1 INTRODUCCIÓN

1.1 Introducción

Actualmente nos encontramos en una época en la que prácticamente todo tratamiento de información que imaginemos se hace mediante un dispositivo con funcionamiento digital. Podemos ver como a lo largo de la última década, se han ido sustituyendo los métodos de adquisición, almacenaje y reproducción de cualquier tipo de información. Así por ejemplo, las cámaras fotográficas, videocámaras, reproductores de música, grabadoras, etc, han sido reemplazadas por sus homólogos digitales con las diferencias que esto conlleva.

La digitalización de señales analógicas aporta considerables ventajas entre las que podemos citar la protección frente a ruidos, la facilidad para encriptar las señales, la posibilidad de procesar digitalmente los datos, el poder enviar las señales digitales a grandes distancias, así como la gran capacidad de almacenamiento que en medios digitales encontramos y que aumenta a velocidad vertiginosa. No obstante, por lo general, la digitalización supone un aumento considerable del ancho de banda. A pesar de esto, el gran desarrollo experimentado por las tecnologías de la información en los últimos tiempos ha provocado, entre otros avances, un incremento de las comunicaciones digitales. El procesamiento digital de señales adquiere un interés especial, puesto que es la base para plataformas tales como Internet, multimedia, televisión digital y sonido digital. La posibilidad de transmitir vídeo digital se presentaba como algo muy lejano hace unos años. Incluso se llegó a pensar que su introducción no se lograría hasta bien entrado el siglo XXI [13].

Lo que hizo cambiar esta percepción fue la aparición de eficaces algoritmos de compresión de vídeo, que reducían de manera significativa el flujo necesario para la transmisión de imágenes.

Para comprender la necesidad de algoritmos compresores necesitaremos primero unas nociones básicas sobre el vídeo digital sin comprimir. Por esto, en el siguiente capítulo haremos una introducción al vídeo digital., hablando de los principales formatos existentes y de las etapas en el proceso de codificación. A continuación, expondremos la necesidad de compresión de una forma razonada. Por último encontraremos las

motivaciones que nos han llevado ha desarrollar este proyecto y los objetivos propuestos en su comienzo. Además, se incluye un apartado con la organización del resto de la memoria.

1.2 Conceptos de vídeo digital

El vídeo es una captura, grabación, almacenamiento, y reconstrucción de una secuencia de imágenes que representan escenas en movimiento. Tanto el cine, la televisión y el vídeo doméstico están basados, entre otros, en un fenómeno conocido como "persistencia de la visión". Dicho fenómeno consiste en una imperfección del ojo humano, por el cual toda imagen que visualice se queda almacenada en la retina durante una fracción de segundo. De esta forma, si al ojo humano se le muestra una rápida sucesión de imágenes, tenderá a unirlas, y si éstas son muy similares, con pequeños cambios y se exponen a una velocidad adecuada, el cerebro las reconoce como imágenes en movimiento.

El vídeo surge como una tecnología íntimamente ligada a la televisión, pues nació como auxiliar de ésta para evitar que toda la programación fuera en directo, facilitando el trabajo de grabación, la planificación de horarios, el almacenaje de programas y la reproducción de los mismos. Por esto actualmente, cualquier formato de vídeo tiene herencias de este sistema.

La innovación en el registro de imágenes visuales y auditivas en este formato, comienza a finales de los años 50 e inicios de los 60. Entre 1965 y 1978 se consolida como un medio con singularidad y aplicaciones propias y empieza a ser independiente de la producción televisiva.

Una fecha importante en la historia de este medio es 1964, durante los Juegos Olímpicos de Tokio, año en el que se hace la primera emisión diferida de la transmisión en directo de este acontecimiento.

En 1965 se efectúa el primer vídeo personal con una intención artística, cuando el coreano Nam June Pail filmó la visita del Papa Pablo VI a Nueva York desde la ventanilla de un taxi.

