

Compresión de Datos, Imágenes y Video

Trabajo Final

Tasa de Distorsión

Profesora: Claudia Russo

Iuliano, Pablo José.....Nº:2122/1

E-mail: piuliano@info.unlp.edu.ar

Introducción

En esta monografía, se provee una visión general de las técnicas de optimización de la tasa de distorsión o rate-distortion y sus aplicaciones prácticas en imágenes y video. Se comenzará con una corta discusión de la teoría clásica de la tasa de distorsión y se mostrara como en muchos escenarios de codificación en la práctica, la alocaación de recursos puede ser puesta en un framework para la tasa de distorsión. Se introducirá las dos técnicas más populares para la alocaación de recursos, llamadas, optimización Lagrangian y optimización dinámica. Luego de una discusión sobre esas dos técnicas, se concluirá con una breve reseña de las recientes literaturas en esta área citando un número de aplicaciones relacionadas a la compresión y transmisión de imágenes y video.

De la teoría de Shannon a la codificación MPEG

Recientemente hemos visto avances significativos en el área de la optimización de la tasa de distorsión o rate-distortion en imágenes y video. En esta sección, primeramente se discutirá la teoría clásica de la tasa de distorsión y las contribuciones fundamentales relacionadas para establecer los límites fundamentales de performance en los sistemas de compresión para fuentes específicas.

Teoría Clásica de la tasa de distorsión

El punto de partida de la teoría clásica de la tasa de distorsión puede ser hallado en los trabajos de Shannon, realizados en 1948. La teoría de la tasa de distorsión llega bajo la premisa de codificar o comprimir una fuente, la cual tiene que ver con la tarea de maximizar la extracción de redundancia de la fuente, sujeta a un criterio de fidelidad. En otras palabras, la teoría de la tasa de distorsión esta vinculada con la tarea de representar una fuente con la menor cantidad de bits posibles para una calidad de reproducción dada.

La representación de la fuente es una proposición demasiado vaga a menos que primero establezcamos que es una “fuente”. Por ejemplo podemos considerar una fuente como un conjunto particular de datos (un archivo de texto, un segmento de audio digitalizado, una imagen o un clip de video). Alternativamente podemos considerar una clase de fuente que esta caracterizada por sus propiedades estáticas (archivos de texto que contienen código de C, segmentos de discursos, imágenes naturales o secuencias de videoconferencia). Cuando uno considera una clase de fuente, se supone que se obtendrá ventajas en la eficiencia de codificación de la fuente por el comportamiento típico de los elementos que se encuentran dentro de esa clase. Por ejemplo, las técnicas que trabajan bien para los discursos pueden no trabajar tan bien para los clips de video. Sin embargo, la entrada que pertenecen a una clase dada pueden presentar variaciones significativas una respecto de la otra (por ejemplo en diferentes escenas en una secuencia de video), y así las

técnicas que permiten un parámetro de selección “entrada por entrada” son comúnmente mejores que aquellas que utilizan la misma codificación para todas la entrada en una clase. Aquí se presentaran técnicas que se esfuerzan por tener en algún sentido lo mejor de los dos mundos para ser aplicado: los esquemas de codificación están basados en la características típicas de una clase de señales, pero los parámetros de codificación, dentro del framework seleccionado para la codificación, son elegidos entrada por entrada para optimizar la compresión particular en la clase estática de interés.

La compresión puede ser alcanzada con técnicas sin perdidas donde la decodificación o descompresión de los datos es una copia exacta del original (en esta clase se pueden nombrar herramienta de Software como el *zip*, *gzip* o *compress*). La compresión sin perdidas es importante cuando uno necesita la reconstrucción perfecta de la fuente. Sin embargo, estos requerimientos hacen que la performace de compresión se vea limitada de alguna manera, especialmente para aplicaciones donde el tamaño de la fuente de información es voluminoso, el ancho de banda o el lugar de almacenaje es severo, y una perfecta rendición de la fuente es excesiva.

En cada escenario, la llamada compresión con perdidas es usada para obtener altas tasas de compresión a costa de una representación imperfecta de la fuente. La relación entre la fidelidad con la fuente y la tasa de codificación es exactamente la tasa de distorsión. Las aproximaciones con perdidas son preferidas para codificar imágenes y video (y son usados en populares algoritmos como el JPEG). La compresión es con perdidas en los casos donde la imagen decodificada no es una copia exacta de la original pero, si las propiedades del sistema visual humano son explotadas correctamente entonces la imagen original y la decodificada serán indistinguibles. En los casos con perdidas uno puede de este modo establecer una relación entre el número de bits en la representación (la tasa) y la fidelidad de la representación (la distorsión). Esta es una relación fundamental que instala la siguiente pregunta: ¿Cuánta fidelidad en la representación estamos dispuestos a dar para reducir el tamaño de almacenamiento (o el número de bits requeridos para transmitir los datos)?.

En la teoría de la tasa de distorsión estamos bajo la suposición de que estamos hablando de codificación de una fuente, pero es importante notar que esta teoría también es aplicable en contextos más generales de transmisión de datos sobre canales con ruido. Esto es debido al famoso principio de separación de comunicación digital de Shannon, donde él demostró la forma óptima de dividir el problema de transmisión de información en forma óptima (óptima en términos del uso más eficiente de los recursos disponibles como poder, ancho de banda, etc.) en eso de representar la información eficientemente y entonces proteger la representación resultante para poder ser transmitida virtualmente libre de perdidas hacia el receptor.

Medidas de Distorsión: El Problema Elusivo

El problema de que medidas de distorsión son más convenientes para los discursos, audio, imágenes o video ha sido objeto de continuo estudio para la representación digital de todas esas señales mencionadas. Claramente, desde que esas fuentes son codificadas y transmitidas para ser, en última instancia, reproducidas o mostradas para un oyente/observador humano, una medida de distorsión debe ser consistente con lo que el

sujeto puede observar u oír. Así las medidas de distorsión que correlacionan bien el impacto percibido de la pérdida deben ser favorecidas. Dejando de lado las obvias diferencias en la percepción entre los individuos, hallar una medida de calidad perceptiva lo suficientemente general, sin mencionar su facilidad para la computación, ha demostrado ser un objetivo elusivo. Así se aplican reglas de diseño de sonido simples y perceptivas, dondequiera que las medidas de calidad perceptiva no están disponibles o son demasiado complejas. Por ejemplo, las características conocidas de la percepción humana dictan que no todas las frecuencias en una señal de audio o una imagen tienen la misma importancia. Con esta regla de diseño en mente, pueden introducirse pesos de frecuencia apropiados al codificador. Ahora que el pesaje perceptivo ha sido realizado, un codificador optimizado puede ser usado para minimizar una medida de distorsión objetiva, como por ejemplo el error cuadrático medio.

Sin embargo, también es importante comprender que las ganancias significativas en calidad objetiva no se pueden trasladar a las ganancias significativas en calidad perceptiva. Desde el éxito de una aplicación final de una codificación particular no hace depender la medida de calidad objetiva, será necesario determinar en la etapa de diseño si alguna aproximación de optimización es aplicable pudiendo ser justificada en términos de la relación entre el costo de implementación y la calidad perceptiva. Esto además enfatiza la necesidad de incorporar al codificador algún criterio de percepción para que cualquier optimización de la codificación tenga que escoger la “perceptually friendly” en los puntos de operación.

Limites en las tasas de distorsión

Uno puede distinguir entre dos clases de límites, aquellos basados en la teoría de Shannon y aquellos derivados de las aproximaciones de la tasa más alta, este último proporciona resultados asintóticos. Para derivar los límites uno necesita, primero caracterizar la fuente y esto puede ser problemático por las diferentes complejidades de las diferentes fuentes, como por ejemplo el video. De hecho, los límites probablemente serán encontrados sólo para los modelos de las fuentes estáticas más simples.

Un interesante y muy usado caso especial en la teoría de tasa de distorsión refiere a el uso de cuantificadores escalares, donde las muestras son cuantificadas una por vez en lugar de cómo una colección o vector. La teoría de la cuantificación escalar óptima (como diseñar cuantificadores escalares con performance lo suficientemente cercana al límite) ha sido ampliamente estudiada. Para estas fuentes simples, cuantificadores prácticos con varios grados de aproximación a los valores óptimos son posibles. El tipo más simple de cuantificador escalar es el cuantificador escalar uniforme, donde los tamaños en los pasos de cuantificación son uniformes. Extensiones más sofisticadas incluyen a los cuantificadores de tasa fija no uniformes (donde cada nivel de cuantificación esta representado con el mismo número de bits), también conocido como los cuantificadores de Lloyd-Max.

Mientras las técnicas mencionadas arriba tratan con estrategias de cuantificación óptimas para una fuente de distribución dada, cuando tratamos con fuentes más complejas como las imágenes y señales de video, surge inmediatamente la siguiente pregunta: ¿Cuál

es la distribución de la fuente más correcta que debe involucrar el modelado exacto de la fuente? Es por consiguiente importante considerar ambos problemas, y optimizar la performance de la codificación de imagen o video, de hecho consiste de dos pasos:

- 1) Dado un tipo particular de datos, digamos un imagen, ¿Cuál es la apropiada probabilística, o modelo para esta fuente?
- 2) Dado un modelo seleccionado, y algún limite aplicable, ¿Cómo podemos acercarnos a un algoritmo practico que tenga una buena performance?

Para codificar imágenes y video ambos pasos son igualmente importantes porque los modelos que pueden capturar adecuadamente la redundancia estática pueden no ser aplicables (o pueden ser muy complejos para poder hallar un limite). Este punto es ilustrado a continuación:

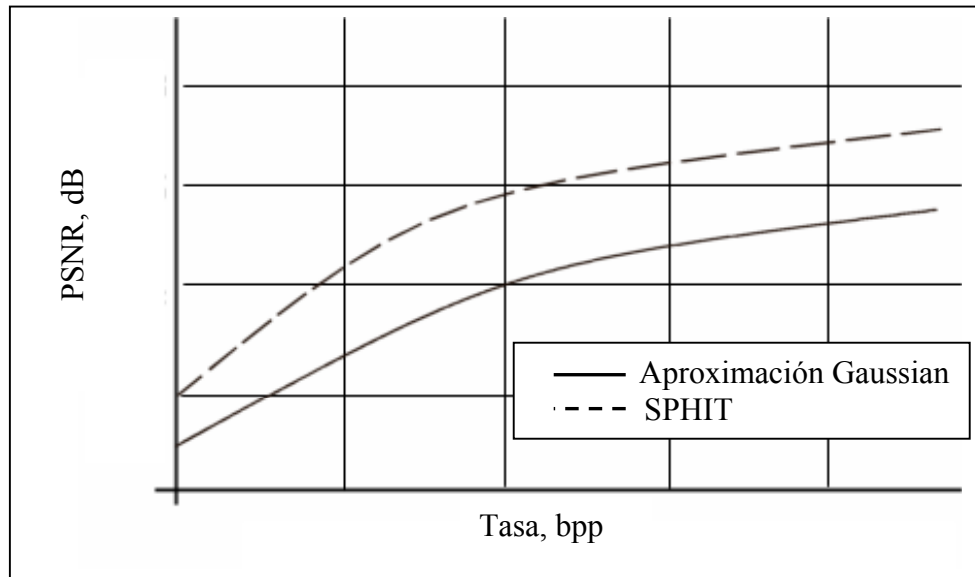
Como se menciono anteriormente, idear un buen algoritmo de codificación de imágenes involucra dos componentes intelectuales importantes:

- i) Seleccionar un modelo o framework operacional seguro
- ii) Esforzarse para optimizar la performance de codificación en el framework seleccionado.

Nos concentraremos en el último componente, pero también nos gustaría dar énfasis brevemente en la importancia del primero. Para este fin, nos enmarcaremos sobre un aventuramiento experimental estático que esperamos pueda ayudar a clarificar las características más sobresalientes de los modelos desarrollados. Nuestra meta es poder visualizar la efectividad de los modelos por sintetizado de imágenes derivado de la realización típica asumida en estos modelos. Las ventajas de cada uno de los ejercicios son difíciles de ver, no pudiendo exponer las virtudes y limitaciones de los diferentes atributos de los modelos de prueba, pero adicionalmente pueden inspirar la creación de un nuevo y mejorado framework que pueda embellecer alguno de los inconvenientes.

