Análisis Práctico de los Parámetros Asociados a la Codificación de Imágenes en Movimiento para Videoconferencia[1]

José Oliver Gil y M.P. Malumbres

Departamento de Informática de Sistemas y Computadores

Universidad Politécnica de Valencia

Camino de Vera, s/n

46071 Valencia Spain

email: joseolivergil@usa.net, mperez@gap.upv.es

Resumen

Se realiza un estudio, usando una plataforma de videoconferencia, de los parámetros que actúan sobre la compresión de imagen en movimiento, y del coste computacional de cada fase de la compresión, intentando mejorar la que resulte más costosa.

1 Introducción

Sin duda la transmisión de imagen digital en movimiento ha sido uno de los temas de investigación que más interés ha despertado en los últimos años. Son múltiples sus aplicaciones dentro del campo de la informática distribuida, aplicaciones como la videoconferencia, el control y monitorización de sistemas robotizados en entornos industriales, la tele-enseñanza (universidades virtuales), la tele-medicina, los canales de difusión de televisión en Internet, los sistemas de vídeo a la carta (*video-on-demand*) y en definitiva un sinfín de aplicaciones que ofrecen servicios interactivos multimedia.

Pero el problema que presenta este tipo de transmisión, por la naturaleza de estos datos, es el alto ancho de banda requerido si no se efectúa ninguna compresión. Así, si se intentase transmitir cuadros (*frames*) sin comprimir en un formato reducido como QCIF (176x144), con una profundidad de 24 bits y una tasa de 10 cuadros/seg, se necesitaría disponer de un ancho de banda de casi 6 Mbps (capacidad de LAN), cantidad excesiva para las líneas disponibles actualmente (64 Kbps en RDSI).

Nuestro objetivo será analizar distintos parámetros de compresión de imagen y buscar alternativas que logren tasas de transmisión adecuadas para videoconferencia bajo redes de banda estrecha. Para lograrlo deberemos:

- Realizar una drástica reducción del volumen de información necesaria para representar los cuadros, estableciendo y modificando los parámetros adecuados.
- Conseguir que la compresión de la imagen se pueda realizar en tiempo real, actuando sobre las fases que sean computacionalmente más costosas. Para esto definiremos nuevas estrategias.

[1] *Este trabajo ha sido financiado por la Generalitat Valenciana (Ref.: GV97-TI-04-37)*

2 Aspectos generales sobre la compresión de imagen para videoconferencia

El estándar JPEG [1], en su versión secuencial con pérdidas, tiene una gran importancia en la compresión de vídeo, ya que es la base de la compresión espacial (aquella que elimina información redundante de una única imagen) de muchos de estos algoritmos.

En primer lugar, JPEG codifica la imagen original en un formato adecuado, se suele usar codificación YUV con cuatro componentes de luminancia por cada una de crominancia (4:1:1). Posteriormente se divide la imagen en bloques del mismo tamaño para aplicar a cada bloque la Transformada Discreta del Coseno, que toma la información representada en el dominio espacial y la convierte al frecuencial. El tamaño de estos bloques es importante ya que si son muy grandes la transformada será computacionalmente muy costosa, y si son pequeños no se conseguirá una gran reducción de la imagen y el bloque resultará poco operativo. Se suele tomar bloques de dimensiones 8x8, que permite una reducción bastante alta con un coste aceptable.

A continuación aplicaremos una cuantización de los bloques dividiendo cada componente por un número entero. En esta fase es donde se realiza la pérdida de información que determinará el índice de compresión de la imagen. Sabemos que la información de altas frecuencias no es apreciable por el ojo humano (o lo es muy poco), que actúa como un filtro pasa bajo, por tanto cuantizaremos más estas componentes preservando aquellas de menor frecuencia. Aquí se define el parámetro de cuantización, que indica el nivel de cuantización que se aplica, y está dentro de un rango de 1 a 30.

Por último se recorre en zig-zag cada bloque aplicando de forma simultánea una codificación *Run Length Encode* (RLE), que agrupa los coeficientes nulos obtenidos tras cuantizar, y una codificación de longitud variable (*Variable Length Coding*, VLC), usando menos bits para los valores más probables.