Posteriormente, en 1968. la *Sony Corporation* produce el "*portpack*", primera cámara portátil comercializada, y en ese mismo año ocurre que Jean Louis Godard graba la

revuelta francesa de estudiantes por la mañana (hecho conocido como el Mayo Francés), y es visto por la noche en una librería francesa [19].

Como hemos dicho, el vídeo y la televisión se relacionan y se pueden considerar medios complementarios. Por esto, para estudiar la transición entre vídeo analógico y vídeo digital tomaremos como ejemplo la señal analógica empleada en televisión.

La señal analógica de televisión está compuesta de imágenes. Como hemos dicho, para crear la sensación de movimiento dichas imágenes deben cambiar cada cierto tiempo de forma que el ojo no lo perciba (25 imágenes por segundo en PAL, y 30 en NTSC). A su vez, cada imagen está formada por líneas (625 líneas para un televisor PAL, 525 para un televisor NTSC), pero una imagen digital está compuesta de píxeles. Las principales características de estos dos sistemas se detallan en la siguiente tabla.

	IMAG/S	CAMP/S	LINEAS/IMAG	T_{LINEA}	F_{LINEA}	T_{IMAGEN}	F_{IMAGEN}	T_{CAMPO}	F_{CAMPO}
PAL B/G	25	50	625	64 s	15625 Hz	40 ms	25 Hz	20 ms	50 Hz
NTSC	30	60	525	63,56 s	15734 Hz	33,33 ms	29,97 Hz	16,66 ms	60 Hz

Tabla 1.1 Características básicas de los sistemas PAL y NTSC

Puesto que más adelante hablaremos del vídeo entrelazado introducimos aquí algunos conceptos. Consiste en realizar una doble exploración entrelazada de las líneas de cada imagen como se muestra en la figura 1.1. La imagen se divide en dos subimágenes o campos explorando alternativamente las líneas pares y las impares. La proximidad entre líneas consecutivas hace que el espectador integre las dos subimágenes y obtenga la sensación de que éstas se están renovando a una frecuencia doble de la real. Con ello se consigue eliminar un fenómeno de parpadeo que aparecía con la exploración normal.

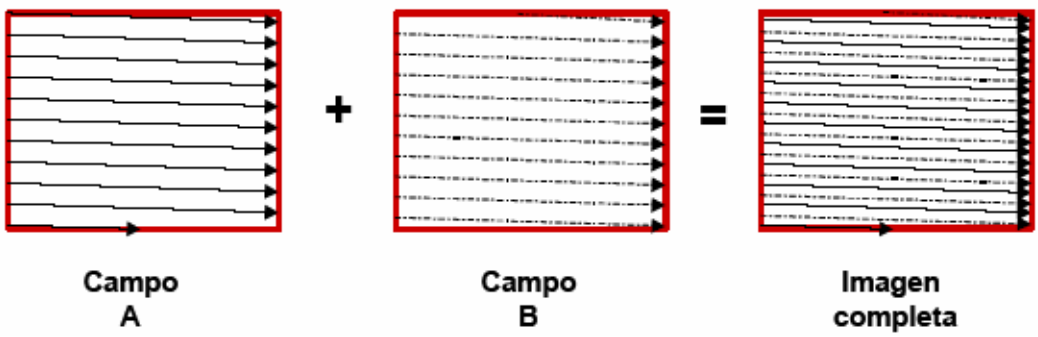

Figura 1.1 División de la imagen en dos campos entrelazados

Para la digitalización de esta señal de vídeo se hace un muestreo de cada una de las imágenes (muestreo temporal) en filas y columnas (muestreo espacial) tal y como se representa en la figura 1.2.