Mientras los algoritmos de codificación prácticos han sido basados en una variedad de framework operacionales, un modelo estático simple ha sido muy popular. Por ejemplo, recientemente el framework state-of-the-art de codificación de las subbandas de una imagen estaba basado sobre modelos i.i.d para las subbandas de una imagen (basado sobre Gaussian o Laplacian p.d.f.s) y técnicas de optimización de alocación de bits para asegurar que los bits estén distribuidos óptimamente entre las subbandas en proporción a su importancia, como su espigamiento a través de las variaciones de su distribución. Un algoritmo de codificación basado en semejante framework sería indudablemente muy eficaz al incorporar la segunda componente mencionada en el párrafo previo, pero esto traería la pregunta obvia acerca de cuan exacto podría ser un modelo i.i.d. Laplacian o un modelo de Gaussian subband. Permítanos intentar encaminar esta pregunta tomando una descomposición en subbandas de una imagen típica, midiendo las varianzas empíricas de las diferentes subbandas, y modelando esas subbandas como distribuciones i.i.d. Gaussian. La figura de abajo muestra la performance de la tasa de distorsión teóricamente obtenible usando complejidad infinita con el estilo Shannon involucrando argumentos asintóticos aleatorios de codificación basados en vectores de muestras infinitamente largos. Aún así, como se ve en la figura, este codificador es hermosamente realizado por un codificador de imágenes de una baja complejidad como el codificador Shapiro's EZW o proveyendo una variante del codificador SPIHT; por ejemplo, para una tasa de codificación de 0.5 BIT por píxel, el codificador SPIHT realiza la i.i.d Gaussian basada en un codificador fuente Shannon por sobre lo 3 dB in SNR. Uno puede concluir que el codificador SPIHT de baja complejidad, a pesar de que ninguna demanda de optimización de la tasa de distorsión fue

formulada, esta basado sobre un más exacto modelo de imagen que el codificador hipotético de Shannon, y esto hace toda la diferencia.



Las curvas de las tasas de distorsión que se alcanzan con (i) el codificador SPIHT y (ii) con el límite de la tasa de distorsión de Shannon correspondiente.

Trataremos de sintetizar la imagen derivada de los parámetros estáticos para el modelo asumido. Como antes, tomaremos una descomposición en subbandas de la imagen, midiendo las varianzas de las subbandas, entonces crearemos una realización aleatoria usando un modelo i.i.d. Laplacian para las subbandas de la imagen, y luego sintetizaremos la imagen basados en la descripción de las subbandas. Si uno asume una muestra aleatoria para las magnitudes de los coeficientes, uno realiza una imagen bizarra. Si se asume que la magnitud esta en distribución Laplacian pero la muestra de la variable aleatoria es conocida, entonces uno sintetiza una imagen de una mejor calidad a la anterior.

Trataremos ahora un modulo más lógico que modele las subbandas de la imagen como las distribuidas con Laplacian, dado por un modelo i.i.d. involucrando vecindad local (digamos ventanas de 3x3). Cada modelo captura el espacio y la caracterización de la frecuencia de la descomposición de la imagen (si bien utiliza un gran número de parámetros). Entonces se pueden representar dos imágenes de forma diferente que muestran la imagen sintetizada basada en los parámetros estáticos derivados de la imagen original correspondientemente al no conocimiento y al conocimiento de la información de la muestra, respectivamente. Esta última, aventaja en una mirada mucho más natural de la imagen y muestra lo promisorio del modelo. De hecho, el codificador EZW, SPIHT y otros pueden ser conceptualizados como el comienzo.

Así mediante la interpretación de los datos como cálculos de “frecuencia-espacial” de la información, ellos pueden derivar las performances significativas alcanzadas contra los codificadores tempranos de subbandas que trataban a los datos solamente como conjuntos de frecuencias de la información.

La tasa de distorsión conoce a la tasa y la distorsión: Las tasas de distorsión Operacionales en el Diseño de Codificadores Prácticos

Como lo antes discutido, la performance de la tasa de distorsión es la preocupación fundamental en el diseño de sistemas de compresión con pérdidas. Nosotros hemos subrayado como la investigación teórica fundamental ha resultado en la computación de los límites de la performance, pero también indicó dos preocupaciones mayores con las referencias teóricas de la tasa de distorsión:

- 1) Complejidad (¿Cuanta memoria, demora (delay) o computación es requerida? Podemos construir un algoritmo práctico para acercarse al límite)
- 2) Errores del modelo (¿Cuan buena son las cosas que asumimos en el modelado? ¿Estas son demasiado simple para caracterizar la fuente en su totalidad?)

Eligiendo los parámetros de un sistema concreto: tasa de distorsión Operacional

Para garantizar que nuestro diseño será práctico, podemos abandonar nuestra búsqueda de la mejor performance sin restricciones de algún sistema. En cambio, comenzaremos por elegir un esquema específico de codificación que capture eficientemente las dependencias estáticas asociadas con la fuente, mientras al mismo tiempo satisface los requerimientos de nuestro sistema de complejidad de codificación, memoria y delay. Entonces podremos buscar los mejores puntos operacionales para este sistema específico.

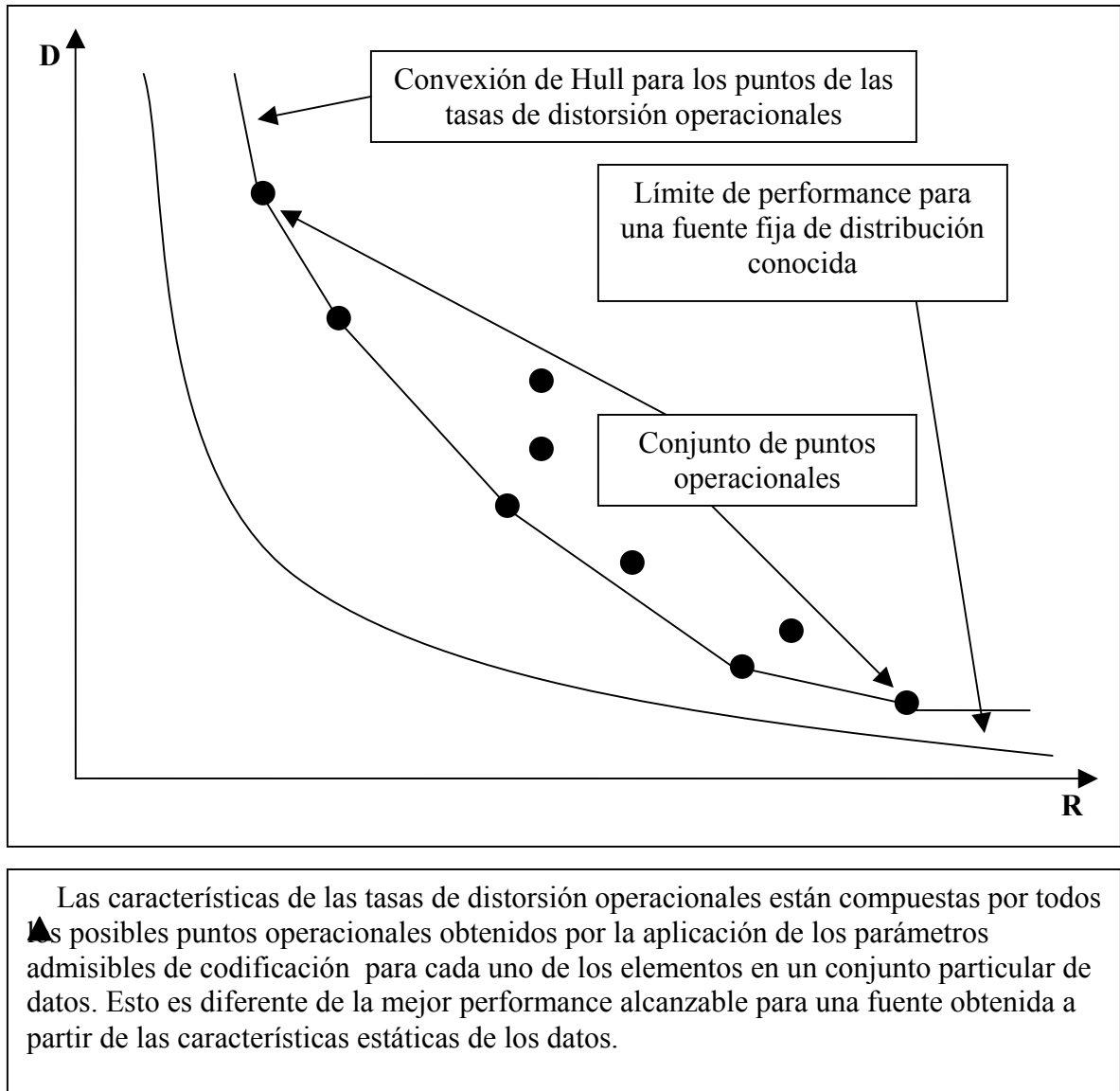
Para un sistema dado y una fuente dada, si podemos considerar todas las posibles opciones de cuantificación, podemos definir una curva operacional de la tasa de distorsión. Esta curva es obtenida por la puntuación para cada tasa de distorsión alcanzada por el diseño de los pasos del codificador/decodificador para la tasa. Hay que notar que los puntos anteriores son operacionales, son directamente alcanzables con la elección de la implementación y con un conjunto de datos de pruebas dados. Este límite permitirá distinguir entre el mejor punto operacional alcanzable y aquellos que son menos óptimos o inalcanzables. Mientras el límite dado por la función de la tasa de distorsión teórica de Shannon no da un procedimiento constructivo para conseguir que la performance óptima, en los casos que se necesita alcanzar una tasa de distorsión operacional, nosotros siempre tratamos con puntos alcanzables.

De ahora en adelante, nosotros consideraremos optimalidad en el sentido operacional, por ejemplo, la mejor performance alcanzable para una fuente dada (como la descrita por un conjunto de entrenamiento para un modelo estático) dándonos a elegir un framework de compresión.

Primero consideraremos en más detalle el problema del diseño óptimo del codificador/decodificador. El diseño del mecanismo básico puede resumirse como sigue: Primero debemos seleccionar un framework de compresión particular, entonces una vez elegido el framework de compresión particular procederemos al diseño del codificador y el decodificador.

En primer lugar, el codificador es optimizado por el decodificador dado, por ejemplo, dando los niveles de reproducción que el decodificador maneja, un codificador es

diseñado para que produzca un mapeo teniendo la mínima distorsión para una tasa dada, para la fuente de entrenamiento especificada. Entonces, el decodificador es optimizado para el codificador dado, por ejemplo, una vez que la entrada ha sido asignada a índices, elegiremos la mejor reproducción para un conjunto particular de índices. Este proceso de diseño itera entre estos dos pasos hasta que converja.



Eligiendo un Buen Modelo: El paradigma de la Transformación-Codificación

En nuestra discusión hasta ahora, “lo mejor” no ha sido definido solamente para un framework particular, esto también es aplicable a la fuente dada, como la especificada por

un modelo probabilístico o por un conjunto de entrenamiento de los datos representativos. Desde algún modelo probabilístico aplicable habremos de derivar un conjunto de entrenamiento a partir de la fuente, sin pérdida de generalidad y asumiendo que las fuentes son representadas por estos conjuntos de entrenamiento.

Como se ha visto que los modelos son propiedades inherentes de las fuentes a ser comprimidas y por lo tanto el diseñador del codificador tiene poca flexibilidad en la selección del modelo. Nada, ha decir verdad. En los hechos una parte fundamental del diseño del codificador es la selección del modelo subyacente y de hecho muchas elecciones son posibles típicamente. Cada una de estas aproximaciones puede modelar el mismo dato original, pero varían ampliamente en los resultados de compresión.

En primer orden de aproximación, un buen sistema de compresión basado sobre complejos modelos tiende a ser más complejo de implementar (pero puede proveer una mejor performance) que un sistema basado sobre modelos simples. Una ilustración simple de esta regla puede ser vista cuando comparamos los cuantificadores escalares contra uno vectorial. Así, la principal dificultad es hallar un modelo que sea:

- Lo suficientemente simple para que un sistema de compresión encaje con ese modelo pudiendo alcanzar una buena performance con costo razonable.
- Pero lo suficientemente complejo para capturar las características principales de la fuente.

Por ejemplo, usando un modelo i.i.d. simple para cada subbanda como base para el diseño de nuestro codificador, resultaría muy pobre la performance que alcanzaría el codificador porque este no explotaría la redundancia espacial existente.

Sin embargo, los (aun simples) modelos que asumen correcta la varianza local y la información de la señal pueden ser vistos para capturar una gran parte de la información de la imagen y de hecho los modelos de este tipo son la razón fundamental de muchos estados de la onda ART para los codificadores de imágenes.

