

Capítulo 2

Análisis de la codificación de vídeo MPEG-I

En este capítulo se presentan las distintas técnicas empleadas para la codificación de vídeo digital y se describe el algoritmo de codificación MPEG-I, el cual se ha revelado como el más eficiente en cuanto al nivel de compresión obtenido para un nivel de calidad dado. Fijando diferentes niveles de calidad de la imagen, objetivos o subjetivos, se desarrolla un detallado análisis de elección de los valores de los parámetros de codificación MPEG-I que maximizan el rango de compresión. Asimismo, se establecen las dependencias entre los valores de los parámetros, la calidad objetiva y la tasa binaria media generada en la codificación. A través del análisis estadístico de la tasa binaria generada por cuadro, para cada modo de codificación del algoritmo MPEG-I, se resaltan las dependencias entre los valores de los parámetros de codificación y su impacto en la tasa binaria resultante.

Se proponen distintos mecanismos de funcionamiento de un codificador MPEG-I y se discuten sus cualidades en cuanto a su adecuación para la transmisión sobre redes ATM y en cuanto a la calidad, objetiva y subjetiva, obtenida. Los mecanismos de funcionamiento se comparan a partir de la tasa binaria y del nivel SNR resultante de la codificación de distintas secuencias. Finalmente, se propone un esquema de codificación MPEG-I para la transmisión de vídeo digital sobre redes ATM. Este esquema se caracteriza por mantener un nivel de calidad subjetiva semi-constante y presenta una gran simplicidad en su funcionamiento.

2.1. Introducción

Las recientes aplicaciones y servicios ofrecidos de vídeo han promovido el desarrollo de nuevos algoritmos de comprensión de vídeo digital que reducen sustancialmente la capacidad de almacenamiento y la tasa binaria de transmisión. De entre los posibles servicios ofrecidos caben destacar los de telefonía [I.F720], videoconferencia [I.F730], distribución de televisión [I.J81], televisión por cable, distribución de televisión de alta resolución [ChiAna94] y vídeo bajo petición [ChaAna94]. El vídeo digital presenta diferentes resoluciones dependientes del servicio o aplicación. Los formatos empleados para los servicios de vídeo parten del formato CCIR-601 [I.R601] especificado para televisión. Así, para videoconferencia y para señal de televisión, con calidad de vídeo doméstico, se emplea el formato CIF (Common Image Format) y en servicios de telefonía el QCIF (Quarter of CIF). Para compatibilizar la señal de vídeo digital proveniente de vídeo NTSC y PAL también se ha especificado el formato SIF (Sequence Intermediate Format) como formato estándar de entrada para los algoritmos de codificación MPEG [MPEGI].

Las técnicas de compresión que emplean los algoritmos para vídeo digital se basan en la explotación de la redundancia espacial y temporal de la señal. El proceso de compresión puede provocar una distorsión o pérdida respecto a la información original, por lo que, aparece un compromiso entre el rango de compresión y la distorsión obtenida. Otras técnicas de compresión no introducen pérdidas pero el rango de compresión resultante suele ser muy inferior.

Las técnicas de compresión se pueden clasificar en función del tipo de explotación de redundancia que realicen. Las técnicas de explotación de la redundancia espacial procesan cada imagen individualmente aprovechando la semejanza entre los pixels de una misma zona, mientras que las técnicas de explotación de la redundancia temporal se basan en el parecido de los pixels situados en una misma posición de un conjunto de campos consecutivos de una secuencia de imágenes.

Las técnicas de compresión basadas en la reducción de la redundancia espacial se pueden clasificar según el tipo de transformación aplicada sobre la imagen en [Ron94]:

Codificación predictiva: Se basa en la codificación del valor diferencial de un pixel respecto al valor estimado a partir de los pixels previamente codificados de su entorno.

Codificación transformacional: Los métodos transformados buscan la extracción de la redundancia de los pixels de una misma zona de la imagen a través de una transformación lineal, de forma que la codificación de los valores obtenidos en el dominio transformado sea inferior a la de los pixels de la imagen. Se ha demostrado que la transformación lineal óptima es la denominada transformada Karhunen-Loeve [Kou95]. Esta transformación se basa en que los pixels de una zona próxima están muy

correlados y en que la distribución de probabilidad de los pixels de una zona es gaussiana. La transformación óptima se puede aproximar por la transformada discreta coseno (DCT), cuando los coeficientes de correlación están próximos a la unidad. En general, las zonas consideradas de la imagen suelen ser bloques rectangulares de pixels.

Codificación en subbandas: Es una descomposición de la señal original utilizando un banco de filtros de distintas bandas frecuenciales y decimando las señales obtenidas adecuadamente para que no aparezca aliasing. El resultado de este esquema crítico de descomposición en subbandas es un conjunto de señales con un número total de muestras igual a la original. Cada una de las señales se codifica independientemente y se pueden recomponer para obtener diferentes resoluciones de la imagen original.

Codificación jerárquica: La imagen original se descompone en una serie de señales de resolución menor hasta llegar a un nivel básico. A diferencia de la codificación en subbandas, la codificación de cada nivel de resolución necesita de los resultados de la codificación de resolución inferior. En el proceso de decodificación, la imagen original se reconstruye paulatinamente con la agregación de los distintos niveles de resolución.

Codificación por segmentación: Esta codificación se basa en la detección de los contornos de los objetos que componen la imagen y una descripción de estos objetos según su textura, luminosidad, etc. Esta técnica, si bien proporciona elevados niveles de compresión, requiere de un alto coste computacional.

Codificación por modelo: Cuando las imágenes que se pretenden comprimir mantienen invariantes los contornos, como un rostro en videotelefonía, basta con detectar en la imagen aquellos parámetros que describen el objeto invariante y, posteriormente, los correspondientes a su textura. De esta forma, se pueden alcanzar elevados niveles de compresión.

La explotación de la redundancia temporal se realiza a través de dos técnicas fundamentales:

Codificación transformacional: De la misma forma que se realizaba sobre una zona de una imagen, se puede aplicar la DCT simultáneamente sobre un grupo de pixels situados en diferentes campos consecutivos, pero en la misma zona espacial de cada campo. De esta forma se obtiene la transformación tridimensional denominada 3D DCT.

Codificación predictiva: En este caso, un bloque de pixels se codifica diferencialmente respecto a otro situado en un campo de referencia temporalmente próximo. En general, esta técnica se aplica buscando el bloque de pixels más similar al que se debe codificar, sobre el campo de referencia. Este mecanismo recibe el nombre de compensación de movimiento (CM), de forma que cada bloque codificado predictivamente va unido a un vector de movimiento o desplazamiento relativo del bloque empleado en el cuadro de referencia.

Codificación por relleno condicional: En este caso, en un campo sólo se codifican aquellos pixels cuyo valor es significativamente diferente de los codificados en el campo previo en la misma localización.

Junto con las técnicas de compresión presentadas también se suelen emplear mecanismos de cuantificación. La cuantificación se puede aplicar a cada muestra del dominio de partida o del dominio transformado (cuantificación escalar) o sobre un grupo de muestras (cuantificación vectorial) a fin de aprovechar la similitud de muestras próximas.

Los algoritmos de codificación suelen conjugar diversas técnicas de las expuestas anteriormente para maximizar el rango de compresión para un nivel de distorsión dado o para una tasa binaria constante. Dentro de los algoritmos de codificación estandarizados, podemos destacar:

- J.80: Codificación diferencial.
- H.120: Codificación diferencial y relleno condicional.
- MJPEG: Codificación DCT.
- H.261: Codificación DCT y CM.
- J.81: Codificación DCT y CM.
- MPEG-I: Codificación DCT y CM.
- MPEG-II: Codificación DCT y CM.
- MPEG-IV: Codificación por modelo.

2.2. Codificación de vídeo MPEG

Dentro de la variedad de técnicas de codificación, el algoritmo de codificación MPEG (Motion Picture Expert Group) se ha revelado como el más adecuado para la transmisión y almacenamiento de secuencias de vídeo [PanZar94]. El algoritmo de codificación MPEG, para vídeo digital, emplea conjuntamente la técnica de compresión transformada, a través de la DCT, y la técnica de compensación de movimiento. La efectividad en la reducción de la tasa binaria para un nivel de distorsión fijado ha sido contrastada con la obtenida por otros algoritmos de compresión basados en las otras técnicas de compresión descritas anteriormente. Este análisis se ha llevado a cabo a través de estudios estadísticos de secuencias de vídeo de larga duración. El algoritmo de codificación de vídeo MPEG inicialmente fue desarrollado para aplicaciones de almacenamiento y recuperación de vídeo comprimido a tasas binarias reducidas, del orden de 1'5 Mbps. Sin embargo, el algoritmo de codificación se ha mostrado muy adecuado para la transmisión de vídeo sobre redes de comunicaciones, dado que reduce

sustancialmente la tasa binaria en transmisión y, por tanto, los recursos necesarios de la red. Dentro de las aplicaciones más comunes donde es empleado este mecanismo de codificación caben destacar: videocorreos electrónicos, videotelefonía, videoconferencia, vídeo juegos, documentos multimedia, distribución de vídeo con calidad VCR (Video Cassette Recorder), vídeo bajo petición y televisión digital de alta definición.

El algoritmo MPEG se adecúa perfectamente para transferir vídeo comprimido sobre redes locales, metropolitanas y de área extensa. Esta adaptación se obtiene configurando el modo de funcionamiento de la codificación de forma que, dependiendo del tipo de red, se puede obtener una tasa binaria constante o variable [TanOku92] [HasRei94]. La diferencia entre la transmisión con tasa binaria constante o variable se manifiesta en que la calidad de la secuencia varía dependiendo de la complejidad y actividad de las imágenes o se mantiene constante. La estandarización del algoritmo de codificación MPEG-I, para el almacenamiento y distribución de vídeo comprimido en el rango de 1 a 2 Mbps con calidad de VCR se realizó en 1991 [MPEGI]. Posteriormente, se ampliaron las aplicaciones del algoritmo de codificación, de forma que se puede emplear para la distribución de vídeo digital con resolución de televisión y televisión de alta definición. A su vez, se incorporaron un conjunto de mejoras para permitir la compatibilidad en la presentación para diferentes resoluciones, disminuir los efectos de las pérdidas en transmisión y permitir la codificación de señales entrelazadas. Esta nueva versión MPEG-II, admite tasas que pueden ir desde 1 a 83 Mbps [MPEGII].

En la recomendación MPEG-I, se aconseja como formato de entrada el SIF ya que tanto para señal proveniente de PAL o NTSC se obtienen tasas binarias entre 1 y 2 Mbps. La secuencia SIF se estructura, como se muestra en la figura 2.2.1, en cuatro niveles de codificación: cuadro, tira o *slice*, macrobloque y bloque. El cuadro es la unidad básica de presentación cuyo número de *pels* (pixels de 8 bits) depende de la resolución. La imagen se estructura en zonas o bloques de 8 x 8 pels donde se aplica la DCT. Las componentes de croma del cuadro se submuestran en una relación 4:2:0 respecto a la componente de luminancia. La agrupación de 4 bloques de luminancia y uno por cada componente de croma se denomina macrobloque. El macrobloque es la unidad básica donde se aplica la técnica de compensación de movimiento. Un conjunto de macrobloques consecutivos horizontalmente se denomina tira o *slice*. La tira es el elemento mínimo donde se puede resincronizar la decodificación en el caso de pérdidas de información. El número de macrobloques consecutivos que forman una tira es seleccionable en el proceso de codificación. En este trabajo se ha considerado la tira como el conjunto de macrobloques que contienen los pixels de 16 líneas consecutivas, es decir, los macrobloques con la misma posición vertical en un cuadro. El presente estudio se ha realizado empleando la resolución estandarizada en la recomendación MPEG-I de 352 x 288 pels, con submuestreo de las componentes de croma tanto vertical como horizontalmente, y 25 imágenes por segundo. Las secuencias analizadas

se han obtenido a través de la digitalización de la señal de vídeo PAL correspondiente a distintos discos láser.

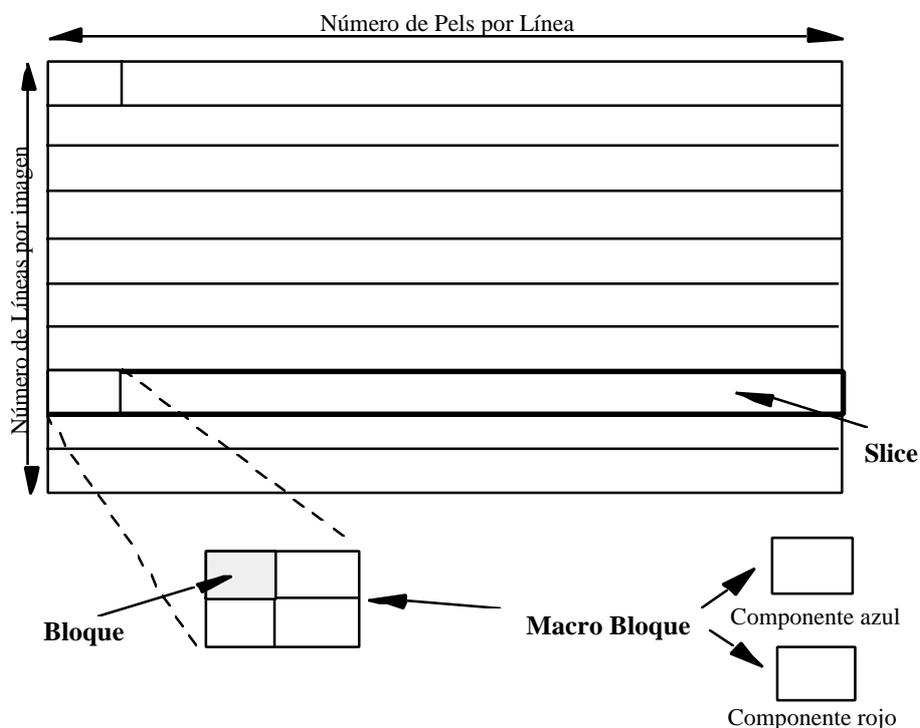


Figura 2.2.1. Estructura de un cuadro SIF

Las cuadros de una secuencia SIF pueden codificarse en tres modos diferentes:

- i) *Intra* (I): son los cuadros codificados empleando únicamente predicción espacial.
- ii) *Predictivo* (P): son los cuadros codificados con predicción temporal hacia atrás, usando como referencia el anterior cuadro I o P, y con predicción espacial
- iii) *Predictivo bidireccional* (B): son los cuadros codificados con compensación de movimiento, empleando como referencias la pasada o futura I o P. La compensación de movimiento se puede realizar sobre los macrobloques de una de las referencias o sobre una semisuma de un macrobloque de cada una ellas. También se aplica predicción espacial en los bloques del macrobloque diferencial obtenido.

El almacenamiento o transmisión de las imágenes de una secuencia se hace de forma que el decodificador pueda procesar la información lo antes posible. Para ello, en el almacenamiento o transmisión, las imágenes de referencia preceden a aquellas que las necesitan para ser decodificadas. Este efecto produce en aplicaciones en tiempo real un retardo de reordenación, dado que el orden de decodificación de los cuadros es distinto al de su presentación. A su vez, el codificador también introduce un retardo de proceso dado que necesita imágenes que temporalmente son posteriores para codificar otras que las preceden. Por ello, no es aconsejable en este tipo de aplicaciones que el número de

imágenes B consecutivas sea superior a 3 [KawChe93].

La secuencia de imágenes transmitida también se estructura en dos niveles, ilustrados en la figura 2.2.2:

- i) *Grupo de imágenes* (Group of Pictures, GoP), compuesto por una imagen I y las imágenes B y P que directa o indirectamente la han utilizado como referencia.
- ii) *Subgrupo de imágenes* (Subgroup of Pictures, SGoP) compuesto por una imagen de referencia I o P y las imágenes B que emplearon la imagen I o P como segunda referencia en su proceso de codificación.

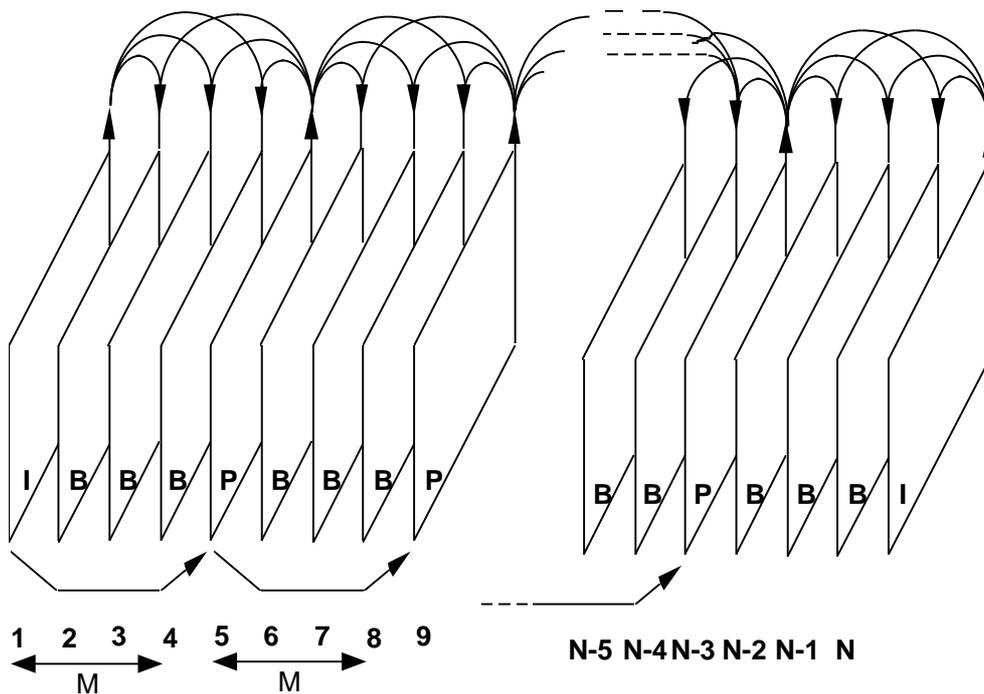


Figura 2.2.2. Estructuración periódica de los modos de codificación de los cuadros en un GoP de N cuadros con M cuadros por SGoP

En la figura 2.2.3 se muestra el proceso seguido por un codificador MPEG-I. La codificación de las imágenes de una secuencia con el algoritmo MPEG se realiza determinando el modo de codificación de cada una de las imágenes de un GoP.

La imagen intra se estructura en bloques a los cuales se les aplica la DCT. Los coeficientes obtenidos son ponderados con una matriz de pesos, ajustada según la sensibilidad de estos coeficientes al sistema visual humano. Las imágenes predictivas son estructuradas previamente en macrobloques con los que se realiza la compensación de movimiento, empleando como referencia la imagen I o P previa. Una vez identificado el macrobloque de la imagen de referencia, cuyo valor de los pixels es más similar al del macrobloque a codificar, se evalúa la conveniencia, o coste en bits, de la codificación en

modo intra y, finalmente, se emplea el método más rentable. En el caso de que se elija el modo intra, se aplica la DCT sobre cada uno de los bloques y se pondera con la matriz de pesos de la codificación I. Si la decisión es codificar en modo predictivo, se determina el macrobloque diferencial, sobre el cual se aplica la DCT en cada bloque con una matriz de pesos especificada para la codificación diferencial. En este caso se aplica la DCT para explotar la ponderación de los coeficientes transformados y no por razones de reducción de la tasa binaria, ya que la señal diferencial carecerá, en general, de redundancia espacial. Las imágenes predictivas bidireccionales se codifican empleando como posibles referencias las imágenes I o P que las preceden o suceden. De forma que las imágenes B, inicialmente se estructuran en macrobloques y se determinan los macrobloques de las imágenes de referencia que más se les asemejan. A continuación se evalúa el coste de codificación al emplear como referencia el macrobloque de la imagen I o P precedente, el macrobloque de la imagen de referencia siguiente, una semisuma de los pixels de los macrobloques de referencia mencionados o el modo de codificación intra. Del mismo modo que se realiza sobre los macrobloques de las imágenes P, se aplica la compensación de movimiento, cuando se emplean las imágenes de referencia, y la DCT.

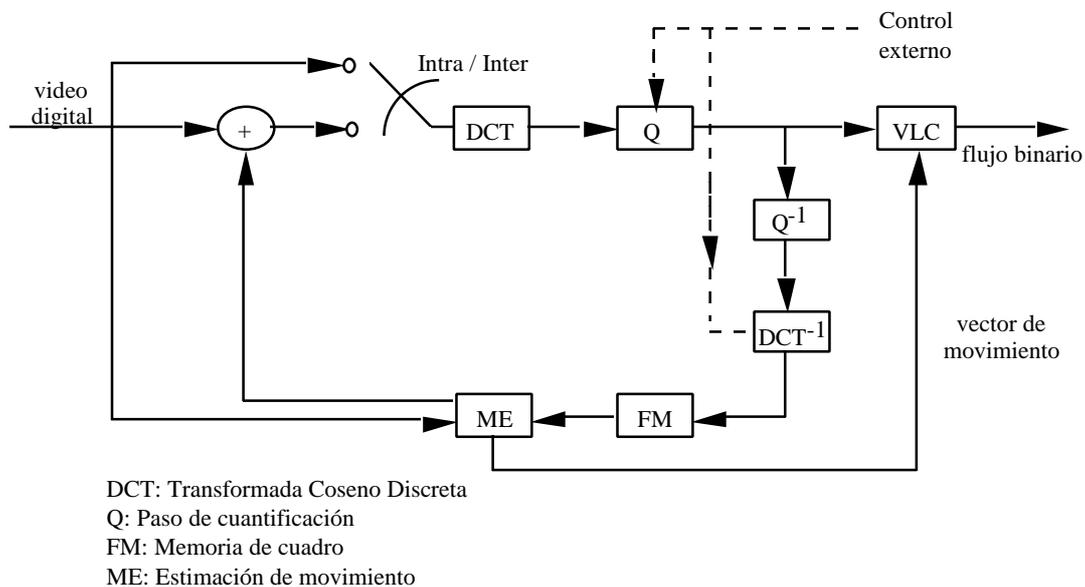


Figura 2.2.3. Esquema de codificación MPEG-I

La codificación de los coeficientes ponderados de la DCT, obtenidos de los bloques de las imágenes, se realizan de maneras diferentes. El coeficiente (0,0), o coeficiente de continua (DC), se codifica diferencialmente en macrobloques consecutivos intra de la misma tira. El resto de coeficientes se codifican independientemente de los coeficientes de otros macrobloques. Dado que los pels de la misma zona suelen tener valores muy

próximos, las componentes de alta frecuencia del dominio transformado son nulas o próximas a cero. Aprovechando esta característica presentada por los coeficientes de la DCT, se realiza un barrido en zig-zag de los coeficientes transformados para aumentar el rango de compresión. Así, empleando la técnica run-length, se asignan códigos de longitud variable en función de la probabilidad de valor del pixel y del número de ceros consecutivos en el barrido. Los códigos de Huffman estandarizados se obtuvieron después del análisis estadístico de múltiples secuencias.

En el proceso de cuantificación se emplea un paso de cuantificación (Q) ajustado según la calidad o la tasa binaria deseada. Los vectores de movimiento obtenidos en la codificación de macrobloques de imágenes P o B también se codifican diferencialmente respecto a los macrobloques precedentes de la misma tira. Esto se debe a que, en general, los objetos que componen una imagen ocupan varios macrobloques, y por lo tanto, el vector de desplazamiento suele ser semejante para macrobloques consecutivos de la misma tira.

