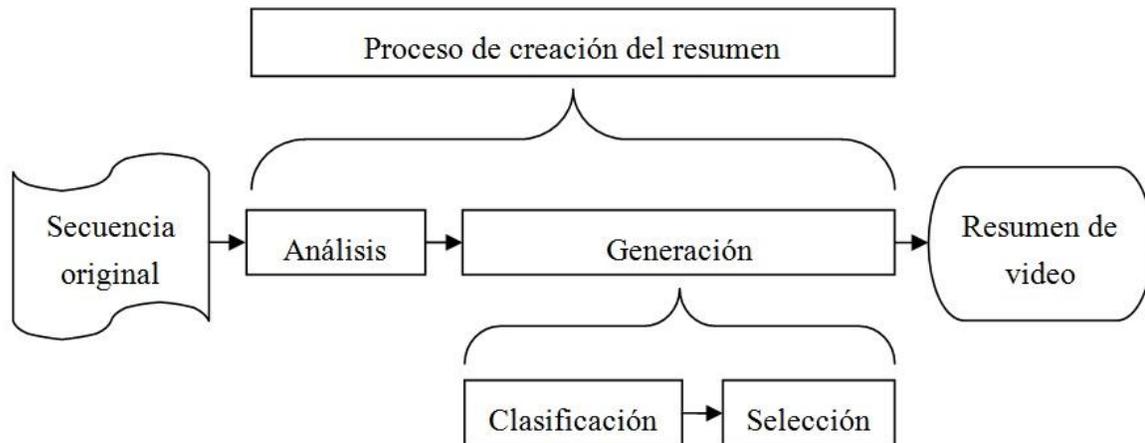


Figura 1. Etapas para la creación de un resumen.

Seguidamente se expone detalladamente cada una de las etapas.

Etapas de análisis

En la etapa de análisis se realizan las operaciones destinadas a extraer los descriptores visuales de la secuencia original con el objetivo de obtener una lista con los datos y los descriptores asociados a las tomas. En esta etapa se decodifica el video y se extraen los fotogramas del mismo. Una vez capturados los fotogramas se procede a la primera fase para segmentar el video, la detección de tomas mediante histogramas (Yinzi, 2010; Díaz, 2011; Boullosa, 2011). Este se divide en cuatro pasos:

1. Calcular el histograma de cada fotograma.
2. Comparar cada fotograma con el contiguo y guardar los resultados.
3. Establecer un umbral entre las comparaciones.
4. Detectar todas las comparaciones que sobrepasen este umbral como un cambio de toma.

En este punto el video ya ha sido segmentado utilizando histogramas de color pero este procedimiento de detección de tomas es muy sensible a los cambios de iluminación. Por esta razón se efectúa la detección mediante bordes a los fotogramas que representan el cambio de una toma a otra (Lienhart, 2001; Bescós *et al.*, 2005; Yinzi, 2010). Esta segunda fase de detección se logra a través de la ejecución de los siguientes pasos:

1. Convertir los fotogramas a escala de grises.

2. Realizar un desenfoque gaussiano.
3. Aplicar una función de detección de bordes.
4. Expandir las líneas resultantes.
5. Comparar los resultados.
6. Establecer un umbral con las comparaciones.
7. Detectar los falsos positivos de la etapa anterior.

Después de haber efectuado la detección mediante histogramas y bordes todavía pueden existir cortes dentro de una toma que no se hayan detectado. Esta es la razón por la cual se utiliza el descriptor SURF (Bay, *et al.*, 2006) para aquellos límites entre tomas que como resultado del análisis de histogramas pertenezcan a un margen dudoso (Berkhin, 2006; Lienhart, 2001; Bescós, *et al.*, 2005; Tuytelaars, *et al.*, 2007). La ejecución de esta tercera fase de detección se logra mediante los siguientes pasos:

1. Extraer las características.
2. Obtener los descriptores.
3. Buscar coincidencias entre fotogramas.
4. Umbralizar el resultado.
5. Detectar falsos positivos de la etapa anterior.

