Capítulo 6

Conclusiones

6.1 INTRODUCCIÓN

El objeto de este último capítulo del proyecto es desarrollar las conclusiones a las que hemos llegado.

Para empezar, mostraremos de forma resumida, usando tablas comparativas, los resultados obtenidos con los algoritmos programados para las secuencias de prueba. En base a la información de estas tablas y comparando con los resultados publicados por otros investigadores, estableceremos las conclusiones finales sobre eficacia y eficiencia de las distintas étapas de nuestro método.

Después, retomando la idea de la introducción y a modo de epílogo, intentaremos establecer un modelo abstracto de complejidad, sobre el que situar el conjunto de todas las funciones, procesos, ficheros de resultado, etc..., de los que consta el proyecto.

Por último, propondremos algunas líneas futuras de trabajo y mejora.

6.2 RESUMEN DE RESULTADOS DE LOCALIZACIÓN DE TRANSICIONES

Antes de mostrar los resultados, nos ha parecido interesante elegir algún parámetro que cuantifique la fiabilidad de la detección de los efectos de transición. En la literatura científica se proponen distintas herramientas, finalmente la métrica elegida ha sido:

$$\tau = \frac{N_R - (N_{FP} + N_{FN})}{N_R} \tag{6.1}$$

donde se definen los parámetros:

- $N_R \equiv$ Número de efectos reales

- $N_{FP} \equiv$ Número de Falsos Positivos. Sabiando que un Falso Positivo (FP) se corresponde con un efecto detectado por el método, pero que realmente no existe.

- $N_{FN} \equiv$ Número de Falsos Negativos. Sabiando que un Falso Negativo (FN) se corresponde con un efecto no detectado por el método, pero que realmente existe.

Para el mejor de los casos (cuando no haya ningún error en la detección), τ deberá valer 1. A medida que aumentan los fallos, el valor de τ disminuye. Para casos muy críticos (los errores son mayores que los efectos reales), se pueden incluso obtener valores negativos. Nos permitirá evaluar la eficiencia del método para cada secuencia, independientemente de la complejidad de ésta (entendiento por complejidad el número de transiciones que tiene).

Podrá aplicarse tanto en la detección de cortes, como en la detección de fundidos y cortinillas.

6.2.1 Valores de τ en las secuencias de prueba

□ Resultados de la detección de cortes

Se han utilizado los siguientes valores para los parámetros configurables:

- Cierre inicial (o Apertura) de tamaño 10.

- Umbral para el módulo de la diferencia, $u_{dc} = 5$.

- Umbral para la correlación, $u_\rho = 0.07$.

Secuencia	**Frames**	N_R	**Detectados**	N_{FP}	N_{FN}	τ
"DRAMA_8.VAL"	3012	11	11	0	0	1
"MOVIE_8.VAL"	3010	14	14	0	0	1
"NEWSB_8.VAL"	1497	5	5	0	0	1
"NEWSA_8.VAL"	1907	16	18	2	0	0.87
"BASKET_8.VAL"	1502	12	13	1	0	0.91
"CYCLING_8.VAL"	1998	1	1	0	0	1
"ALCOI_8.VAL"	783	9	9	1	1	0.77
"MALVA_8.VAL"	1782	5	7	2	0	0.6
"TARON_8.VAL"	4584	43	44	3	2	0.88
"CHAPL_8.VAL"	2713	10	25	17	2	-0.9
"CHAPL_8C.VAL"	2713	10	17	9	2	-0.1

Tabla 6.1: Resultados de la detección de cortes.

En la tabla se ha querido diferenciar los resultados correspondientes a las secuencias con calidad de imagen buena (de "DRAMA_8.VAL" a "CYCLING_8.VAL") de las procedentes de películas antíguas (resto).

Para las primeras, los valores obtenidos para τ, son del orden de 0.95. Analizando más en profundidad, se observa que la efectividad del método se reduce por los Falsos Positivos o Falsas Alarmas, pero en ningún caso se dan Falsos Negativos. Es decir, todos los cortes que tiene la secuencia son detectados, y además se detectan otros que no lo son. En la teoría de segmentación, esto se conoce como sobresegmentación. Al final del capítulo de cortes, ya vimos que estos FP son producidos por la inclusión dentro de la secuencia, de fotogramas cuyas imágenes se han desplazado espacialmente o por variaciones muy bruscas del brillo de la imagen (debidas por ejemplo, a la aparición de un flash de luz de una cámara). Como la probabilidad de aparición de estos efectos es baja, también lo es la de los FP, y por consiguiente la sobresegmetación es mínima.