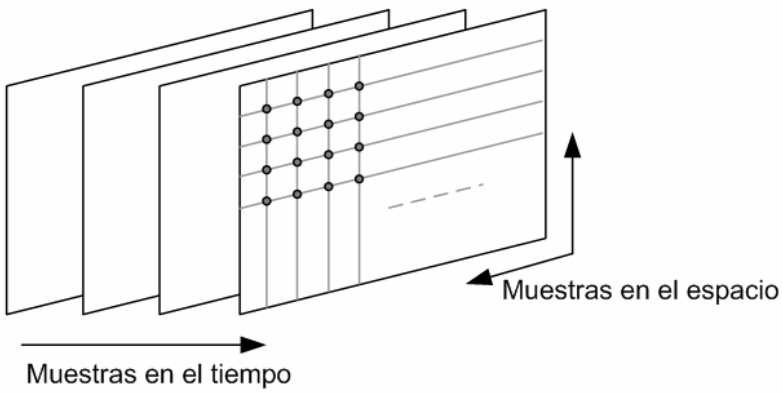

Figura 1.2 Muestreo espacial y temporal de la secuencia de vídeo

Si consideramos una señal analógica en blanco y negro (solo tiene información de la luminancia, es decir, del nivel de gris en cada punto) y tomamos muestras equiespaciadas en cada una de sus líneas, se obtendrá un muestreo de la imagen con estructura rectangular (cada cuadro de televisión guarda una relación de aspecto rectangular 4/3), en la que cada muestra representa los valores de luminancia obtenidos como se representa en la figura 1.3.

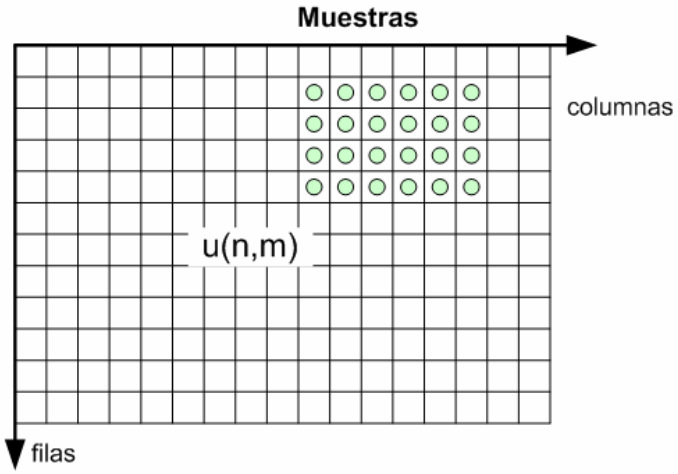

Figura 1.3 Retícula de muestreo rectangular para la digitalización de las imágenes

De este modo, la imagen digital puede tratarse como una matriz u[n,m]. Cada una de estas muestras serán los llamados elementos de imagen o píxeles. El número de muestras que se toman en cada línea de la imagen debe guardar una relación con el número de líneas de cada imagen con objeto de que la resolución en ambas direcciones tenga valores parecidos. Así pues, como primera aproximación, en un sistema como el europeo (PAL) que utiliza 575 líneas activas, el número de muestras tomadas en cada línea debería ser de unas (4/3) x 575 = 766 muestras, donde hemos multiplicado por la relación de aspecto para tener en cuenta que la imagen no es cuadrada. En la figura 1.4 se muestran distintas versiones de la misma imagen muestreada con distintas resoluciones (en este caso cuadradas).

Figura 1.4 Versiones de una misma imagen con distintas resoluciones

Cada una de estas muestras de luminancia deberá representarse con un número de bits o una profundidad. La experiencia empírica demuestra que el número mínimo de niveles a partir del que no se aprecia ninguna mejora en la calidad de las imágenes se sitúa entre los 45 y 60 por lo que bastaría con 6 ó 7 bits para codificarlos. Debido a que toda la estructura de las memorias está basadas en palabras de 8 bits, se decidió cuantificar los niveles de luminancia con 8 bits (256 niveles posibles).