Hay muchas aproximaciones para alcanzar este objetivo dual de “encajar sin complejidad excesiva”, la mayoría de ellos basados en un principio simple: “reemplazar un único modelo por una multitud de modelos simples”. Aquí se detallan un caso particular de este principio, llamado transformación-codificación.

El paradigma transformación-codificación dicta que se debe descomponer la fuente en sus frecuencias componentes usando bloques transformadores como la transformada discreta del coseno (DCT) o codificación de subbandas usando los populares filtros wavelets. Así, considerar la relación de la tasa de distorsión en el dominio de la transformada. Del punto de vista del modelado este tiene la principal ventaja de permitirnos usar modelos simples.

Esta aproximación es extremadamente usada pudiendo verificarse por como esta ha sido usada ampliamente en los estándares de codificación de JPEG2000 Y MPEG1/2/4.

Cuando consideramos fuentes como el video, necesitamos herramientas adicionales que nos permitan explotar completamente la redundancia entre frames consecutivos en una secuencia de video. La compensación de movimiento es la aproximación más popular para alcanzar este objetivo. Los codificadores procesan los parámetros de movimiento (por ejemplo los vectores de movimientos en MPG) y el decodificador usa estos en la reconstrucción del frame original. Un framework particular especifica como la información del movimiento es transmitida y como esta debe ser interpretada por el decodificador.

Un ejemplo del paradigma de la Transformación-Codificación

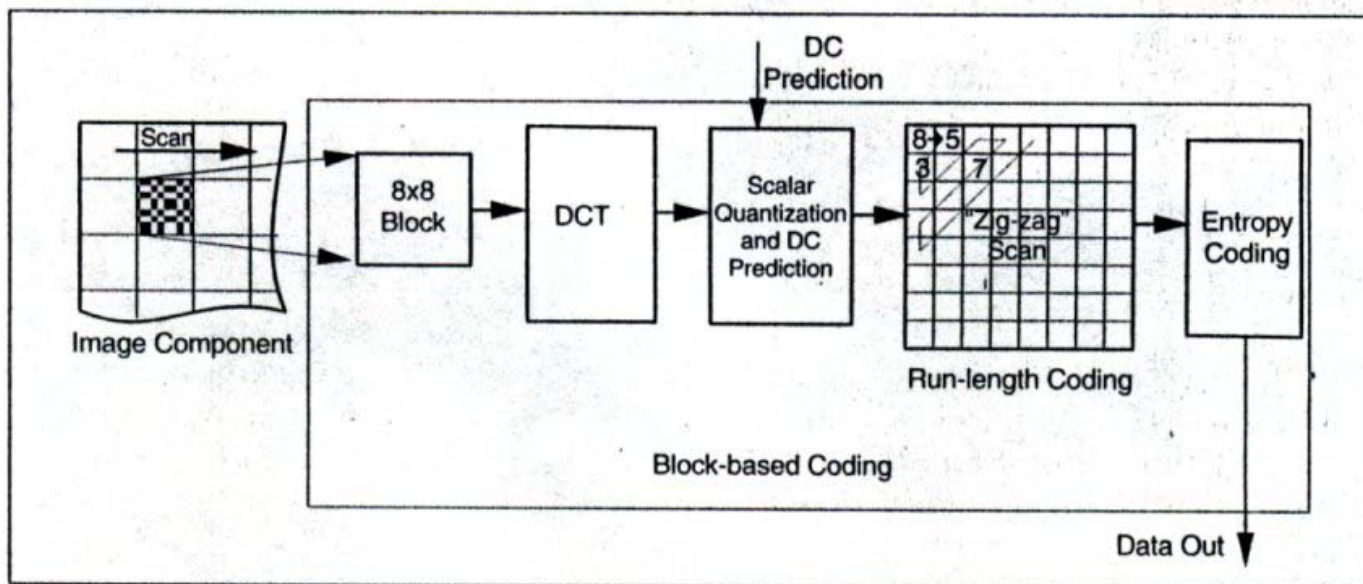
El termino transformación codificación generalmente describe técnicas de codificación donde la fuente es primero descompuesta usando una transformación lineal y donde cada una de las frecuencias componentes obtenidas de la descomposición son entonces cuantificadas.

Un codificador típico de imágenes basado en realizar transformaciones comprende una transformación lineal front-end seguida por una etapa de cuantificación escalar y una codificación entropica. La transformación cumple un rol dual:

- i) Compactación de energía, así el grueso de la energía de la señal es aislada en una pequeña fracción de los coeficientes de la transformada
- ii) Decorrelacion de la señal, así hay poca perdida de performace debido a la cuantificación escalar simple: esto es posible porque el conjunto de todos los coeficientes de la transformada representan una frecuencia dada pudiendo ser, a priori, modelada como una fuente que ocupa menos memoria para la cual se pueden hallar cuantificadores simples y eficientes.

La cuantificación escalar es la parte del framework donde se registran perdidas y confina la representación a un conjunto discreto de índices que corresponden a niveles discretos de cuantificación, mientras que la última etapa, la codificación entropica, remueve la redundancia en el stream de los índices de la cuantificación.

Comercialmente los estándares de compresión de imágenes y video están basados en la transformada discreta del coseno (DTC). La figura a continuación provee un ejemplo del más popular modo de operación, llamado “baseline”, con el cual opera el estándar de compresión JPEG.



A continuación se detalla una breve descripción del algoritmo de codificación JPEG:

La imagen se descompone en bloques de 8×8 con el propósito de emprender el proceso de transformación, cuantificación, y por último la codificación entropica. Los bloques son scaneados y transformados independientemente unos de otros usando un bloque DCT. Luego, cada bloque de 8×8 es cuantificado usando una cuantificación escalar uniforme. La duración del paso de cuantificación es definida por cada uno de los 64 coeficientes de frecuencias usando una matriz de cuantificación de 8×8 . Típicamente, una sola tabla de cuantificación es usada para cada una de las componentes de color; sin embargo, hay casos donde se pueden utilizar arriba de cuatro tablas diferentes si es necesario. En este último paso, es donde se registra la pérdida de información y por lo tanto la información codificada no podrá ser recuperada perfectamente por el decodificador. Sin embargo, esta operación de cuantificación permite alcanzar altas tasas de compresión pagando el precio de alguna degradación en la calidad.

El primer coeficiente de frecuencia cuantificado es llamado coeficiente DC, representa el promedio de los valores en el bloque y es precedido a partir del bloque previo codificado. Solo la diferencia con el coeficiente DC anterior es codificada, la cual ocupa mucho menos espacio que el valor absoluto del coeficiente. Los restantes 63 coeficientes de frecuencia, llamados coeficientes AC, son codificados usando solamente los valores correspondientes del bloque.

La codificación entropica constituye el segundo componente básico en la relación con la tasa de distorsión, como esta determina el número de bits que serán usados para una imagen en particular y el seteo de la cuantificación. La codificación entropica es con pérdidas y esta mapea cada uno de los varios índices de la cuantificación en códigos dados. Una manera simple de compactar el stream de índices es asumir que el modelo ocupa menos memoria para los índices y comprimirlos llevando al stream al primer orden de entropía. Sin embargo, asumir menos memoria para los índices es típicamente una mala idea, y la ganancia significativa puede ser obtenida explotando la memoria en el stream de bits cuantificado. Una manera simple de explotar esto es a través de codificar la cantidad de ceros. JPEG usa una codificación basada en una entropía bidimensional sobre la cantidad de ceros y la magnitud de los coeficientes distintos de cero. Esto se logra utilizando la codificación Huffman o la codificación aritmética. A pesar de la aparente rigidez de la sintaxis JPEG, es sorprendente la cantidad de espacio que se obtiene de ganancia con la brillante optimización de la codificación. La sintaxis permite a la matriz de cuantificación y la tabla de codificación entropica ser adaptadas en cada una de las imágenes procesadas para obtener la tasa de compresión deseada. Una posible opción más sutil puede ser aplicada en la codificación para engañar la óptima decodificación en cuanto a la tasa de distorsión mientras mantiene total compatibilidad en la sintaxis. Un ejemplo de esto es tener unos pequeños valores distintos de cero que rompen una potencial cadena enorme de ceros, esto es muy caro en términos de la tasa de compresión con su relativa contribución a reducir la tasa de distorsión. Si el codificador se puede “mentir” al decodificador acerca de esos coeficientes y entonces a esos valores distintos de cero los puede hacer pasar como valores iguales a ceros. Luego la tasa de distorsión obtenida no es la más mala de todas, pero la performance es significativamente incrementada.

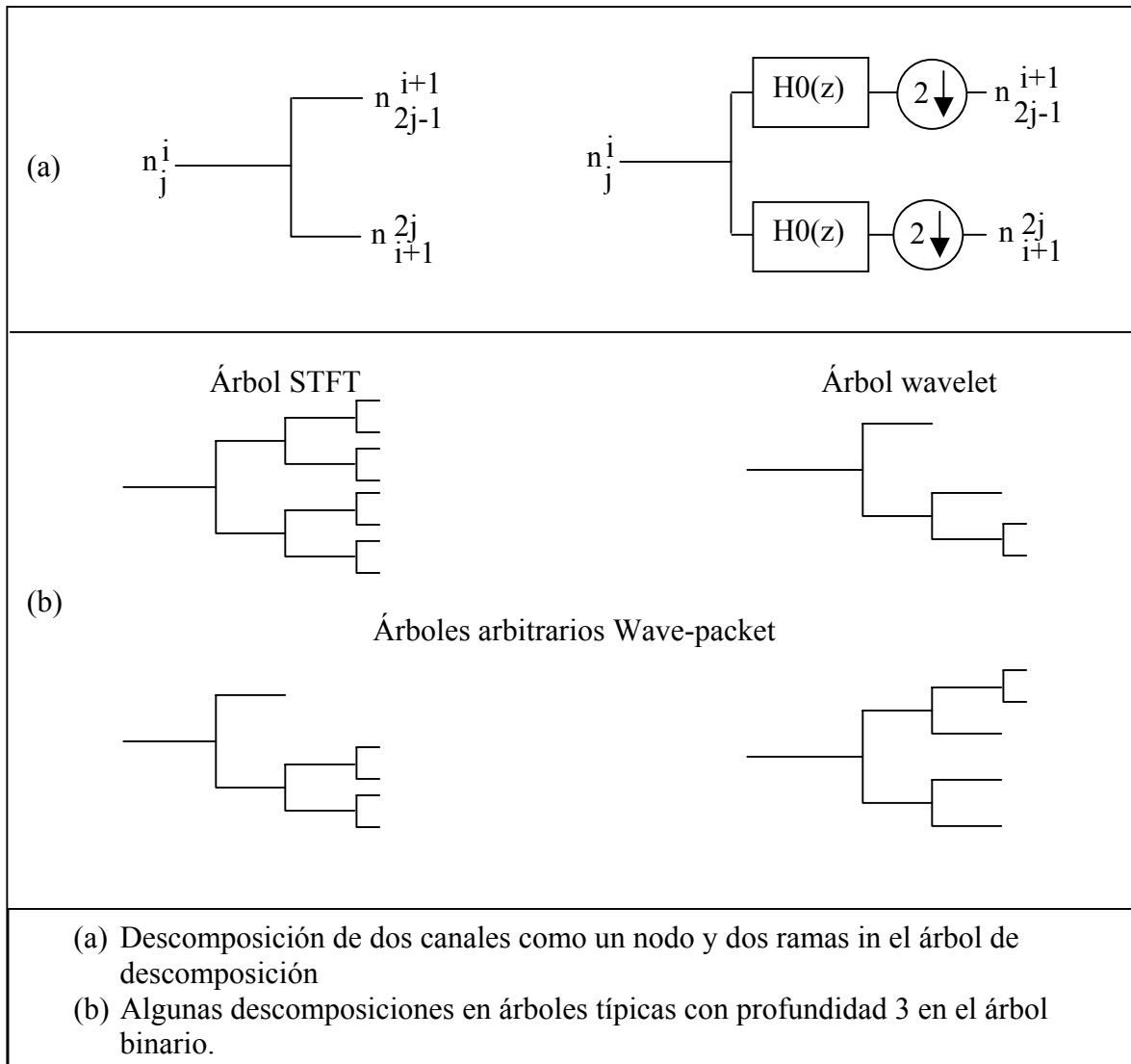
Transformaciones adaptativas basadas sobre expansiones wavelet

Adaptación en procesamiento de señales es una de las cosas más poderosas y deseables, y las aplicaciones de compresión no son la excepción. Aquí perfilaremos el rol de tener transformaciones adaptativas en codificaciones de imágenes wavelet. Mientras la DCT es la transformación elegida en los estándares comerciales de compresión como JPEG, MPEG, etc. La transformación discreta wavelet (DWT) ha emergido recientemente como una alternativa superior y con miras a reemplazar a la DCT.