El proceso de codificación de las imágenes P y B no se realiza sobre las imágenes de referencia originales sino sobre la imagen decodificada, a fin de que el codificador sea conocedor del resultado de la decodificación del receptor. Esto obliga al codificador a realizar el proceso de decodificación de la imagen para evitar la deteriorización progresiva de las imágenes. El codificador debe ser ajustado para que la elección de la codificación intra en los macrobloques de las imágenes P o B se decida cuando se alcance un nivel de distorsión determinado. El umbral de decisión deberá depender del cuantificador seleccionado, de forma que, cuanto más fino sea el paso de cuantificación antes se adopta la codificación intra. Se debe buscar una relación de compromiso entre la relación de compresión y la reducción de la tasa binaria para que las imágenes P o B no tengan un nivel de distorsión mucho más elevado que las imágenes I.

El estándar de codificación MPEG-II amplía las aplicaciones a las que estaba dirigido el MPEG-I. Las principales mejoras introducidas en el MPEG-II son: la posibilidad de operar con imágenes entrelazadas al emplear compensación de movimiento sobre macrobloques de 16 x 8 pels, aumenta la precisión de los coeficientes de DC a 10 bits, permite la cuantificación no lineal, mejora el control frente a errores en su sintaxis e introduce el concepto de escalabilidad. La escalabilidad permite que, con el mismo flujo de bits, se puedan emplear diferentes resoluciones espaciales y temporales junto con la posibilidad de disponer de diferentes niveles de calidad. La escalabilidad espacial y temporal consigue compatibilizar la decodificación sobre diferentes elementos de presentación. La escalabilidad de diferentes niveles de calidad, o escalabilidad SNR (signal to noise ratio), permite dividir el flujo de información del codificador, de forma que se asegura una calidad mínima en transmisión (flujo base) y una calidad añadida (flujo secundario). Esta separación de flujos es especialmente interesante en sistemas de comunicaciones con probabilidad de pérdida no nula, ya que permite gestionar, de

manera diferente, la información básica de la complementaria.

El algoritmo de codificación MPEG tiene dos modos de operación, configurables según el tipo de aplicación para la cual se emplea la compresión. En transmisiones sobre circuitos de capacidad fija, el algoritmo se configura para generar una tasa binaria constante, modo CBR (Constant Bit Rate). En el caso de que el sistema de comunicaciones soporte servicios de tasa variable, el algoritmo se puede configurar en modo VBR (Variable Bit Rate).

El modo de funcionamiento VBR presenta la ventaja, respecto al CBR, de poder mantener una calidad, subjetiva u objetiva, constante en toda la secuencia codificada de imágenes [SimRos93]. La calidad subjetiva se entiende como un nivel de percepción humano en la calidad de la imagen, mientras que la calidad objetiva es una cuantificación que intenta ponderar el error, o distorsión, de la imagen decodificada respecto a la original. En general, la medida empleada en este caso es el PSNR (Power Signal to Noise Ratio) [Wan94],

$$\text{PSNR} = 10 \log \left(\frac{255^2 R}{\sum_{i=1}^R (p'(i) - p(i))^2} \right) \quad (\text{dB}) \quad (2.2.1)$$

donde R es el número de pels en la porción de imagen a analizar, p(i) es el valor del pel original y p'(i) es el valor del pel decodificado.

Las variaciones de la tasa binaria generada en la codificación se deben a razones tanto intrínsecas, debidas al algoritmo de codificación, como extrínsecas debidas a la complejidad y actividad de la secuencia a codificar. Las razones intrínsecas están relacionadas, fundamentalmente, con los modos de codificación aplicados sobre las imágenes. Así, las imágenes I necesitan un número superior de bits a las imágenes P o B al emplear únicamente la técnica transformada DCT. Asimismo, las imágenes P suelen generar mayor número de bits que las B, dado que sólo emplean compensación de movimiento respecto a las imágenes de referencia anteriores. Dentro de la codificación de las imágenes, otro factor que provoca variaciones de la tasa binaria generada es la explotación de la entropía a través de tablas de códigos de longitud variable, según el tamaño y posición de las ráfagas (run-length) de los coeficientes de la DCT.

Las razones extrínsecas que provocan variaciones en la tasa binaria dependen del contenido de las imágenes a codificar. Las imágenes con mayor grado de detalle o mayor relieve tienen un nivel de complejidad superior y reducen la efectividad de la explotación de la redundancia espacial. Las secuencias de gran actividad, con movimientos rápidos de cámara, zooms y cambios de plano, impiden el empleo de la técnica de compresión predictiva, por lo que, también provocan aumentos en la tasa binaria respecto a secuencias de menor actividad.

Para conseguir una tasa binaria constante en el modo de funcionamiento CBR es preciso intercalar, entre la salida del codificador y el canal, una memoria tampón, o buffer, que absorba las variaciones de la tasa binaria generada en la codificación. Los bits almacenados en el buffer son extraídos a velocidad constante, mientras que el codificador llena el contenido del buffer de forma irregular. Para controlar el retardo introducido por la inserción del buffer, se dimensiona éste con una capacidad limitada y se regula la tasa de generación en el proceso de codificación dependiendo del nivel de ocupación del buffer.

La tasa de generación y el nivel de distorsión de la imagen se pueden controlar a través de varios parámetros que intervienen en el proceso de codificación. Los parámetros controlables que afectan a la generación de la tasa binaria son la resolución espacial y temporal de la secuencia, el número de imágenes por segundo generadas, el número N o M de imágenes que componen un GoP o un SGoP, la cuantificación de las tiras de imagen o la cuantificación de los macrobloques individualmente. Los parámetros no controlables son los extrínsecos como el contenido estadístico de la secuencia y la actividad de la escena. En general, los parámetros de resolución y frecuencia de las imágenes se acuerdan al principio de la sesión y no se modifican en su transcurso.

El parámetro más adecuado para controlar la tasa de generación o el nivel de calidad de la imagen es el paso de cuantificación para un macrobloque o para el conjunto de macrobloques de una tira de imagen. Es el más apropiado dado que no introduce una sustancial sobrecarga de señalización y tiene una rápida respuesta temporal sin añadir un elevado coste computacional. También se puede utilizar, como parámetro de control, la variación del número de elementos que componen un GoP o SGoP. Este mecanismo no se puede emplear de forma sostenida cuando aumenta la complejidad de la secuencia, ya que, como se analizará posteriormente, el exceso de imágenes B provoca automáticamente un aumento de la tasa binaria debido al incremento de macrobloques codificados en modo intra.

2.3. Elección de parámetros de codificación de vídeo MPEG

Para determinar la combinación de parámetros de funcionamiento del algoritmo de compresión MPEG se ha realizado un intenso estudio con diferentes secuencias de larga duración. En este análisis se ha buscado establecer los valores de los parámetros que dan un nivel de calidad subjetiva u objetiva de la imagen con un mayor rango de compresión. Este estudio se ha realizado con secuencias SIF provenientes de la digitalización de la señal de vídeo PAL con resolución de 352 x 288 pels de luminancia

y submuestreo vertical y horizontal de las componentes diferenciales azul y rojo de croma. El formato de las imágenes es progresivo con 25 cuadros por segundo.

El análisis de la elección de parámetros se ha realizado observando la influencia de los parámetros controlables en el algoritmo de codificación en la calidad subjetiva y objetiva. Tras este estudio se comprobará que empleando ambos criterios se llega a la misma elección.

2.3.1 Análisis de la calidad subjetiva en la codificación de vídeo MPEG-I

La metodología empleada para el análisis subjetivo de los parámetros de codificación se basó en el estudio de diversas secuencias que presentaban diferentes niveles de complejidad, movimiento y cambios de imagen. Este estudio se ha llevado a cabo con el empleo de estaciones de trabajo SUN Sparc20 equipadas para la visualización y captura de secuencias de vídeo. Los parámetros controlables en la codificación MPEG analizados son: el número de imágenes que componen un GOP (N), el número de imágenes de los SGoP (M) y el paso de cuantificación (Q). Se han utilizado las secuencias enunciadas en la tabla 2.3.1.

Secuencia	Número de cuadros
Geografía de Catalunya-I	3600
Geografía de Catalunya-II	15000
Miss America	300
Flower-garden	500
Live in Central Park	4200
Jurassic Park	35000

Tabla 2.3.1. Secuencias empleadas en el análisis subjetivo de la codificación de vídeo MPEG-I

Las secuencias de "Geografía de Catalunya", "Live in Central Park" y "Jurassic Park" presentan escenas de alto y bajo movimiento junto con variaciones de complejidad de las imágenes. Las secuencias "Miss America" y "Flower-garden", por el contrario, tienen bajo movimiento y complejidad. Parte de la contribución de este estudio fue presentada en [MatSal94] y [MatSal95]. Otro interesante análisis de calidad subjetiva puede encontrarse en [GilRos95] donde se comparan diferentes secuencias codificadas con MPEG-I, MPEG-II y MPEG-IV, para diversos tamaños de monitor y diferentes tasas binarias.

En la figura 2.3.1 se muestra la serie temporal bits por cuadro, correspondiente a la codificación de las imágenes I, P y B de la secuencia "Live in Central Park". A través de la inspección de la evolución de la serie temporal se puede relacionar el tipo de escena y su complejidad.

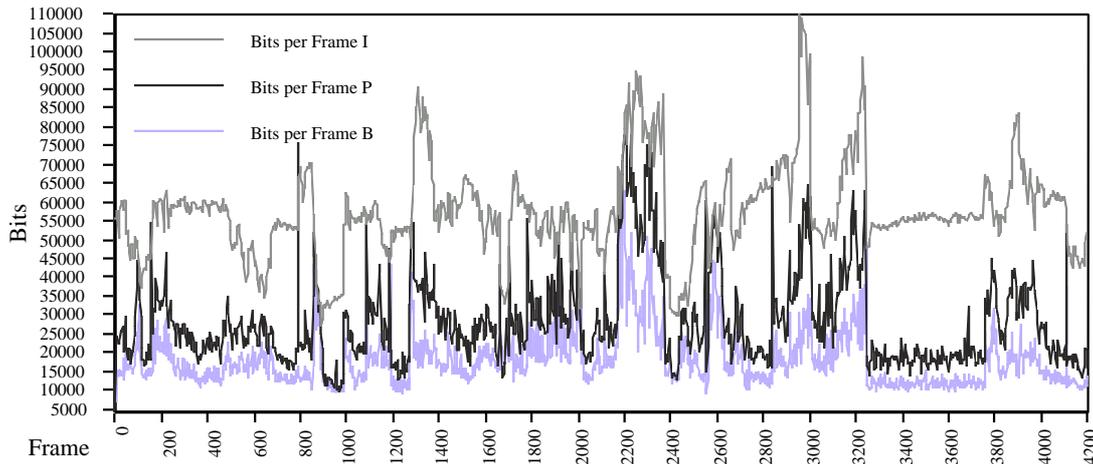


Figura 2.3.1. Series temporales bits por cuadro de la codificación ($Q=9$, $M=2$, $N=6$) de la secuencia "Live in Central Park" para los modos de codificación MPEG-I

La codificación intra de los cuadros, con un paso de cuantificación constante, permite determinar el nivel de complejidad de las imágenes y escenas. También se pueden identificar los movimientos de cámara y los cambios de escena a partir de las rápidas transiciones de la serie temporal I. Cabe destacar que se observa una relación directa entre el número de bits empleados en la codificación de una imagen I y su complejidad. Las series temporales correspondientes a las curvas P y B tienen dos comportamientos diferenciados, asociados a la actividad de las escenas. Cuando se producen cambios de escena o existen movimientos rápidos de la cámara, la curva P está muy próxima a la I. Este efecto es debido a que la técnica de compresión predictiva no se puede aplicar, obligando a emplear el modo intra en la mayoría de los macrobloques de los cuadros P. En estos casos se puede notar que la curva B también se comporta como las I y P pero con un nivel de proporcionalidad en el número de bits empleados en la codificación. Las curvas P y B muestran un comportamiento notablemente diferente cuando las escenas tienen un grado de actividad reducido. En este caso, la tasa binaria generada en las imágenes P y B decrece rápidamente hasta alcanzar una tasa moderada, manteniendo las curvas P y B una relación constante y próxima en la tasa binaria generada. Este mismo comportamiento es observado en todas las ternas (Q , M , N) evaluadas que aparecen en las tablas 2.3.2 y 2.4.1.

El parámetro de codificación M afecta directamente a las imágenes P y B. Para las imágenes P marca la distancia, en cuadros, de la imagen P a codificar respecto a la imagen de referencia. Dependiendo de la posición de la imagen B a codificar, dentro del SGoP, el parámetro M define la distancia a la que se encuentran las imágenes de referencia para la codificación de la B.

El número de imágenes de un GoP (N) determina la distancia entre dos imágenes I. La elección del valor del parámetro N permitirá maximizar la compresión y determinará

el tiempo mínimo, en recepción, necesario para recuperar las pérdidas de información producidas en la transmisión.

El análisis subjetivo se ha realizado con diferentes niveles de calidad de las distintas secuencias, fijados a través del paso de cuantificación Q . A su vez, se han variado los valores de N y M a fin de observar su influencia en la calidad resultante en la decodificación. En el análisis subjetivo de la codificación MPEG-I se observa que para valores de M inferiores a 4 la calidad subjetiva se mantiene. Cuando M sobrepasa este valor, se detecta una degradación de la imagen y es más apreciable conforme aumenta M . Esto es debido a que muchos de los macrobloques codificados en las imágenes P y B están próximos al umbral de decisión de la codificación del macrobloque en modo intra, por lo que, cuando se codifican diferencialmente están en el nivel máximo de distorsión admisible en el paso de codificación fijado. Este efecto es perceptible cuando un conjunto de imágenes consecutivas mantienen un nivel bajo de calidad y es más pronunciado cuanto mayor es el número de imágenes consecutivas. Asimismo, analizando el rango de compresión obtenido, este efecto se manifiesta en la aparición de macrobloques codificados en modo intra, por lo que el nivel de compresión disminuye respecto a valores de M próximos a 3. Esto es también observable en la serie temporal correspondiente a las B dado que aparece un pequeño rizado periódico de la tasa binaria por cuadro.

Respecto al parámetro N , es notable el efecto de frescos o restauración de la calidad cuando se emplean valores superiores a 24. Esto es debido al mismo efecto anterior. Las imágenes P utilizan como referencia las anteriores imágenes P . Así, cuanto mayor es la distancia entre cuadros P y mayor es su número en un GoP la calidad de estas imágenes disminuye paulatinamente, provocando que gran cantidad de macrobloques alcancen valores próximos al umbral de decisión del modo de codificación intra. Como consecuencia, las imágenes B de los SGoP disponen de referencias más deterioradas por lo que el efecto de distorsión aumenta en estas imágenes de forma apreciable. A su vez, un elevado número de imágenes en un GoP nos lleva a una pérdida de compresión, debida al empleo de un exceso de codificaciones en modo intra de los macrobloques de las imágenes B , que introducen un aumento de la tasa binaria y no ayudan en el aumento de la compresión del resto de las imágenes. Sin embargo, el aumento de la tasa binaria en los cuadros P , debido a la codificación intra de los macrobloques, es aprovechado por las imágenes P y B que las emplean como referencia.

Tras este estudio se ha propuesto un procedimiento sistemático de selección de los parámetros de codificación para el ajuste de la calidad subjetiva con la obtención de un adecuado grado de compresión. Este procedimiento queda ilustrado en la figura 2.3.2.

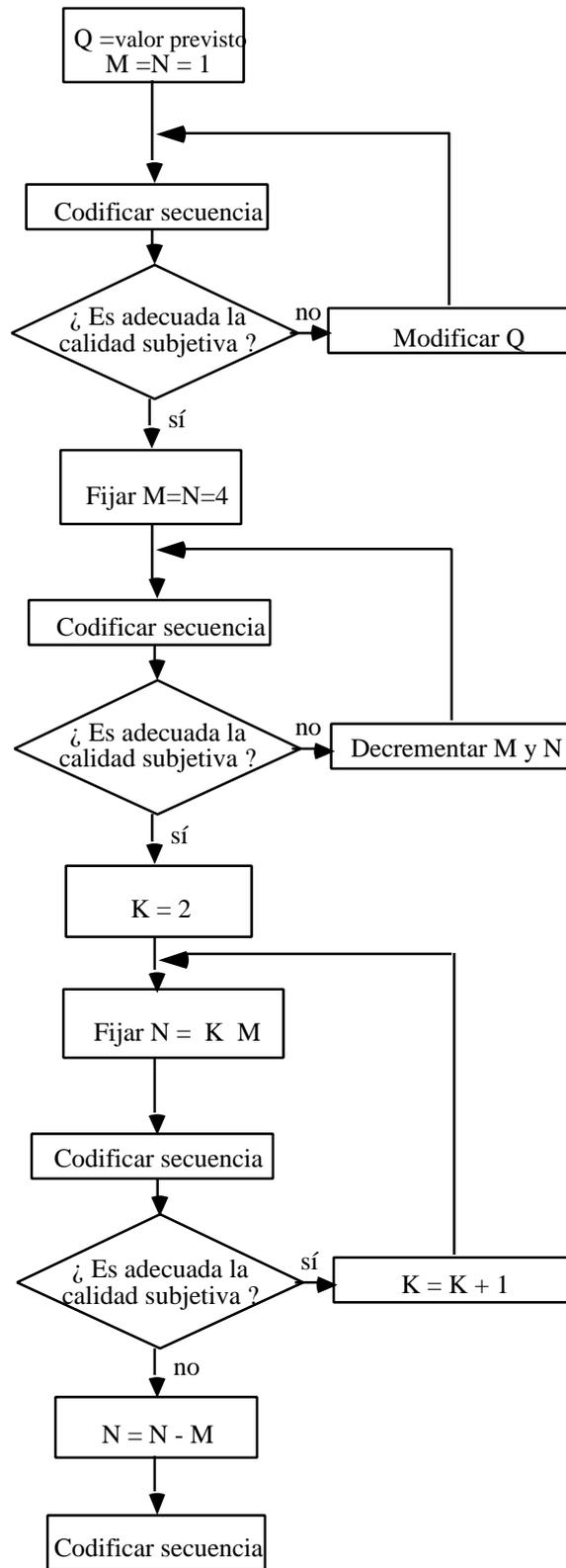


Figura 2.3.2. Diagrama de flujo de la selección de parámetros del algoritmo de codificación de vídeo MPEG-I

En un principio se ajusta el codificador con $M=1$ y $N=1$ y se varía el parámetro Q hasta alcanzar el nivel de calidad deseado. De forma práctica se ha comprobado que un nivel alto de calidad se obtiene con Q entre 4 y 6, una calidad media se alcanza con niveles de Q próximos a 10 y una calidad mínima aceptable se ajusta en valores de Q próximos a 16. Asimismo, si se dispone de un conocimiento de las características de la secuencia a codificar, como factor de corrección de los valores recomendados, es aconsejable decrementar el valor de Q cuando las escenas de la secuencia contengan gran actividad o complejidad y, por el contrario, se debe actuar de forma opuesta cuando se disponga de secuencias con bajo movimiento o alberguen gran cantidad de primeros planos. Una vez ajustado el valor deseado de la calidad, se decrementan simultáneamente los valores de M y N a partir del valor 4. Se selecciona el mayor valor de M para el cual no se perciben distorsiones de la secuencia y finalmente, se incrementa N en múltiplos del valor de M , hasta que aparezca el efecto de frescos en la presentación de la secuencia. Se fija el valor de N al valor inmediatamente anterior al último analizado.

2.3.2. Análisis cuantitativo de la calidad en la codificación de vídeo MPEG-I

De la misma forma que se realizó con la calidad subjetiva se ha llevado a cabo un análisis cuantitativo basado en el PSNR medio, obtenido en la codificación de las secuencias para diferentes ternas (Q, M, N). En la tabla 2.3.2 se muestran los resultados de algunas codificaciones de una secuencia de 4200 cuadros de la grabación "Live in Central Park". Las conclusiones derivadas de este análisis han sido contrastadas con la codificación, para diferentes ternas (Q, M, N), de tres secuencias de 12000 cuadros cada una obtenidas de la grabación de "Geografía de Catalunya". Cabe destacar que los resultados presentados se han agrupado en cuatro niveles de calidad según el valor de Q elegido. Así, se asocia $Q=4$ a un nivel alto de calidad, $Q=6$ a un nivel medio alto, $Q=9$ un nivel medio y $Q=20$ a un nivel bajo. Los valores de M y N se han fijado adecuadamente para observar las tendencias de la PSNR y del rango de compresión en función de su variación. Se puede comprobar que los rangos de compresión dependen fuertemente del Q elegido y son muy insensibles a la variación de N y M , siempre y cuando, N y M estén en los intervalos fijados ya en el análisis de la calidad subjetiva.

El rango de compresión (R) se ha definido como el cociente entre la tasa media resultante de la codificación MPEG-I y la tasa media de transmisión necesaria para transmitir la secuencia SIF sin compresión. En la tercera columna de la tabla 2.3.2 se especifica la tasa media generada por el codificador. Se puede observar que el rango de compresión aumenta rápidamente cuando se incorporan, en un GoP, imágenes codificadas en modo P y B. Cuando la calidad seleccionada es alta el aumento de la

compresión sólo es significativo hasta $M=3$ y $N=6$. Sin embargo, se observa que la elección más aconsejable es $M=2$ y $N=4$ o $N=6$. Para valores medios y bajos de calidad M y N pueden alcanzar los valores 4 y 12 respectivamente. En la figura 2.3.3 se muestran los rangos de compresión para los valores más representativos de N y M en función del paso de cuantificación Q . En la figura 2.3.4 se pone de manifiesto la variación del PSNR en función de Q para diferentes valores de M y N .

Conjugando los resultados obtenidos tanto en la evaluación subjetiva de las codificaciones como en el análisis cuantitativo podemos concluir que una adecuada elección de parámetros N y M sería $N=4$ y $M=2$ o $N=6$ y $M=2$. Sólo en casos donde la actividad de las escenas que componen la secuencia sea muy reducida y los niveles de complejidad sean bajos parece conveniente aumentar estos valores a $N=9$ y $M=3$ o $N=12$ y $M=4$. Estos valores también pueden ser empleados cuando la calidad fijada en la secuencia es baja y la probabilidad de pérdida o error de los medios de transmisión empleados sea reducida.