En el caso en que se trate de imágenes en color, puede optarse por muestrear las 3 componentes de color RGB o bien la luminancia y las señales diferencia de color. En el primer caso, la resolución de las tres imágenes debe ser la misma. En el segundo, es habitual muestrear las señales diferencia de color con una retícula de muestreo que incluya un menor número de muestras espaciales aunque con el mismo número de bits.

1.2.1 Formatos de vídeo digital

La primera norma que surge para la televisión digital será la CCIR 601, que se encarga del muestreo de la señal, sin llevar a cabo ningún tipo de compresión. Puesto que se hará referencia a este formato en el futuro, considero oportuno introducir algunos conceptos.

Recomendación 601

En 1982, el ITU-R (anteriormente CCIR), en su recomendación 601, desarrolló un conjunto de especificaciones para señales de televisión digital en estudios de televisión o de producción de vídeo.

El objetivo de esta recomendación era facilitar el intercambio de programas a escala internacional. Las recomendaciones definen muchos parámetros comunes entre el formato americano (NTSC) y europeo (PAL) con el objeto de que los fabricantes puedan incluir varios módulos comunes en equipos para 525 y 625 líneas. El uso de este formato permite la interconexión entre distintos equipos digitales.

Existen básicamente dos variantes básicas conocidas como 4:2:2 y 4:4:4, donde los dígitos indican la proporción entre las muestras dedicadas a la luminancia (Y) y las dedicadas a las componentes de color (Cr y Cb, componentes diferencia).

Las dimensiones de la componente de luminancia en los formatos 4:4:4 y 4:2:2 son de 720x576 en los sistemas de 625/50 y de 720x480 en los de 525/30.

El sistema más utilizado es el 4:2:2, en el que están codificadas la luminancia y las señales diferencia de color. El formato 4:4:4 puede utilizarse también con componentes RGB.

En la figura 1.5 se muestra la disposición de las muestras tomadas en el formato 4:2:2. Como se puede ver, las señales diferencia de color tienen la mitad de muestras en el sentido horizontal y el mismo que la luminancia en el sentido vertical.

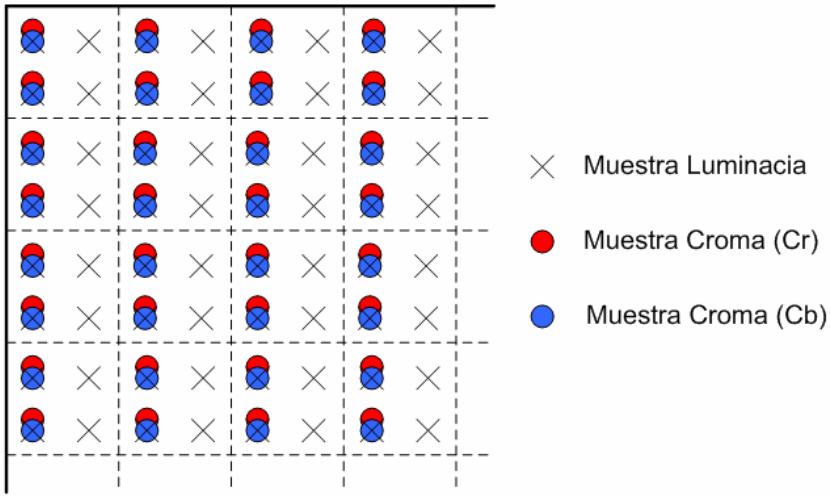

Figura 1.5 Distribución espacial de las muestras de luminancia y croma en el formato 4:2:2

Además de las dos variantes básicas 4:2:2 y 4:4:4, destaca también el formato 4:2:0 que es una simplificación del 4:2:2. Se obtiene reduciendo a la mitad la frecuencia de muestreo de las componentes de croma en el sentido vertical. Con ello, se iguala la densidad de muestras de croma en las dos direcciones. Las muestras de croma se obtienen a partir de las muestras del formato 4:2:2, promediando dos filas consecutivas. En la figura 1.6 se muestra un ejemplo del muestreo de este formato.