Dependiendo de la posterior utilización de la información de cortes detectados, podría ser más interesante no detectar ningun corte erróneo, aún a costa de perder alguno de los que si están presentes. Es decir, preferir la aparición de Falsos Negativos. En este caso hablaríamos de subsegmentación. Sería sobre todo para aplicaciones donde fuera imprescindible minimizar los cortes encontrados. Para lograrlo, habría varias alternativas:

- Aumentar el valor de los umbrales u_{dc} y u_{ρ}.

- Aumentar el tamaño del cierre (o apertura) inicial.

La primera esta clara, si aumentamos el valor del umbral para cualquiera de las métricas, se reducirá el número de picos detectados, y sólo lo harán los más altos (cortes muy evidentes). Y la segunda también tiene una explicación sencilla. Recordemos que con un cierre inicial, de un determininado elemento estructurante, limitamos el número de picos que puede haber en el intervalo temporal correspondiente al tamaño del elemento estructurante. Al tomar un valor muy alto, solo detectaremos un corte dentro del intervalo de ese tamaño, que además se corresponderá con el corte más evidente.

Siguiendo con este mismo parámetro de cierre inicial, su elección correcta dependerá del tipo de secuencia. En secuencias de cine, el cierre recomendado de tamaño 10 (un corte máximo cada aproximadamente medio segundo) da muy buenos resultados. Para otro tipo de secuencias, que incluyan muchos cortes (deportes rápidos como el baloncesto, películas con escenas de mucha acción, etc...), se podría tomar una valor de 5. En cualquier caso, será habitual conocer a priori si el contenido de la secuencia requiere unos valores más altos o no de cierre, y en caso de desconocimiento, es recomendable comenzar con valor bajo (del orden de 5) e ir refinando si se optienen muchos Falsos Positivos.

Los resultados obtenidos para las secuencias procedentes de películas antiguas, no son tan buenos. La efectividad en la detección de cortes es directamente proporcional a la calidad de imagen; de esta forma, con secuencias de mala calidad, como "CHAPL_8.VAL", los resultados pueden llegar a considerarse muy malos: se obtienen gran número de Falsos Positivos, que además por el efecto del cierre, originan la aparición de Falsos Negativos.

La mayoría de los errores son producidos por variaciones del brillo o por desplazamientos temporales de la imagen (vibraciones). Una primera alternativa para intentar mejorar, consiste en la ya comentada compensación de brillo y contraste. Como se observa en la tabla, para la secuencia compensada, "CHAPL_8C.VAL", se reducen parte de los Falsos Positivos. Pero, aun así, los resultados son malos. Más adelante propondremos una alternativa mejor.

□ Resultados de la detección de fundidos

Se han utilizado los siguientes valores para los parámetros configurables:

- Umbral para la correlación fotogramas $n-1$, n y $n+1$, $u_{\rho_{fundidos}} = 0.3$.
- Umbral para el error en la estimación de la varianza, $u_{\sigma^2_{fundidos}} = 350$.

Secuencia	**Frames**	N_R	**Detectados**	N_{FP}	N_{FN}	τ
"NEWSB_8.VAL"	1497	1	1	0	0	1
"NEWSA_8.VAL"	1907	1	1	0	0	1
"CYCLING_8.VAL"	1998	3	2	0	1	0.66
"ALCOI_8.VAL"	783	4	15	11	0	-1.75
"MALVA_8.VAL"	1782	1	1	1	1	-1
"CHAPL_8.VAL"	2713	2	100	100	2	↓↓
"CHAPL_8C.VAL"	2713	2	20	20	2	↓↓

Tabla 6.2: Resultados de la detección de fundidos.

□ Resultados de la detección de cortinillas

Secuencia	**Frames**	N_R	**Detectadas**	N_{FP}	N_{FN}	τ
"NEWSA_8.VAL"	1907	2	1	0	1	0.5
"MALVA_8.VAL"	1782	1	1	0	0	1

Tabla 6.3: Resultados de la detección de cortinillas.

Al igual que para los cortes, la calidad de la imagen influye en los resultados obtenidos en la detección de los fundidos y las cortinillas. Centrándonos en los fundidos, para secuencias con muy mala imagen, los resultados pueden llegar a ser desastrosos, como ocurre por ejemplo para "CHAPL_8.VAL". Ni siquiera mediante la compensación de brillo y contraste se mejora mucho.