La descomposición en subbandas o wavelet consiste de una descomposición de frecuencias en un octavo de banda. Esto es implementado por un banco de filtros con estructura de árbol consistiendo de una cascada de filtros lowpass ($H_1(z)$) y filtros highpass ($H_0(z)$) seguidos por el por el paso de los diezmadores (ver figura inferior (a)). Por división recursiva de las salidas de las ramas lowpass, se realiza una descomposición de frecuencias de un octavo de banda asociada con una estructura de árbol logarítmica de los bancos de filtros (ver figura inferior (b)). Esta propiedad de la descomposición logarítmica de frecuencias de los wavelet da una buena selectividad de frecuencias en las frecuencias bajas y una buena selectividad temporal (o espacial) en las altas.

El problema de alocação de recursos para el framework de codificación de transformada adaptativa involucra seleccionar el punto de operación de la combinación de la transformación, cuantificación y codificación entropica en orden para realizar la mejor relación con respecto a la tasa de distorsión. Dependiendo de la flexibilidad del framework, una o todas las componentes funcionales antes mencionadas pueden ser optimizadas todas juntas o no. La aproximación tradicional en la compresión es usar una transformación fija (como DCT o la DWT) y luego seleccionar una estrategia de cuantificación que concuerde con las propiedades de la entrada del proceso y la transformación fija. La estrategia de cuantificación esta basada típicamente en un modelo para las funciones de densidad de probabilidad caracterizando los coeficientes de la transformación, o en la ausencia de un modelo analítico, se basa en el entrenamiento sobre una gran clase de señales posibles consideradas “típicas”.

Como primer paso hacia el logro de una transformación adaptativa, es claro que una mejora puede ser implementada cuando buscamos que conjunto de árboles binarios utilizar para un conjunto particular de filtros, en lugar de usar un árbol fijo de la transformación wavelet (ver figura). Un algoritmo rápido, conocido como algoritmo de “árbol simple”, halla el mejor árbol juntamente con la mejor cuantificación y estrategia de codificación entropica, la idea es buscar el mejor árbol (para una función de costo de la tasa de distorsión apropiada para la compresión) para la señal, esta búsqueda se realiza sobre una librería de árboles wavelets.



Codificación basada en estándares: La Sintaxis necesaria para optimizar la tasa de distorsión

Hasta ahora, hemos considerado el diseño completo de un sistema de compresión en el cual para un conjunto dado de construcciones, buscamos algoritmos de codificación y decodificación que se ajusten a nuestras necesidades. Ahora, asumamos que el decodificador ha sido seleccionado; esto significa que tenemos una completa especificación del lenguaje que puede entender el decodificador, con una descripción que acompaña a la salida que es producida de una entrada dada.

Este escenario es precisamente el asumido por los más recientes estándares de compresión internacionales. Motivados por el deseo de maximizar la interoperatividad, estos estándares proveen un acuerdo sobre la sintaxis del bitstream que algunos

decodificadores puede usar para proveer la señal de salida. El acuerdo en cada estándar permite codificar/decodificar productos de diferentes proveedores.

Aún, esto no está claro a priori, cuánta flexibilidad puede gozar el codificador a la hora de seleccionar sus modos de operación, si es restringido a un decodificador particular. La mayoría de los estándares son diseñados para dotar al codificador con una gran flexibilidad y creatividad en sus selecciones de los parámetros del sistema, y esto es un gran problema en cuanto a la performance entre la mejor opción y la peor. En todas las aplicaciones basadas en estándares el codificador puede seleccionar ciertos parámetros que resultaran en varios niveles de performance de la tasa de distorsión. Esto nos lleva a una situación donde el número de puntos de operación es discreto, y por lo tanto los límites operacionales de la tasa de distorsión es determinada por la convexión de Hull del conjunto de todos los puntos de operación. Por ejemplo, cuando se usa compresión JPEG se puede optar por diferentes tasas de calidad para una imagen en particular y aún se garantiza que esta podrá ser decodificada.

Así, la tarea del codificador de seleccionar el mejor punto de operación de un conjunto discreto de opciones acordadas anteriormente, a priori por una regla fija de decodificación, es usualmente referida como optimización de la sintaxis necesaria. La opción de operación seleccionada es comunicada por el codificador al decodificador como información interna, y típicamente como parte del header.

Si consideramos que un framework general de codificación ha sido seleccionado, el cual puede acomodar diferentes tipos de fuentes, y así los parámetros específicos pueden ser seleccionados para cada imagen (y para cada tasa deseada) y enviarlos al decodificador como overhead. Entonces nuestra meta es:

Formulación 1: optimización discreta de la tasa de distorsión Selección de parámetros par una entrada dada.

Dado un framework específico de codificación donde el decodificador está totalmente definido, optimizando la codificación de una imagen particular, o secuencia de video en orden de alcanzar algunos objetivos de tasa de distorsión.

Notar que aquí asumimos conocimiento determinístico de la entrada y nuestra meta es optimizar la selección de parámetros para esa entrada. No buscamos más optimalidad sobre un conjunto de entradas, pero esto nos confina bastante a hacer lo que consideramos lo mejor para la entrada dada, teniendo impuesta la necesidad que dicta el framework de codificación. Este es un escenario muy real, como en la mayoría de los estándares de compresión de imágenes y video definidos hasta ahora, donde la selección del modo de codificación puede ser optimizada para cada entrada. La ganancia registrada utilizando estas técnicas es significativa.

Sin embargo, la selección del framework inicial de codificación es la clave para la performance del sistema. No importa cuán sofisticadas sean las técnicas de optimización utilizadas, si el framework de codificación es limitado o defectuoso, no hay muchas mejoras que se puedan realizar. Si nos situamos en un framework con tasa de distorsión operacional y nos limitamos solamente a aquellos puntos operacionales de tasa de distorsión entonces el framework inicial es realizable.

En el espíritu de la tasa de distorsión operacional, definimos la solución óptima como aquella que alcanza la mejor función sobre todos los posibles puntos de operación. Notar que definimos conjuntos finitos de opciones de codificación para el codificador y por

ende existe una opción de parámetros optima realizable: si todos fallan, uno puede hacer una comparación exhaustiva de todos los posibles puntos de operación y elegir el mejor. Obviamente, nuestra meta será hallar aquellos puntos de operación sin realizar una búsqueda de fuerza bruta.

Transmisiones con demora (o delay) restringida y control de buffer

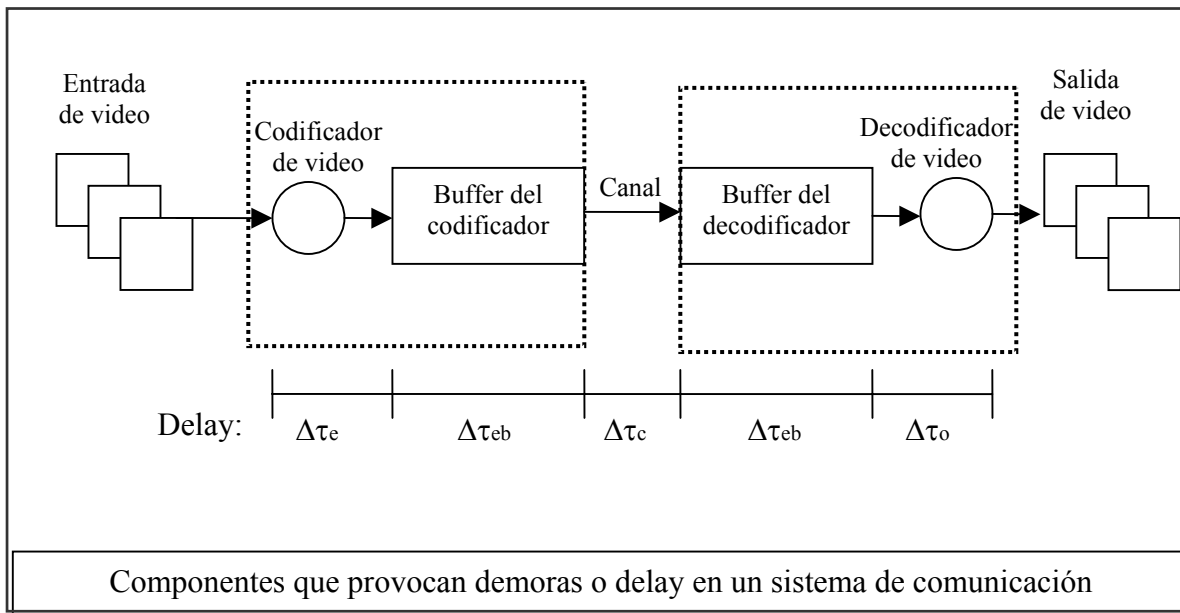
La mayoría de los algoritmos de compresión de video explotan la redundancia espacial y temporal a través de la transformación codificación y estimación de movimiento, respectivamente. Sin embargo, el grado de redundancia y por consiguiente la resultante tasa de distorsión, puede fluctuar ampliamente de escena a escena. Por ejemplo, escenas con alto contenido de movimiento requerirán más bits que las escenas estacionarias.

Si consideramos una transmisión típica en tiempo real, como la que ilustra la figura de abajo. Como se describió, los frames del video requerirán una tasa de bits variable y por consiguiente será necesario tener buffers para los codificadores y decodificadores para suavizar las variaciones en la tasa de bits. Asumiendo que tenemos dispositivos de captura para la entrada de video y la salida, los frames se muestran a una tasa constante, y no hay pérdida de frames durante la transmisión, es fácil ver que la demora o delay de fin-a-fin en el sistema permanecerá constante.

Si llamamos $\Delta\tau$ al delay de fin-a-fin: un frame codificado en un tiempo t tiene que ser decodificado en un tiempo $t + \Delta\tau$. Esto impone una restricción sobre la tasa que puede ser usada para cada frame (esta tiene que ser lo suficientemente baja para que la transmisión pueda garantizar que este dentro del delay impuesto).