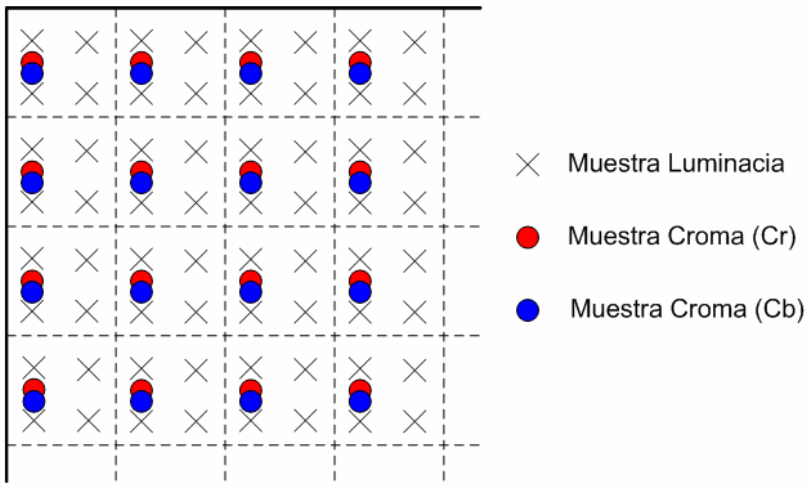

Figura 1.6 Distribución de las componentes de luminancia y croma en el formato 4:2:0

Formatos reducidos

En algunas aplicaciones donde no resulta necesaria una excesiva calidad de la imagen de vídeo digital suelen emplearse reducciones sobre el tamaño de la imagen. Con ello, se

consigue una importante reducción respecto a la tasa de bits original, que puede hacer factible la codificación de la señal en soportes de baja densidad o canales de reducido ancho de banda.

Uno de los formatos de vídeo reducido más populares es el SIF (*Source Intermediate Format*) que, esencialmente, consiste en un submuestreo de factor dos del formato 4:2:0 que se aplica tanto a las componentes de luminancia como de croma. Los tamaños de las matrices de luminancia para el formato SIF son de 360x288 para el estándar de 625 líneas y 360x240 para el de 525 líneas. Las matrices de croma también se submuestrean en un factor 2 en cada dirección respecto las matrices de croma del 4:2:0 (180x144 para 625 líneas y 180x120 para 525). También se realiza una reducción de la frecuencia de imagen a 25 Hz para el sistema europeo y a 30 Hz para el americano. Con ello, las imágenes resultantes no son entrelazadas. En estas condiciones, se obtiene una calidad equivalente al formato de vídeo analógico VHS.

El formato CIF (*Common Intermediate Format*) es un compromiso entre el formato SIF para 625 y 525 líneas. Utiliza 360x288 muestras de resolución de luminancia (europeo) y una frecuencia de refresco de 30 Hz (americano).

Los formatos QSIF y QCIF (*Quater*) se obtienen reduciendo de nuevo la resolución espacial en un factor 4 (factor 2 en cada dirección) y la resolución temporal en un factor de 2 o 4. Estos formatos suelen utilizarse para la transmisión de señales de vídeo telefonía con el estándar de compresión H.261 o para la transmisión de vídeo en directo por Internet. Los submuestreos espaciales se realizan filtrando las señales SIF (o directamente la 4:2:0). Las posiciones espaciales de las componentes de luminancia y croma resultantes son parecidas a las del formato SIF.