Otra posible causa de Falsos Negativos en los fundidos, es la duración de éstos. Para fundidos muy largos (muchos fotogramas intermedios), el método falla. El problema se encuentra en que cuando el fundido se prolonga temporalmente a través de muchos fototogramas, las variaciones entre un fotograma y el siguiente serán muy pequeñas (valores bajos de $d_\rho^{fundidos}$). Podría pensarse en reducir el umbral de detección de posibles fundidos, pero el problema que surge inmediatamente es la aparición de gran número de Falsos Positivos (debidos a movimientos de la imagen y de objetos grandes en ella). La elección adecuada del valor de $u_{\sigma^2_{fundidos}}$ también influye en gran medida. Por todo esto, la detección de fundidos largos tiene una difícil solución.

Las cortinillas son un efecto de transición poco utilizado; en todas las secuencias con las que hemos trabajo solo hemos encontrado tres o cuatro casos. El método programado solo funciona con cortinillas verticales y horizontales que empiecen en uno de los extremos, aunque esto no es un gran problema: la inmensa mayoría de las cortinillas son así (la cortinilla no detectada de 'NEWSA_8.VAL" es una excepción, se trata de una cortinilla horizontal que comienza en el centro de la imagen). El método de detección de cortinillas, tal y como esta programado, aunque efectivo, es un método lento y laborioso; y su utilización debería limitarse a las secuencias donde a priori se sepa que contienen alguna. Debido al procesado que se realiza sobre las imágenes "tira" utilizadas en la detección de las cortinillas, las variaciones de brillo no influyen de manera negativa en los resultados, como ocurre con los cortes y los fundidos.

6.2.2 Comparación con los resultados publicados de otros métodos

La información sobre los valores de τ obtenidos por otros métodos aparece recopilada en Joly [14]. En este artículo, se presentan cada uno de los algoritmos publicados y los valores medios de τ que se pueden obtener.

En base a estos datos, se puede decir que:

- Para los cortes, se obtienen resultados que van desde 0.75 para los más sencillos (algunos incluso no pasan del 0.6) hasta el 0.85 o el 0.99 (esto último, para un método muy sofisticado, basado en el error residual de energía después de una transformación afín). Los métodos más usados (se corresponden con los más eficientes computacionalmente, es decir, los más rápidos), se mueven en torno al 0.8, lo que demuestra que nuestro método se encuentra a un buen nivel. Además, en ningún caso se especifica si se han probado con imágenes de mala

calidad, donde probablemente todos los métodos fallarían, como también lo hace el nuestro.

Otra cuestión a favor de la utilización de $d_c(X,Y)$ y $d_\rho(X,Y)$, es su sencillez. Como ya describimos, se trata de métodos computacionalmente sencillos y eficientes (se realizan restas y sumas de pixels), frente a otros métodos que obtienen resultados similares, pero que implican procesamiento complejo (obtención de histogramas, procesado estocástico, transformaciones afines, etc...).

- En el caso de las transiciones graduales, fundidos y cortinillas, la mayoría de las propuestas que hay en la bibliografía parten de una idea básica: son efectos de escasa aparición y además difíciles de localizar. Por tanto, no se debe desperdiciar mucha computación en su localización. Incluso se da el caso que efectos de incrustación, como las cortinillas, raramente son tenidos en cuenta. Al igual que para los cortes, los valores de τ obtenidos, dependen de la complejidad del método. Los mejores valores se mueven en torno al 0.8 (para un método basado en el calculo de la distancia de Hausdorff entre histogramas calculada para contornos de imagen dividida en bloques, se llega al 0.92). Podemos decir, por consiguiente, que nuestros métodos para fundidos y cortinillas también se encuentran dentro de lo aceptable.

6.3 MODELO CONCEPTUAL DE COMPLEJIDAD PARA LA INDEXACIÓN DE VÍDEO

En el capítulo de conceptos básicos, mostramos un esquema de los pasos llevados a cabo en la indexación de vídeo, que reproducimos en la figura 6.1.

De las tres étapas necesarias: segmentación en planos, obtención de fotogramas clave y análisis de la imagen, a lo largo de este proyecto se han cubierto las dos primeras. En las distintas fases de esta étapas, se han procesado varios tipos de información: desde valores de los pixels de la imagen para cada fotograma, hasta planos completos de la secuencia. Pasamos por ejemplo, de un número en coma flotante correspondiente a la varianza de un fotograma, hasta un fichero de texto donde se describe el contenido de cada plano.

A través de todo el trabajo realizado, nos ha parecido interesante incluir en las conclusiones del proyecto un esquema que lo resuma desde este punto de vista. Se trata de un modelo de cuatro niveles:

Nivel 1: **Procesado fotograma a fotograma** Es el nivel más básico conceptualmente. Se corresponde con la fase de análisis temporal de la secuencia, abarcando los procesos de cálculo de: varianza, métricas para los cortes y métrica para los fundidos.