Triplet	SNRav	R	Kbps	Triplet	SNRav	R	Kbps
				N=1, M=1, Q=9	37.5302	21.20	1435
N=1, M=1, Q=4	40.1582	10.91	2788	N=2, M=1, Q=9	37.0801	30.14	1009
N=2, M=1, Q=4	40.0306	12.45	2441	N=2, M=2, Q=9	36.9667	32.64	932
N=2, M=2, Q=4	39.5401	14.08	2160	N=3, M=1, Q=9	36.8601	34.62	878
N=3, M=3, Q=4	39.3161	15.06	2018	N=3, M=3, Q=9	36.6993	38.34	793
N=4, M=1, Q=4	39.8758	13.17	2310	N=4, M=1, Q=9	36.7048	37.24	817
N=4, M=2, Q=4	39.4719	14.80	2054	N=4, M=2, Q=9	36.6613	40.06	759
N=4, M=4, Q=4	39.2052	15.31	1985	N=4, M=4, Q=9	36.5214	41.06	740
N=6, M=1, Q=4	39.7767	13.26	2294	N=5, M=1, Q=9	36.5912	38.93	781
N=6, M=2, Q=4	39.4217	14.95	2034	N=6, M=1, Q=9	36.4986	40.01	760
N=6, M=3, Q=4	39.274	15.36	1979	N=6, M=2, Q=9	36.5056	42.88	709
N=8, M=2, Q=4	39.3832	14.92	2038	N=6, M=3, Q=9	36.4556	43.54	698
N=9, M=3, Q=4	39.2392	15.31	1982	N=8, M=2, Q=9	36.3969	44.15	689
N=12, M=4, Q=4	39.1517	15.29	1987	N=8, M=4, Q=9	36.3198	44.58	682
N=16, M=8, Q=4	39.0463	14.84	2042	N=9, M=3, Q=9	36.3273	45.00	675
				N=12, M=2, Q=9	36.2435	45.31	671
	SNRav	R	Kbps	N=12, M=4, Q=9	36.2107	45.42	669
				N=16, M=8, Q=9	36.051	43.58	695
N=1, M=1, Q=6	38.8541	15.70	1936		SNRav	R	Kbps
N=2, M=1, Q=6	38.5208	20.41	1486				
N=2, M=2, Q=6	38.2019	22.79	1334	N=1, M=1, Q=20	34.6725	34.64	878
N=3, M=3, Q=6	37.9339	25.80	1178	N=2, M=1, Q=20	34.1974	53.64	567
N=4, M=1, Q=6	38.2373	23.68	1284	N=2, M=2, Q=20	34.4354	54.34	560
N=4, M=2, Q=6	38.0042	26.16	1162	N=3, M=1, Q=20	33.9536	64.77	470
N=4, M=4, Q=6	37.7772	27.00	1125	N=3, M=3, Q=20	34.2617	65.10	467
N=6, M=1, Q=6	38.0779	24.70	1231	N=4, M=1, Q=20	33.7766	72.10	422
N=6, M=2, Q=6	37.9	27.28	1114	N=4, M=2, Q=20	34.0226	71.49	425
N=6, M=3, Q=6	37.7924	27.89	1090	N=4, M=4, Q=20	34.1118	70.92	429
N=8, M=2, Q=6	37.827	27.68	1098	N=6, M=1, Q=20	33.5468	80.83	376
N=8, M=4, Q=6	37.6679	28.27	1075	N=6, M=2, Q=20	33.8068	79.02	385
N=9, M=3, Q=6	37.7145	28.34	1071	N=6, M=3, Q=20	33.8813	78.29	388
N=12, M=4, Q=6	37.6052	28.46	1069	N=9, M=3, Q=20	33.6808	82.90	366
N=16, M=8, Q=6	37.4546	27.35	1108	N=12, M=4, Q=20	33.5806	83.75	363

Tabla 2.3.2. Análisis de la influencia de los parámetros de codificación del algoritmo MPEG-I en la calidad y rango de compresión

Con el presente estudio también se han podido ajustar las curvas que relacionan el

nivel de calidad, seleccionado mediante el paso de cuantificación (Q), con el rango de compresión (R) y la PSNR. La bondad del ajuste de estas curvas se muestra en las figuras 2.3.5 y 2.3.6 respectivamente. Se puede comprobar que el ajuste es válido tanto para la elección M=2 y N=4 como M=2 y N=6. La curva que relaciona el paso de cuantificación y el rango de compresión se ha denominado factor de escala empírico [LucNeu94]. Obsérvese que la curva mostrada en la figura 2.3.5 está normalizada respecto al valor del rango de compresión resultante de la codificación con Q=9. Como derivación del ajuste de ambas curvas se puede obtener a su vez la curva tasa-distorsión en la codificación MPEG-I, mostrada en la figura 2.3.7, cuando la selección de parámetros es M=2 y N=4 o N=6. De forma que:

$$R = 7.2Q^{0.774} \quad (2.3.1)$$

$$PSNR = 45.879 \cdot Q^{-0.1} \text{ (dB)} \quad (2.3.2)$$

se puede relacionar directamente:

$$PSNR = 6.32v^{0.13} \quad (2.3.3)$$

siendo v la tasa binaria media en bits por segundo.

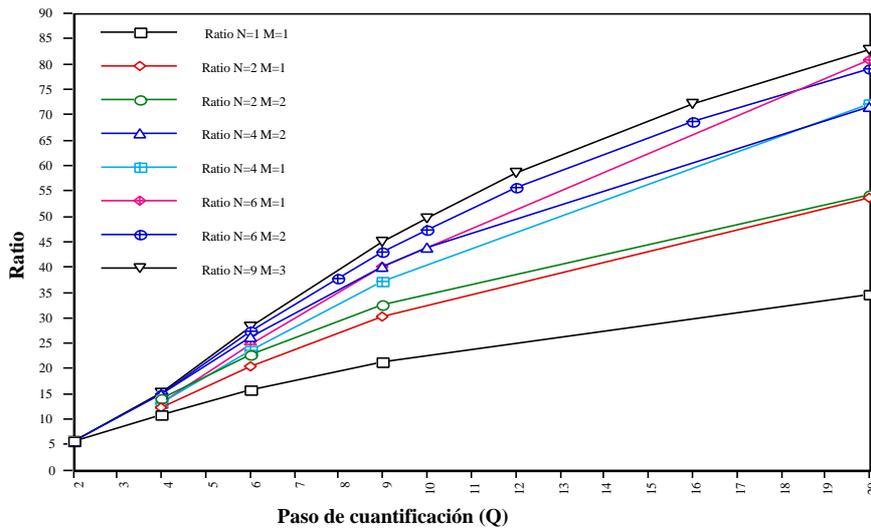


Fig. 2.3.3. Rango de compresión para la codificación MPEG-1 de la secuencia "Live in Centra Park" en función de los parámetros Q , M y N

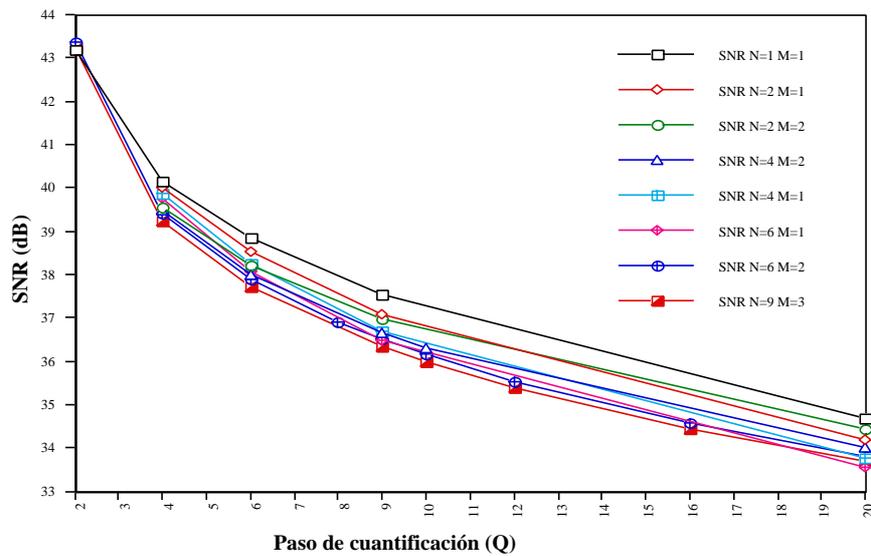


Fig. 2.3.4. Valor medio de la PSNR para la codificación MPEG-1 de la secuencia "Live in Centra Park" en función de los parámetros Q , M y N

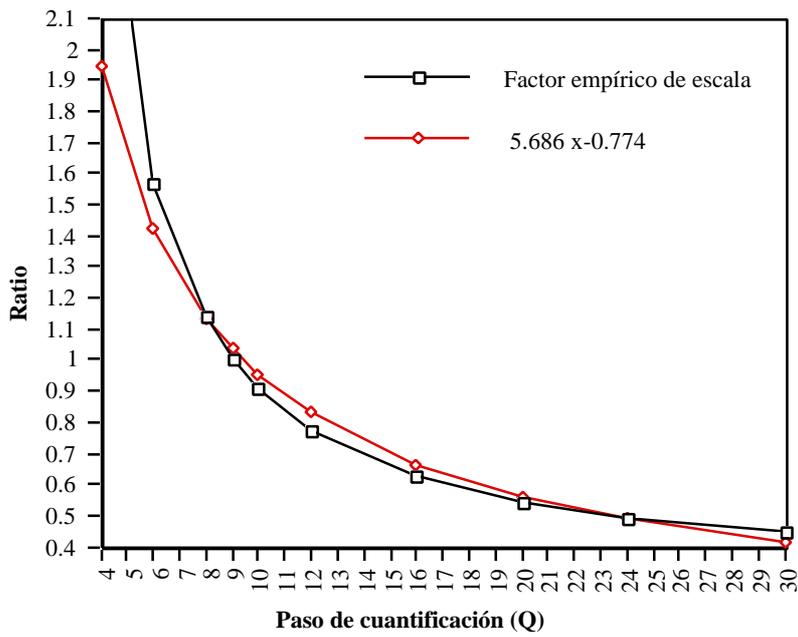


Fig. 2.3.5. Factor de escala empírico para la codificación MPEG-1 ($M=2$, $N=4$) normalizado para $Q=9$

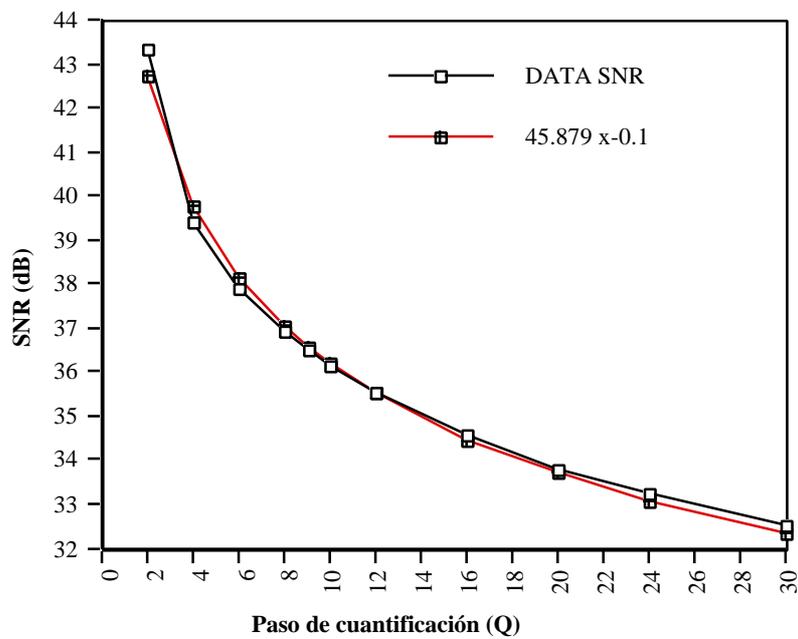


Fig. 2.3.6. Relación entre la PSNR y el paso de cuantificación (Q) para la codificación MPEG-1 con ($M=2$, $N=4$)

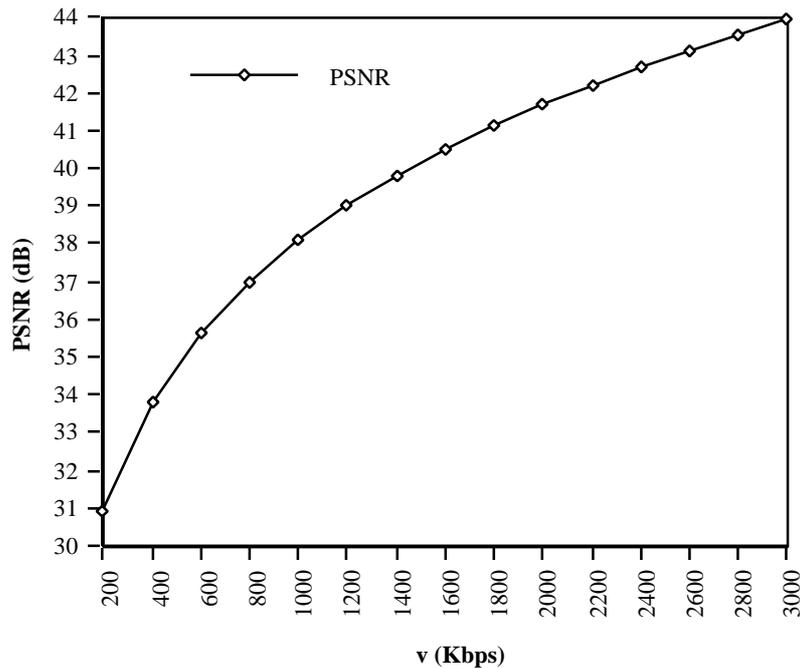


Figura 2.3.7. Curva de distorsión para la codificación MPEG-I con parámetros $N=4$ o $N=6$ y $M=2$

2.4. Análisis estadístico de la tasa binaria del algoritmo de codificación MPEG-I para vídeo

El estudio estadístico del comportamiento de la tasa binaria en la codificación permite llevar a cabo un análisis de la sensibilidad del algoritmo de compresión MPEG-I a la variación de sus parámetros [VerPin89] [SooGoo89]. Este análisis estadístico se ha realizado sobre las secuencias "Live in Central Park" y "Geografía de Catalunya" a partir de la consideración de las tasas binarias generadas en las imágenes I, P y B como tres variables aleatorias. Asimismo, se ha observado el comportamiento de la tasa binaria generada por el codificador, sin distinguir el modo de codificación empleado en las imágenes.

En primer lugar se ha realizado un análisis de los dos primeros momentos estadísticos de las variables aleatorias. Un conjunto significativo de resultados, para diferentes valores de las ternas (Q, M, N) , se muestra en la tabla 2.4.1. Se comprueba que tanto la media como la desviación estándar de la variable aleatoria I, adoptan valores muy próximos para los mismos pasos de cuantificación Q. En el caso de las P, se produce un pequeño aumento conforme se incrementa el valor del número de imágenes en un SGoP para un mismo valor de Q. Esta diferencia se acentúa para pasos de cuantificación más reducidos, aunque, la PSNR es prácticamente constante en todos

los casos. El incremento del número de cuadros de un SGoP provoca un aumento de la tasa binaria de las imágenes P. No obstante, este efecto queda contrareestado por la aparición de más imágenes B cuya tasa binaria suele ser más reducida. A su vez, las imágenes B también incrementan su tasa binaria cuando los cuadros de referencia están más alejados. Los datos presentados en la tabla confirman que la variable aleatoria B, efectivamente, aumenta su valor medio conforme se incrementa el valor de M. Sin embargo, no se manifiestan variaciones importantes en su desviación típica y en la PSNR. El tráfico conjunto presenta pequeñas variaciones en media y desviación estándar para los mismos valores de Q, salvo para los casos donde se emplea un número de imágenes en el GoP inferior a 3. En estos casos el valor medio es muy superior y la desviación típica es inferior, dado que, sólo se emplean uno o dos modos de codificación.

La relación temporal entre la tasa de codificación y el valor de los parámetros que controlan el algoritmo de codificación MPEG-I se ha analizado a través de la función de autocovarianza normalizada, o índice de autocorrelación. El índice de autocorrelación se define de forma genérica para una variable aleatoria discreta $X[n]$ como [KruSas95]:

$$r[k] = \frac{E\{(X[n] - E\{X[n]\})(X[n+k] - E\{X[n]\})\}}{\text{Var}\{X[n]\}} \quad (2.4.1)$$

donde $E\{X[n]\}$ es el valor esperado de $X[n]$ y $\text{Var}\{X[n]\}$ es la varianza de $X[n]$. En este caso se asocia la información generada en un intervalo de tiempo con el valor de $X[n]$, siendo n el número de intervalo considerado.

Q	M	N	I			P			B			Total	
			SNR	mean	std dev	SNR	mean	std dev	SNR	mean	std dev	mean	std dev
6	1	1	38.9	77476.6	17738.9							77476.6	17738.9
9	1	1	37.5	57396.1	12491.3							57396.1	12491.3
20	1	1	34.7	35120.8	5945.73							35120.8	5945.73
6	2	4	38.9	77490.1	17897.6	38.2	47484.7	17864.2	37.5	30520.2	12431.6	46507.6	24565.8
9	2	4	37.5	57398.3	12613	36.6	28071.6	12117.6	36.3	17989.1	7159.96	30365	19024.2
20	2	4	34.7	35124.8	6023.64	33.6	12993.2	6348.16	33.9	9970.08	3103.05	17016.2	11599.6
6	4	4	38.9	77490	17897.6				37.4	34225.2	14581.8	45049.1	24279.5
9	4	4	37.5	57398.4	12613				36.2	20366.7	8863.45	29631.2	18849.7
20	4	4	34.7	35125	6023.66				33.9	11158.3	4015.41	17154.3	11343.9
6	1	6	38.9	77391	17828.2	37.9	43622.7	17422.9				49250.8	21516.8
9	1	6	37.5	57349.1	12563.1	36.2	25014.6	11245.8				30403.6	16619
20	1	6	34.7	35106	6028.98	33.3	11038.1	5455.29				15049.4	10538.2
6	2	8	38.9	77542.8	18050.7	38.0	48828.7	18527.5	37.4	31870.7	12709.7	43941.9	21710.2
9	2	8	37.5	57430.5	12750.3	36.4	29195.2	12612.5	36.1	18845.1	7327.36	27551.7	16013.7
6	4	8	38.9	77542.6	18050.8	38.2	54064.2	19927.1	37.4	35435.9	14671.4	43033.2	21382.1
9	4	8	37.5	57430.6	12750.2	36.6	33510.9	14144.4	36.1	21222.6	8918.53	27288.9	15813.2
6	4	12	38.9	77560.4	18107.6	38.1	54690.8	19984.3	37.4	36212.7	14762.1	42742.7	20299.8
9	4	12	37.5	57455.4	12780.1	36.5	34076.4	14048.6	36.0	21754	8957.5	26786.4	14557.9
20	4	12	34.7	35150	6186.49	33.4	16643.7	7743.26	33.5	11760.4	4051.95	14525.4	8191.58
6	8	16	38.8	77738.2	18770.4	38.2	59834.1	20455.4	37.3	41008.6	16898.1	44486.6	19753.7
9	8	16	37.5	57562.5	13297.9	36.6	38485.4	14747	35.9	25036.6	10696.6	27914.8	13887.5

Tabla 2.4.1. Análisis de la media y desviación típica de la codificación MPEG-I para diferentes ternas (Q, M, N)

En las figuras 2.4.1, 2.4.2 y 2.4.3 se muestran las funciones de autocovarianza para las variables aleatorias asociadas a los bits por imagen I, P o B, considerando un conjunto representativo de valores de las ternas (Q, M, N).

En la figura 2.4.1 se puede observar que el decaimiento de la autocovarianza es ajustable por una función exponencial. El parámetro N indica en este caso el valor de submuestreo respecto a la curva con N=1. Esto se manifiesta en el factor de decaimiento exponencial, ya que, aumenta proporcionalmente con el valor de N. Las curvas de autocorrelación de I son muy insensibles a la variación del paso de cuantificación como muestran las ternas (Q=6, M=4, N=8) y (Q=9, M=4, N=8).

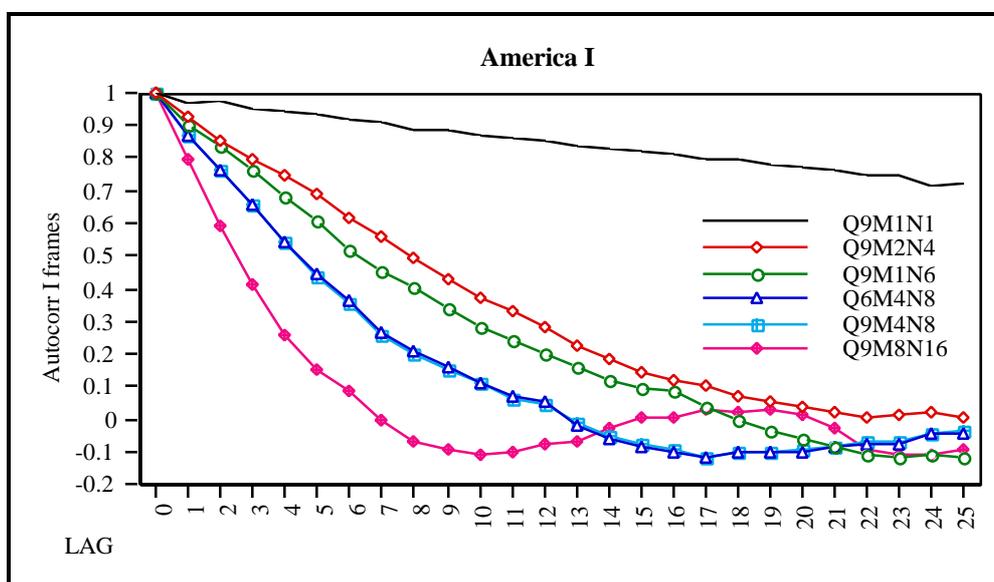


Figura 2.4.1. Índice de autocorrelación del número de bits por cuadro para el modo de codificación intra, en función de los parámetros de codificación (Q, M, N)

Las curvas de autocovarianza de las P no muestran un ajuste exponencial en sus primeros valores, aunque, su decaimiento se puede aproximar a partir de un decalado de 3 imágenes. En este caso, el factor de decaimiento viene relacionado con el incremento de M y no se manifiesta una notable dependencia con el número de imágenes del GoP. A su vez, los cuadros P mantienen su insensibilidad respecto al paso de cuantificación.

La autocovarianza de la tasa binaria de las imágenes B tampoco presenta un decaimiento exponencial en sus primeros valores. Para valores superiores a 6, la autocovarianza de las B se puede aproximar exponencialmente. Esta autocovarianza es más sensible al parámetro Q, aunque, tampoco tiene una significativa dependencia con su valor. Cuanto menor es el valor de M el decaimiento de las curvas es más rápido. El factor de decaimiento presenta la misma dependencia con el parámetro M que el de los cuadros P. Cuando M toma valores superiores a 4 es perceptible un rizado en la serie

temporal de las B, el cual queda reflejado en la función de autocovarianza en oscilaciones de período M-1 cuadros. Este rizado se debe a la ineficiencia de la codificación predictiva cuando los cuadros de referencia están muy alejados de las imágenes B. Estos incrementos periódicos de la tasa binaria reducen sustancialmente el rango de compresión y a su vez incrementan la distorsión, lo cual hace inapropiado el empleo de valores de M superiores a 4.

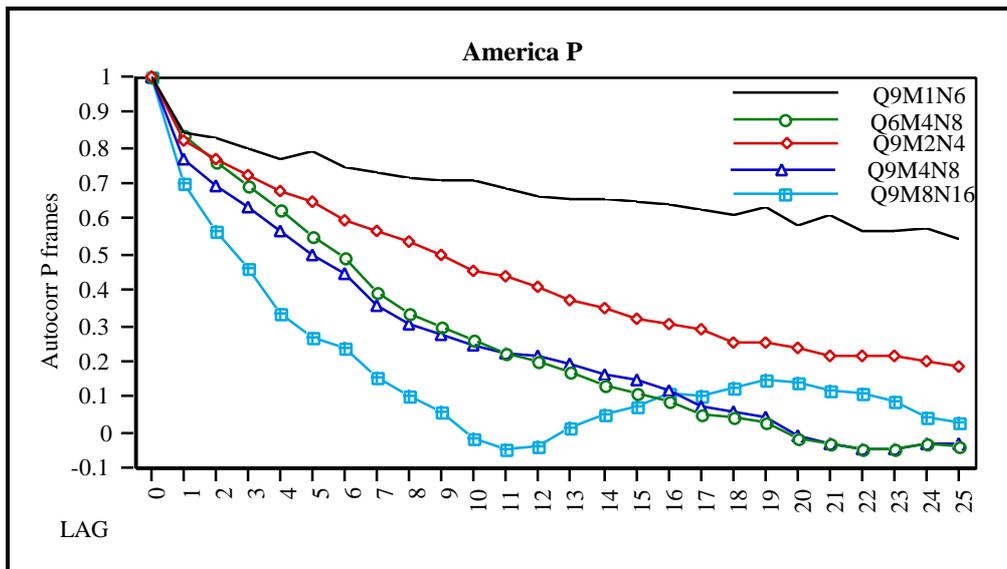


Figura 2.4.2. Índice de autocorrelación del número de bits por cuadro para el modo de codificación predictivo, en función de los parámetros de codificación (Q, M, N)

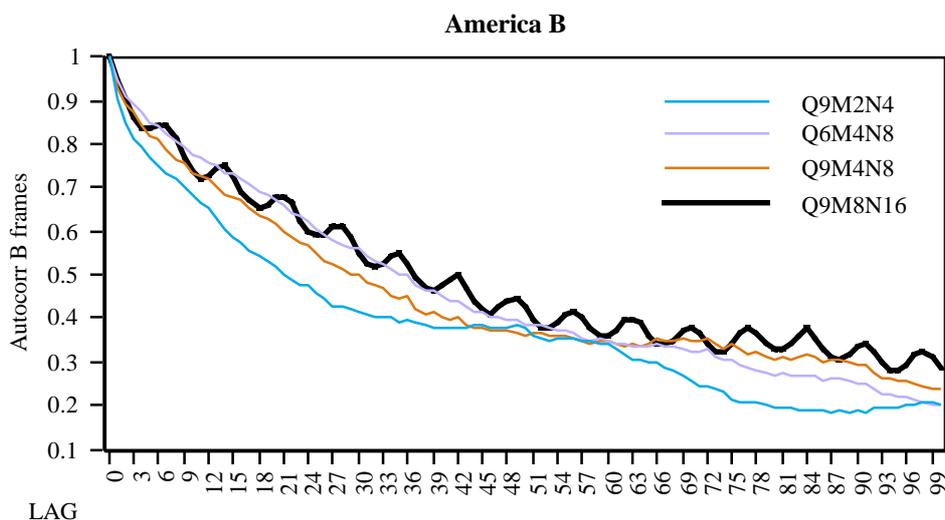


Figura 2.4.3. Índice de autocorrelación del número de bits por cuadro para el modo de codificación predictivo bidireccional, en función de los parámetros de codificación (Q, M, N)

Las funciones de distribución de probabilidad de las variables aleatorias I, P y B se han estudiado para diferentes valores de los parámetros de codificación. Todos los histogramas obtenidos presentan las mismas características. Dependiendo del valor del paso de cuantificación están centrados sobre los valores medios mostrados en la tabla 2.4.1. En todos los casos, los histogramas I, P y B presentan forma de campana cuya anchura se mantiene bastante constante. En las figuras 2.4.4, 2.4.5, 2.4.6 y 2.4.7 se presentan los histogramas correspondientes a la codificación de la secuencia "Live in Central Park" para la terna ($Q = 9, M = 2, N = 4$).