Formatos de píxel cuadrado

En aplicaciones informáticas es conveniente que los píxeles tengan una relación de aspecto cuadrada, ya que las tarjetas de visualización y los programas de tratamiento de imagen asumen esta propiedad. Por ello, es conveniente redefinir unos formatos alternativos para poder gestionar las señales de vídeo digital mediante ordenador. Los formatos utilizados en entornos informáticos se suelen denominar formatos de píxel cuadrado y son equivalentes a los formatos que hemos considerado hasta ahora, con la salvedad de que se modifica el número de elementos por fila para forzar que la relación de aspecto del píxel sea cuadrada. Así, en el estándar de 625 líneas, el tamaño de una imagen 4:2:0 será de 768x576 píxeles (768 = (4/3) x 576) mientras que en el estándar de

525 líneas tendremos un tamaño de imagen de 640x480 píxeles (640 = (4/3) x 480). Este formato es conocido como VGA y, como vemos, proviene del estándar americano. El resto de formatos SIF, CIF y QCIF de píxel cuadrado se obtienen dividiendo las resoluciones espaciales por factores de dos. En la tabla 1.2 se comparan los tamaños de cada formato en las versiones de 625 líneas, 525 líneas y formatos cuadrados.

	Europeo TV	***Americano TV***	***Eur. Cuadrado***	***Amer. Cuadrado***
4:2:0	720x576	720x480	768x576	640x480
SIF	360x288	360x240	384x288	320x240
CIF	360x288	360x288	384x288	384x288
QSIF	180x144	180x120	192x144	160x120

Tabla 1.2 Comparación entre formatos digitales

1.2.2 Codificación de la señal de vídeo digital

En la figura 1.7 se representan las distintas etapas conceptuales que intervienen en el proceso de codificación de la señal de vídeo digital.

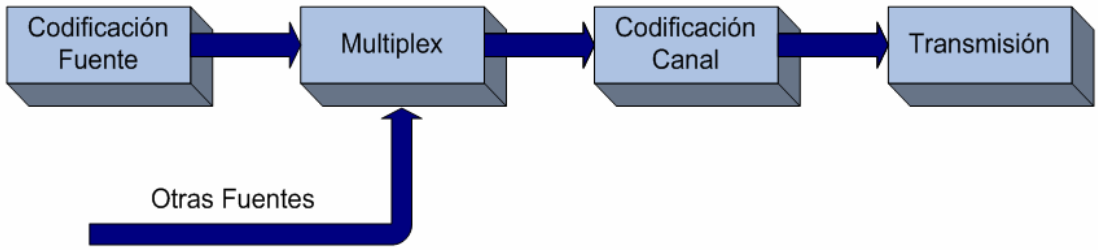

Figura 1.7 Etapas en la codificación y transmisión de una señal de vídeo digital

- La codificación de fuente consiste en extraer toda la redundancia posible en la señal, reduciendo el número de bits con el que se representa la información sin que ello suponga una pérdida aparente de calidad para el sistema visual humano. Los métodos y estrategias empleados para la codificación de fuente son muy variados, y resulta difícil establecer un procedimiento de compresión óptimo puesto que los resultados dependen, en gran medida, de las características de las señales. Generalmente los métodos de compresión avanzados combinan distintas estrategias simultáneamente para codificar la señal.

- La multiplexación de distintas señales básicas es necesaria para formar una única trama binaria que transporte la información y contenidos de todo el canal definido por la plataforma operadora.
- La codificación de canal se utiliza para adecuar la forma de onda de las señales a las características del canal. En esta etapa se introducen códigos de protección frente a errores, códigos de encriptación de la información y se adecua la forma de onda de las señales para que puedan ser directamente utilizadas por la etapa de modulación o transmisión. Esta última etapa se encarga de adaptar los datos a las características del medio de transmisión, para garantizar una correcta recepción de las señales.
- El medio de transmisión condiciona el tipo de modulación que se utiliza y la codificación de canal. Así, en un medio como el cable, los sistemas de detección y corrección de errores pueden ser más simples que en el caso de la comunicación vía satélite, debido a que la relación señal a ruido es mucho mayor y, por tanto, se reduce la probabilidad de error.