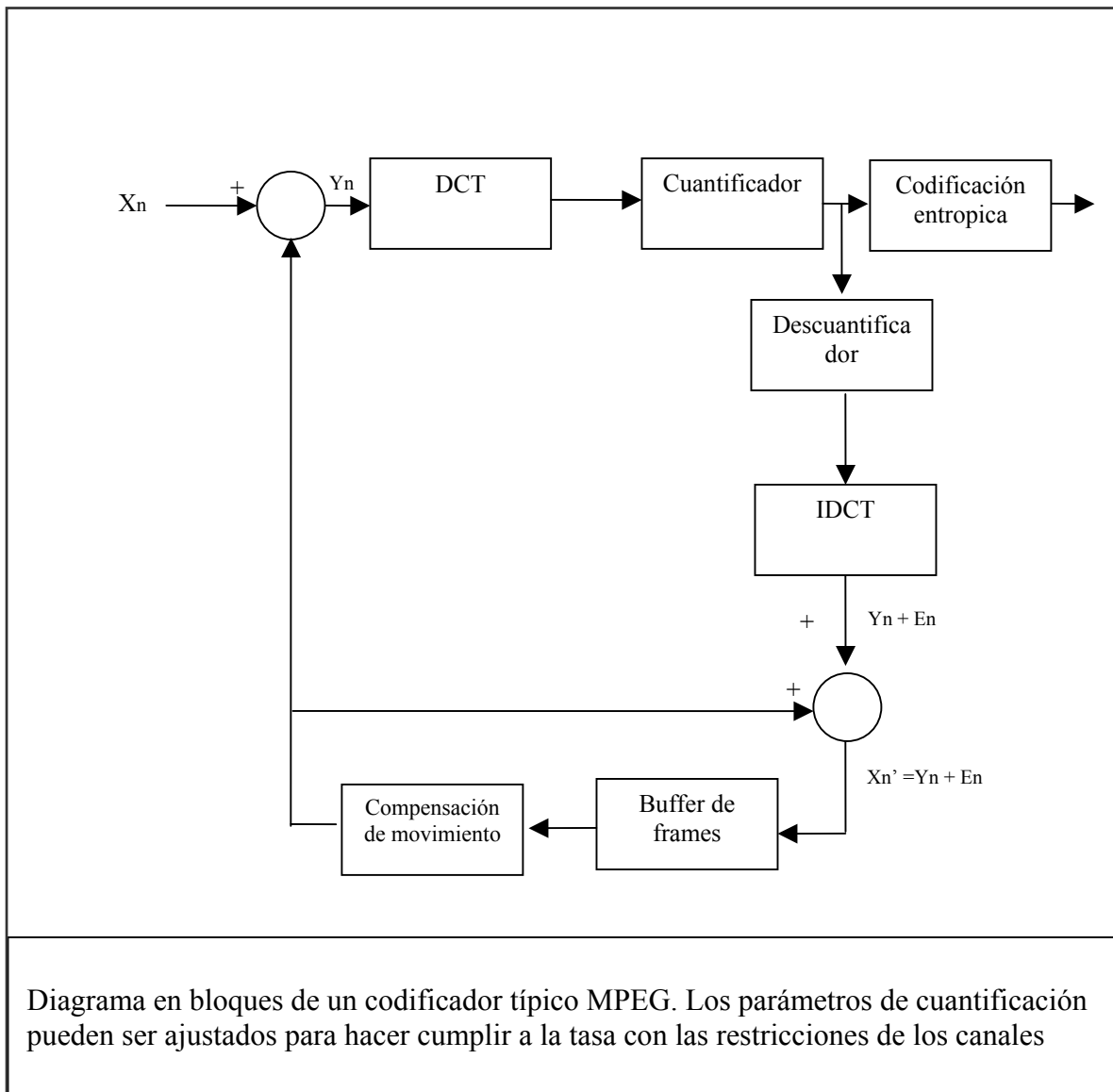
Considerando el caso cuando la transmisión toma lugar sobre un solo canal CBR. De los componentes de delay de la figura solamente $\Delta\tau_{eb}$ y $\Delta\tau_{ed}$ (el tiempo insumido en la codificación del buffer y la decodificación del mismo, respectivamente) serán ahora variables. Si consideramos, por ejemplo $\Delta\tau_{eb}$, este delay será a lo sumo B_{\max} / C , donde B_{\max} es la capacidad del buffer para el codificador y C es la tasa del canal en bits por segundo. Es claro que B_{\max} ha de ser menor que $\Delta\tau \cdot C$ o de lo contrario podríamos tener almacenamiento de frames en el buffer, con lo cual experimentaríamos demasiado delay.

Si consideramos la transmisión de una secuencia donde la tasa cambia varias veces cerca de un orden de magnitud cuando el tamaño del paso de cuantificación se mantiene constante, tenemos dos opciones (i) tendremos que usar buffers muy grandes (y la correspondiente gran demora de fin-a-fin que genera) o (ii) tendremos que ajustar la tasa de la fuente, y así resignar calidad, para hacer posible el uso de un buffer mas pequeño (con una demora pequeña) sin perder ningún dato. Los delays requeridos para transmisiones de este tipo podrán ser excedidos por mucho, por lo tanto, en las aplicaciones practicas es necesario realizar un control de la tasa para ajustar los parámetros de codificación y respetar las restricciones con respecto a la demora permitida.



Como se muestra en la figura inferior, es posible ajustar la tasa de video (y la calidad) para modificar el tamaño de los pasos de cuantificación usados para cada frame. Es por lo tanto fácil de ver que el problema del control de tasa puede ser interpretado como un problema de asignación de recursos donde el objetivo es determinar cuantos bits usar en cada parte de la secuencia de video y así maximizar la calidad entregada al usuario final.

Notar que en los casos donde la transmisión es realizada sobre canales VBR o donde la secuencia es precodificada y almacenada (como en un DVD), también es necesario realizar una asignación de la tasa. Por ejemplo, para almacenar una película completa en un DVD es necesario preanalizar la película y entonces alocar apropiadamente la tasa destino de las partes de nuestra película. De esta manera, alocamos más bits para las escenas más demandantes y unos pocos para las escenas menos demandantes, lo cual resulta en una calidad global uniforme.



Técnicas de tasa de distorsión para codificación conjunta de la fuente y el canal

Mientras las técnicas basadas en minimizar la tasa de distorsión han tenido un mayor impacto en los frameworks de compresión de imágenes y video, su utilidad se extiende más allá de la compresión y alcanza un framework mucho más grande que es el de los sistemas de transmisión de imágenes y video. El problema de transmitir señales de imágenes y video involucra naturalmente la codificación de la fuente y la codificación del canal. La fuente de imagen o video tiene una tasa de distorsión característica asociada que cuantifica la relación óptima entre la compresión en forma eficiente y la distorsión resultante. El objetivo clásico de la codificación de la fuente es operar lo más cercanamente posible de este límite fijado para la tasa de distorsión. Luego sobreviene la tarea de realizar

una transmisión fiable de la fuente codificada sobre un canal con ruido que es caracterizado por una capacidad de canal que cuantifica la tasa máxima a la cual la información puede ser transmitida con fidelidad a través del mismo. El objetivo clásico de la codificación del canal es entregar la información a una tasa que sea lo más cercana a la capacidad del canal. Para comunicaciones punto-a-punto sin restricciones de demora (o delay), uno puede teóricamente separar las tareas de codificación de la fuente y el canal conllevando a no tener pérdidas de performance alguna. Esta importante información teórica, nos lleva al celebrado principio de separación de Shannon, con lo cual permite diseños separados de los esquemas de compresión/descompresión de una fuente y los esquemas de codificación/decodificación de un canal. El principio de separación es ilustrado en la figura inferior.

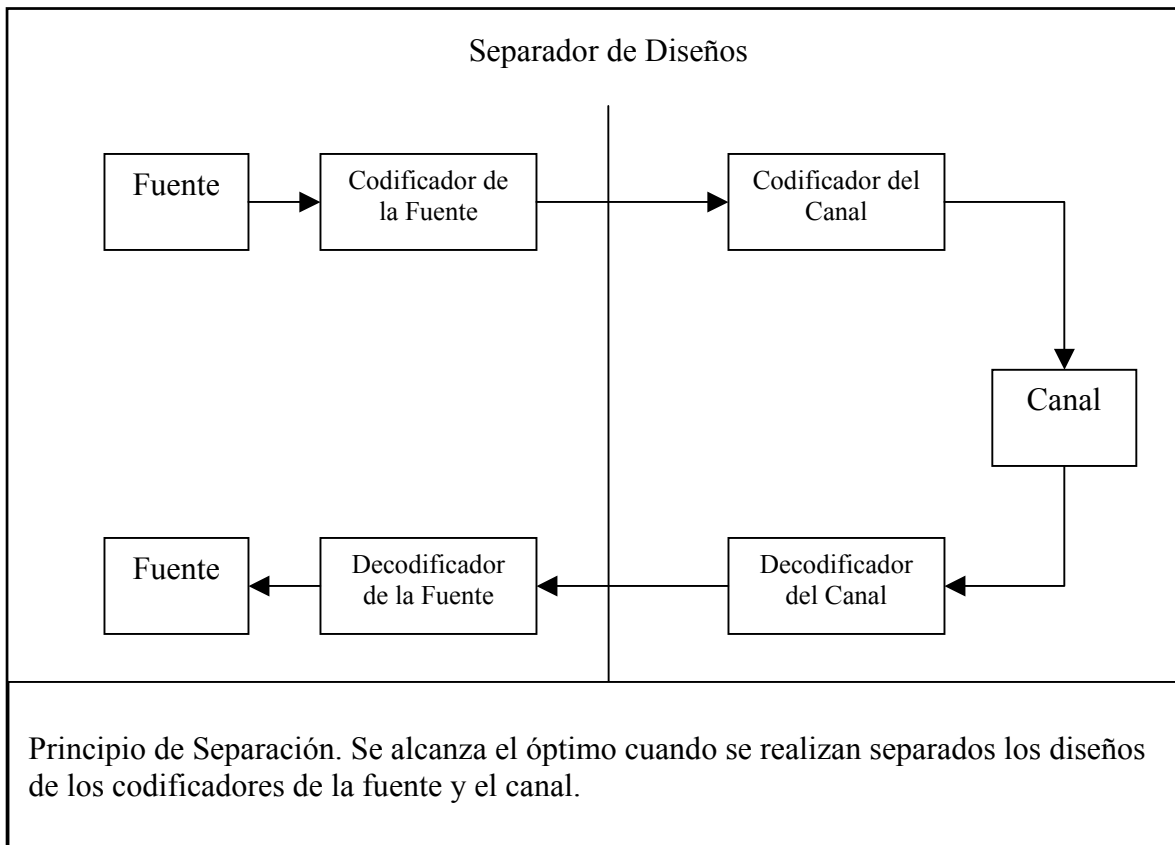
Es importante recalcar que la información teórica realiza dos importantes supuestos:

- i) El uso de longitudes de bloques arbitrariamente largos para codificar la fuente y el canal
- ii) Capacidad de recursos computacionales arbitrariamente altos (y asociados con el delay)

Es obvio que en la práctica tales supuestos no están presentes, debido a que se fijan restricciones de demoras y límites prácticos de los recursos de computación. Para ciertos escenarios de comunicación multiusuario como el broadcast y el multicast el teorema de la separación no es aplicable y se necesita una interacción más cercana entre los componentes de codificación de la fuente y el canal.

Formularemos el problema básico de la codificación conjunta de la fuente y el canal. Es importante notar que hay muchos aspectos para este problema que dependen del número de elementos que son diseñados juntos y que están involucrados en el problema. Si solo el codificador de la fuente y el dispositivo de corrección de errores están involucrados, los recursos a ser alocados son utilizados por la tasa total de bits entre el codificador de la fuente y el canal. Si el modulador-demodulador (MODEM) está incluido en el cuadro de optimización, entonces el poder o energía de transmisión puede tomarse como una restricción.

La función de costo del sistema, es típicamente la distorsión de fin-a-fin de la copia de la fuente entregada: debido a la naturaleza probabilística del canal, uno tiene que abandonar las métricas determinísticas de la distorsión que fueron formuladas con anterioridad y considerar las medidas esperadas de distorsión. La distorsión es generada por la cuantificación de la fuente, lo cual tiene un carácter determinístico de un conjunto discreto fijo de opciones de cuantificación, y el ruido del canal, el cual es obviamente de naturaleza probabilística. Esto contrasta con el uso de los métodos de tasa de distorsión para la codificación de la fuente, donde la cuantificación era la única fuente de distorsión.



Problemas típicos de Alocaación

Ahora se nos presentaran una serie de problemas genéricos que nos presagiaran alguna posible restricción que tendrá el codificador cuando realice la selección de los parámetros. Esta descripción de los problemas esta dividida en dos clases dependientes, de si la compresión es orientada al almacenamiento o a la aplicación de transmisión. Seguido presentaremos estos problemas:

- Selección de la unidad básica de codificación: hasta ahora hemos considerado una relación genérica de la tasa de distorsión donde la unidad de codificación, sea esto un ejemplo, un bloque de una imagen o una imagen, es codificada con cierta distorsión cuando una tasa particular es seleccionada. En un escenario practico, esto será necesario para decidir a que nivel o niveles de granularidad se optimiza el proceso de de codificación. Por ejemplo, es posible considerar frames de video como las unidades básicas de codificación en un ambiente de codificación de video, así la tasa de medida de frames y de distorsión por cada frame en la secuencia y entonces decidir los puntos de operación para los frames. Alternativamente uno puede operar a un nivel más fino y considerar opciones de codificación para un solo frame con la unidad básica de codificación siendo, por ejemplo, los bloques de 8x8 píxeles usados en JPEG.