Finalmente como resultado de esta etapa se obtiene una lista con los datos relativos a la toma:

- Posición inicial.
- Longitud.
- Fotograma clave.
- Umbral del Histograma de color.
- Umbral del SURF

La etapa de análisis constituye la etapa crítica para todo el proceso, pues la lista resultante es almacenada y constituye la entrada para la etapa de generación. Además el análisis no se ejecutará nuevamente pues aunque se requiera generar otro resumen esto se garantiza con los datos almacenados. Se analiza una vez y se genera tantas veces como sea necesario, es decir la etapa solamente se realizará la primera vez que se requiera un resumen, luego se garantiza la escalabilidad ejecutando la etapa de generación varias veces. Seguidamente se explica la generación en detalles.

Generación escalable

En la etapa de generación radica la mayor complejidad para la creación del resumen de video. Por esta razón se ha dividido esta etapa en dos: clasificación y selección (Valdés *et al.*, 2008). Durante la clasificación se proporciona una calificación a cada toma, atendiendo a las características analizadas anteriormente. En la etapa de selección se determina cuál o cuáles de las tomas, deben incluirse en el resumen de video a generar, y cuál o cuáles pueden obviarse atendiendo a la clasificación realizada previamente.

En el trabajo se asumen los criterios de escalabilidad expuestos en (Herranz, *et al.*, 2010; Herranz, 2010) donde se define que un resumen escalable es aquel que se constituye mediante un grupo de sumarios embebidos, $SS = \{S_1, S_{1+1}, \dots, S_{L-1}, S_L\}$ donde $l \in \mathbb{N}$ denota la escala del resumen y L la longitud del mismo. A su vez se hace necesario el cumplimiento de la siguiente restricción $S_1 \subset S_{1+1} \subset S_{L-1} \subset S_L$, considerando que cada sumario de menor escala es embebido en el sumario de escala superior. Como se puede apreciar el sumario de mayor longitud se conforma mediante la agrupación de los sumarios de menor longitud, por lo que se debe lograr una jerarquía en la que el sumario de menor longitud garantice la mayor representatividad de la secuencia original.

En esta etapa se persigue obtener, las tomas más representativas del video, buscando eliminar la redundancia ofreciendo la mayor cantidad de información posible. Con los resultados del análisis de histograma de color y SURF se confecciona una matriz de proximidad (Berkhin, 2006), esta matriz es utilizada posteriormente para crear un grupo de K clúster $\{C_0, C_1, \dots, C_{k-1}\}$ según las distancias entre cada par de tomas $d(t_i, t_j)$ agrupándose las tomas con características similares. Luego para cada clúster C_k las tomas más cercanas al centroide, determinadas por una puntuación establecida según su duración y variabilidad en el grupo, se consideran las más representativas del grupo, lo que les establece mayor prioridad para incluir en el resumen, conformándose la jerarquía necesaria para lograr la escalabilidad (Dumont, *et al.*, 2008; Xu *et al.*, 2009).

Para la generación final del resumen se asumen las dos modalidades para la representación de los resúmenes de video que se pueden observar comúnmente:

1. Resúmenes construidos con imágenes estáticas, que constituyen una representación de los fotogramas relevantes, extraídos de la secuencia de video original para mostrar el contenido de la misma a través de un guión gráfico estático, lo que comúnmente se conocen como *storyboard*. Esta modalidad es definida en (Truong, *et al.*, 2007) de la siguiente forma: $R = A_{\text{Keyframe}}(V) = \{f_1, f_2, \dots, f_k\}$ donde A_{Keyframe}

constituye el procedimiento de extracción de los fotogramas claves al video V , obteniéndose una representación R constituida por los fotogramas $\{f_1, f_2, \dots, f_k\}$.

2. Resúmenes que se componen de pequeños segmentos de video. Esta modalidad de resumen consiste en un conjunto de segmentos de video que se extraen del video original, los cuales se agrupan ya sea por un corte o a través de un efecto gradual de transición para obtener un video de menor tamaño que el original. En este caso se constituye un pequeño video conocido como *skimming* de video. El *skimming* de video es definido en (Truong *et al.*, 2007) de la siguiente forma: $K = A_{Skim}(V) = E_{i1} \cup E_{i2} \dots \cup E_{ik}$ donde A_{Skim} constituye el procedimiento de generación del *skim* del video V , $E_i \in V$ es el fragmento i ésimo a incluir en el video *skim* y \cup la operación de integración obteniéndose una representación K , constituida por la integración de los segmentos $E_{i1}, E_{i2}, \dots, E_{ik}$. Esta integración se realiza siguiendo el flujo temporal del video original y la operación de integración utilizada es una transición de desaparición simple.