1.3 Necesidad de compresión

Para tener una idea general del volumen de datos binarios que supone la digitalización de una señal de vídeo, y por ello, la necesidad de compresión, consideremos como ejemplo el caso del estándar 4:2:2. La luminancia se muestrea a una frecuencia de 13.5 MHz tomando 8 bits por muestra, lo que da un total de 108 Mbps. Además, cada una de las señales de diferencia de color se muestrean a 6.75 MHz, nuevamente con 8 bits por muestra, lo que produce un nuevo flujo de 108 Mbps. En total tenemos un flujo de 216 Mbps asignados exclusivamente a la señal de vídeo digital, sin tener en cuenta las componentes de audio o de información adicional que pueden desear incluirse en la señal de televisión [6]. Este flujo de datos, si bien puede ser soportado en la transferencia de datos entre equipos de un estudio, es excesivo para su transmisión directa al usuario, ya que requeriría utilizar un considerable ancho de banda.

Por ello, en el momento que haya que transmitir o almacenar la señal de vídeo digital es fundamental aplicar procedimientos y tratamientos específicos de los datos que permitan comprimir la cantidad de información que debe transmitirse al canal.

En el caso de señales de vídeo analógicas resultaba suficiente un ancho de banda de 5 MHz para mantener una resolución espacial y tasa de refresco de imágenes aceptable. Evidentemente, aunque se usen códigos y modulaciones digitales de gran eficiencia espectral, el ancho de banda necesario para transmitir una tasa de bits de 216 Mbps es muy superior a los 5 MHz que requiere el sistema analógico.

Por esta razón, los formatos digitales derivados del ITU 601 se han mantenido como estándares digitales para el intercambio de información entre equipos terminales en estudios de grabación y producción de vídeo y no se han extendido a sistemas de transmisión y almacenaje de la señal de vídeo.

Tomando estos formatos digitales como punto de partida, es necesario aplicar un proceso de reducción de la tasa de bits que permita la transmisión o almacenaje de las señales sin una pérdida aparente de calidad o sin hacer uso excesivo de los medios de los que se dispone. Este proceso es conocido como compresión o codificación y consiste básicamente en la reducción del número de parámetros requeridos para representar la señal, manteniendo una buena calidad de imagen.

1.4 Motivaciones

En un primer momento, el proyecto iba a consistir en un estudio y comparación de distintos algoritmos de detección de cambios de plano en vídeo comprimido. La idea consistía en programar en MATLAB algunos de los algoritmos más característicos y comparar su bondad con respecto a distintos parámetros. Se pensó en basarse en el estándar de vídeo comprimido MPEG-2. Para ello, en primer lugar, me centré en la documentación de estos algoritmos, seleccionando tres de ellos. Cuando comencé la parte técnica y llegó la hora de programar, resultó que no existía ningún código abierto en MATLAB que analizara la estructura del vídeo comprimido. Esto era necesario, ya que cualquier algoritmo de cambio de plano compara determinados parámetros (presentes en el bitstream de vídeo) en las sucesivas imágenes, de forma que cuando detectan una diferencia mayor a un umbral, indicarán que se ha producido el cambio.

Ya que no existía en código abierto una aplicación que hiciera lo que necesitábamos, nos planteamos desarrollarla en este proyecto. Además, en aras a servir a futuras investigaciones y como un aporte más a la comunidad de la información, se ha decidido publicar la aplicación desarrollada en el “*File Exchange*” de *Mathworks*.

La razón de haber escogido el estándar de vídeo comprimido MPEG-2, y no otra de las versiones de MPEG, es principalmente el gran número de aplicaciones en las que se emplea este estándar, desde sistemas de almacenamiento digital como el DVD (en una versión modificada) hasta la televisión de alta definición.

1.5 Objetivos

Aunque el principal objetivo de este proyecto es el desarrollo de un decodificador de vídeo MPEG-2 en MATLAB, su consecución necesariamente se ha tenido que basar en la realización de varias tareas previas.