- Complejidad: la complejidad es un factor determinante, que su importancia esta en ascenso en varias técnicas de optimización de la tasa de distorsión. Las dos mayores fuentes de complejidad pueden ser identificadas. Primero, los datos de la tasa de distorsión puede ser medidos de las imágenes y entonces muchas elecciones de codificación/decodificación pueden tener que ser realizadas para determinar los valores de la tasa de distorsión (Notar que esto no es consistente con muestra asunción de un framework de tasa de distorsión operacional, donde la meta es seleccionar el mejor de los puntos de operación de los que están disponibles). La segunda fuente de complejidad viene de la búsqueda, aun si los datos de la tasa de distorsión son conocidos o han sido adecuadamente modelados, tendremos que buscar el mejor de los puntos de operación y que en si misma puede ser una tarea complicada. En resumen, la complejidad no depende solamente sobre el número de operaciones requeridas, sino también del retraso de la computación de la solución óptima y, relacionado con esto, el almacenamiento requerido para el algoritmo de búsqueda. Obviamente, algoritmos más complejos pueden ser aplicados en aplicaciones de codificación off-line, en cambio en aplicaciones de tiempo real o on-line habrá que limitar el retraso en la codificación hasta hacerlo admisible. Algoritmos complejos pueden ser justificados si se busca mejorar la calida en escenarios donde la codificación es realizada una sola vez, pero la decodificación se realiza varias veces. Desde que los estándares como MPEG proveyeron un framework común de decodificación, es posible desarrollar codificadores que poseen un rango de escenarios, de alta calidad, de codificación profesional de alta complejidad a bajo costo, productos para el consumidor de baja calidad.
- Función de costo: la distorsión y la tasa pueden ser parte de las funciones objetivo a ser optimizadas. Las funciones objetivos pueden ser fácilmente computadas por cada unidad de codificación, pero cuando nuestro problema involucra la decisión sobre la aloación de un conjunto de unidades de codificación, definir la función de costo total requiere algo más. Por ejemplo si asumimos que la distorsión es nuestra función objetivo entonces habrá un montón de alternativas para definir la medida de distorsión total dando las medidas de distorsión individuales para cada unidad de codificación. Por ejemplo uno puede asumir que minimizar la distorsión promedio es un objetivo deseable. Pero consideremos una secuencia larga de video, ¿Es realmente apropiado tomar el promedio como medida de distorsión? Esta pregunta es valida y ella justifica que sea necesario reconsiderar el objetivo anterior.
- Notación: consideremos N unidades de codificación donde cada una tiene M diferentes puntos de operación posibles. Por cada unidad de codificación i tenemos información acerca de sus tasa t_{ij} y de sus distorsiones d_{ij} cuando usamos el cuantificador j . No haremos asunciones de ninguna estructura en particular para el r_{ij} y d_{ij} ; simplemente usaremos la convención que los índices de cuantificación son listados en orden ascendiente. Ninguna otra asunción será hecha. Consideremos aquí que los datos de la tasa de distorsión son conocidos, esto posibilita reemplazar las medida r_{ij} y d_{ij} con los valores que son estimados en base a los modelos, pero esto no modifica, ni afecta a algoritmo que se propone.

Restricciones de almacenamiento: asignación de presupuesto restringido

En la primera clase de problemas a considerar es la restricción de la tasa por algún tipo de impedimento sobre el número máximo de bits que se pueden usar. Este número de bits aplicables, o presupuesto R_t ha de ser distribuido a través de diferentes unidades de codificación con la meta de minimizar alguna medida de distorsión total. Por ejemplo nosotros podríamos querer usar JPEG para comprimir las imágenes en una base de datos de imágenes y también que todas ellas quepan en un disco de computadora. Este problema puede ser re planteado como sigue:

Formulación 3 – asignación de presupuesto restringido

Hallar el cuantificador óptimo, o punto de operación, $x(i)$ para cada unidad de codificación i tal que

$$\sum_{i=1}^N r_{ix(i)} \leq R_t \quad (1)$$

Y alguna medida $F(d_{1x(1)}, d_{2x(2)}, \dots, d_{nx(n)})$ que esta minimizada.

Por ejemplo, si nosotros estamos interesados en una distorsión promedio mínima (MMSE), tenemos que:

$$F(d_{1x(1)}, d_{2x(2)}, \dots, d_{nx(n)}) = \sum_{i=1}^N d_{ix(i)}$$

Alternativamente, una aproximación MiniMax (MMAX) puede ser tal que:

$$F(d_{1x(1)}, d_{2x(2)}, \dots, d_{nx(n)}) = \max_{i=1}^N d_{ix(i)}$$

Finalmente, la aproximación del óptimo lexicográfico (MLEX) ha sido recientemente propuesta como una extensión de la solución MiniMax. El MLEX compara dos soluciones por ordenamiento lexicográfico de sus distorsiones o sus índices de cuantificación. Por simplicidad, asumimos los índices del cuantificador que son usados en las comparaciones con $j = 1$ siendo el mejor cuantificador. Entonces, para comparar dos soluciones, nosotros ordenamos los índices de cuantificación de todas las unidades de codificación de mayor a menor: entonces comparamos la lista ordenada resultante y entonces decimos que el representado por el menor número es el mejor en el sentido del MLEX. Por ejemplo consideremos 4 unidades de codificación que reciben la siguientes dos asignaciones (1, 3, 4, 4) y (3, 2, 3, 2). Luego ordenamos y obtenemos (4, 4, 3, 1) y (3, 3, 2, 2) y dado que tenemos $3322 < 4431$, la segunda asignación es mejor que la primera en el sentido MLEX.

Alocación bajo múltiples restricciones parciales de presupuesto

Una visión más general del problema de la formulación 3 puede surgir en muchas situaciones donde no solamente estamos limitados sobre la tasa total, sino que también sobre la tasa posible para los subconjuntos de unidades de codificación. Asumiendo, por ejemplo que un conjunto de imágenes han sido ubicados en un dispositivo de almacenamiento que esta físicamente particionado (por ejemplo un arrays de discos) y que es imposible (o indeseable por razones de performance) dividir las imágenes a través de uno o mas dispositivos. En este caso, tendremos que tratar con restricciones parciales sobre el conjunto de imágenes asignando a cada dispositivo en particular.

Consideremos el caso donde dos dispositivos de almacenamiento cada uno de tamaño $R_t / 2$, son usados. Tendremos entonces la siguiente restricción:

$$\sum_{i=1}^{N_1} r_{ix(i)} \leq R_t / 2$$

Donde N_1 es el número de unidades de codificación que son almacenadas en el primer dispositivo de almacenamiento. N_1 , en si mismo no puede ser dado y puede tener que ser determinado.

Demora (delay) restringida en la alocación y buffering

Una restricción de almacenamiento simple como en la formulación 3 no puede abarcar situaciones donde las unidades de codificación (por ejemplo, una serie de frames de video) son encaminadas a través de un link o una red de comunicación hacia un receptor. En esta situación, cada unidad de codificación esta sujeta a restricciones de delay.

Por ejemplo, si tenemos una unidad de codificación a ser codificada a un tiempo t y asumimos que esta tendrá que estar disponible para el decodificador en un tiempo $t + \Delta t$, donde Δt es el delay de fin-a-fin del sistema. Si cada unidad de codificación tarda t_u segundos, entonces el delay de fin-a-fin puede ser expresado como $\Delta N = \Delta t / t_u$ en unidades de codificación. Por ejemplo, si un codificador de video comprime 30 frames por segundo y el sistema opera con un delay de de fin-a-fin de $\Delta t = 2$ segundos, entonces el decodificador tardara 2 segundos para descomprimir y mostrar el primer frame (asumimos que no hay delays en el canal de transmisión). El codificador de video tendrá que asegurarse que la tasa de selección por cada frame sea tal que los frames no arriben demasiado tarde al decodificador.

Dadas las restricciones de delay para cada unidad, nuestro problema se transforma:

Formulación 4: Restricción de delay en la alocación

Buscar el conjunto óptimo de cuantificadores $x(i)$ tal que (i) cada unidad de codificación i codificada en un tiempo t_i es recibida por el codificador antes de un

determinado “deadline” $t_i + \delta_i$ y (ii) dada una medida de distorsión, por ejemplo una de las usadas en la formulación 3, que esté minimizada.

Este puede ser un problema fácil de resolver sino se tienen restricciones sobre el ancho de banda en la transmisión.

La complejidad del problema de aloación depende de las características del canal. Específicamente necesitamos conocer si el canal provee una restricción en la tasa de bits (CBR) o una tasa de bits variable (VBR), si el delay o demora del canal esta restringida, si el canal es de confianza, etc. Por simplicidad, en lo que sigue asumiremos que tenemos $\delta_i = \Delta t$ para todas las i .

En ambos casos CBR y VBR los datos serán almacenados en buffers en el codificador y en el decodificador. Asumimos una tasa variable en el canal $c(i)$ durante el i -ésimo intervalo de codificación. Entonces tendremos que el estado del buffer del codificador en el tiempo i es

$$B(i) = \max (B(i-1) + r_{ix(i)} - c(i), 0)$$

Con $B(0) = 0$ siendo el estado inicial del buffer.

Ahora consideraremos que restricciones necesitan ser aplicadas a los estados del codificador. Primero, el estado $B(i)$ no puede crecer indefinidamente porque el buffer físico es finito. Si B_{\max} es la memoria física disponible entonces necesitaremos garantizar que $B(i) \leq B_{\max}$ todo el tiempo. En resumen, las restricciones de delay de la formulación 4 no pueden ser violadas necesitando garantizar que los datos que corresponden a la unidad de codificación i son transmitidos antes de $t_i + \Delta t$; esto significa, la transmisión ha sido completada durante los próximos ΔN intervalos de la unidad de codificación. Intuitivamente, todo lo que necesitamos asegurar es que las futuras tasas del canal, en las próximas ΔN unidades de tiempo, son suficientes para transmitir todos los datos almacenados en el buffer.

Definamos el tamaño efectivo del buffer, $B_{\text{eff}}(i)$, como:

$$B_{\text{eff}}(i) = \sum_{k=i+1}^{i+\Delta N} c(k)$$

Esto es la suma de las tasas futuras del canal en los próximos ΔN intervalos. Entonces es fácil ver que la correcta transmisión es garantizada si $B(i) \leq B_{\text{eff}}(i)$, $\forall i$.

Llamamos tamaño efectivo porque este define una restricción impuesta sin reparar en el tamaño físico del buffer. En general, la restricción aplicable será impuesta por el menor de los $B_{\text{eff}}(i)$ y el B_{\max} . Asumiendo que hay suficiente almacenamiento físico disponible en el buffer (B_{\max} es siempre más grande que $B_{\text{eff}}(i)$) nuestro problema se convierte en:

Formulación 5: Restricción de asignación del buffer

Hallar el conjunto óptimo de cuantificadores $x(i)$ para cada i tal que el buffer este ocupado:

$$B(i) = \max (B(i-1) + r_{ix(i)} - c(i), 0),$$

y es tal que

$$B(i) \leq B_{\text{eff}}(i)$$

y alguna medida $F(d_{1x(1)}, d_{2x(2)}, \dots, d_{nx(n)})$ que esta minimizada.

Es interesante notar que las restricciones dependen de las tasas del canal. Cuando las tasas del canal pueden ser elegidas por el usuario (en una transmisión sobre una red), esto genera una serie de interesantes preguntas sobre ¿Cuál es la mejor combinación de tasas de la fuente y las del canal dadas las restricciones para la tasa del canal? En escenarios donde el canal es inseguro, nosotros no podríamos determinísticamente saber que tasas futuras tendremos, pero es posible, si los modelos del canal esta disponibles y reemplazamos las tasas del canal en la formulación 5 por los valores estimados que se encuentran en el modelo.

Diferentes herramientas para optimizar la tasa de distorsión

En esta sección describiremos en más detalle algunas de las técnicas básicas que pueden ser aplicadas al problema que acabamos de describir. Nuestra meta es explicar estas herramientas en términos genéricos y proveer una presentación de trabajos específicos donde versiones modificadas de estos métodos han sido exitosamente aplicadas en una variedad de escenarios de compresión.