Esta etapa es ejecutada para obtener resúmenes con la longitud deseada, por tanto, se puede repetir todas las veces que sea necesario, para garantizar las condiciones de escalabilidad.

Resultados y discusión

Pruebas de compatibilidad de formatos

Con el objetivo de comprobar el procedimiento propuesto se procede a probarlo en videos con características distintas, para esto se utiliza una muestra intencional no probabilística de videos confeccionada por el autor seleccionando videos con distintas codificaciones. Se realiza esta prueba para validar la compatibilidad del procedimiento desarrollado con videos codificados en diferentes formatos. Entre las características de estos videos se encuentran el contenedor, el codificador, el *bitrate*¹, los fotogramas por segundo (FPS) y la resolución.

A continuación se muestra el resultado de las pruebas de soporte de formato para la ejecución del procedimiento con cuatro videos con distintas codificaciones.

Tabla 1. Prueba de formato 1.

Nombre del video	Stay
------------------	------

¹ Frecuencia de bits por segundo.

Contenedor	AVI
Codificador	MPEG4-Xvid
Bitrate(KB/s)	128
FPs	29.97
Resolución	800 x 600
Valoración: Se procesa el video correctamente	

Tabla 2. Prueba de formato 2.

Nombre del video	Girlfriend
Contenedor	WMV
Codificador	WMV2
Bitrate(KB/s)	2400
FPs	30
Resolución	720 x 480
Valoración: Se procesa el video correctamente	

Tabla 3. Prueba de formato 3.

Nombre del video	Campaign Cinematics 07
Contenedor	MPG
Codificador	MPEG1
Bitrate(KB/s)	1150
FPs	25
Resolución	720 x 480
Valoración: Se procesa el video correctamente	

Tabla 4. Prueba de formato 4.

Nombre del video	Where Have You Been
Contenedor	ogg
Codificador	Theora
Bitrate(KB/s)	1596
FPs	29
Resolución	352 x 240
Valoración: Se procesa el video correctamente	

Tabla 5. Prueba de formato 5.

Nombre del video	Dance Again
Contenedor	webm
Codificador	Vp8
Bitrate(KB/s)	1596
FPs	29
Resolución	352 x 288
Valoración: Se procesa el video correctamente	

Como se puede apreciar (ver tablas 1-4) se obtienen resultados satisfactorios para la generación de resúmenes con videos con distintas codificaciones, este resultado evidencia mayor eficacia con respecto al trabajo realizado por (Herranz, 2010) pues demuestra que el procedimiento desarrollado se puede utilizar en videos con una codificación que no se base en los estándares MPEG.

Pruebas de la etapa de análisis

En la etapa de análisis se garantiza la obtención de todos los datos necesarios para la posterior generación del resumen, es por este motivo que se determina validar esta etapa de forma independiente, debido al proceso de segmentación que se realiza en la misma, constituye una sección crítica en el procedimiento propuesto. Las pruebas desarrolladas a continuación se realizan con el objetivo de validar la detección de tomas durante el análisis utilizando las medidas *precision*² y *recall*³. Para este análisis se emplea la base de datos de videos creada por Bescós.

² Medidas de evaluación para determinar la fracción de cambios de tomas del video que son detectados.

Tabla 6. Prueba de detección de tomas.

Nombre del video	Número de Fotogramas	Cortes	Cortes Detectados	Fp	Fn	Recall	Precision
CNN	30 715	54	51	4	3	94%	93%
Días contados	27 643	13	12	1	0	100%	92%
Friends	90 341	21	21	2	0	100%	91%
La vita e bella	114 997	127	125	6	2	98%	96%

Como se puede apreciar (ver tabla 5) en la etapa de análisis del procedimiento desarrollado para la detección de tomas se garantizan índices de *precision* por encima del 90% en todos los casos, lo que se puede considerar aceptado para estos procedimientos (Mohanta *et al.*, 2010; Davis *et al.*, 2006).