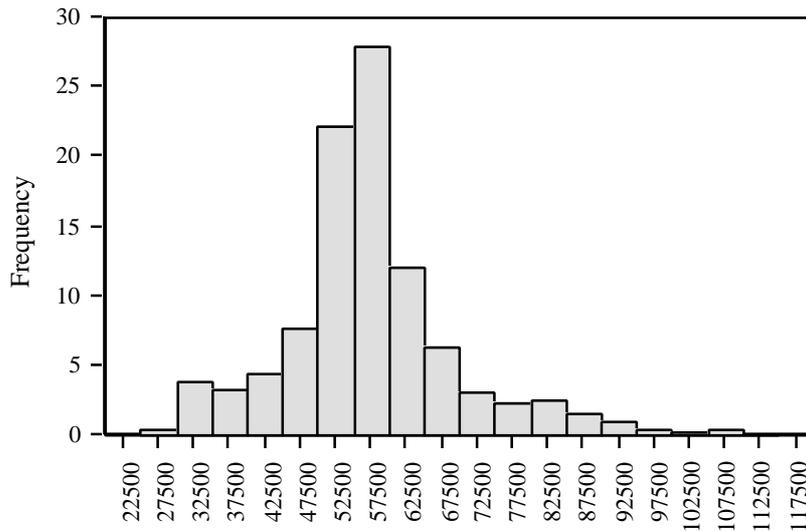


Figura 2.4.4. Histograma del modo I para la codificación ($Q=9, M=2, N=4$) de la secuencia "Live in Central Park"

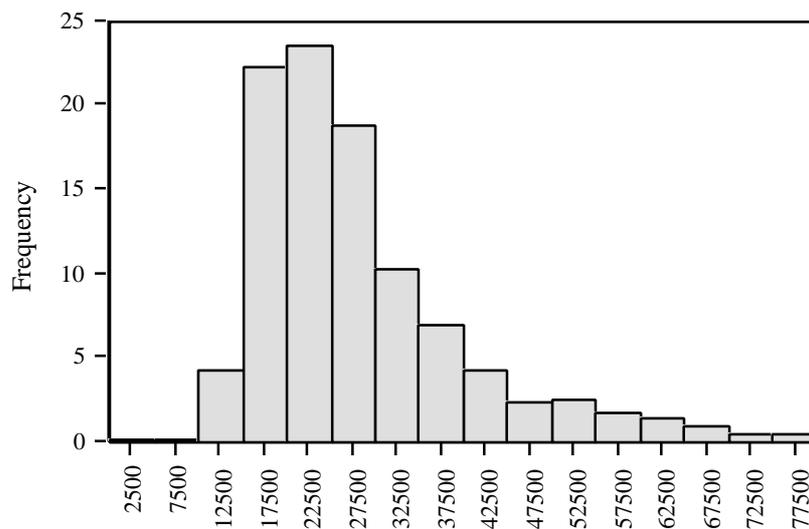


Figura 2.4.5. Histograma del modo P para la codificación ($Q=9, M=2, N=4$) de la secuencia "Live in Central Park"

El histograma de los cuadros I se relaciona con la probabilidad de la complejidad de las imágenes de la secuencia. Para esta secuencia, compuesta por escenas variadas, se observa que la función de distribución de probabilidad se corresponde aproximadamente con una distribución binomial. Los histogramas de las series temporales de los modos P y B se ajustan mejor a una distribución gamma, puesto que disponen de una larga cola de valores altos. El histograma conjunto se presenta en la figura 2.4.7 mostrando dos máximos relativos, uno correspondiente a la agregación de las variables aleatorias P y B, y el otro a la variable aleatoria asociada a la generación de los cuadros I. Obviamente, la distribución conjunta no será ajustable con una distribución de probabilidad simple.

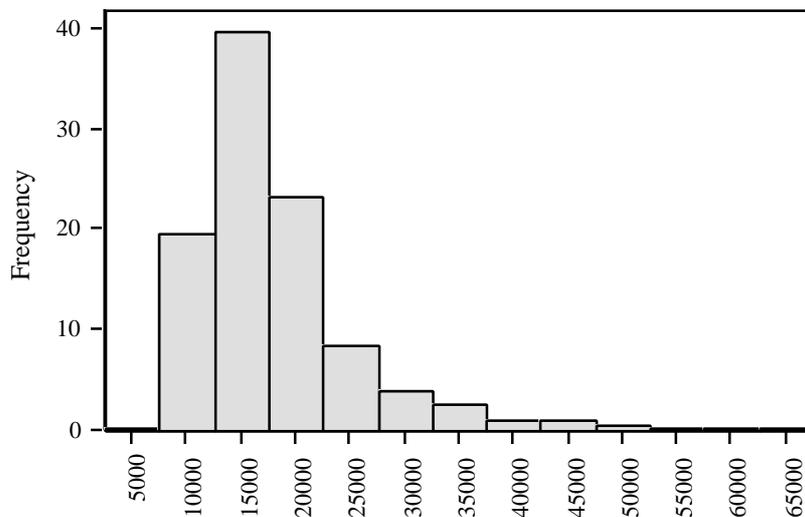


Figura 2.4.6. Histograma del modo B para la codificación ($Q=9, M=2, N=4$) de la secuencia "Live in Central Park"

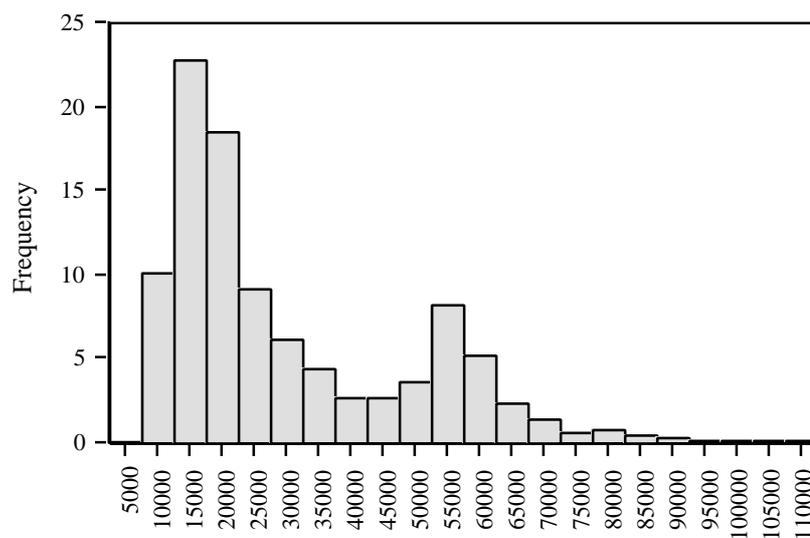


Figura 2.4.7. Histograma de la tasa binaria para la codificación ($Q=9, M=2, N=4$) de la secuencia "Live in Central Park"

2.5. Estudio de los modos de funcionamiento del algoritmo MPEG-I

Los codificadores de vídeo admiten dos modos de funcionamiento: el modo CBR (Constant Bit Rate) y el modo VBR (Variable Bit Rate). En modo de funcionamiento CBR regula el proceso de codificación para obtener una tasa de generación constante en el enlace de entrada a la red. Para ello, se debe intercalar un buffer entre el flujo de bits generado en el proceso de codificación y el canal de transmisión. El propósito del buffer es absorber los cambios intrínsecos de la tasa binaria generada en el proceso de codificación y monitorizar el nivel de actividad de la escena, según el cual, se ejerce una regulación sobre la tasa media generada, lo cual incide directamente en la calidad de la imagen. En la especificación del algoritmo de codificación MPEG-I se ilustra la gestión del buffer y los parámetros de regulación del algoritmo que permiten controlar su nivel de ocupación.

Los codificadores se pueden configurar también en el modo de funcionamiento de velocidad variable. La ventaja que ofrece el modo VBR respecto al modo CBR es que permite mantener una calidad objetiva o subjetiva semiconstante. En modo VBR, se inyecta directamente el flujo de bits generados en la codificación sobre el canal de comunicaciones. La tasa generada en modo VBR variará, dependiendo de la complejidad y actividad de la escena. Las fluctuaciones de la tasa binaria deberán ser soportadas por el sistema de comunicación empleado.

En este trabajo se han desarrollado cinco tipos de codificación de tasa variable para ser posteriormente analizados y comparados junto con el modo de funcionamiento CBR. Estos tipos de funcionamiento se han denominado:

i) Codificación VBR Q constante (Q_Fix). Este esquema de funcionamiento mantiene fijo el paso de cuantificación en la codificación, de forma que, se obtenga una calidad subjetiva semiconstante a lo largo de la codificación.

ii) Codificación VBR Q constante con distribución intra (Dist_I). Es una técnica de funcionamiento cuya operación se basa en el modo Q_Fix pero realizando una distribución de la codificación intra de un GoP en los macrobloques de los cuadros P del GoP. Como ventaja presenta una reducción de la tasa binaria máxima entregada a la red y la facilidad del suavizado del tráfico generado por el codificador.

iii) Codificación VBR con objetivo PSNR y control por tira (SL_PSNR). Los esquemas de funcionamiento con objetivo PSNR realizan una corrección del paso de cuantificación en una unidad estructural de la codificación a fin de mantener el valor de la PSNR en un nivel fijado a priori. En este caso se realiza un ajuste del paso de cuantificación en cada tira a partir de la experiencia obtenida en la codificación de las tiras precedentes del mismo cuadro.

iv) Codificación VBR con objetivo PSNR y control por macrobloque (MB_PSNR). Este esquema de funcionamiento con objetivo PSNR ejerce la corrección del paso de cuantificación sobre los macrobloques del cuadro a partir del resultado obtenido en macrobloques situados en la misma posición espacial de cuadros precedentes.

v) Codificación VBR con objetivo PSNR y control híbrido espacio-temporal por macrobloque (MB_ST_PSNR). Para conjugar las ventajas de la regulación espacial y temporal se ha propuesto un mecanismo de funcionamiento híbrido entre el SL_PSNR y el MB_PSNR. Este modo aplica una corrección del paso de cuantificación sobre los macrobloques del cuadro en función de los valores obtenidos de PSNR en el macrobloque de igual posición espacial del cuadro precedente y de los macrobloques codificados anteriormente del mismo cuadro con posiciones adyacentes.

2.5.1 Modo de funcionamiento CBR

La ventaja de este modo de funcionamiento del codificador es la simplicidad presentada en la ubicación de recursos requeridos en el sistema de transmisión, puesto que, el ancho de banda necesario es constante. Obviamente, este es el único modo de funcionamiento que se adecúa para la transmisión de vídeo en redes de conmutación de circuitos, como la RDSI, donde el ancho de banda asignado para los servicios con requerimientos temporales es fijo. El modo de funcionamiento CBR presenta como desventajas: la introducción de un retardo adicional en la transmisión, debido al tiempo de almacenamiento en el buffer y la alta variabilidad de la calidad de la imagen decodificada cuando la secuencia presenta diferentes niveles de complejidad y actividad. En el presente estudio se ha empleado un mecanismo de regulación de tasa generada semejante al presentado en [Wan94], aunque, éste se proponía para un codificador DPCM/DCT. En este caso, el control se ha ejercido sobre el paso de cuantificación de las tiras horizontales de las imágenes, en función del nivel de ocupación del buffer, fijándose los parámetros de codificación en $M=2$ y $N=6$.

La estrategia utilizada se basa en considerar el número de bits empleados en la codificación del anterior GoP para ajustar los que se deben generar en el actual GoP. Los valores correspondientes para las imágenes I, P y B se denominan valores objetivos de codificación de la tira. Para determinar el paso de cuantificación a emplear en cada tira se habilita un contador que se interpreta como buffer virtual cuya capacidad, o valor máximo que puede alcanzar, es igual a la capacidad del buffer insertado en el codificador. Este buffer virtual se intenta mantener a $1/4$ de su capacidad. El buffer virtual se incrementa o decrementa después de cada tira codificada en la cantidad diferencial entre los bits generados y el objetivo fijado. Dependiendo del valor del contador, se modifica el paso de cuantificación de la siguiente tira de la imagen.

Para mantener el buffer virtual a un 1/4 de su capacidad, la regulación se ha diseñado de forma que el paso de cuantificación se corresponda con el valor esperado de la tasa generada en la codificación. Para ello, se han utilizado los resultados de la sección anterior que relacionaban la tasa media binaria con el paso de cuantificación. Así, la relación obtenida entre la tasa media binaria (v) y el paso de cuantificación (Q) queda expresada por la ecuación:

$$v = \frac{4.2 \cdot 10^6}{Q^{0.774}} \quad (\text{bps}) \quad (2.5.1)$$

aproximada por la relación:

$$v \approx \frac{4.2 \cdot 10^6}{Q} \quad (\text{bps}) \quad (2.5.2)$$

finalmente, para corregir esta aproximación se ha ajustado la relación, obteniendo:

$$Q \approx \frac{6 \cdot 10^6}{v} \quad (2.5.3)$$

Se puede comprobar la bondad del ajuste en la tabla 2.5.1

v (Kbps)	Q fijado	Q estimado
2034	4	3.95
1114	6	6.38
709	9	9.46
385	20	16.58

Tabla 2.5.1. Bondad del ajuste de la aproximación del paso de cuantificación como función de la tasa binaria de generación

Nótese también que, la variabilidad del paso de cuantificación dependerá del tamaño del buffer empleado, dado que los valores que puede tomar el cuantificador dependen del tanto por ciento de ocupación del buffer. Si denominamos B_{virt} al nivel de ocupación del buffer virtual y B_{max} al tamaño del buffer, la relación entre el valor de B_{virt} y el de Q queda expresada como:

$$Q = 1 + \left[\frac{4 B_{\text{virt}}}{B_{\text{max}}} Q_{\text{priori}} \right] \quad (2.5.4)$$

donde Q_{priori} , se obtiene de la expresión 2.5.3. Con esta relación se obliga a que el valor

máximo alcanzado por el cuantificador sea 4 veces el predeterminado, lo cual, es una sustancial reducción de la tasa binaria. A su vez, se consigue minimizar la variabilidad del cuantificador. Para iniciar el proceso, se han deducido unos valores objetivo t_I , t_P y t_B para la codificación de las tiras I, P y B respectivamente. El valor de t_I se ha obtenido heurísticamente y los valores de t_P y t_B se han prefijado a partir de la relación, recomendada por el estándar de codificación de vídeo MPEG_I, entre las tasas de codificación de los distintos modos de codificación. Estos valores iniciales quedan expresados como:

$$\begin{aligned}
 t_I &= 4000 \left(\frac{n^\circ \text{ bloques del cuadro}}{396} \right) + 700 \\
 t_P &= \frac{t_I}{3} \\
 t_B &= \frac{t_I}{5}
 \end{aligned}
 \tag{2.5.5}$$

Obsérvese que se ha tenido en cuenta la resolución de la imagen, y se han considerado para una secuencia de 25 cuadros por segundo.

2.5.2 Modo de funcionamiento VBR Q_Fix

Este es el mecanismo de funcionamiento de la codificación MPEG-I más simple. El parámetro Q se fija a un valor constante a lo largo de toda la secuencia, con lo que, la tasa variable generada por el codificador depende únicamente de la complejidad y actividad de las escenas. En otros trabajos [GilRos95], se ha denominado a este modo de funcionamiento codificación con calidad quasi-constante, ya que la valoración objetiva de la calidad, a través de la PSNR, sufre pequeñas fluctuaciones. En cuanto a los parámetros de codificación, N y M, se han fijado en N=6 y M=2, según las conclusiones obtenidas en la sección 2.3.

2.5.3 Modo de funcionamiento VBR Dist_I

El modo de funcionamiento Dist_I es prácticamente igual al modo Q_Fix, salvo que, la codificación intra se ha distribuido sobre las imágenes de un GoP. Este mecanismo de funcionamiento fue propuesto en el trabajo [KawChe93]. En este esquema no se fija el modo de codificación intra en ninguna de las imágenes del GoP, así, el cuadro que debería tratarse en modo intra se codifica en modo P. La distribución del modo intra se realiza sobre todas las imágenes P, de forma que, un número de macrobloques, alineados verticalmente, son forzados al modo de codificación I. La distribución vertical

se realiza de manera que en un GoP, todas las posibles posiciones de los macrobloques dentro de los cuadros P sean codificadas en modo intra al menos una vez. Obviamente, se intenta uniformizar el número de macrobloques forzados al modo intra en las imágenes P.

Una ilustración de este mecanismo distribuido intra aparece en la figura 2.5.1. En la configuración de parámetros $N=6$ y $M=2$, se realiza una distribución vertical en los tres cuadros P de los macrobloques forzados intra. En el presente estudio, los cuadros tienen una resolución de 352×288 pels, por lo que, la distribución vertical intra se ha realizado en 8 macrobloques de cada tira para los primeros dos cuadros P del GoP y los 6 últimos macrobloques de cada tira de la última P. La distribución de la codificación intra dentro del GoP se realiza verticalmente con el fin de distribuir el incremento de la tasa binaria de la forma más uniforme posible dentro de la codificación del cuadro, o sea, en las tiras de éste. El efecto provocado por la localización de estos macrobloques intra, al encontrarse en posiciones horizontalmente consecutivas y en la misma situación vertical, no es perceptible en ningún momento en la decodificación, a pesar del posible incremento periódico de la calidad en la zona de la imagen correspondiente. Este efecto es inapreciable al igual que no lo era el efecto de refresco del cuadro intra, cuando N tomaba un valor reducido.

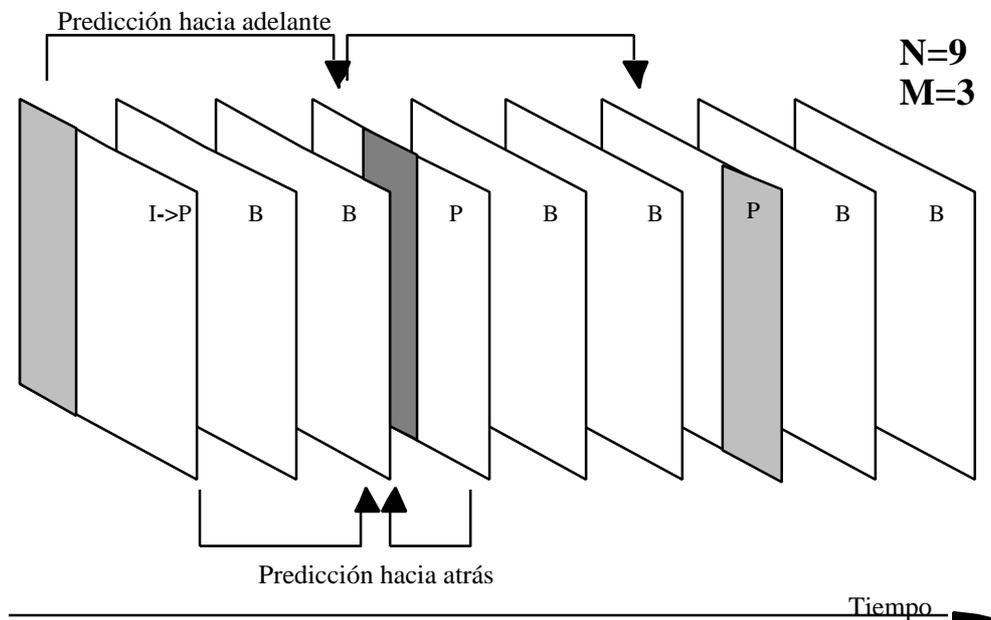


Figura 2.5.1. Esquema de codificación Intra distribuido para el modo de funcionamiento VBR del algoritmo MPEG-I

La desventaja que presenta la distribución intra es el aumento de la complejidad en la decodificación cuando se pretende hacer una visualización rápida de la secuencia, aunque, gracias a la actual velocidad de proceso de los circuitos especializados se podría realizar una decodificación selectiva de los macrobloques de las imágenes P y dada la redundancia temporal, recomponer imágenes intra con los macrobloques codificados en este modo dentro de las P del GoP. Los resultados obtenidos serían equivalentes a la decodificación selectiva de imágenes I del caso estándar. La ventaja de la codificación VBR Dist_I es la reducción de la variabilidad de la tasa binaria y la reducción de la tasa máxima generada, lo cual permite reducir la ubicación de recursos en redes que admiten fuentes de tráfico VBR. Además, esta reestructuración del GoP también puede ser empleada en el modo de funcionamiento CBR, aumentando la facilidad de la regulación de la tasa binaria. Esta estrategia de codificación, aplicada sobre el modo de funcionamiento CBR, también tendría incidencia en la reducción del retardo de almacenamiento, puesto que podría reducir el tamaño del buffer.

2.5.4 Modo de funcionamiento VBR SL_PSNR

En los modos con objetivo PSNR se pretende mantener constante la variación de la calidad objetiva a través del control de la codificación. La codificación VBR SL_PSNR realiza un control semejante al elegido en el modo de funcionamiento CBR. Esta regulación se basa en la alta correlación entre pixels correspondientes a tiras consecutivas de un cuadro. Se aprovecha la experiencia obtenida en las tiras anteriores para ajustar el paso de cuantificación de la tira a codificar. Así, el paso de cuantificación Q se modifica para cada tira de la imagen, dependiendo del valor PSNR obtenido en la anterior tira codificada. En particular, para la primera tira del cuadro, el ajuste se calcula en función del valor medio de la PSNR resultante de la codificación del último cuadro del mismo modo.

La curva de regulación mostrada en la figura 2.5.2 relaciona la corrección del paso de cuantificación y el error respecto al objetivo PSNR. Esta relación se ha obtenido heurísticamente a través de un conjunto de ensayos. La curva de regulación presenta una zona muerta, o zona donde el regulador no modifica el valor del paso de cuantificación, y dos zonas que relacionan la variación del paso de cuantificación y el error positivo o negativo. Obsérvese que, para errores negativos el decremento de Q es más suave que para valores positivos. Esta consideración está acorde con el comportamiento de la pendiente de la curva que relacionaba la PSNR y el Q , mostrada en la figura 2.3.6. En la figura 2.5.3 se muestra el diagrama de flujo de la codificación SL_PSNR. Obsérvese que la función f se deriva de la curva de regulación ilustrada en la figura 2.5.2.