Las recomendaciones H.261 [2], pertenecientes al conjunto de estándares H.320 del ITU (antiguo CCITT), contemplan tanto la supresión de redundancia espacial como temporal (aprovechando la similitud entre cuadros consecutivos). Define dos tipos de cuadro: I y P.

Un cuadro de tipo I se codifica básicamente como en JPEG, mientras que el de tipo P puede presentar una codificación diferencial (compresión temporal). En este caso se agrupan cuatro bloques en un macrobloque (tamaño 16x16), que puede ser inter o intra (según exista o no codificación diferencial), y se realiza una estimación de movimiento para buscar en el cuadro anterior (llamado cuadro de referencia) la sección de 16x16 (macrobloque de referencia) más parecida a cada macrobloque del cuadro que estamos comprimiendo. Un parámetro que estudiaremos será la extensión de la ventana donde se busca el macrobloque de referencia, centrada en el macrobloque a codificar.

Un macrobloque se codifica como inter si se encuentra otro de referencia suficientemente parecido, en este caso almacenamos el error cometido al tomar ese macrobloque como referencia del actual. En otro caso, almacenaremos el macrobloque como intra, donde se trata como si fuese de un cuadro I.

Para más información sobre estos estándares y otros de compresión de vídeo como MPEG [3], adecuado para codificación de vídeo de calidad, se aconseja consultar las referencias bibliográficas [4], [5] y [6].

3 Implementación de una plataforma para videoconferencia

Nuestro trabajo se ha centrado en desarrollar una plataforma de videoconferencia sobre Windows 9x, en la que se pueda analizar tanto los parámetros descritos en el punto anterior, como el coste computacional de cada una de las fases de la compresión de vídeo, para intentar optimizar los parámetros y plantear alternativas para las etapas de mayor coste.

El entorno de videoconferencia desarrollado está basado en los protocolos TCP/IP, en la clásica arquitectura cliente-servidor y en el API multimedia que proporciona Windows para la captura de imágenes y procesado de ficheros de intercambio AVI.

En cuanto a los codificadores de vídeo, se pretende integrar las mejores características de los algoritmos de compresión estudiados, usando la implementación oficial del estándar MPEG-2 [7] y las recomendaciones H.261. Cuando examinamos estos algoritmos, los costes de cada paso de la codificación de cuadros de tipo I están descompensados, claramente uno de ellos es mucho más costoso que el resto: el cálculo de la Transformada Discreta del Coseno ocupa el 60% del tiempo de codificación de cada cuadro. Lo mismo se observa al codificar cuadros P, siempre que no se realice una estimación de movimiento (tomando una ventana de búsqueda nula).

Al intentar acelerar la fase del cálculo de la Transformada del Coseno podemos plantear la siguiente pregunta ¿por qué calcular todos los coeficientes frecuenciales cuando la mayor parte son eliminados o fuertemente cuantizados en la fase siguiente? Realmente no tiene mucho sentido cuando se trabaja en videoconferencia, donde se va a aplicar altos parámetros de cuantización. De esta forma definimos un nuevo componente al que llamamos minibloque, que es sencillamente un subbloque de coeficientes perteneciente al bloque de 8x8 que se utiliza habitualmente. Así, al aplicar la Transformada Discreta del Coseno a un bloque completo, obteniendo un minibloque, lo que estamos haciendo es calcular un menor número de coeficientes, y por lo tanto reducir considerablemente el coste computacional. Las dimensiones del minibloque, que indica el número de coeficientes calculados, será un nuevo parámetro que podremos estudiar para lograr realizar la compresión en tiempo real.

4 Estudio de los parámetros que actúan sobre la compresión

Para este apartado hemos usado como servidor un ordenador con procesador Pentium-166 MMX. Se utiliza una secuencia de vídeo AVI (160x120 puntos, 24 bpp., 10 cuadros/seg. y 8.4 seg.) recogida de una videocámara de sobremesa. La secuencia parte de una situación estática para realizar un desplazamiento de cámara que se detiene para recoger a los personajes que aparecen en ella.