Pruebas de generación escalable

A continuación, se muestra una tabla con los resultados de las pruebas realizadas al procedimiento propuesto, para la generación de varias secuencias sintetizadas del mismo video, teniendo en cuenta la variación de la longitud de dichas secuencias. Se establecen varias escalas tanto para la generación de *storyboard* como para la generación de *skimming* de video. Para este análisis se han establecido las siguientes escalas: 20, 30 y 40 cantidad de fotogramas de la secuencia original, para la generación de un *storyboard* y del 10, 20 y 30 por ciento de la secuencia original para el caso del *skimming* de video. En la tabla se muestran los tiempos en segundos que demora el procedimiento para la generación del resumen para las escalas establecidas.

³ Medidas de evaluación para determinar la fracción de cambios de tomas del video que se detectan correctamente.

Tabla 7. Pruebas para la generación de resúmenes escalables.

Nombre del video	Generación					
	Storyboard(#)			Skimming(%)		
CNN	20	30	40	10	20	30
	0.67 (s)	0.74 (s)	0.81 (s)	0.98 (s)	1.36 (s)	1.64 (s)
Días contados	20	30	40	10	20	30
	0.68 (s)	0.76 (s)	0.79 (s)	1.02 (s)	1.43 (s)	1.87 (s)
Friends	20	30	40	10	20	30
	0.67 (s)	0.72 (s)	0.79 (s)	0.99 (s)	1.33 (s)	1.63 (s)
La vita e bella	20	30	40	10	20	30
	0.69 (s)	0.73 (s)	0.82 (s)	0.97 (s)	1.29 (s)	1.61 (s)

Igual que en la etapa de análisis, para las pruebas de generación escalable se empleó la base de datos de videos creada por Bescós. Los resultados evidencian que el procedimiento propuesto brinda la posibilidad de crear resúmenes de varias escalas, es decir, se obtienen adecuadamente resúmenes con distintas longitudes y en tiempos de respuesta aceptables. Nótese que se duplica la cifra de escalabilidad requerida y no necesariamente se duplica en correspondencia el tiempo requerido, esto se debe a que la etapa de análisis para una escala mayor no se ejecuta. Aún obteniéndose estos resultados no debe descartarse la posibilidad de utilizar otros descriptores visuales que permitan mejorar la eficacia del procedimiento lográndose resultados superiores.

Validación de usabilidad de los resúmenes

Para garantizar la usabilidad de un resumen de video, se debe crear un sumario de video que posea un lenguaje audiovisual que cumpla con dos condiciones esenciales: cobertura semántica y agrado visual. La primera de estas condiciones se refiere a que el resumen creado debe preservar la mayor cantidad de información representativa posible, descartándose la mayor cantidad de información redundante, con el objetivo de minimizar el tiempo requerido para su visualización. La segunda de estas condiciones se refiere a que el resumen no solamente debe ser informativo, sino que debe poseer características visuales agradables para el usuario que finalmente lo visualiza (Truong, *et al.*, 2007; Ren, *et al.*, 2010).

Para validar la usabilidad del resumen se ha comprobado la capacidad para suministrar al usuario una comprensión global de la información contenida en el video original, capacidad para proporcionar una síntesis de la información

del video original haciendo prevalecer los contenidos representativos sobre los verbales y la capacidad para lograr un encadenamiento en mosaico en el que las tomas o imágenes que componen, en conjunto el resumen, tengan sentido. Para validar el cumplimiento de estas dos condiciones se aplicó una encuesta como instrumento de diagnóstico. Se selecciona una técnica de muestreo aleatorio simple (Bay *et al.*, 2006) considerando una población constituida por 90 usuarios de SIAV. Para determinar el tamaño de la muestra se utiliza el software Stats⁴ considerando un nivel de confianza de 95%, estableciéndose aplicar la encuesta a 30 usuarios. La encuesta aplicada consta de cuatro preguntas en las que el usuario puede asignar un número real entre 1 y 10 estableciendo su valoración sobre la usabilidad del resumen final.