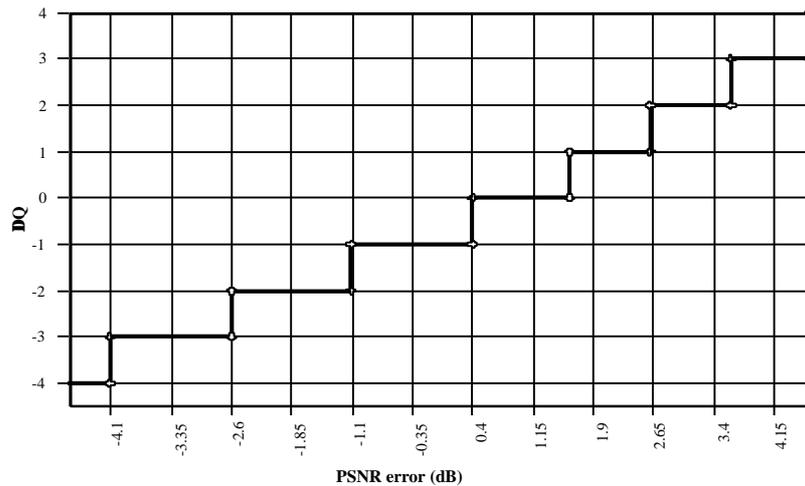


Figura 2.5.2. Curva de regulación aplicada a los modos de funcionamiento VBR con objetivo PSNR

2.5.5 Modo de funcionamiento VBR MB_PSNR

En este caso también se busca una codificación que mantenga la calidad objetiva a través del valor del PSNR. Para la regulación se ha empleado la variación del paso de cuantificación sobre la unidad mínima controlable, es decir, el macrobloque. A diferencia del SL_PSNR este mecanismo de regulación se basa en la redundancia temporal y no en la espacial. Se considera que los macrobloques con la misma posición espacial, en diferentes cuadros próximos, tendrán una alta similitud, y por tanto, a partir de la experiencia de la codificación de un macrobloque situado en la misma posición espacial en una imagen previa, se puede adecuar la cuantización del actual macrobloque, a fin de ajustar el PSNR a un valor próximo al objetivo.

Para llevar a cabo la codificación MB_PSNR es necesario estructurar las imágenes de forma matricial teniendo como elemento el macrobloque. En este caso se ha distinguido entre imágenes I, P y B. Para los modos de codificación I y P, separadamente, se mantiene un control del paso de cuantificación para cada macrobloque de la imagen. El paso de cuantificación a aplicar se determina en base al valor previo empleado y a la PSNR obtenida anteriormente. La curva de actuación empleada es la misma que se utiliza para el modo SL_PSNR cuya ilustración aparece en la figura 2.5.2.

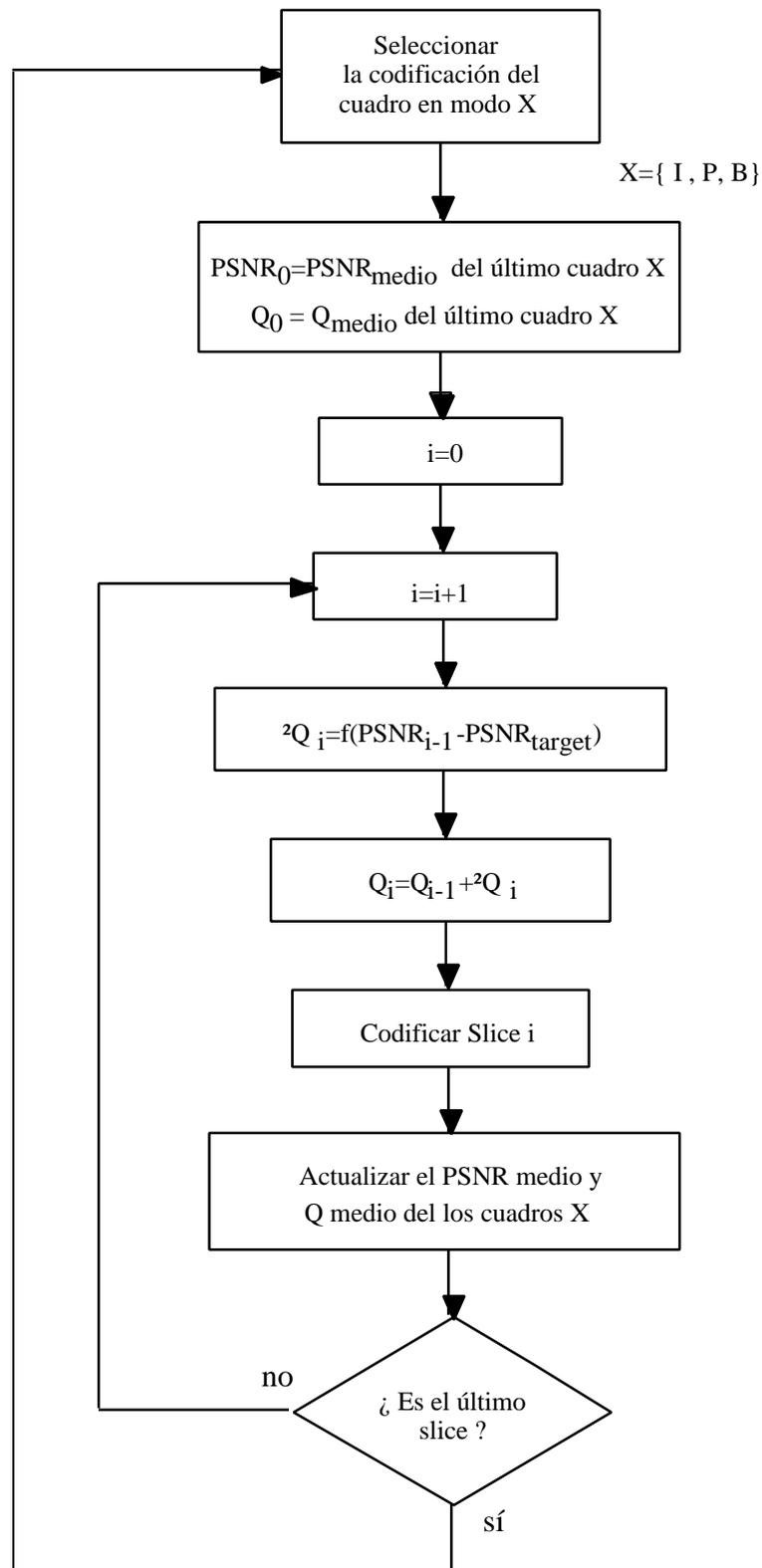


Figura 2.5.3. Diagrama de flujo del algoritmo del modo de funcionamiento SL_PSNR

A diferencia de las imágenes I o P, las imágenes B utilizan como referencia imágenes I o P pasadas y futuras para su codificación. Esto implica que no existe una dependencia directa entre la codificación de imágenes consecutivas B. Aunque se estudió el empleo de la regulación en las imágenes B, de igual forma que para las I y P, los resultados revelaron que, por el propio mecanismo de codificación predictivo bidireccional, el macrobloque de referencia empleado en la codificación de los macrobloques de las imágenes B, de la misma posición, variaba significativamente dependiendo de la posición de la imagen B en el SGoP. Esto conducía, en un elevado número de ocasiones, a una regulación inadecuada de los macrobloques B. Se observó que lo más apropiado para la codificación de los macrobloques B era emplear un valor de cuantificación relacionado con la imagen de referencia predecesora o una semisuma de los valores empleados en las dos referencias posibles. Además, sobre este valor de cuantificación deducido se comprobó que era conveniente reducir el paso de cuantificación, porque, en general, el valor de la PSNR obtenido en los macrobloques B era siempre algo inferior al objetivo. Con estas consideraciones, el valor de cuantificación ajustado en los macrobloques B se obtiene de la relación 2.5.6, donde se han acotado los valores máximos y mínimos a 4 y 20 respectivamente.

$$Q_B = \min \left(\max \left(4, \left\lfloor \frac{Q_{ref.1} + Q_{ref.2}}{2} \right\rfloor - 4 \right), 20 \right) \quad (2.5.6)$$

En la figura 2.5.4 se muestra el diagrama de flujo de la codificación MB_PSNR, donde, las funciones f y g señaladas se corresponden con la aplicación de la curva de regulación y la expresión 2.5.6 respectivamente. Nótese que:

- i) La matriz $Q1(X,i)$ se ha empleado para el almacenamiento del paso de cuantificación del macrobloque i-ésimo del modo X, donde X puede ser I o P.
- ii) La matriz $Q2(a,i)$ contendrá los valores de los pasos de codificación empleados en las imágenes de referencia en las que se apoyarán los cuadros B del SGoP.
- iii) Los valores de la PSNR resultantes de la codificación de los macrobloques de los modos X se guardan en la matriz $PSNR(X,i)$. X podrá tomar valores I o P.

La ventaja que aporta este mecanismo de codificación MB_PSNR sobre el SL_PSNR es la uniformidad del PSNR en toda la imagen. En el MB_PSNR se intentan regular todos los elementos controlables de la imagen para mantener su PSNR mientras que el control por tira provoca que puedan existir macrobloques de la tira con valores notablemente distintos de PSNR. Por el contrario, el control SL_PSNR tiene una respuesta mucho más rápida cuando aparecen cambios de escena ya que el reajuste o aprendizaje se ejerce en el mismo cuadro.

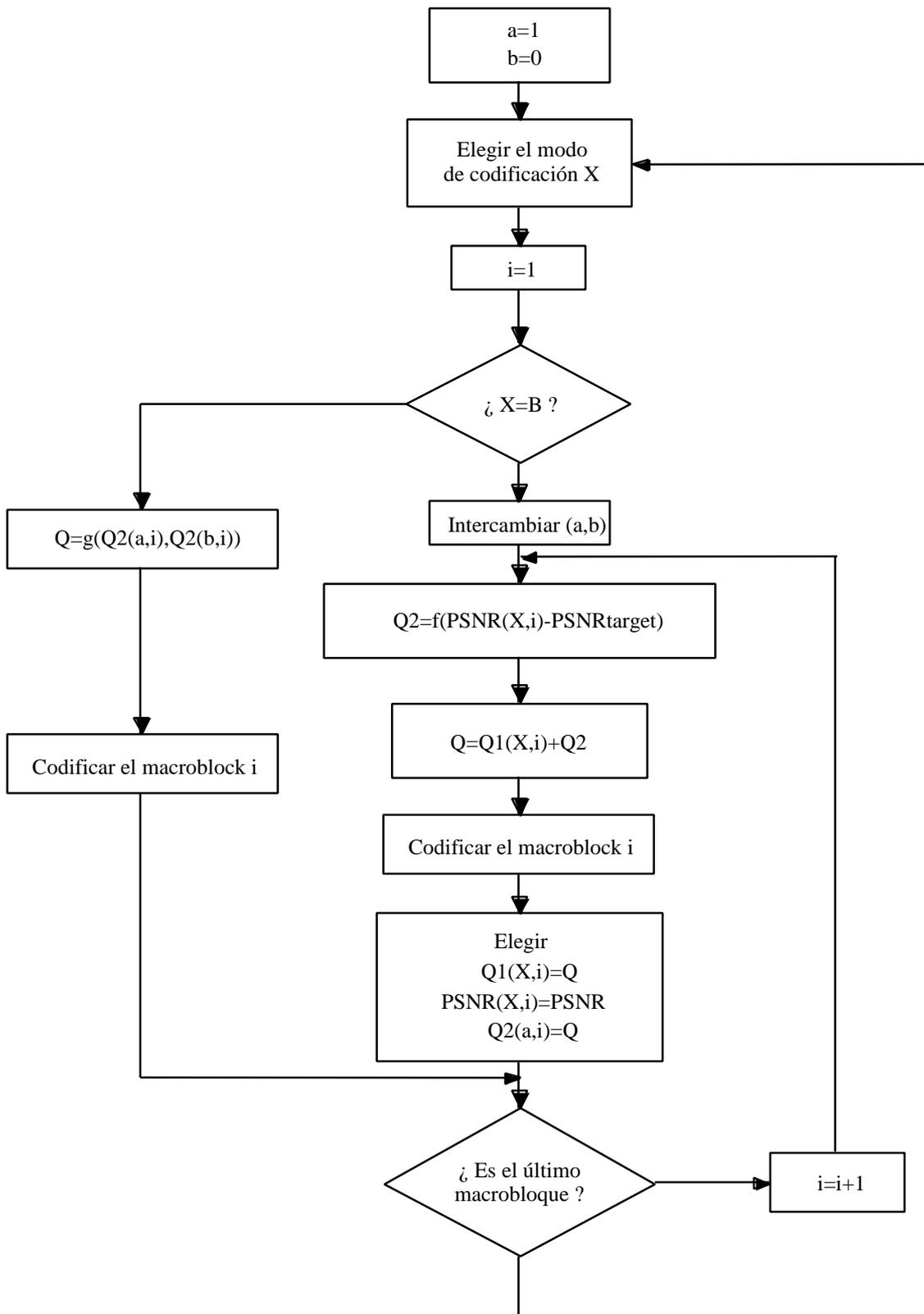


Figura 2.5.4. Diagrama de flujo del algoritmo del modo de funcionamiento MB_PSNR

2.5.6 Modo de funcionamiento VBR MB_ST_PSNR

El modo de funcionamiento VBR con objetivo PSNR controlado a nivel de macrobloque y con regulación basada en el conocimiento espacio-temporal (MB_ST_PSNR), emplea la misma regulación temporal que el MB_PSNR. Para aumentar el tiempo de respuesta se ha considerado a su vez el resultado de la codificación de los macrobloques adyacentes del mismo cuadro. Así, el valor de Q utilizado en la codificación del macrobloque se obtiene a partir del calculado para el modo MB_PSNR (Q_{MB-SNR}) y del ajustado para cada macrobloque colindante. La corrección espacial se traduce finalmente en un valor (Q_{MB-SP}) derivado de la ponderación de los valores corregidos de los macrobloques adyacentes. Luego, el valor de la codificación del macrobloque en el modo de funcionamiento MB_ST_PSNR se puede expresar como:

$$Q_{MB_ST_SNR} = (1 - \alpha)Q_{MB-SNR} + \alpha Q_{MB-SP} \quad (2.5.7)$$

donde $\alpha \in [0,1]$. El valor de α determina la ponderación de la corrección espacial y de la corrección temporal. Obviamente, si $\alpha=0$ la codificación es la realizada en el modo MB_PSNR. Para calcular el valor del término Q_{MB-SP} se debe tener en cuenta la posición espacial del macrobloque. Cabe distinguir tres tipos de situaciones espaciales como se ilustra en la figura 2.5.5.

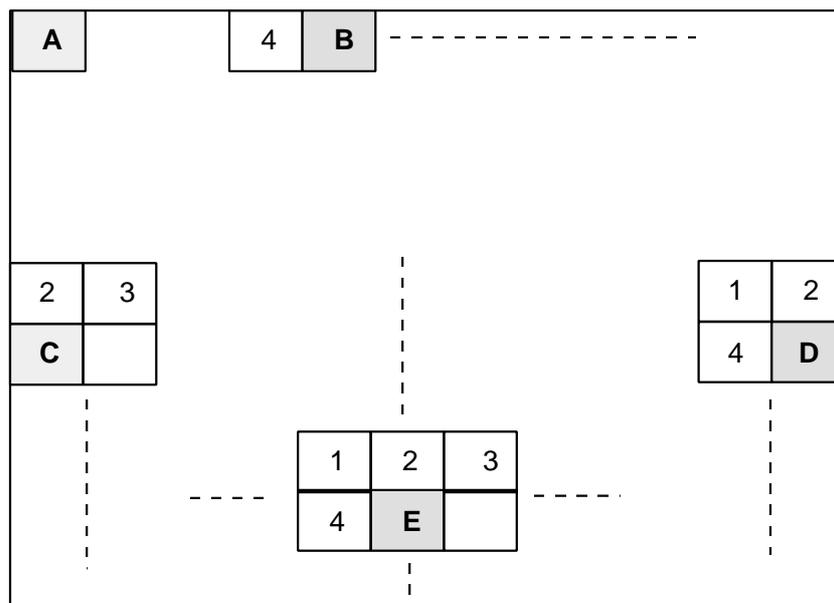


Figura 2.5.5. Posiciones de los macrobloques y sus adyacentes previamente codificados

Se puede observar que como máximo un macrobloque podrá emplear el resultado de la codificación de hasta cuatro macrobloques adyacentes previamente procesados. Como caso particular, el primer macrobloque del cuadro sólo podrá emplear el valor derivado del ajuste temporal, dado que no se dispone de ninguna experiencia espacial. Considerando un caso genérico, podemos expresar:

$$Q_{MB-SP} = \alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3 + \alpha_4 Q_4 \quad (2.5.8)$$

donde los Q_i , con $i \in [1,4]$, son los valores resultantes del ajuste del paso de cuantificación utilizado en la codificación de los macrobloques adyacentes. Asimismo, se ha ponderado cada uno de estos valores en función de la posición del macrobloque a codificar. En la tabla 2.5.2 se muestran los valores de α y α_i para cada uno de los cinco casos.

Posición del Macrobloque	a	a₁	a₂	a₃	a₄
A	0	-	-	-	-
B	0.4	0	0	0	1
C	0.4	-	0.6	0.4	-
D	0.4	0.2	0.4	-	0.4
E	0.4	0.2	0.3	0.3	0.2

Tabla 2.5.2. Factores de ponderación de la codificación MB_ST_PSNR

2.5.7 Comparación de los modos de funcionamiento de la codificación MPEG-I

Con la aparición de las redes ATM se han intensificado los estudios sobre las ventajas que ofrecía la transmisión VBR frente a la CBR y las dificultades que surgen en la gestión y ubicación de una tasa binaria variable [COST211] [TanOku92]. En este trabajo se realiza un análisis comparativo de los mecanismos de funcionamiento del MPEG-I, propuestos en la sección anterior, a partir de una secuencia de más de 25 minutos de duración extraída del último tramo de la película "Jurasic Park". Esta secuencia se caracteriza por disponer de escenas muy variadas, resultando muy adecuada para el análisis de los diferentes modos de funcionamiento.

Los tres parámetros estudiados en el análisis han sido:

- i) El factor de cuantificación (Q).
- ii) La relación señal a ruido (PSNR).
- iii) La tasa binaria.

Los resultados se han presentado gráficamente, mostrando los valores medios resultantes de la tasa binaria, el paso de cuantificación y la PSNR. Los valores medios obtenidos se han calculado, o bien por cuadro, o bien por GoP, en función del nivel de detalle que se pretendiera analizar. La presentación por GoP se ha adoptado cuando el interés del estudio se centraba más en la evolución temporal de la codificación de diferentes tipos de escena, mientras que la elección del análisis por cuadro se ha efectuado cuando se pretendía observar el comportamiento de las codificaciones por tipos de escena o en las transiciones entre éstas.

En primer lugar se presenta en la figura 2.5.6 una serie temporal del PSNR y en la figura 2.5.7 una serie temporal de la tasa binaria de los modos CBR, Dist_I y Q_Fix. Cabe resaltar que la tasa binaria mostrada para el modo de funcionamiento CBR no es fija ya que se ilustra la tasa binaria media generada en el GoP lo cual aporta información sobre la intervención que realiza el mecanismo de regulación. En estas gráficas se puede observar que los modos de funcionamiento Q_Fix y Dist_I presentan prácticamente los mismos valores de tasa binaria y PSNR para todos los GoP de la secuencia. Es decir, el comportamiento es equivalente en ambos modos de funcionamiento. Como contraste, el modo de funcionamiento CBR mantiene la tasa binaria entre cotas muy próximas asegurando un nivel medio constante a corto plazo. Esta regulación provoca diferencias proporcionales de hasta 4 entre la tasa binaria Q_Fix y la tasa binaria CBR. La PSNR en el Q_Fix se mantiene centrada casi en el mismo valor que la del CBR pero sus valores permanecen mucho más acotados que en el CBR. Obsérvese que, cuando el modo Q_Fix emplea muchos más bits en los GoP que el modo CBR, éste reduce sustancialmente su PSNR. Por el contrario, cuando la tasa binaria del modo Q_Fix es inferior a la del modo CBR existe un aumento del PSNR respecto al PSNR del Q_Fix.

En las figuras 2.5.8 y 2.5.9 se ilustra el comportamiento de la tasa binaria y el PSNR para los modos de funcionamiento Q_Fix, SL_PSNR, MB_PSNR y MB_ST_PSNR. Se observa que, dependiendo de la actividad de la escena la tasa binaria, puede incrementarse proporcionalmente en un factor 2 en los modos SL_PSNR, MB_PSNR y MB_ST_PSNR respecto al modo Q_Fix. Los niveles alcanzados por el modo SL_PSNR llegan a ser superiores. Esto es debido a que la regulación del modo de funcionamiento SL_PSNR permite mantener mucho más constante la PSNR que en el modo MB_PSNR aunque este mantiene mucho más acotado los valores de PSNR que el Q_Fix. A través de las series temporales de PSNR se muestra que en los cambios de escena, o escenas con mucha actividad, el modo SL_PSNR reacciona mucho más rápido en la corrección del PSNR que el modo MB_PSNR, efecto que se corresponde con la propia naturaleza del mecanismo de regulación en cada caso. El modo MB_ST_PSNR se comporta de forma muy similar al MB_PSNR aunque se observa su mejora en el tiempo de respuesta, aproximándose al comportamiento del SL_PSNR.

Para realizar una comparación más exhaustiva de los modos de funcionamiento se

han analizado dos partes de esta secuencia correspondientes a los intervalos de GoP (800,1300) y (2000,2250). Las figuras 2.5.10 y 2.5.11 muestran los valores de tasa binaria, PSNR y Q para cada modo de funcionamiento en cada intervalo especificado. En las gráficas se aprecia con mayor detalle el mantenimiento del PSNR por parte de los modos SL_PSNR, MB_PSNR y MB_ST_PSNR, con una mayor respuesta temporal del primero. El coste de esta regulación de calidad objetiva es el incremento de la tasa binaria respecto a los otros modos de funcionamiento. Los modos de funcionamiento Q_Fix y Dist_I mantienen valores prácticamente idénticos ofreciendo tasas binarias más reducidas y la PSNR acotada en torno del valor central. El modo CBR presenta variaciones de calidad inaceptables en escenas de alta complejidad o movimiento.

Un análisis más pormenorizado de los modos de funcionamiento, observando cada modo de codificación, se presenta para las imágenes correspondientes a los intervalos de GoP (2020,2050) y (2140,2180). En las figuras 2.5.12 a 2.5.21 se muestran los resultados comparativos de la tasa binaria, el PSNR y el paso de cuantificación para cada modo de funcionamiento respecto al Q_Fix.

A través de las figuras 2.5.12 y 2.5.13 se observa que la codificación Q_Fix presenta mayor diferencia en los valores de PSNR obtenidos para cada modo de codificación que en la CBR. Queda patente que las imágenes B tienen un nivel de calidad inferior a las I y P y que la regulación en el modo CBR solventa parcialmente esta diferencia. Las variaciones del PSNR del modo de funcionamiento CBR son mucho más amplias que en el modo Q_Fix. La tasa binaria en el modo CBR se mantiene en valores muy próximos mientras que en Q_Fix varía según el grado de complejidad de las imágenes.