Problemas independientes

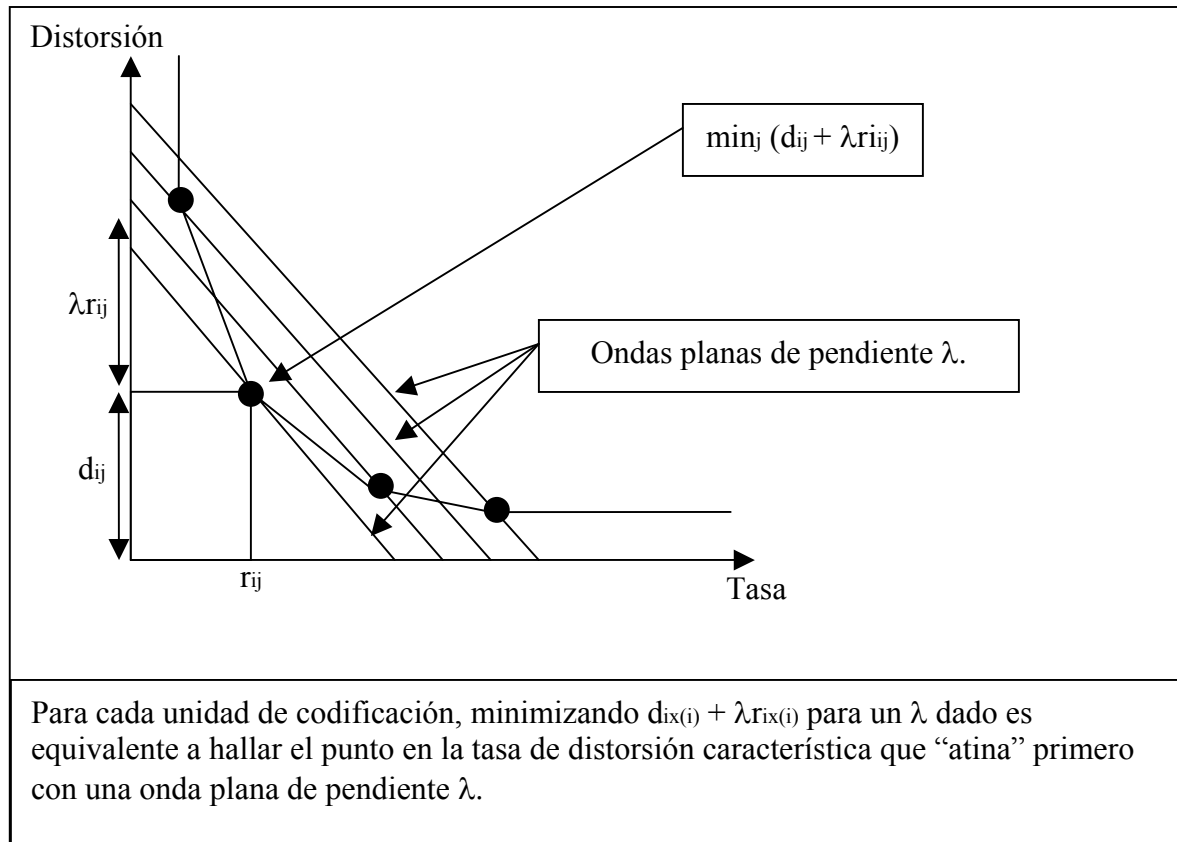
Consideremos primero el caso donde la tasa r_{ij} y la distorsión d_{ij} pueden ser medidas independientemente para cada unidad de codificación, la tasa de distorsión para la unidad de codificación i puede ser computada sin requerir que otra unidad de codificación sea codificada. Un ejemplo de este escenario es la asignación de bits para los diferentes bloques en una codificación DTC, donde los bloques son independientemente cuantificados y entropicamente codificados (usando predicción INTER bloque). Otro ejemplo puede ser la asignación de bits para frames de un video codificados con MPEG en modo INTRA solamente o usando motion JPEG.

Esto es también útil para escenarios que involucran algún tipo predicción o codificación basada en el contexto, aunque son por naturaleza “dependientes” pero algunas veces pueden ser aproximados usando estrategias de asignación “independientes” con alguna perdida de performance pequeña. Aun si hiciéramos aproximaciones independientes resultaría en perdida de performance y los efectos de dependencia son usualmente ignorados para acelerar la computación. Por ejemplo, es común considerar la asignación de

bits para una secuencia de video como si se pudiera tratar independientemente, sin embargo, debido al loop de la estimación de movimiento, la asignación de bits para un frame tiene la potencialidad de afectar los frames subsecuentes.

Optimización Lagrangian

La solución clásica para el problema de la formulación 3 esta basada en la versión discreta de la optimización Lagrangian, introducida primeramente por Everett.



La idea básica de esta técnica es la siguiente, introducir un multiplicador Lagrange $\lambda \geq 0$, este es un número real no negativo, y considerando el costo Lagrangian $J_{ij}(\lambda) = d_{ij} + \lambda r_{ij}$. La figura de arriba refiere la interpretación grafica del costo Lagrangian. Como el índice de cuantificación j se incrementa (la tasa decrece y la distorsión se incrementa) tendremos una relación entre la tasa y la distorsión. El multiplicador Lagrange nos permite seleccionar un punto específico de la relación. Minimizando el costo Lagrangian $d_{ij} + \lambda r_{ij}$ cuando $\lambda = 0$, es equivalente a minimizar la distorsión seleccionando el punto más cercano al eje Y en la figura. Consecuentemente, minimizar el costo Lagrangian cuando λ se convierte en arbitrariamente grande es equivalente a minimizar la tasa, y así hallamos el punto más cercano al eje X en la figura. Los valores intermedios de λ determinan puntos operacionales intermedios.

Entonces tenemos:

Teorema 1: Si el mapeo $x'(i)$ para $i:1,2,...,N$ minimizan

$$\sum_{i=1}^N d_{ij} + \lambda r_{ij} \quad (2)$$

entonces esto también es una solución óptima para el problema formulado en la formulación 3, para el caso particular donde:

$$R_t = R(\lambda) = \sum_{i=1}^N r_{ix'(i)} \quad (3)$$

y

$$D(\lambda) = \sum_{i=1}^N d_{ix'(i)} \leq \sum_{i=1}^N d_{ix(i)} \quad (4)$$

para algún x que satisfice la Eq. (1) y con R dada por la Eq. (3)

Si removemos la restricción de presupuesto de la Eq. (1), para una “calidad” operacional dado λ , entonces la Eq. (2) puede ser re-escrita como:

$$\text{Min} \left(\sum_{i=1}^N d_{ix(i)} + \lambda r_{ix(i)} \right) = \sum_{i=1}^N \min (d_{ix(i)} + \lambda r_{ix(i)}) \quad (5)$$

el mínimo puede ser computado independientemente para cada unidad de codificación. Notar también que para cada unidad de codificación i , el punto en la tasa de distorsión característica que minimiza $d_{ix(i)} + \lambda r_{ix(i)}$ es aquel punto en el cual la línea de pendiente λ es tangente a la convexión de Hull de la tasa de distorsión característica (ver figura). Por esta razón normalmente nos referimos a λ como la pendiente, y es la misma para todas las unidades de codificación en la secuencia, podríamos referirnos a este algoritmo como una “optimización de pendiente constante”.

La explicación intuitiva del algoritmo es simple. Por considerar los puntos de operación en una pendiente constante, estamos haciendo que todas las unidades de codificación operen en la misma correspondencia marginal para tener un bit extra en la relación de la tasa de distorsión. Así la reducción del MSE al usar un bit extra para una unidad de codificación dada podría ser igual al incremento del MSE incurrido al usar un bit menos para otra unidad. Por esta razón no hay alocaión que sea más eficiente para un presupuesto particular. Esta técnica es bien conocida en problemas de optimización donde las funciones de costo y objetivo son continuas y diferenciables. Everett contribuyo para demostrar que la técnica de Lagrangian puede ser usada también para problemas de

optimización discretos con ninguna pérdida de optimización si la solución existe con el presupuesto requerido, esto es que existe un punto en la convexión de Hull que reúne el presupuesto requerido.

La propiedad del método del multiplicado de Lagrangian es muy atractivo en términos de computación. Buscar el mejor cuantificador para un λ dado es fácil y puede ser hecho independientemente de cada unidad de codificación (Notar que se esta considerando que los datos de la tasa de distorsión ya han sido computados y estamos discutiendo sobre la complejidad de la búsqueda. Buscar los datos de la tasa de distorsión puede en si mismo requerir de una complejidad substancial). Aun, se ha de buscar el λ “correcto” para alcanzar la solución óptima para la tasa requerida, hallar λ tal que $R(\lambda)$, como fue definido anteriormente, es cercano o igual al presupuesto especificado. Hallar el λ correcto puede ser hecho usando una búsqueda por bisección. Notar que el número de iteraciones requeridas en la búsqueda de λ puede ser mantenido bajo mientras no queramos tener una concordancia exacta con la tasa presupuestada. En escenarios como en la codificación de video, donde pueden ser realizadas alocaiones en frames sucesivos que tienen características similares, es posible inicializar el multiplicador Lagrange para un frame con el valor en el cual la convergencia fue alcanzada para el frame previo, el cual reducirá otra vez el número de iteraciones requeridas, proveer un buen supuesto inicial del λ conducirá a reducir la complejidad.

Optimización Lagrangian Generalizada

Los problemas de alocaión con restricciones múltiples, como los mencionados en la formulación 4, puede también ser resueltos usando técnicas Lagrangian. Esta aproximación esta basada sobre la generalización de los métodos Lagrangian. La idea básica es introducir un multiplicador Lagrange para cada una de las restricciones, las cuales pueden así ser relajadas. El problema ahora es que la solución puede ser hallada solamente para un vector correcto de multiplicadores Lagrange $\Lambda = \{\lambda_1, \dots, \lambda_n\}$ y la búsqueda en un espacio multidimensional no es sencilla como cuando se usa un multiplicador Lagrange simple.

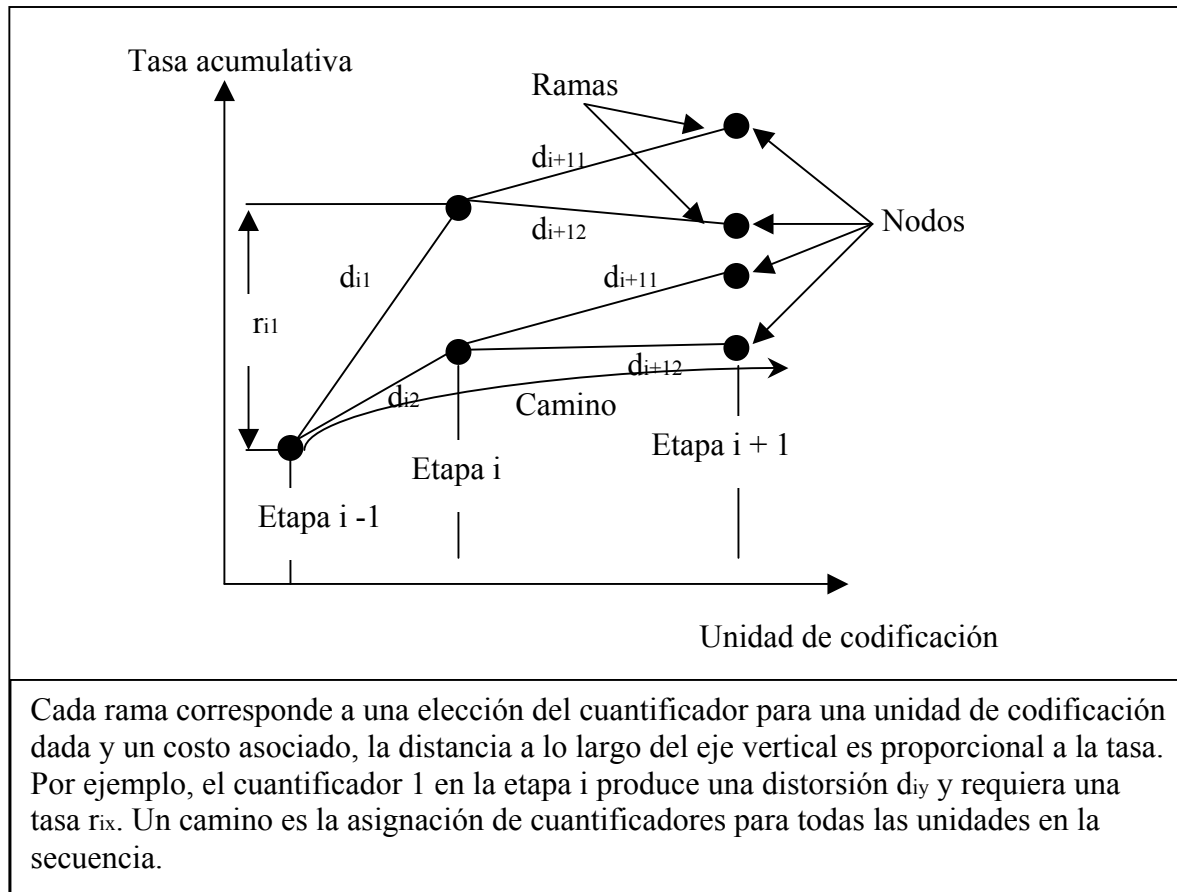
Típicamente el problema considerado involucra una estructura significativa en cuanto a restricciones se refiere, y que puede guiar la búsqueda de la solución para el vector de multiplicadores de Lagrange. Por ejemplo, en algunos casos, estas restricciones están embebidas. Si son N unidades de codificación, tenemos una serie de c restricciones donde la restricción k es una restricción de presupuesto que afecta las unidades de codificación 1 a n_k . Las otras restricciones igual afectan los bloques 1 a $n_1, n_2, \dots, n_c = N$, respectivamente. En aquellos y en casos similares, una estrategia de búsqueda puede ser derivada para hallar el vector óptimo Λ de modo iterativo.