Para el análisis de los resultados se procesa la encuesta promediando los valores obtenidos en cada una de las preguntas por los 30 usuarios obteniéndose los resultados que se muestran en la siguiente tabla.

Tabla 8. Valores promedio obtenidos con la encuesta.

Preguntas	1	2	3	4
Valores promedio	6.1	5.3	4.7	6.6

Luego promediando los valores obtenidos se obtiene un nivel de usabilidad de 5.67. Para determinar el nivel de usabilidad del resumen en una escala valores numéricos y su equivalente nominal que se muestra en la tabla.

Tabla 9. Escala para establecer el nivel de usabilidad.

Escala	Numérica	0-1.9	2-3.9	4-5.9	6-7.9	8-10
	Nominal	Muy bajo	Bajo	Medio	Alto	Muy alto

Los resultados, demuestran que el procedimiento cumple medianamente con las condiciones de usabilidad establecidas. Se debe continuar refinando el mismo, pues en la capacidad de brindar cobertura semántica y agrado visual al usuario final, los resultados del procedimiento demuestran que aún persisten ineficiencias. Como consecuencia de este resultado se recomienda redefinir la etapa de generación del procedimiento propuesto.

4 Software que permite a los investigadores determinar el tamaño de una muestra aleatoria simple y generar aleatoriamente números de casos necesarios para la muestra

Conclusiones

El procedimiento propuesto permite la generación eficiente de resúmenes escalables de video, abordando la duración de las secuencias sintetizadas como requisito de escalabilidad. Además, admite la generación de resúmenes de videos que no están codificados en estándares MPEG.

Las etapas necesarias para la generación de un resumen se establecen correctamente, obteniéndose buenos resultados en la etapa de análisis, donde la detección de los cambios de tomas muestra índices de *precision* y *recall* por encima del 90% en todos los casos. Además en esta etapa se garantiza la obtención de los datos necesarios para la generación escalable, sin necesidad de analizar el video original nuevamente. Por su parte en la etapa de generación se logra la obtención de resúmenes con longitudes distintas, garantizando la escalabilidad deseada y consiguiéndose la representación de los resúmenes en las modalidades de *storyboard* y *skimming* de video.

Las principales dificultades del procedimiento propuesto se evidencian en la capacidad para lograr adecuadamente cobertura semántica y agrado visual para el usuario final, lo que implica la necesidad de redefinir la etapa de generación establecida.

Recomendaciones

Se recomienda estudiar la posibilidad de trabajar con otros descriptores visuales lo que puede mejorar los resultados obtenidos. A pesar de que los resultados demuestran la necesidad de continuar trabajando en el procedimiento propuesto, se considera que puede ser utilizado en sistemas de gestión, procesamiento y transmisión de contenidos audiovisuales, minimizando las limitaciones de estos sistema para obtener varias secuencias sintetizadas con la información significativa de un mismo video lo que beneficiará el proceso de catalogación y pre-visualización de materiales audiovisuales.

Referencias

- BAY, H., *et al.* SURF: Speeded Up Robust Features. en: *Computer Vision – ECCV 2006*. LEONARDIS, A.;BISCHOF, H.*et al*, Springer Berlin Heidelberg, 2006. 3951: 404-417.p.
- BENINI, S., *et al.* *Extraction of significant video summaries by dendrogram analysis*. IEEE International Conference on Image Processing, Atlanta, GA, IEEE, 2006a. 133–136 p.
- *Hierarchical Summarization of Videos by Tree-Structured Vector Quantization*. 2006 IEEE International Conference on Multimedia and Expo, Toronto, Ont., IEEE. ICME, 2006b. 969 - 972 p.