En las figuras 2.5.14 y 2.5.15 se puede observar que el comportamiento de los modos Q_Fix y Dist_I son prácticamente iguales. La ausencia de imágenes completamente codificadas en modo I en el caso Dist_I provoca que, en el mismo GoP, sólo existan dos tipos de tasas, unas correspondientes a las P, superiores a las P del Q_Fix dado que incorporan un alto número de macrobloques en modo intra, y las B que son prácticamente iguales a las del Q_Fix.

La evolución del modo de funcionamiento SL_PSNR se muestra en las figuras 2.5.16 y 2.5.17. En este caso se observa que el valor de PSNR es muy estable y la regulación de cambios de escena se produce en el tiempo de un GoP. A su vez, la variabilidad del PSNR en un GoP aumenta cuando la complejidad de la imágenes se incrementa. Se comprueba que la tasa binaria aumenta sustancialmente respecto al modo Q_Fix cuando la PSNR del Q_Fix se reduce unos pocos decibelios respecto al valor central. De esto se desprende que, el mecanismo SL_PSNR es excesivamente activo, dado que pequeños aumentos en la complejidad de la imagen incrementan sustancialmente la tasa binaria.

La comparación entre el modo de funcionamiento MB_PSNR y el Q_Fix puede ser observada en las figuras 2.5.18 y 2.5.19 los niveles de PSNR en las imágenes B del modo MB_PSNR respecto a los niveles de I son inferiores a los obtenidos en el modo

Q_Fix. Esto se debe a que el paso de cuantificación, cuando se hace un ajuste a nivel de macrobloque, debe ser bastante más fino que el de las referencias. Este efecto contribuye sustancialmente a la PSNR de las imágenes B. Como implicación observamos que la tasa binaria se reduce frente al modo Q_Fix en las imágenes B, a pesar del incremento de señalización que implica la variación continuada del paso de cuantificación en cada macrobloque. En cuanto a los cambios de escena del modo de funcionamiento MB_PSNR, observamos que tiene un tiempo de respuesta para la regulación del PSNR del orden de 4 GoPs.

Finalmente, el comportamiento del modo de funcionamiento MB_ST_PSNR se ilustra en las figuras 2.5.20 y 2.5.21. Este modo de funcionamiento presenta menos diferencia en la PSNR de los modos de codificación, respecto al Q_FIX, que el modo MB_PSNR. A su vez, también se reduce considerablemente el tiempo de respuesta en los cambios de escena. Como notable mejora respecto al modo SL_PSNR, se observa que no se introduce una corrección excesiva cuando aumenta moderadamente el nivel de complejidad de la escena, por lo que no se alcanzan niveles tan elevados de la tasa binaria.

En la tabla 2.5.4 se muestran los valores medios y la desviación estándar de la tasa binaria, PSNR y Q para los distintos modos de funcionamiento. Cabe resaltar que, aunque todos muestran un valor central de PSNR próximo, su variabilidad es muy diferente. A su vez, el decremento de la varianza del PSNR tiene un coste importante en la tasa binaria generada.

	Bit Rate (bits per frame)		PSNR (dB)		Quantizer Step	
	mean	std deviation	mean	std deviation	mean	std deviation
CBR	39832	18136	38.42	3.22	6.23	2.8
Q_Fix	39844	27955	38.65	2.2	6	0
Dist_I	40455	26458	38.56	2.1	6	0
SL_PSNR	64597	58971	38.78	0.91	8.1	6.5
MB_PSNR	51526	45679	38.46	1.82	10.4	6.27
MB_ST	59538	43855	38.72	1.55	10.58	5.5

Tabla 2.5.4. Resultados estadísticos de los parámetros del algoritmo MPEG-I en la codificación de la secuencia "Jurassic Park" para los distintos modos de funcionamiento del codificador de vídeo

Si bien el estudio cuantitativo de los modos de funcionamiento puede contribuir a la comprensión del comportamiento de la codificación en función de las técnicas regulación aplicadas, es importante, asimismo, el análisis subjetivo de las codificaciones resultantes. Se ha llevado a cabo un detenido examen subjetivo de las secuencias codificadas con los distintos modos de funcionamiento. Como resultado se ha observado que:

i) El tiempo necesario para la percepción de un cambio de escena es superior a 2 segundos.

ii) En escenas de gran actividad la percepción de la calidad es más reducida y, por tanto, los desniveles de la PSNR dentro de la imagen pasan desapercibidos.

iii) No existe una correspondencia directa entre nivel de PSNR y calidad subjetiva. Así, se pueden dar casos donde escenas con un valor medio superior de PSNR resulten menos nítidas que otras con menor PSNR. Asimismo, la cuantificación PSNR es más orientativa respecto al nivel subjetivo cuanto mayor sea la resolución de la imagen.

En el examen de las secuencias resultantes de los modos de funcionamiento propuestos se revela que:

i) El modo de funcionamiento CBR presenta notables diferencias en la calidad de la imagen, mientras que en el resto de modos de funcionamiento casi no son perceptibles. Además, no existe una correspondencia entre el nivel de PSNR y la calidad subjetiva de la secuencia.

ii) En ninguno de los modos de funcionamiento son detectables las diferencias de calidad introducidas por los modos de codificación. Esto se debe al carácter integrativo del sistema visual humano y a la adecuada elección de parámetros N y M planteada en este capítulo.

iii) El tiempo de respuesta en todos los modos de funcionamiento con objetivo PSNR es adecuado y, para escenas de gran actividad, puede resultar contraproducente que sea excesivamente rápido. Esto se traduce en regulaciones erróneas cuando los objetos se desplazan rápidamente, cambiando bruscamente su posición dentro de la imagen.

Como resultado del análisis subjetivo concluimos que se puede considerar válido el modo de funcionamiento Q_FIX para la obtención de una calidad subjetiva semiconstante. Aunque, para homogeneizar la calidad en toda la imagen sería más adecuado el empleo de una regulación temporal. En cualquiera de los dos casos sería aconsejable una regulación del paso de cuantificación. Teniendo en cuenta la percepción visual humana y los errores de corrección producidos en las escenas de gran actividad, la regulación debería tener una respuesta temporal muy suave. Asimismo, en la regulación temporal se debe asegurar que el paso de cuantificación aplicado sobre posiciones espaciales cercanas no debe presentar grandes diferencias. La regulación espacial sólo es adecuada para el ajuste de la tasa binaria generada, es decir para el modo de funcionamiento CBR. El modo DIST_I presenta buenas cualidades para la transmisión, control de tasa binaria y predicción de la tasa binaria por cuadro. Para algunas aplicaciones es más adecuado el modo DIST_I, puesto que, este modo proporciona los mismos resultados que el Q_FIX.

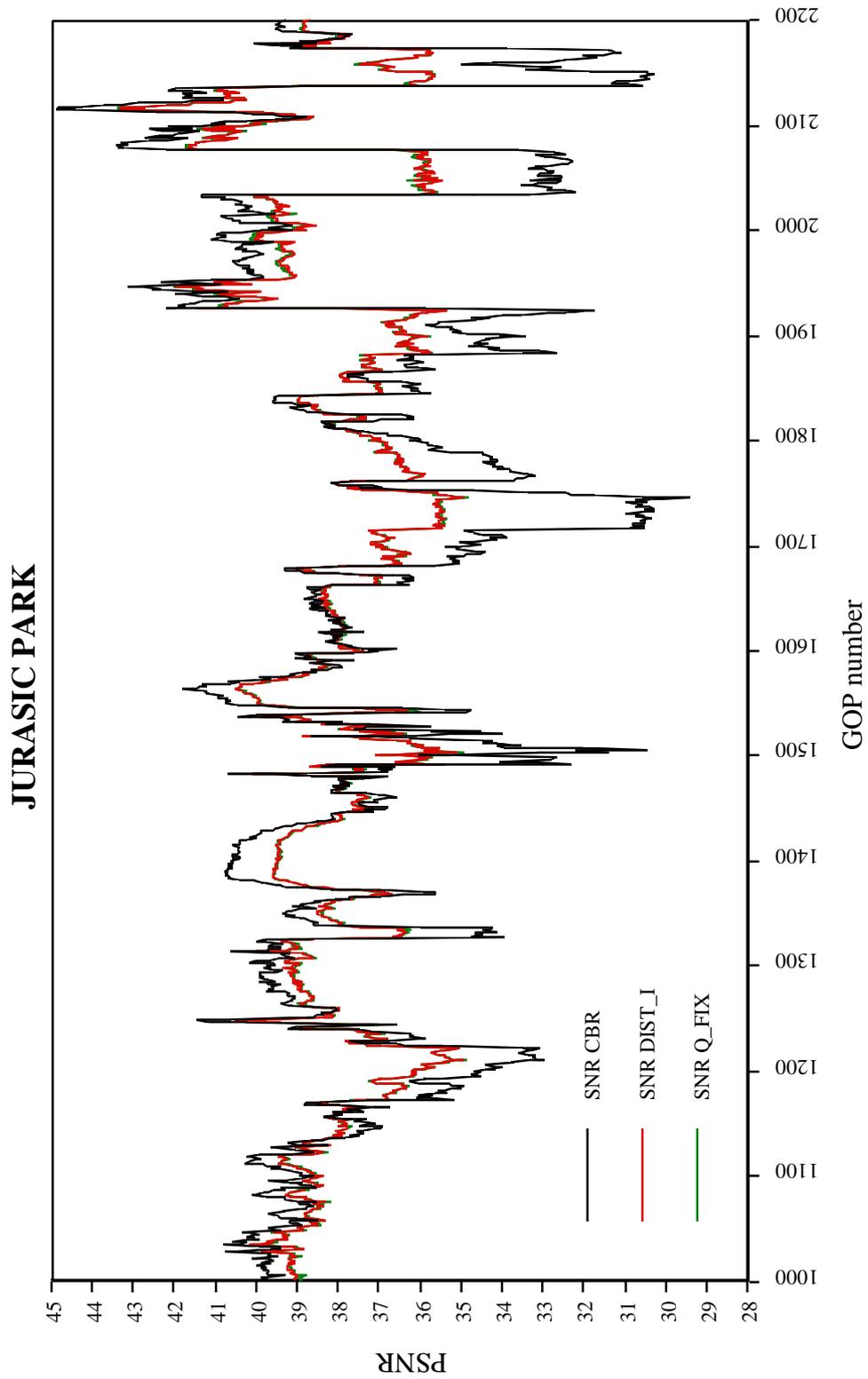


Figura 2.5.6. Comparación del PSNR de los modos de funcionamiento CBR, DIST_I y Q_FIX

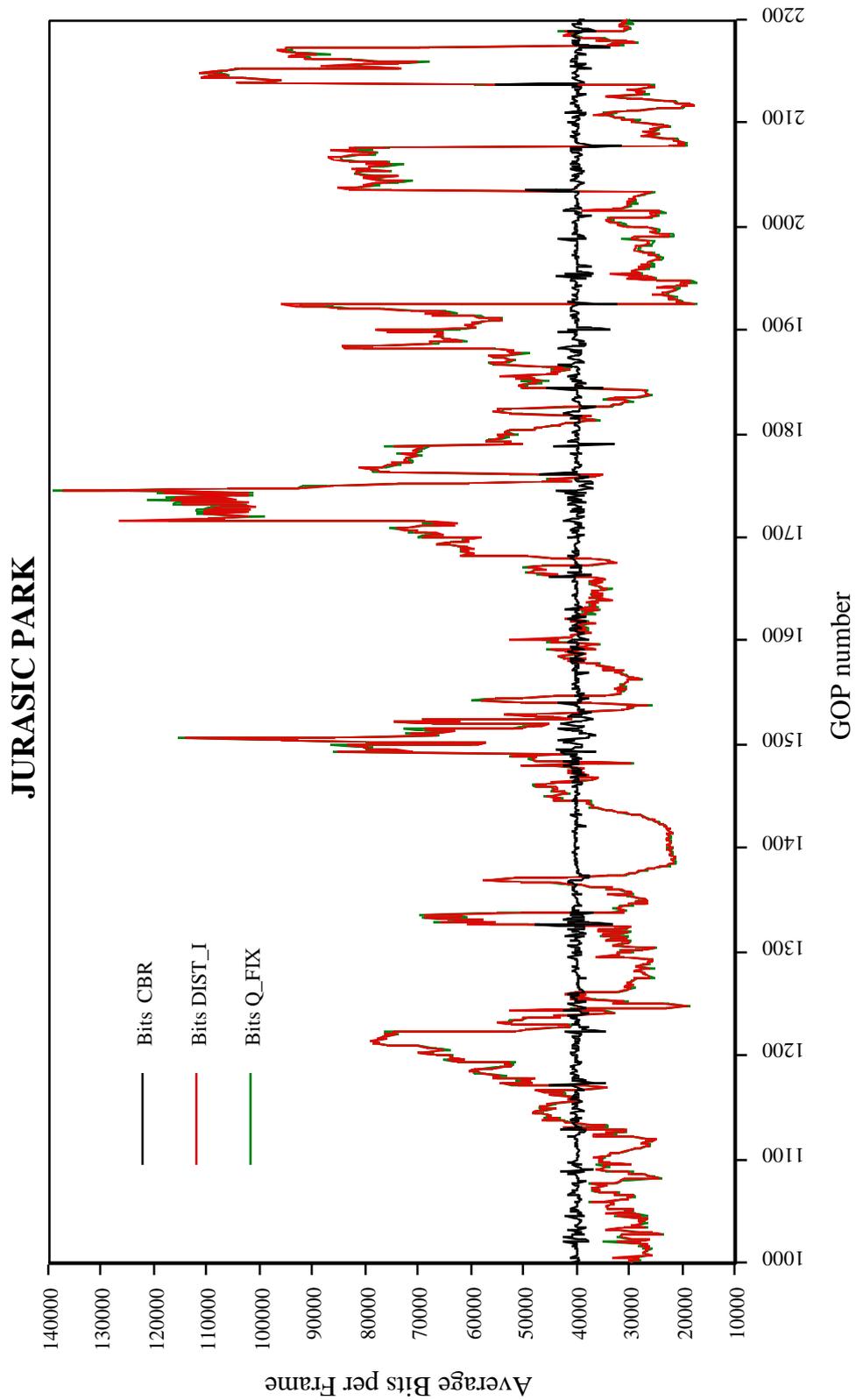


Figura 2.5.7. Comparación de la tasa binaria de los modos de funcionamiento CBR, DIST_I y Q_FIX

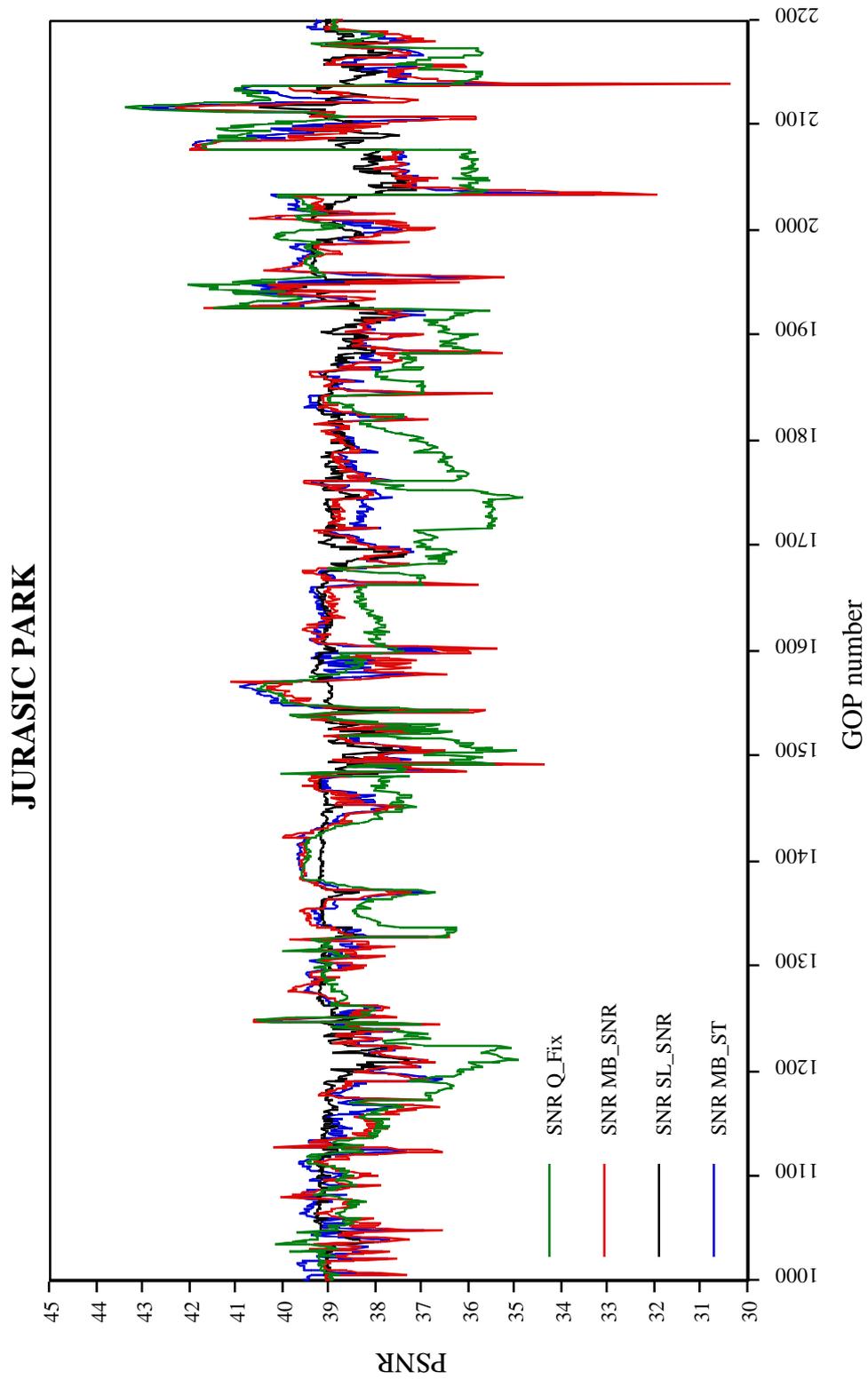


Figura 2.5.8. Comparación del PSNR de los modos de funcionamiento SL_PSNR, MB_PSNR, MB_ST_PSNR y Q_FIX

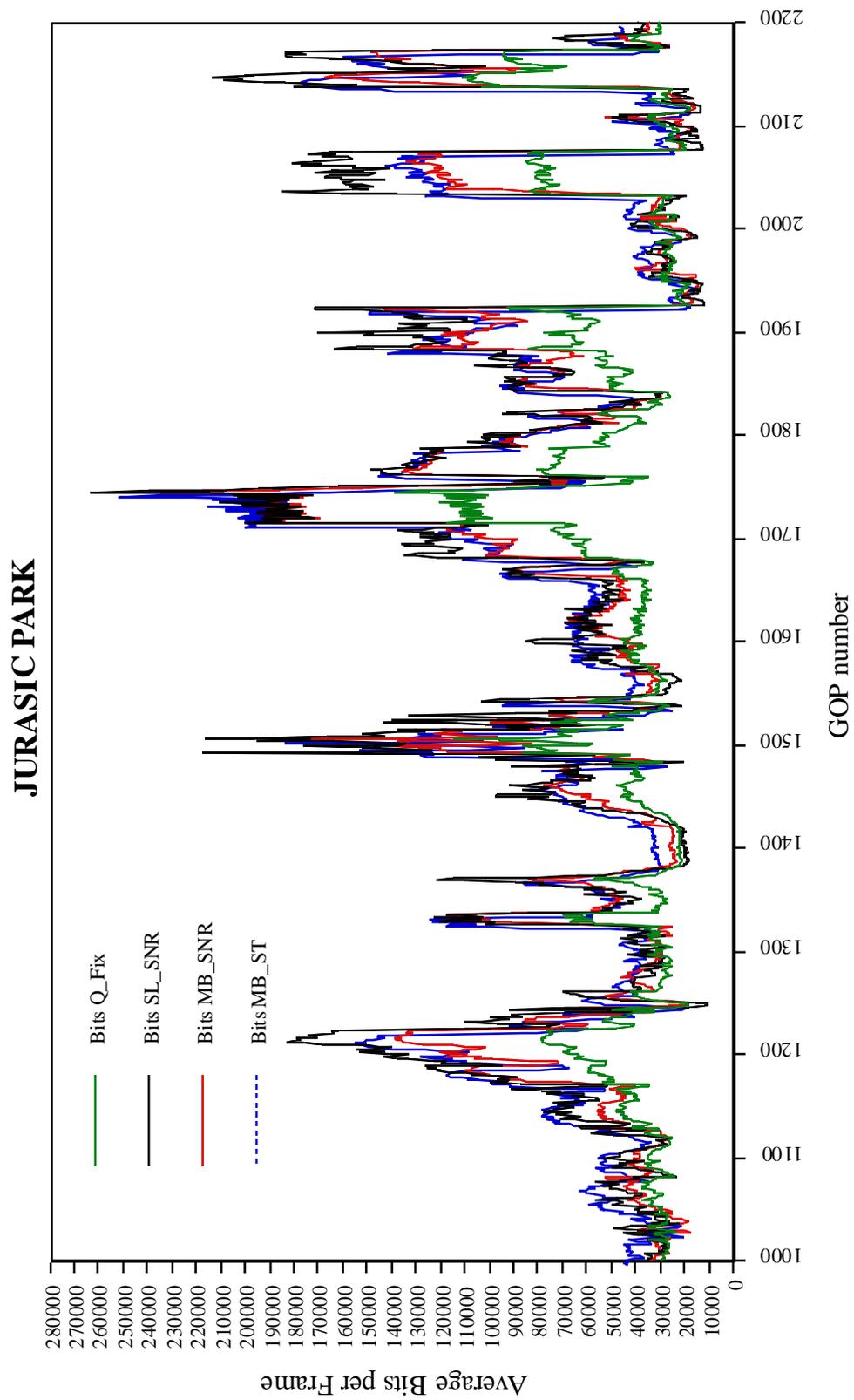


Figura 2.5.9. Comparación de la tasa binaria de los modos de funcionamiento SL_PSNR, MB_PSNR, MB_ST_PSNR y Q_FIX