En primer lugar se debía entender por qué es necesaria la compresión al mismo tiempo que estudiar las técnicas empleadas para realizarla. Estas técnicas, basadas en la compresión empleada en imágenes estáticas, deben conseguir un flujo de bits óptimo para las aplicaciones en las que se usará la información de vídeo. Hasta este punto, muchos de los distintos estándares de compresión existentes se basan en procesos equivalentes.

Centrándonos en el estándar elegido, MPEG-2, la primera tarea a realizar fue un estudio exhaustivo del estándar. Aparte de las técnicas de compresión empleadas en este caso particular, debíamos comprender las distintas estructuras existentes en su bitstream. Al ser un estándar abierto, la recomendación proporciona simplemente una sintaxis que deben cumplir los vídeos comprimidos con este estándar, lo cual deja gran libertad al diseñador del codificador o decodificador. Las distintas estructuras que forman un bitstream con formato MPEG-2 y las relaciones entre ellas vienen recogidas en la recomendación ISO/IEC 13818-1, denominada Capa de Sistema. Además, como la información que se iba a descomprimir era la de vídeo, se realizó el estudio pertinente de la Capa de Vídeo, recogida en la recomendación ISO/IEC 13818-2.

En la parte de desarrollo técnico se fijaron dos objetivos. El primero de ellos era realizar un programa capaz de demultiplexar la información de vídeo, para su posterior descompresión y análisis. El segundo era el decodificador propiamente dicho. Para la realización de este programa se debía emplear como patrón la sintaxis recogida en la Capa de Vídeo del estándar, con los mismos algoritmos y, en la medida de lo posible, la misma información. De esta forma se conseguiría que cualquier persona ajena al proyecto, siempre que tenga en cuenta la recomendación, pudiera comprender e incluso modificar el código para cualquier otra aplicación.

Aunque no se fijó como objetivo minimizar la velocidad de descompresión, se llevará a cabo un estudio de tiempos empleados en el desarrollo de la aplicación para detectar cuál serían las partes mejorables del algoritmo.

1.6 Organización de la memoria

La memoria está dividida en ocho capítulos, a su vez subdivididos en partes y temas concretos que relacionamos a continuación.

El primer capítulo sirve de introducción o de preámbulo, mostrándose algunos conceptos de vídeo digital y la necesidad de comprimir este tipo de información. Incluyéndose también las motivaciones que nos han llevado a realizar este proyecto y los objetivos fijados.

En el segundo capítulo se exponen las técnicas usadas en los estándares de compresión de vídeo. Se da por tanto, una visión general, sin particularizar ni definir ningún estándar en concreto.

El tercer capítulo nos detalla la información relacionada con el estándar MPEG-2, en cuya primera parte, se habla de su historia y de la organización de la recomendación ISO/IEC 13818. Posteriormente, y con un nivel más detallado que en el capítulo segundo, se presentan las técnicas de compresión empleadas.

En el cuarto capítulo hablamos de la Capa de Sistema (ISO/IEC 13818-1), que contiene la forma de multiplexar la distinta información de vídeo, audio y de otros tipos. Además se explican las dos formas de agrupar dicha información.

En el quinto capítulo damos un repaso general a las técnicas de compresión de audio y de una forma algo más superficial que en el caso de vídeo hablamos del proceso de codificación empleado en MPEG-2.

En el sexto capítulo llegamos a la información de vídeo. Presentando la estructura de datos de vídeo y los principales parámetros que definen este tipo de información.

La información del desarrollo se recoge en el capítulo siete. En primer lugar se explican las características fundamentales de las aplicaciones diseñadas, haciendo un estudio de tiempos. Por último, se muestra un ejemplo del bitstream de vídeo y su interpretación correspondiente.

Finalmente y como colofón, en el octavo capítulo se incluyen las conclusiones, los objetivos cumplidos y las posibles vías futuras de desarrollo a partir de este proyecto.