Los parámetros que vamos a estudiar son: el parámetro de cuantización, las dimensiones del minibloque y la ventana de búsqueda. Sin embargo, también se permite seleccionar el tipo de dato con el que se desea que se opere en el cálculo de la Transformada del Coseno (entre float y double), el tipo de recorrido en el paso de RLE/VLC (entre el planteado en MPEG1 –zigzag- y en MPEG2, en diversas pruebas resultó mejor el primero), el número de cuadros P entre dos cuadros I y la condición que permite determinar cuándo dos macrobloques son suficientemente parecidos en la estimación de movimiento.

Para estudiar el error cometido al efectuar la compresión podemos usar diversos métodos. Nosotros hemos calculado el Error Cuadrático Medio (MSE, Mean Square Error), para tener una referencia absoluta del error, y la Relación Señal Ruido (SNR, Signal to Noise Ratio) media por cuadro, con la que se estudia en qué medida afecta el ruido (error cometido) a la imagen, y que suele expresarse en decibelios (calculados como 20·log(SNR)). De esta forma, si las dimensiones de la imagen son MxN y llamamos F a la matriz que representa la imagen original y G a la reconstrucción de la misma, podemos definir las ecuaciones de la figura 1.

$$MSE = \frac{\sum_{j=1}^{M}\sum_{k=1}^{N}(f(j,k)-g(j,k))^2}{M \cdot N} \qquad SNR = \frac{\sum_{j=1}^{M}\sum_{k=1}^{N} f^2(j,k)}{\sum_{j=1}^{M}\sum_{k=1}^{N}(f(j,k)-g(j,k))^2}$$

Fig. 1: *Ecuaciones para el cálculo del error cometido*

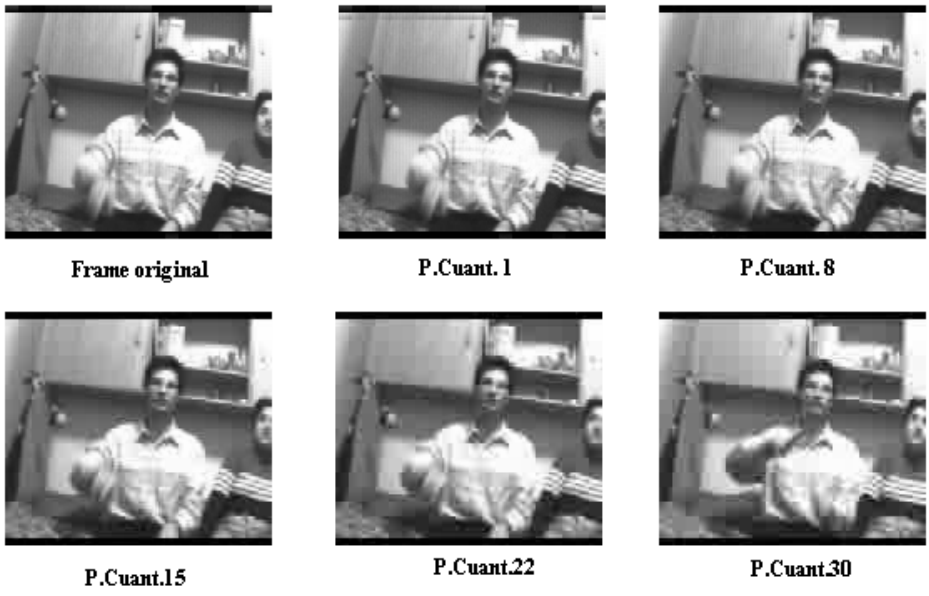

Fig. 2: *Cuadros con distintos parámetros de cuantización*

Estos parámetros tienen la ventaja de que permiten determinar de una forma objetiva la calidad de una imagen comprimida, sin embargo ha sido discutida múltiples veces su correspondencia con la idea de calidad de imagen que el ojo humano aprecia. Cuando una persona observa una imagen, cataloga su calidad con expresiones tan ambiguas como "no existe distorsión apreciable" o "es una imagen aceptable", que resultan menos precisas pero más prácticas. Por eso también hemos incluido unos cuadros, pertenecientes a la secuencia de ejemplo, con los que, a partir de un cuadro de referencia, se puede apreciar de forma visual el efecto de variar los diversos parámetros que vamos a estudiar.