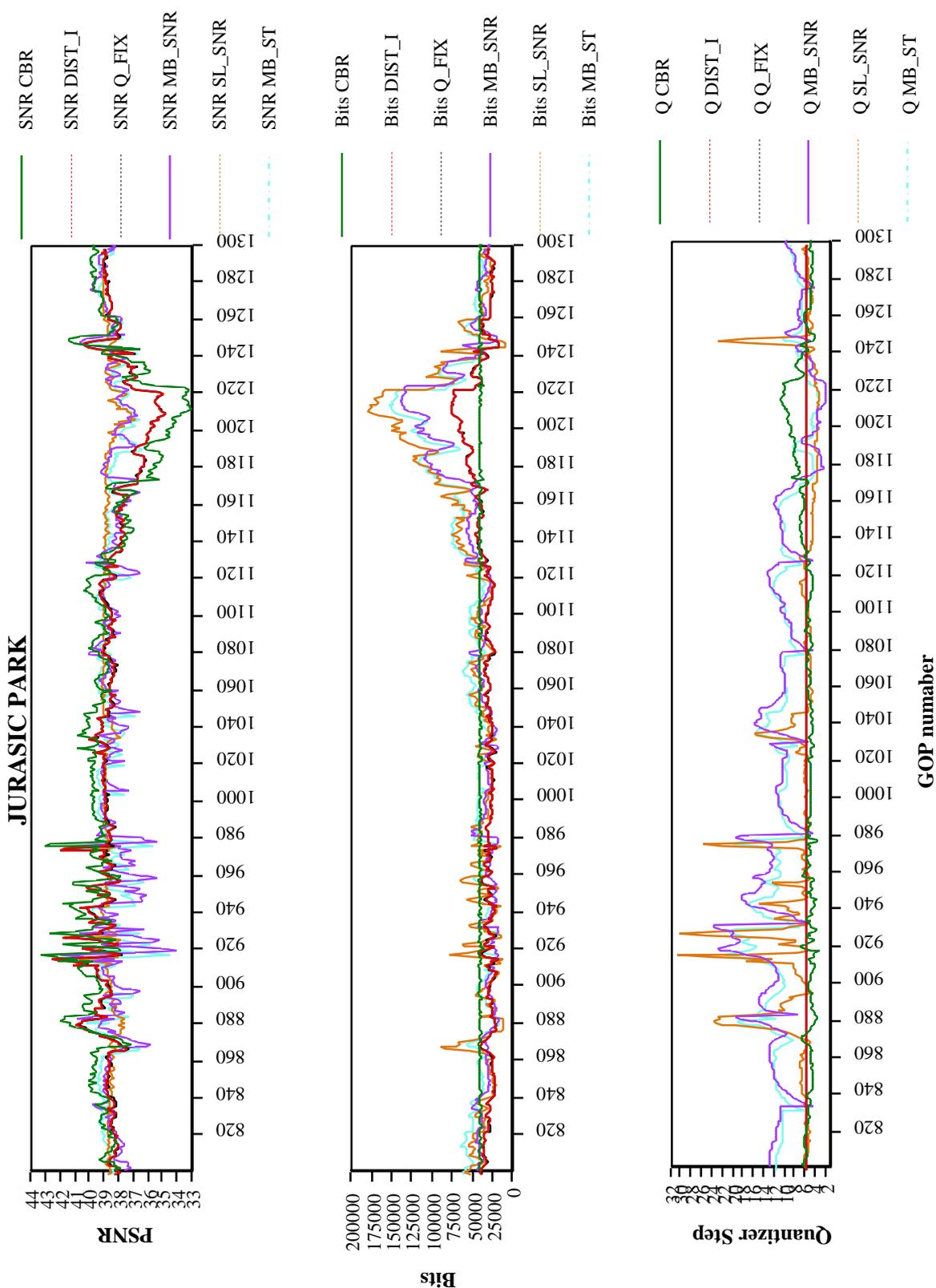


Figura 2.5.10. Comparación de los modos de funcionamiento de la codificación MPEG-I en el intervalo de GoPs (800, 1300)

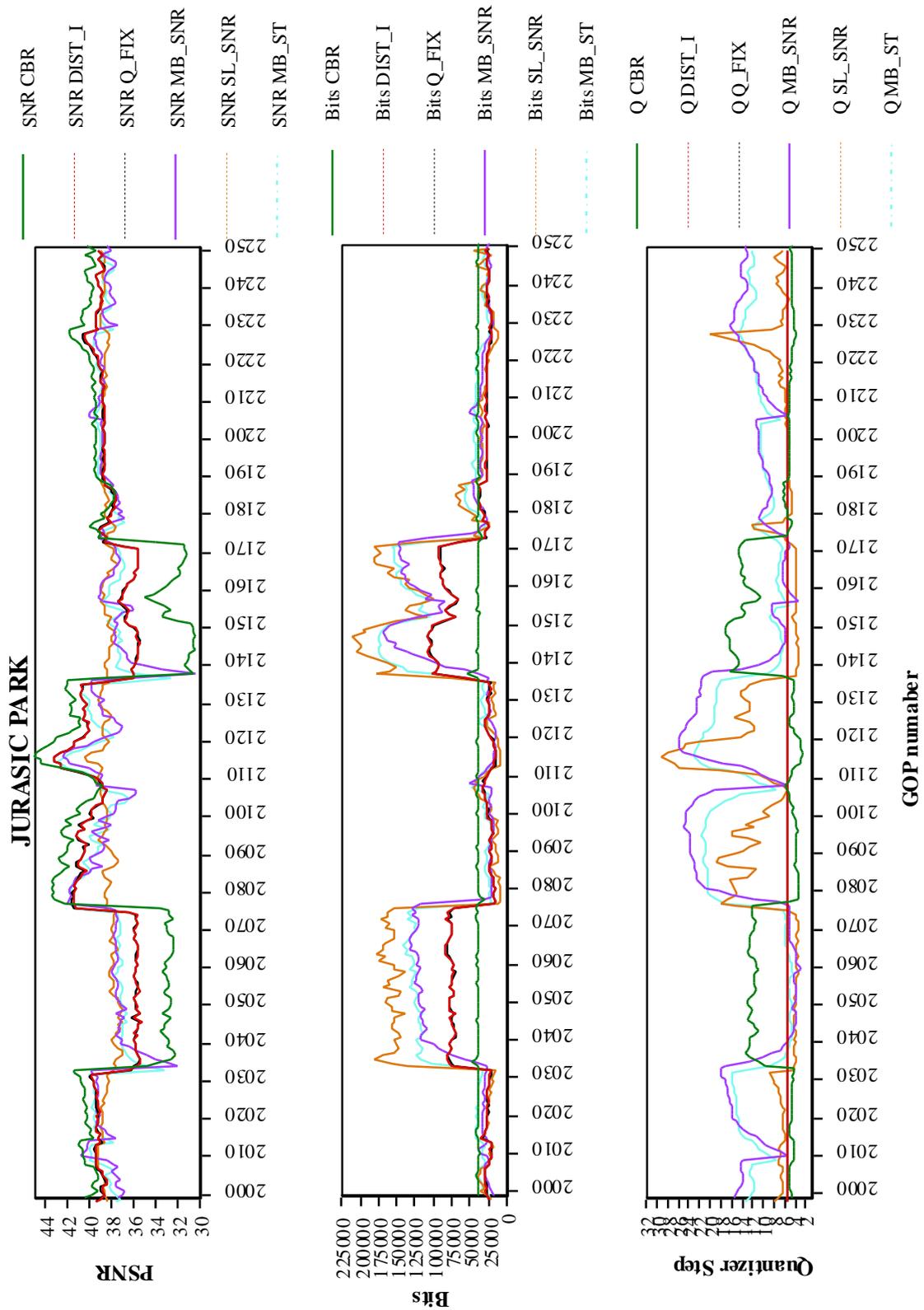


Figura 2.5.11. Comparación de los modos de funcionamiento de la codificación MPEG-I en el intervalo de GoPs (2000, 2250)

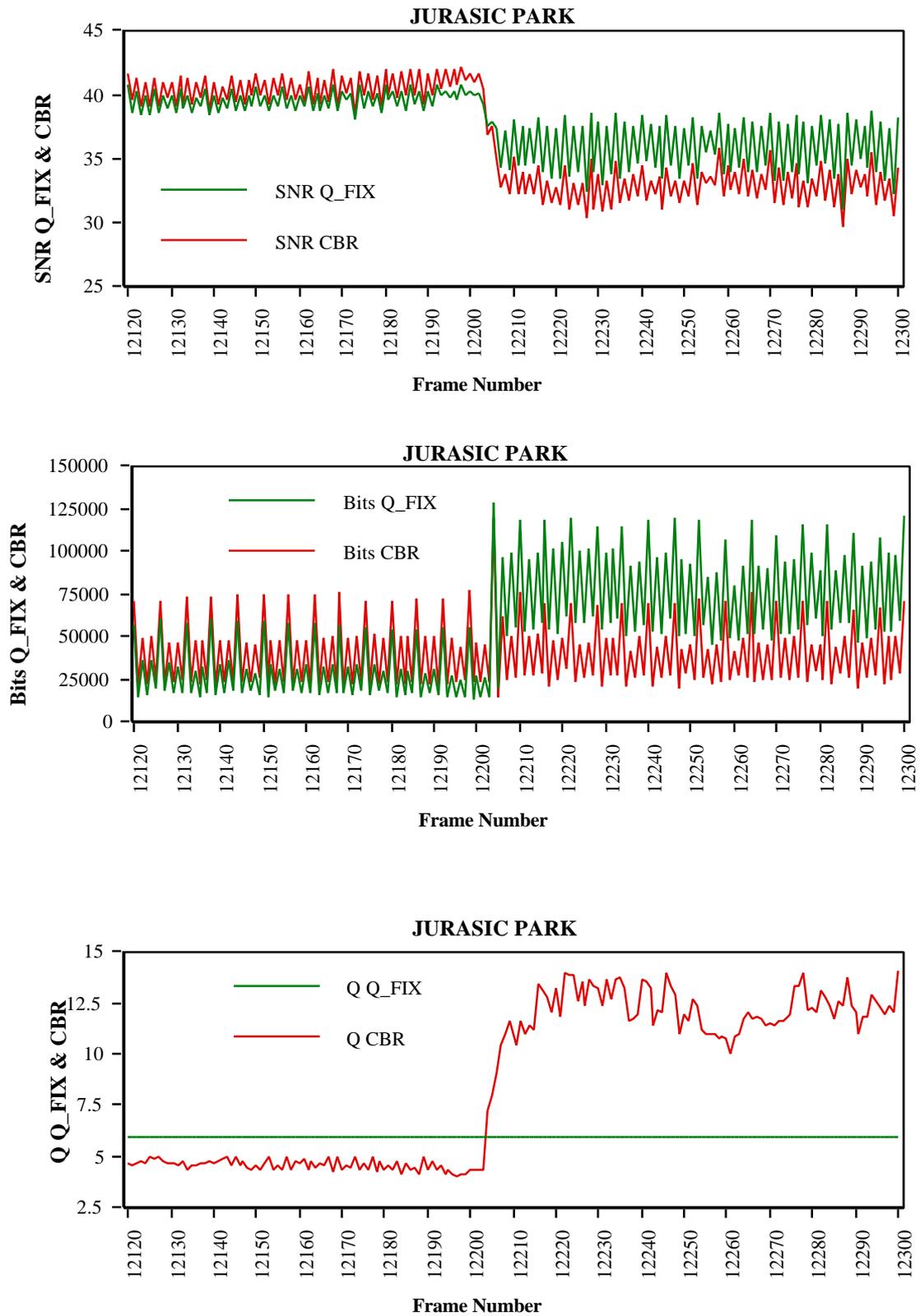


Figura 2.5.12. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y CBR para las imágenes comprendidas en el intervalo (12120, 12300)

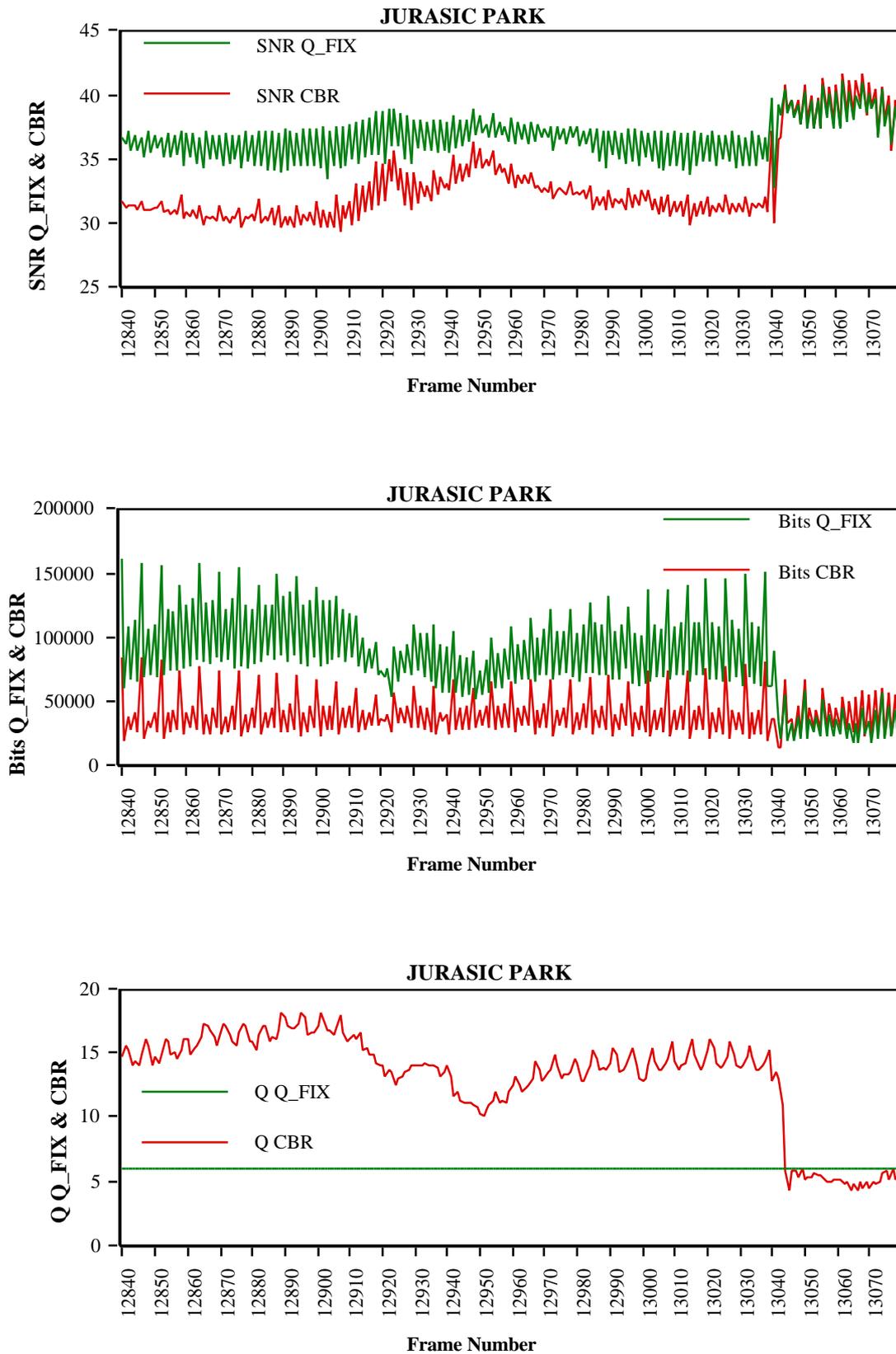


Figura 2.5.13. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y CBR para las imágenes comprendidas en el intervalo (12840, 13070)

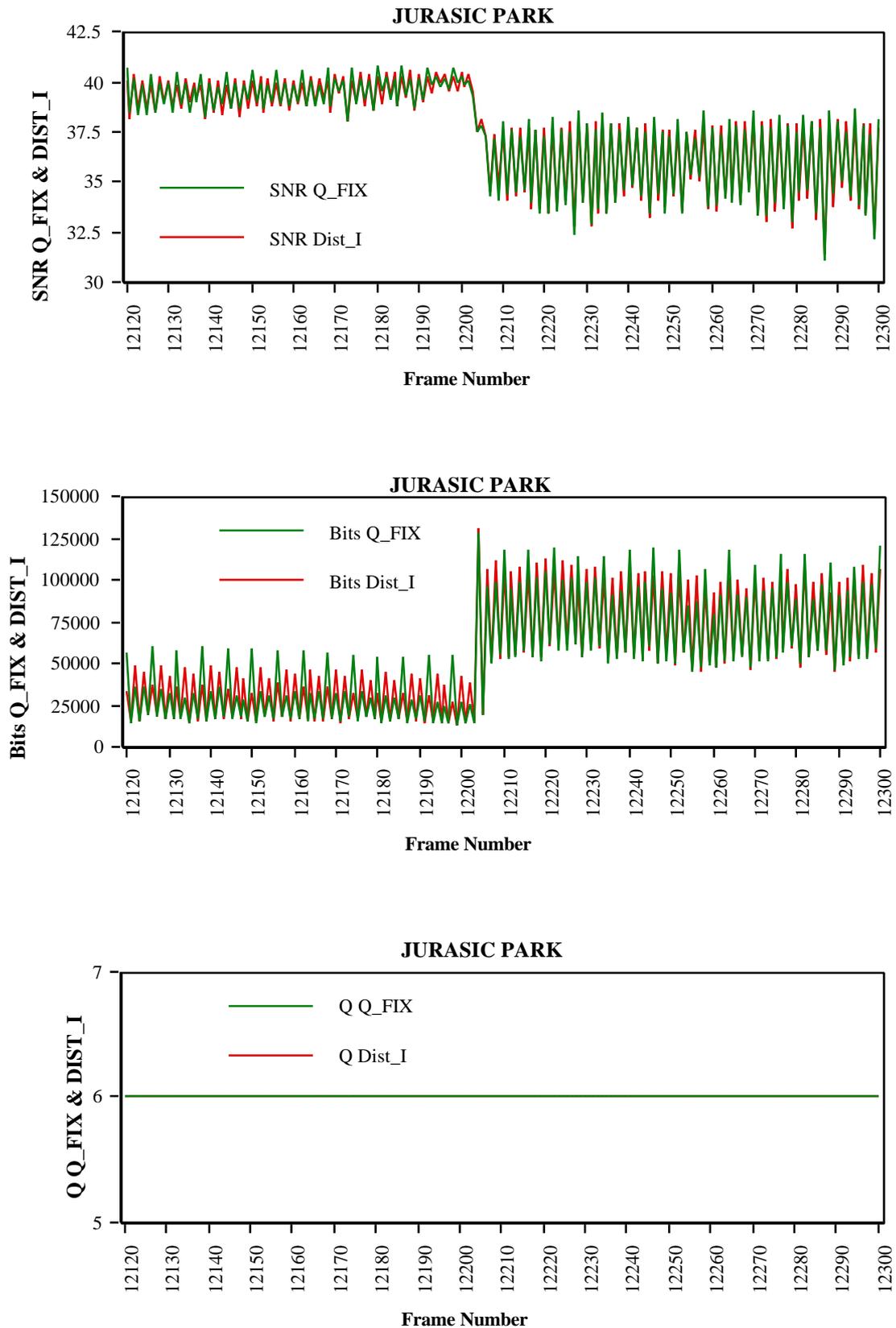


Figura 2.5.14. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y DIST_I para las imágenes comprendidas en el intervalo (12120, 12300)

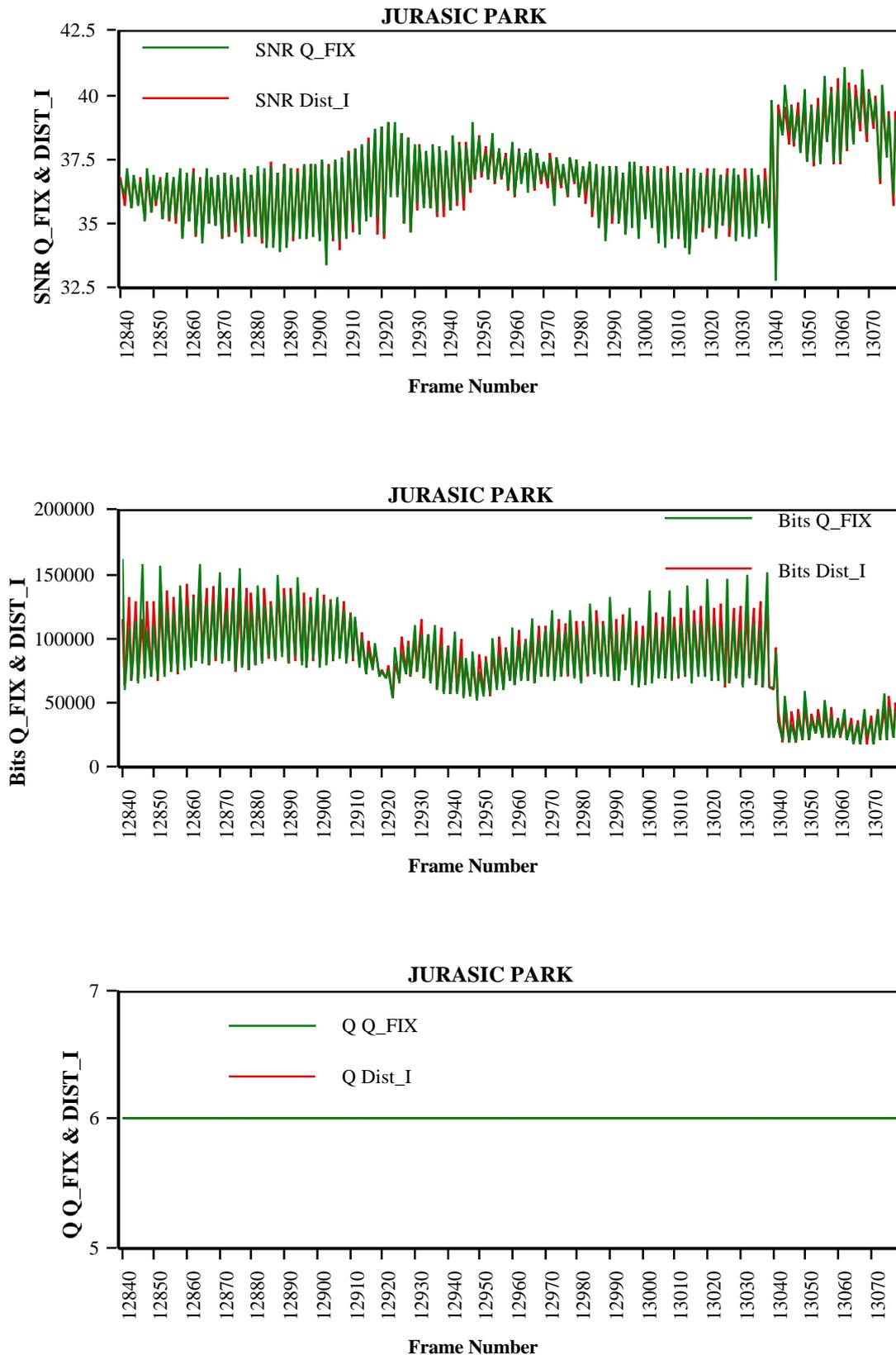


Figura 2.5.15. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y $DIST_I$ para las imágenes comprimidas en el intervalo (12840, 13070)

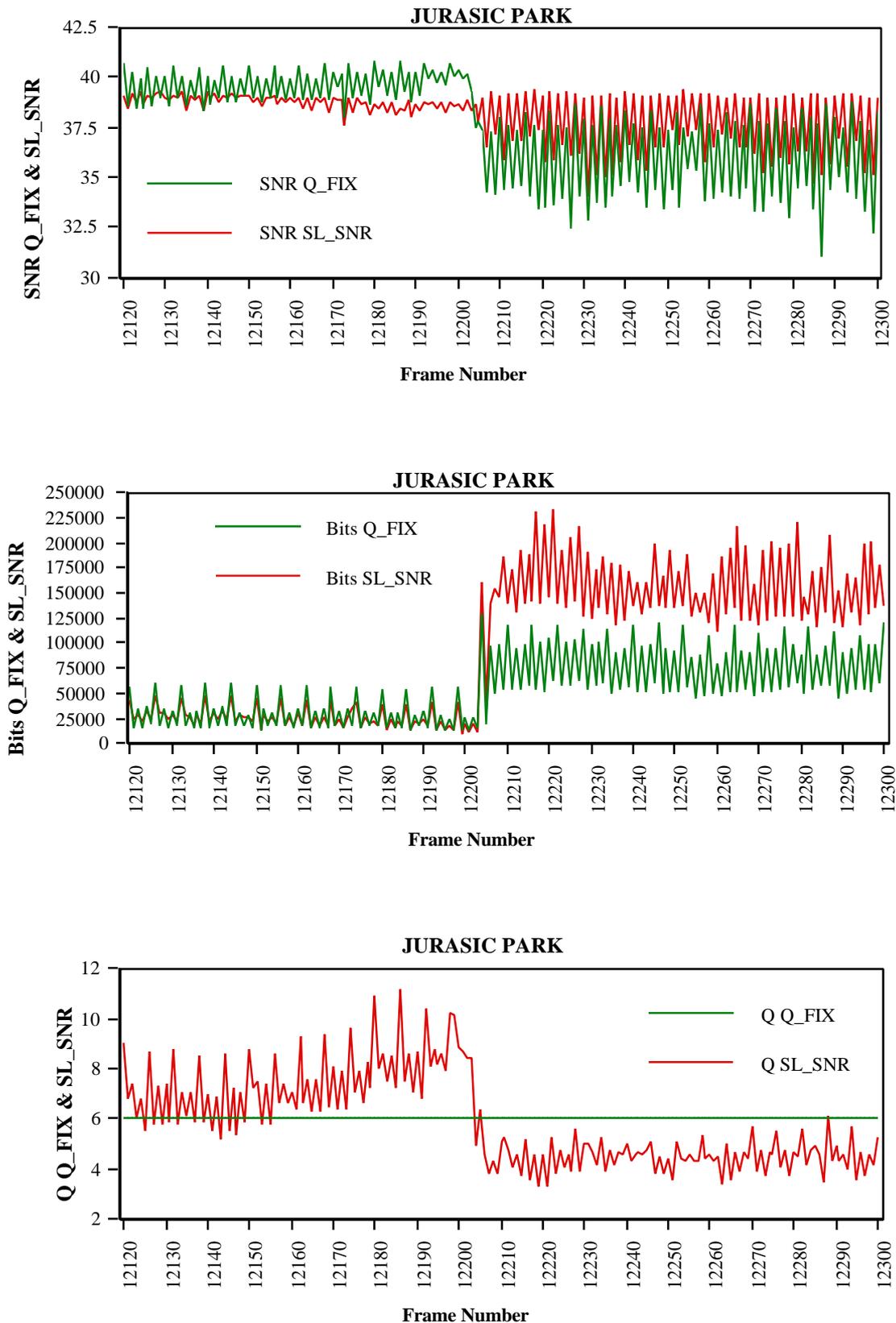


Figura 2.5.16. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y SL_PSNR para las imágenes comprendidas en el intervalo (12120, 12300)