- BERKHIN, P. A Survey of Clustering Data Mining Techniques. en: *Grouping Multidimensional Data*. KOGAN, J.;NICHOLAS, C.*et al*, Springer Berlin Heidelberg, 2006. 25-71.p.
- BESCÓS, J., *et al*. A Unified Model for Techniques on Video-Shot Transition Detection *IEEE Transactions on Multimedia*, 2005, 7(2).
- BOULLOSA GARCÍA, Ó. *Estudio comparativo de descriptores visuales para la detección de escenas cuasi-duplicadas*. Video Processing and Understanding Lab. Departamento de Ingeniería Informática. Madrid, Universidad Autónoma de Madrid. Escuela Politécnica Superior, 2011. p.
- CISCO. *Cisco Visual Networking Index: Forecast and Methodology, 2011–2016*, 2012.
- DAVIS, J., *et al*. *The relationship between Precision-Recall and ROC curves*. *Proceedings of the 23rd international conference on Machine learning*. Pittsburgh, Pennsylvania, ACM, 2006. 233-240.
- DÍAZ BERENGUER, A., *et al*. *Sistema de gestión y transmisión de contenidos audiovisuales.: 6to Congreso Internacional de Tecnologías, Contenidos Multimedia y Realidad Virtual. Informática 2013*. La Habana, Cuba, 2013. 5.
- DÍAZ ESPINOZA, D. A. *Implementación y comparación de descriptores para búsquedas en video.:* Facultad de Ciencias Físicas y Matemáticas. Departamento de Ciencias de la Computación. Santiago de Chile, Universidad de Chile, 2011. p.
- DUMONT, E., *et al*. *Redundancy removing by adaptive acceleration and event clustering for video summarization*, Klagenfurt University, Austria., 2008. p. 978-0-7695-3130-4/08
- GRAMSCI, S. K. *A Scalable Video Streaming Approach using Distributed B-Tree*. Computer Science. Bacouver, The University of British Columbia, 2011. p.
- HERRANZ ARRIBAS, L. *A Scalable Approach to Video Summarization and Adaptation*. Escuela Politécnica Superior de Ingeniería Informática. Madrid, Universidad Autónoma de Madrid. , 2010. 224. p.
- HERRANZ ARRIBAS, L., *et al*. *A Framework for Scalable Summarization of Video* *IEEE Transactions on Circuits and Systems for Video Technology*, 2010, 20(9): 1265-1270.

- LIENHART, R. Reliable Transition Detection in Videos: A Survey and Practitioner’s Guide *International Journal of Image and Graphics*, 2001, 1(3).
- LÓPEZ VIDALES, N., *et al.* Jóvenes y televisión en 2010 un cambio de hábitos. *Revista de estudios de comunicación*, 2011, 30: 97-113.
- MAASS, M., *et al.* De memorias y tecnologías radio, televisión e Internet en México *Estudios sobre las culturas contemporáneas*, 2005, (22): 193-220.
- MITCHELL, J. L., *et al.* *MPEG Video Compression Standard*. Nueva York, EE.UU., Kluwer Academic, 1996. p. ISBN 0-412-08771-5
- MOHANTA, P. P., *et al.* A heuristic algorithm for video scene detection using shot cluster sequence analysis. *Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing*. Chennai, India, ACM, 2010. 464-471.
- OVER, P., *et al.* *The TRECVID 2008 BBC Rushes Summarization Evaluation*, Vancouver, Canada, 2008. p. 978-1-60558-309-9/08/10
- REN, J., *et al.* Activity-driven content adaptation for effective video summarization *J. Vis. Commun. Image R*, 2010, 21: 930–938.
- TRUONG, B. T., *et al.* Video Abstraction: A Systematic Review and Classification *The ACM Transactions on Multimedia Computing, Communications, and Applications.*, 2007, 3(1): 37.
- TUYTELAARS, T., *et al.* Local Invariant Feature Detectors: A Survey *Foundations and Trends in Computer Graphics and Vision*, 2007, 3(3): 177–280.
- VALDÉS, V., *et al.* *On Video Abstraction Systems’ Architectures and Modelling*. Madrid, Video Processing and Understanding Lab. Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain, 2008.
- WAN, T., *et al.* A New Technique for Summarizing video sequences through histogram evolution *IEEE*, 2010, 6(10).
- XU, R., *et al.* *Clustering*. Wiley, 2009.

- YINZI, C. *A Temporal Video Segmentation and Summary Generation Method Based on Shots' Abrupt and Gradual Transition Boundary Detecting* Second International Conference on Communication Software and Network, 2010. p.
- ZHU, X. Exploring video content structure for hierarchical summarization *Multimedia Systems*, 2004, 10: 98-115.