En la figura 2 mostramos el resultado de aplicar distintos parámetros de cuantización a una imagen de la secuencia (rotulada como imagen de referencia) sin aplicar estimación de movimiento y sin minibloques.

Parámetro de cuantización	Error Cuadrático Medio (MSE)	SNR en decibelios (dB)
1	3.90678	86.77950
8	19.92804	59.75699
15	45.19672	52.66352
22	70.26506	48.85046
30	98.51935	45.93886

Tab. 1: *Error cometido según la cuantización aplicada*

Lógicamente, la calidad de la imagen va decreciendo a medida que la cuantización aumenta, tal y como se aprecia en la tabla 1 y en la propia imagen. A su vez, en la tabla 2 hemos resumido los principales valores que nos muestran las estadísticas relacionadas con la compresión en función de la cuantización.

Parámetro de cuantización	Tasa de Compresión	Tiempo de compresión (seg)	Cuadros por segundo	Ancho de banda requerido (Kbps)
1	80.209 %	0.24386	3.22	313
8	97.527 %	0.21212	4.451	54
15	98.668 %	0.20796	4.612	30
22	99.051 %	0.20658	4.655	21
30	99.238 %	0.20588	4.673	17

Tab. 2*: Estadísticas según la cuantización aplicada*

Se observa como, a efectos visuales, con un parámetro de cuantización de 8 la calidad de la imagen es prácticamente la misma que con 1 y, sin embargo, presenta una tasa de compresión mucho mayor, un 97 % frente a un 80 %. Realmente este nivel de compresión sí que repercute de forma importante en el MSE y, sobre todo, en la Relación Señal Ruido, donde se pasa de 86.77dB a 59.75dB, sin embargo ya hemos comentado la dudosa correspondencia que existe entre la calidad que aprecia una persona y la que se observa con estos parámetros de estimación de error cometido ya que, a simple vista, las dos imágenes resultan muy parecidas.

El tiempo de compresión es prácticamente el mismo en todos los casos y se reparte principalmente en un 51% en el cálculo de la Transformada, un 14% en cuantización, un 13% en conversión YUV, un 12% en codificación RLE/VLC, un 3% en calcular el error cometido con el macrobloque de referencia y un 7% en otros cálculos.

Como ya hemos explicado en el punto anterior, una forma de acelerar el cálculo de la Transformada consiste en no calcular aquellos componentes de mayor frecuencia. En la figura 3 se aprecia que con minibloques de 4x4 la imagen presenta resultados aceptables, y que a partir de ese tamaño la calidad se degrada de forma notable. Si comparamos las tablas

Fig. 3: *Cuadros con distintos tamaños de minibloque*

1 y 3 se puede observar que el nivel de error que se produce al tomar minibloques de 4x4 es muy parecido al que se registraba con bloques completos y una cuantización de 22, y que en

Tamaño del minibloque	Error Cuadrático Medio (MSE)	SNR en decibelios (dB)
4x4	72.49720	48.58230
3x3	151.26217	42.19385
2x2	260.80246	37.43812
1x1	1005.03558	25.71948

Tab. 3: *Error cometido según el tamaño de minibloque*

Tamaño del minibloque	Tasa de Compresión	Tiempo de compresión (seg)	Cuadros por segundo	Ancho de banda requerido (Kbps)
4x4	98.188 %	0.14563	6.27	54
3x3	98.551 %	0.13020	6.94	48
2x2	98.995 %	0.11592	7.66	37
1x1	99.436 %	0.10342	8.48	23

Tab. 4: *Estadísticas según el tamaño del minibloque*

todo caso provoca una calidad visual aceptable.

La tabla 4 presenta un resumen de los resultados al tomar diversos tamaños de minibloques, manteniendo un parámetro de cuantización de 8 y sin estimación de movimiento. Podemos observar como con un tamaño de minibloque de 4x4, hemos pasado de transmitir 4.4 cuadros/seg., en el caso anterior, a 6.27 cuadros/seg. El coste de cada paso de compresión, con minibloques de 4x4, pasa a ser del 36% en cálculo de la Transformada, 23% en la conversión YUV y 22% para realizar la cuantización, con lo que se observa un mayor equilibrio.