La información de entrada son los fotogramas de la secuencia, y más concretamente, se accede al valor de cada uno de los pixels de la imagen correspondiente. Como resultado, se obtienen una serie de ficheros [1] de texto. Su contenido son números en coma flotante que corresponden al valor de los parámetros para cada fotograma (o para cada pareja de fotogramas).

En adelante, la secuencia será identificada por un conjunto de valores numéricos resultado del procesado de los pixels de sus imágenes. En este caso, el usuario no puede fijar ningún parámetro o umbral, ya que el cálculo es independiente del tipo de secuencia en todos los aspectos (Grises o RGB, mucho o poco movimiento, etc...). También de cara al usuario, podemos decir que los datos de este nivel no le aportan ninguna información, es necesario su procesado posterior.

Nivel 2: **Procesado efectos de transición** La información de partida son los ficheros de texto procedentes del Nivel 1, y como salida obtendremos otros ficheros de texto, pero cuyo contenido serán los números de fotogramas que contienen efectos de transición (fotogramas frontera).

Se corresponde con la fase de detección de cortes, detección de fundidos y localización de cortinillas. En este caso, el usuario sí deberá configurar unos parámetros,

[1] Para una descripción detallada del contenido y formato de estos ficheros, como de los del resto del proyecto, remitimos al lector al apéndice D.

de cuya elección depende en buena medida la efectividad del método.

El contenido de los ficheros resultado ya es comprensible por el usuario, no contienen datos numéricos sin sentido como en el Nivel 1, sino que se corresponden con números de fotogramas. Sin embargo, este usuario ha de saberlo, ha de conocer el formato de estos ficheros.

La información sobre cada efecto esta aislada (hay un fichero para cada uno) y será la base utilizada en la segmentación.

Nivel 3: **Procesado planos** El objetivo de este nivel es combinar la información de los distintos efectos de transición, buscando la correcta segmentación de la secuencia en sus planos temporales. El usuario podrá seleccionar qué efectos desea utilizar.

Su entrada son los ficheros del anterior nivel, y la salida serán las fronteras (entendiendo por frontera el primer fotograma y el último) de los planos. También se incluiría en este nivel la fase de extracción de los fotogramas clave de cada plano (previamente se deberán tener definidos los planos de la secuencia).

Se ha decidido que los ficheros de texto, con los resultados, no tengan un nombre general como en los niveles anteriores, sino que incluyan el nombre de la secuencia. Directamente puede considerarse que la información aportada por este nivel, es la segmentación definitiva de la secuencia (por eso lo de incluir su nombre).

Nivel 4: **Procesado semántico** Hasta el Nivel 3, todo el proceso es automático: el usuario solo debe configurar algunos parámetros, y el programa se encarga del resto. Podría ser visto como un modelo de caja negra, donde la entrada se corresponde con la secuencia, y la salida con la identificación de sus planos.

Sin embargo, hemos añadido la posibilidad de un análisis subjetivo de la imagen. Ayudado por algunas herramientas programadas, y haciendo uso del resultado de la segmentación, el usuario puede describir textualmente el contenido visual de cada uno de los planos. Y el resultado será un fichero de texto, asociado a la secuencia, cuyo contenido será totalmente comprensible para cualquier individuo, aunque no sea el usuario del programa.

En el esquema actual de indexación automática de vídeo, el análisis de la imagen no se realiza por contenido semántico. Resulta imposible, si no es mediante la incorporación de un modelo de inteligencia artificial. Lo que se hace es realizar aproximaciones al contenido mediante el análisis de sus objetos, análisis de color, forma, movimiento, etc.

Al pasar de un nivel a otro, se reduce la cantidad de datos disponibles, al mismo tiempo que aumenta la información útil para el usuario. Conceptualmente, podemos decir que se reduce la complejidad. Entendiendo por complejidad la dificultad de

comprensión de la información por parte del usuario. Y este ha sido el objetivo de nuestro proyecto.