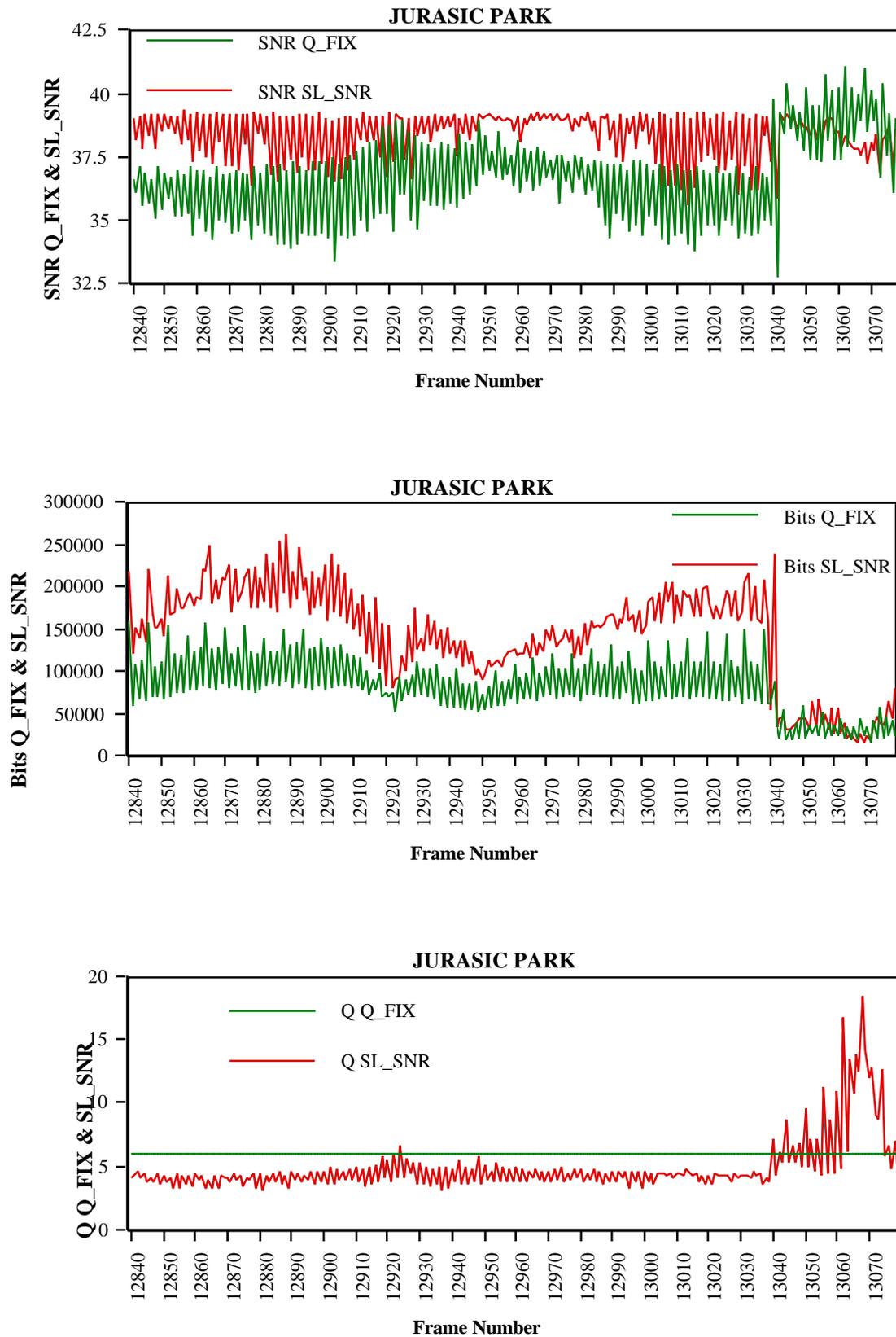


Figura 2.5.17. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y SL_PSNR para las imágenes comprendidas en el intervalo (12840, 13070)

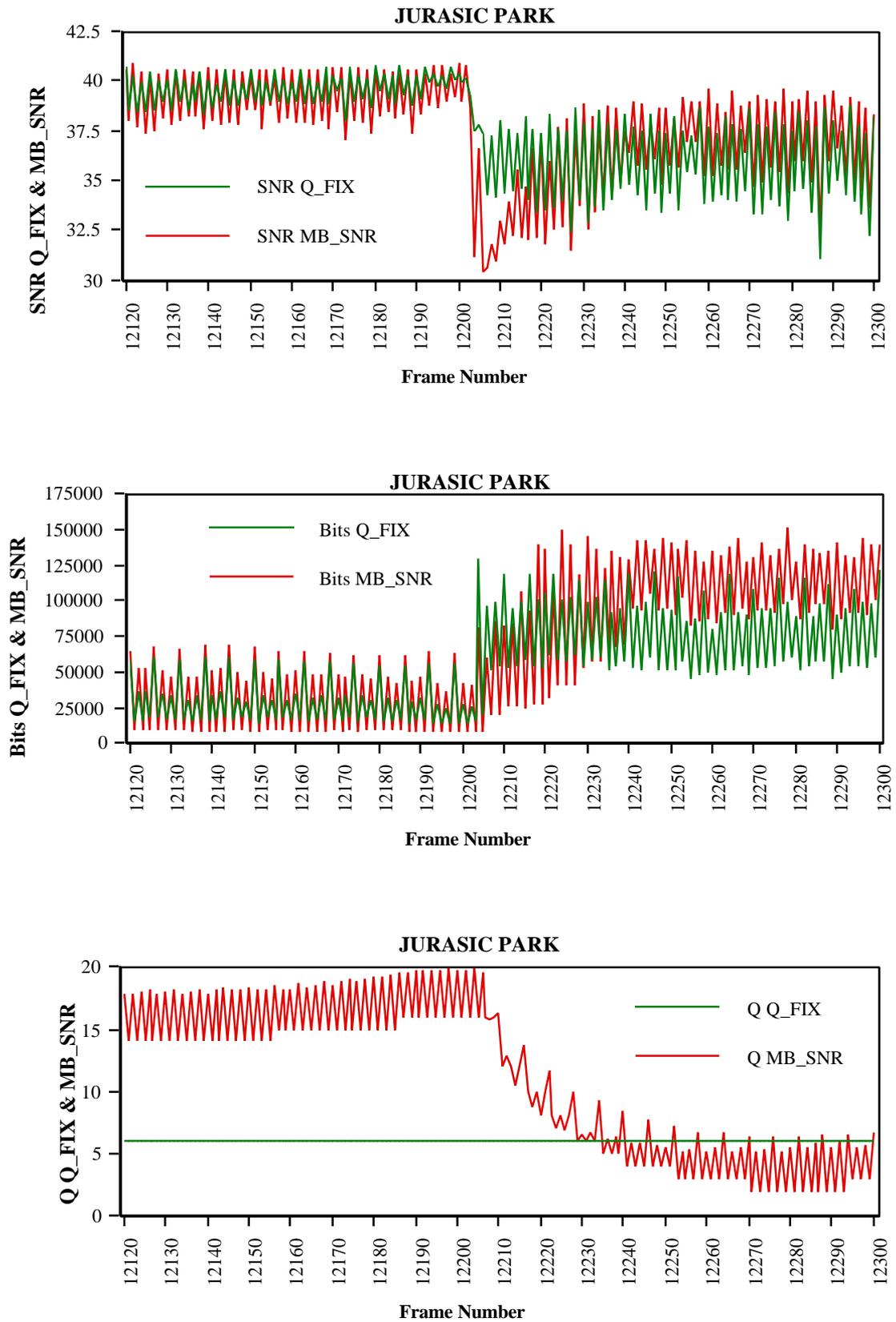


Figura 2.5.18. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y MB_PSNR para las imágenes comprendidas en el intervalo (12120, 12300)

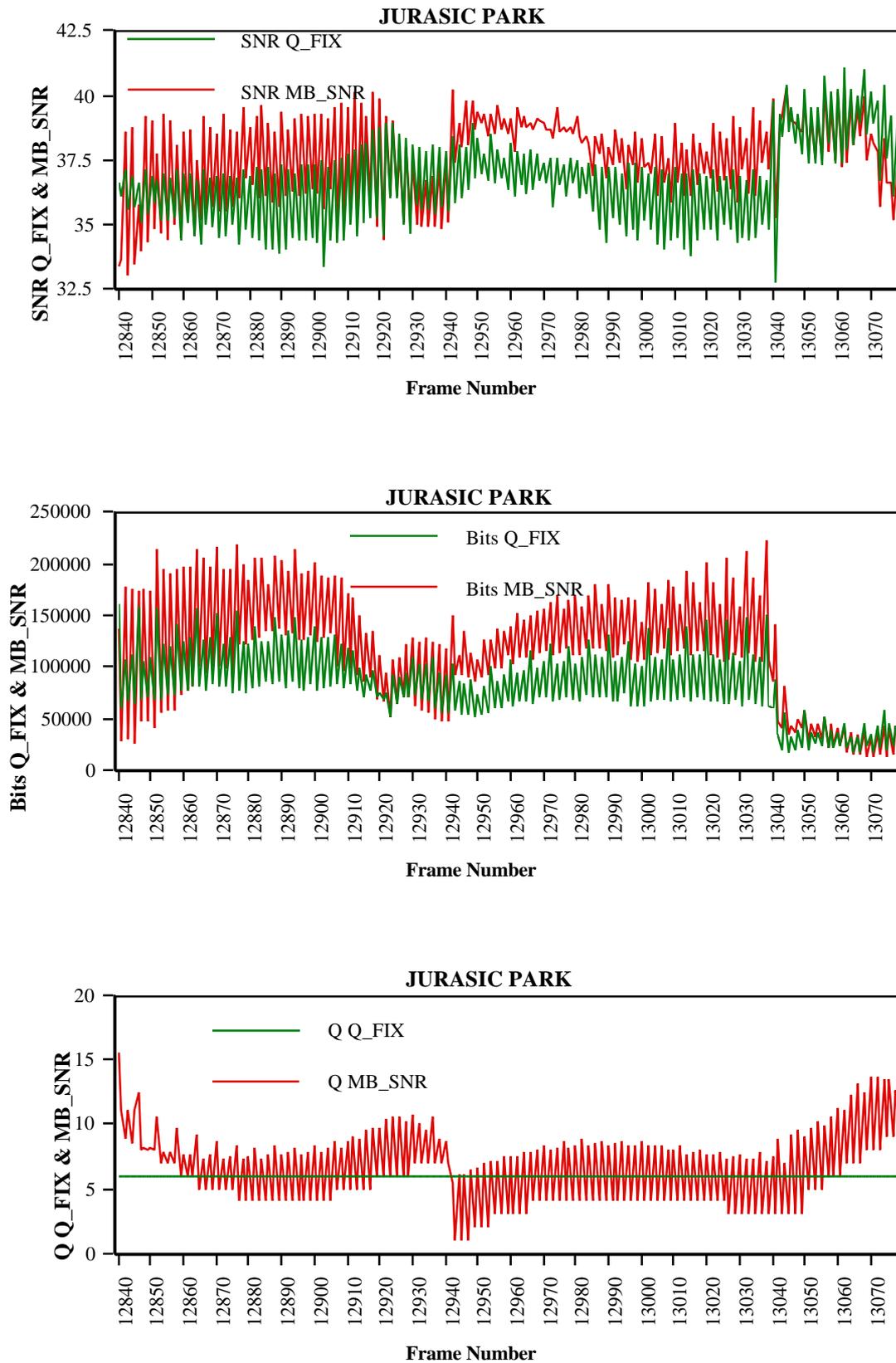


Figura 2.5.19. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y MB_PSNR para las imágenes comprendidas en el intervalo (12840, 13070)

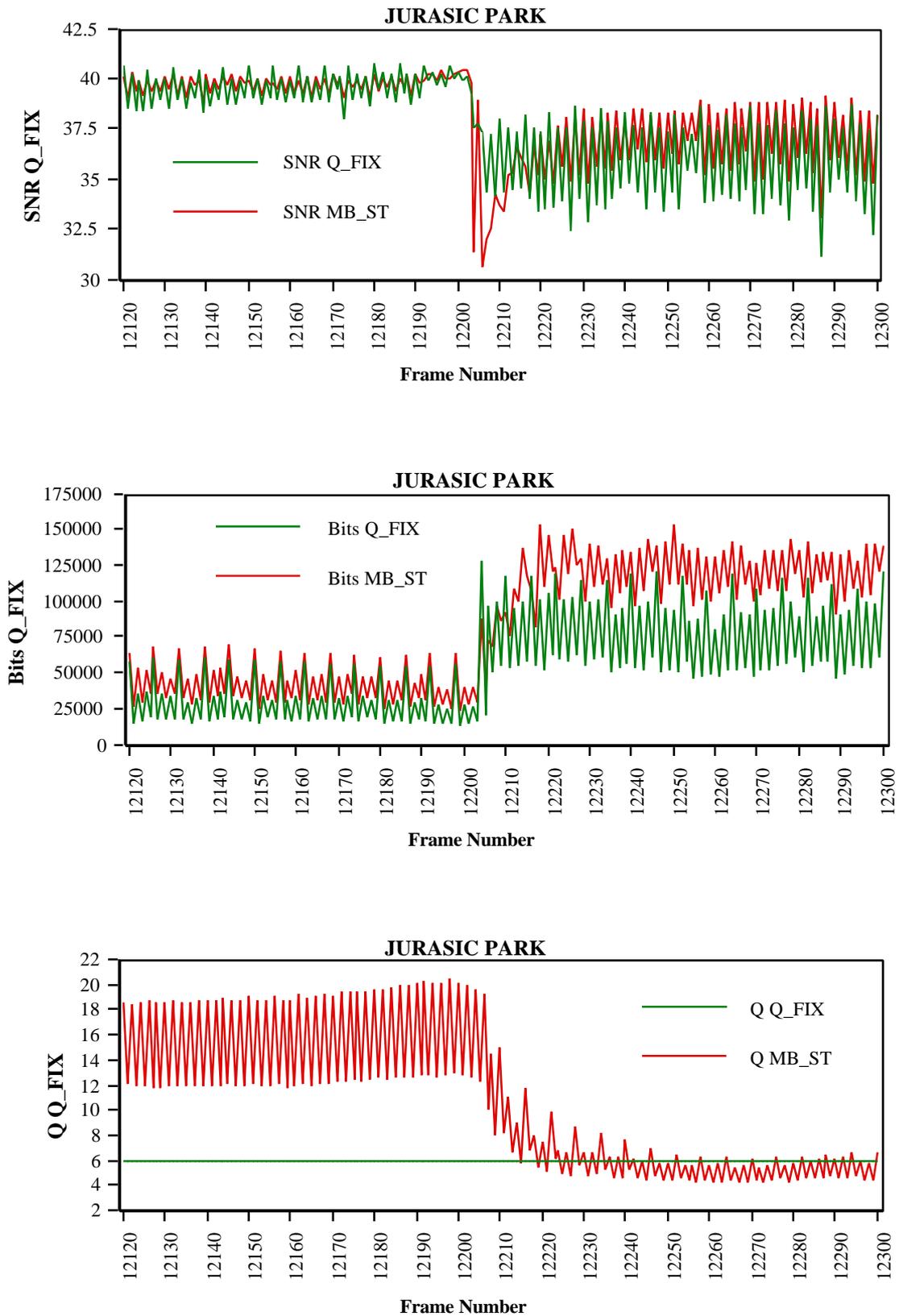


Figura 2.5.20. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y MB_ST_PSNR para las imágenes comprendidas en el intervalo (12120, 12300)

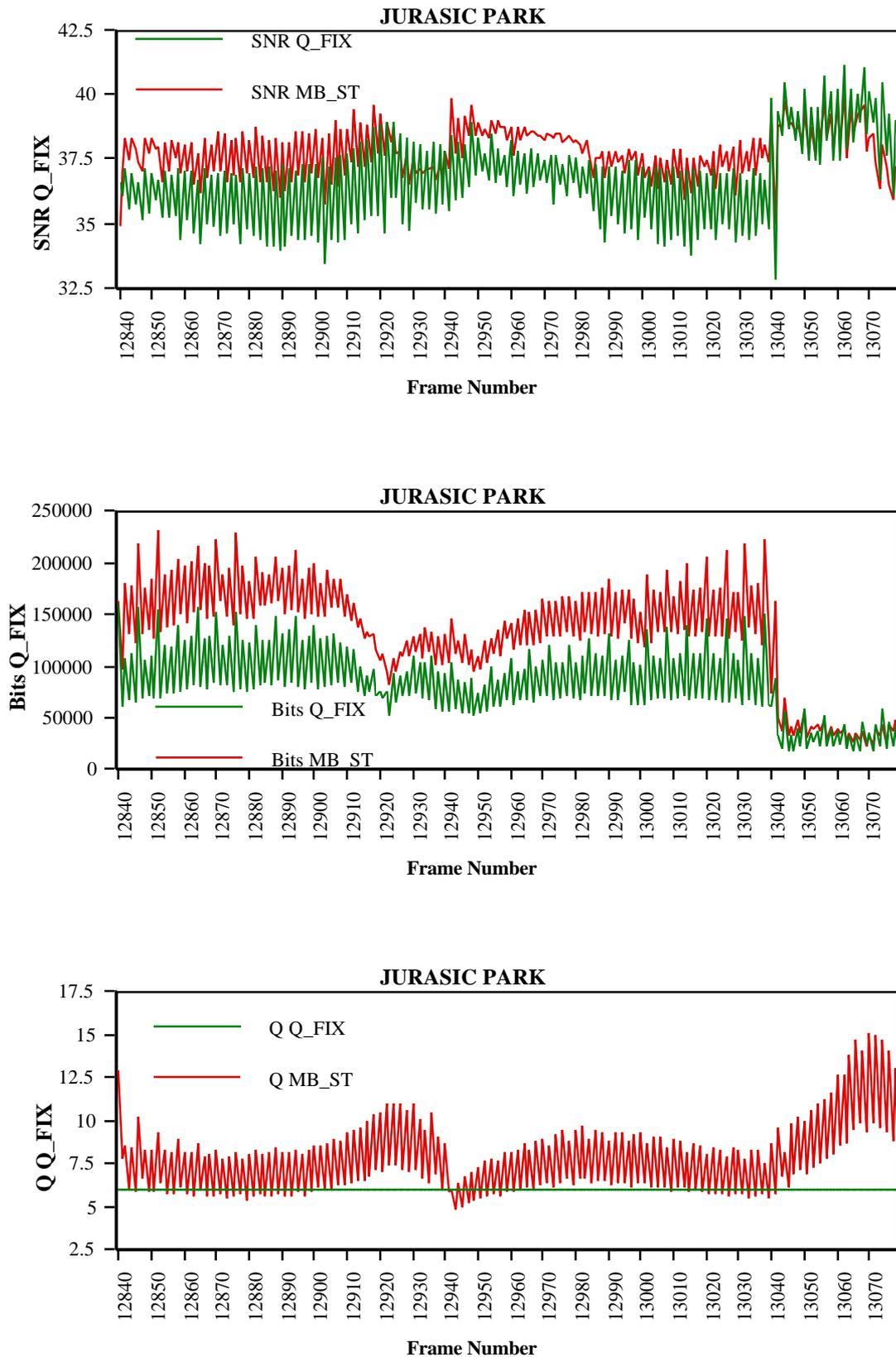


Figura 2.5.21. Comparación a nivel de cuadro de los modos de funcionamiento Q_FIX y MB_ST_PSNR para las imágenes comprendidas en el intervalo (12840, 13070)

2.6. Conclusiones

En este capítulo se ha desarrollado un análisis de la transmisión de vídeo digital sobre redes ATM, empleando el algoritmo de codificación MPEG-I. Se ha establecido un procedimiento de elección de los parámetros del codificador en base al nivel de calidad subjetiva deseado y a la maximización del rango de compresión. El análisis cuantitativo de la tasa binaria por cuadro ha conducido, igualmente, a los mismos criterios de elección de los valores de los parámetros. Asimismo, se ha llevado a cabo el estudio estadístico de la tasa binaria por cuadro a través de la evaluación de los dos primeros momentos, las funciones de autocovarianza e histogramas de las tasas generadas en cada modo de codificación. La función de autocovarianza ha permitido resaltar los efectos temporales observados en la elección de los parámetros aparecidos en el análisis subjetivo y objetivo de la codificación.

Se han propuesto distintos modos de funcionamiento del codificador MPEG-I y se ha discutido su comportamiento mediante la codificación de distintas secuencias. El análisis del comportamiento de los modos de codificación se ha realizado a partir de la observación de los resultados obtenidos de la PSNR, la tasa binaria y el paso de cuantificación. Asimismo, se han relacionado estos resultados con la calidad subjetiva obtenida y las ventajas presentadas en la transmisión sobre redes ATM.

Tras este estudio, se propone como esquema válido de transmisión a calidad semi-constante, con un nivel de complejidad reducido, el empleo del modo de funcionamiento DIST_I con objetivo PSNR con una regulación suave del paso de cuantificación. La regulación se lleva a cabo a partir del valor medio de la PSNR obtenido en el anterior cuadro codificado en el mismo modo y el paso de cuantificación será el mismo para todos los macrobloques de la imagen. En la figura 2.6.1 se ilustra este mecanismo de funcionamiento.

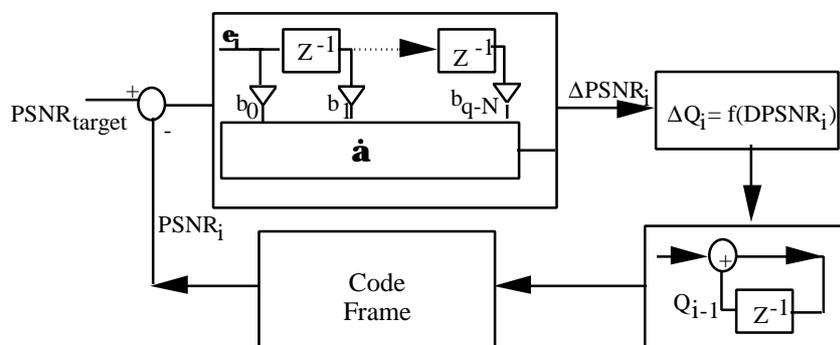


Figura 2.6.1. Esquema de funcionamiento con objetivo PSNR basado en el modo DIST_I para la transmisión con calidad semi-constante.