Ventana de búsqueda	Tasa de Compresión	Tiempo de compresión (seg)	Cuadros por segundo	Ancho de banda requerido (Kbps)
0x0	80.209 %	0.24386	3.22	313
5x5	80.301 %	0.28587	2.85	276
10x10	80.310 %	0.35757	2.37	229
15x15	80.310 %	0.45531	1.88	182

Tab. 5: *Estadísticas según las dimensiones de la ventana de búsqueda*

Por último vamos a estudiar el efecto producido al introducir estimación de movimiento, y su impacto en los tiempos de cómputo. En la tabla 5 tenemos información sobre las estadísticas obtenidas al variar el tamaño de la ventana de búsqueda, manteniendo una cuantización de 1 y un tamaño de minibloque completo (8x8).

Observando los resultados podemos concluir que la estimación de movimiento no parece ser factible cuando se pretende realizar videoconferencia por software (hace claramente asimétricos el codificador y decodificador). Además, el beneficio obtenido es muy pequeño, pasando de una tasa de compresión de 80.209%, sin estimación de movimiento, a otra de 80.310% en el caso de mayor búsqueda, mientras que el tiempo empleado para codificar cada cuadro pasa de 0.24489 a 0.45531 segundos por cuadro. De hecho la fase de estimación de movimiento ha pasado a ser la más costosa, con un 48% de tiempo consumido frente al 24% de la Transformada.

5 Conclusiones y trabajo futuro

Hemos hecho un análisis de los parámetros que influyen en una codificación de vídeo, especialmente orientado a videoconferencia, basándonos en los estándares más populares de compresión de vídeo. Tras este estudio, parece que la estimación de movimiento no es aconsejable en un entorno de videoconferencia por su elevado coste. Sin embargo, podemos aligerar el cálculo de la Transformada usando minibloques y consiguiendo mayores tasas de cuadros/seg. para un mismo ancho de banda, sin afectar excesivamente a la calidad de la imagen.

Con un parámetro de cuantización de 8 y un tamaño de minibloque de 4x4 se obtiene una buena calidad de imagen (48.58dB), con una tasa de cuadros/seg. aceptable (más de seis cuadros/seg) y un ancho de banda necesario de 55 Kbps, apto para enlaces RDSI de 64Kbps. Si se desea utilizar un menor ancho de banda podemos realizar una cuantización de 16 con el mismo tamaño de minibloque, para lo que se necesitaría tan sólo 33 Kbps, obteniendo una SNR de 46.59dB.

Por otra parte, estamos estudiando codificadores de vídeo adaptativos que permitan variar dinámicamente los parámetros de codificación en función de los retardos extremo a extremo, ancho de banda disponible, etc. Para esto estamos implementando una versión del protocolo RTP [8] que nos permitirá realizar una realimentación de las características de la transmisión, desde el cliente (descompresor) hacia el servidor (compresor).

Referencias

[1] *Digital Compression and Coding of Continuous Tone Images (Part 1: Requirements and Guidelines)*. ISO/IEC 10918-1, 1992.

[2] *Video Codec for Audiovisual Services at px64 Kbit/s. CCITT Recommendation H.261*. International Telegraph and Telephone Consultative Committee (CCITT), 1990.

[3] *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s*. ISO/IEC 11172-2: Video, 1993.

[4] B. Furht. *A Survey of Multimedia Compression Techniques and Standards (Part I, Part II)*. Journal of Real-Time Imaging, Vol.1, N°1 pág. 49-67 y N°5 pág. 319-338, 1995.

[5] J.L.Mitchell, W.B.Pennebaker, C.E.Fogg, D.J.LeGall. *MPEG Video Compression Standard*. Digital Multimedia Standards Series, Chapman&Hall, 1997.

[6] Raymond Westwater, Borko Furht. *Real-Time Video Compression: Techniques and Algorithms*. Multimedia Systems and Applications, Kluwer Academic Publishers Group, 1997.

[7] MPEG Software Simulation Group. *MPEG-2 Encoder / Decoder* v1.2, 1996.

[8] H. Schulzrinne, et al. *RTP: A Transport Protocol for Real-Time Applications*. RFC 1889, Enero 1996.