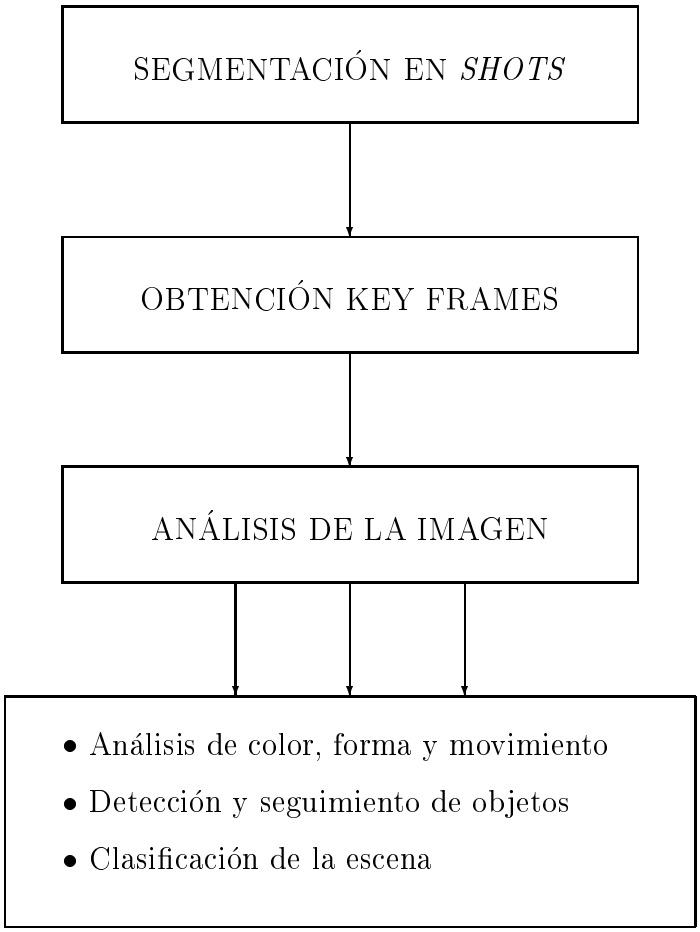

Figura 6.1: Esquema con los pasos llevados a cabo en la indexación de vídeo.

6.4 PROPUESTAS DE TRABAJO FUTURO

Este ha sido el primer proyecto sobre Indexación de Vídeo realizado dentro del GPIV. Se trata, por tanto, de una primera aproximación al problema de la segmentación temporal y de su posterior procesado. Ya que los resultados obtenidos han sido considerados satisfactorios, pero siendo factible la mejora de muchos aspectos, proponemos una serie de posibles líneas de continuación:

- Hemos comentado en repetidas ocasiones la dificultad de segmentar de forma correcta las secuencias de vídeo procedentes de películas antíguas. Se ha intentado mejorar el método utilizando algunas alternativas, si bien la mejora no suele ser notable.

  Ya que el objetivo de la segmentación para estas películas, es poder aplicar los algoritmos de restauración de forma separada a cada uno de sus planos, nos parece que podría probarse con la siguiente estrategia: realizar una primera restauración a la secuencia, intentando sobre todo eliminar las variaciones bruscas de brillo y las "vibraciones" más destacables (ambos fenómenos son los principales responsables de las falsas alarmas). Sobre esta primera aproximación a la secuencia restaurada, ya se podría intentar la segmentación en sus planos, y proceder con los algorimos avanzados de restauración para eliminar los artefactos e imperfecciones de las imágenes, así como eliminar más finamente las vibraciones, todo ello plano a plano. Esto podría verse como un proceso iterativo, y si es necesario, podrían incorporarse más "pasadas".

- El tema de la detección de cortes consideramos que ha quedado lo suficientemente estudiado y que con los resultados obtenidos no sería necesario profundizar más. Sin embargo, para el caso de las transiciones graduales, debería intentarse alguna estrategia alternativa. El tema de las cortinillas no ha quedado totalmente resuelto, y utilizando las mismas ideas como base, se podría mejorar. De igual forma para los fundidos, e intentando centrarse sobre todo en los fundidos largos (con muchos fotogramas intermedios), difíciles de localizar en este proyecto.

- En línea con lo anterior, también consideramos interesante incluir aspectos de detección y compensación de movimiento como parte del proceso de segmentación. Irían enfocados a la detección de efectos *intra-frame*: zoom, planos largos, panorámicas, barridos en horizontal y vertical de escenas, etc...; y también a la detección de los mencionados fundidos de larga duración.

- La extracción de fotogramas clave también es susceptible de mejora. Sobre todo, debería investigarse algún método de "clustering" que utilice una combinación de la mayor parte de información disponible.

- En este proyecto, el análisis final del contenido de la secuencia se ha realizado de forma textual, demostrando que se puede utilizar las segmentación para facilitar la tarea. Podría por tanto pensarse en ofrecer esta utilidad a alguna T.V. o entidad que realice indexado textual de vídeo, adaptandolo a las necesidades propias.

- Para terminar, aunque se trate de un tema muy amplio, debería intentarse el análisis automático de contenido de los distintos planos de la secuencia mediante el análisis del contenido de sus fotogramas clave. Se podría probar con características como: color, objetos, formas. Esto implicaría un uso extendido de herramientas de segmentación de imágenes.