Programación Dinámica

Como se menciono, las técnicas Lagrangian tienen el defecto de que no es posible alcanzar puntos que no residen en la convexión de Hull de la tasa de distorsión

característica. Una formulación alternativa, es formular la asignación como un problema determinístico de programación dinámica.

En este caso, nosotros creamos un árbol que representara todas las posibles soluciones. Cada etapa en la creación del árbol corresponde a una de las j unidades de codificación y cada nodo del árbol en una etapa dada que representa la utilización de una tasa acumulativa posible. Por ejemplo, como en la figura de abajo, para la tasa acumulativa en el bloque $i-1$ sumando la correspondiente tasa para cada posible elección de cuantificación, así generamos nuevos nodos con la apropiada tasa acumulativa.



Cada rama tiene un costo de distorsión que corresponde a un cuantificador particular, y por ello como viajamos de la raíz a las hojas, podemos calcular la distorsión acumulada para cada solución. Lo que debe quedar claro es que esta es de hecho una manera de representar todas las posibles soluciones, ya que al atravesar el árbol obtenemos asignaciones sucesivas para cada una de las unidades de codificación.

Ahora si consideramos que pasa si dos caminos convergen en un nodo, o sea dos soluciones alternativas proveen la misma tasa acumulativa. Intuitivamente se ve que la solución que tiene la distorsión más alta en aquel punto debe ser removida (podada del árbol), ya que en esta etapa ambas soluciones tienen los mismos bits sobrantes en uso. Aquellos caminos que son perdidos hasta ahora, serán todos perdidos. Esto es la esencia del principio de optimalidad introducido por Bellman y como este se aplica a este problema en

particular. Esta rama particular de la programación dinámica (DP), la cual maneja funciones de costo determinísticas y nos ayuda a hallar el camino mas corto (en el sentido del costo de la rama) en un grafo, es también conocido como el algoritmo Viterbi o algoritmo de Dykstra para el camino mínimo.

Será fácil incorporar elementos adicionales, para el crecimiento del árbol así podremos resolver los problemas de las formulaciones 3 y 5. Por ejemplo, introducir una restricción de presupuesto total, es suficiente para podar las ramas que exceden la tasa total de alocación deseada (el árbol no puede crecer sobre el “techo” especificado por la restricción de presupuesto). Similarmente, si una restricción sobre el buffering como las especificadas en la formulación 5 es introducida, entonces necesitaremos podar las ramas que exceden la capacidad máxima de buffer en cada etapa.

El algoritmo puede ser resumido informalmente como sigue. En cada etapa i , para todos los nodos sobrevivientes, sumamos las ramas correspondientes a todas las posibles cuantificaciones en esa etapa (la tasa r_{ij} la determina el nodo final y la distorsión d_{ij} es sumada del camino de distorsión). Podar las ramas que exceden la restricción de la tasa y entonces para cada nodo remanente en la etapa $i + 1$, se mantiene solamente la rama de costo mínimo.

Uno podría concluir que la optimización Lagrangian se prefiere por las ventajas que ofrece en cuanto a complejidad. Sin embargo la aproximación Lagrangian tiene un inconveniente y es que solamente los puntos en la convexión de Hull de la tasa de distorsión característica global operacional pueden ser alcanzados. Este no es un problema si la convexión de Hull es lo “suficientemente densa”, sin embargo, en algunas aplicaciones esto puede resultar en una reducción importante de optimalidad.

Problemas dependientes

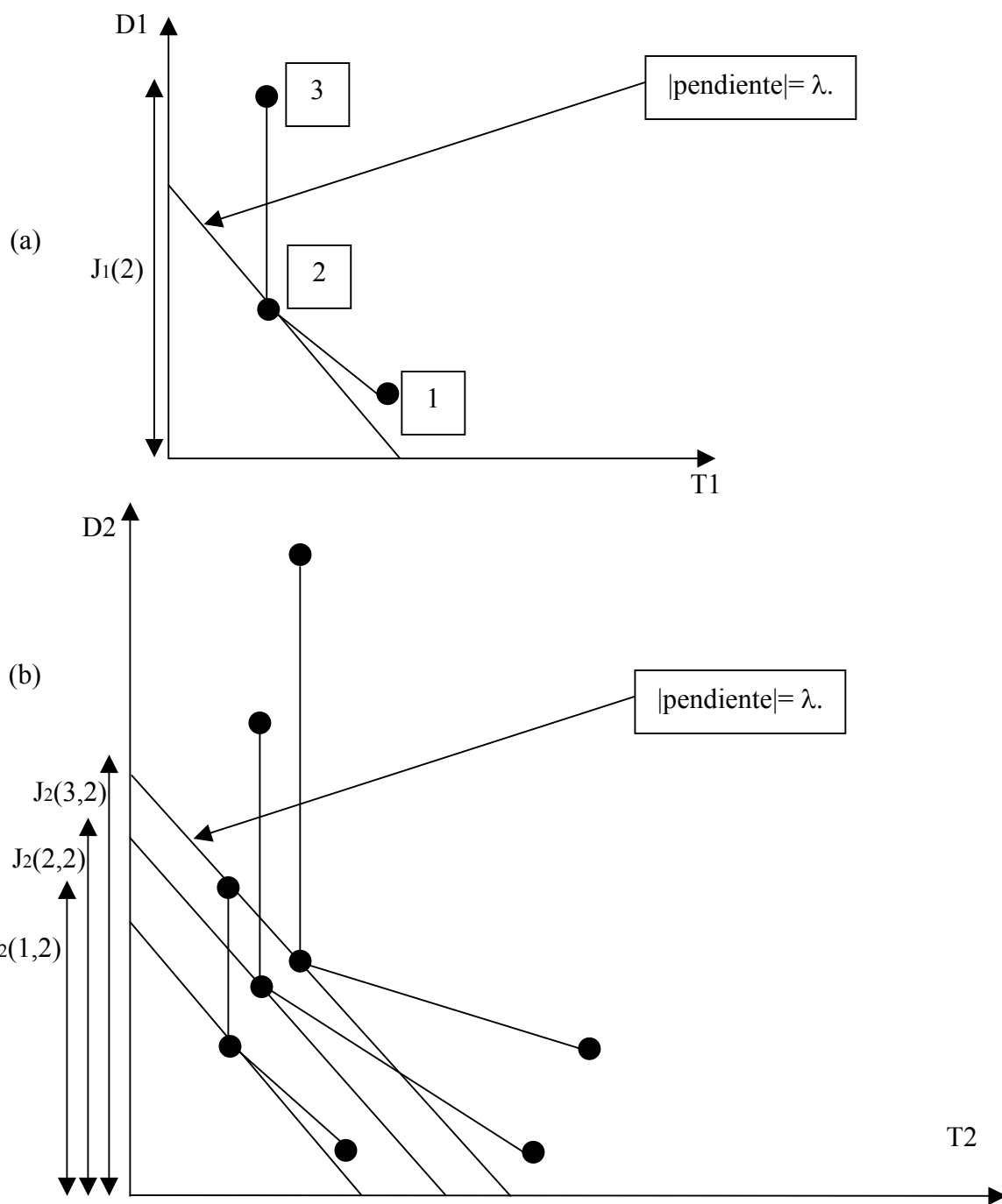
Hasta ahora hemos asumido que la selección del modo de codificación puede ser hecha independientemente de cada unidad de codificación sin afectar a las otras. Existen, sin embargo, escenarios donde esta asunción no es valida.

Esto es típicamente, el caso en esquemas de codificación basados en predicción. Por ejemplo, asumimos que cada unidad de codificación i es predicha de la precedente unidad de codificación $i - 1$. El predictor es construido usando los datos cuantificados pasados y así la codificamos $X_i - P(X_i^{?}-1)$, o sea, el error de predicción. Como usamos datos cuantificados, el error de predicción y así los puntos operacionales de la tasa de distorsión admisible para i dependerá de nuestra elección del cuantificador para $i - 1$. Cada opción $X(i - 1)$ resultara de características diferentes.

Un ejemplo de este escenario es ilustrado por la figura de abajo, donde todas la posibles opciones de la tasa de distorsión para 2 frames de video donde cada uno puede ser codificado usando 3 configuraciones diferentes de cuantificación y donde el frame 2 es predicho a partir del frame 1 (notar que hay 9 posibles opciones para el frame 2, ya que las opciones para el frame 1 afectarán los valores resultantes para el frame 2). Debe ser notado que un algoritmo que considere los 2 frames independientemente puede seleccionar el (para una pendiente λ dada) mismo cuantificador para ambos frames, o sea que puede incurrir en un costo $J_i(2)$ para el frame 1 y entonces, dado que el cuantificador 2 fue seleccionado al

seleccionar el cuantificador del frame 1, puede elegirse el mínimo $J_2(2,x)$ de todos, lo cual sería $J_2(2,2)$. Sin embargo, en este ejemplo particular, la aproximación “avara” puede ser llevada a cabo, la cual aloca primero para el frame 1 y luego para el 2. La mejor de todas las performances posibles es alcanzada cuando el cuantificador 1 es usado por el primer frame y el cuantificador 2 es usado para el segundo. Aun si $J_1(2) < J_1(1)$, se tiene que $J_1(2) + J_2(2,2) > J_1(1) + J_2(1,2)$.

Muchos tipos de escenarios de dependencias pueden ser identificados. Consideraremos 2 ejemplos concretos dentro del framework de codificación MPEG que ilustran diferentes formas de dependencia.



Tasa de distorsión operacional característica de dos frames en un framework de codificación dependiente, donde el frame 2 depende del frame 1.

- (a) curva de la tasa de distorsión de los frames tratados independientemente
- (b) curva de la tasa de distorsión de los frames tratados en forma dependiente

Dependencia basada en el enrejado: la selección de la cuantificación a nivel macrobloque en un stream de video MPEG es un problema de dependencia, porque la tasa r_{ij} para el macrobloque i y el cuantificador j dependen de la elección del cuantificador para el macrobloque $i - 1$. Esto es porque la codificación entropica predictiva de los índices de cuantificación es usada para incrementar la eficiencia de la codificación. En esta situación es posible representar todas las selecciones posibles para un macrobloque dado, con cada etapa del enrejado correspondiente a un macrobloque. La programación dinámica puede entonces ser usada para hallar el camino de costo mínimo en este enrejado, donde el costo de la rama es típicamente definido como el costo de Lagrangian. Como en el ejemplo de la figura de arriba tomamos la dependencia dentro las cuentas evitando hacer selecciones “avaras” de los parámetros de codificación, donde la asignación del cuantificador es optimizado para la unidad actual de codificación solamente.

En general, las dependencias basadas en el enrejado surgen en casos donde la estructura subyacente es tal que la memoria en el sistema es finita (las elecciones de codificación para i dependen solamente de un conjunto finito de unidades de codificación previas) y el numero de casos posible es también finito. En otras palabras, en este caso, los parámetros de codificación disponibles para una unidad de codificación dada dependen del “estado” del sistema -el conjunto finito de parámetros que determina completamente los valores alcanzables. Para este tipo de dependencias, uno puede usar una aproximación de programación dinámica, donde el estado corresponde al estado del sistema y las ramas (cada una corresponde a una elección de cuantificación) tienen asociada un costo Lagrangian que combina la tasa y la distorsión para las elecciones de los parámetros dados.

Dependencia basada en árboles: Un segundo ejemplo de dependencia puede ser visto cuando se analiza los efectos de la compensación de movimiento en un framework MPEG. Luego de la compensación de movimiento, el codificador transmite la diferencia entre el frame previo y el actual. Esta diferencia es comprimida y usada para reconstruir la versión decodificada del frame actual. Es fácil ver que tendremos un loop de predicción recursivo, y así el frame residual dependerá de la elección de los parámetros de codificación para todos los frames anteriores. En este caso podemos observar que todas las posibles combinaciones generadas por las elecciones sucesivas de cuantificación pueden ser representadas como un árbol con el número de ramas creciendo exponencialmente con el número de niveles de dependencia.

La principal conclusión es que el crecimiento exponencial en el número de combinaciones hace que la solución exacta sea muy compleja. Sin embargo, es posible hacer aproximaciones que simplifican la búsqueda de la solución óptima. Una buena heurística incluye el uso de asunciones monolíticas (un predictor de cuantificación mas exacto resulta típicamente en un menor error de predicción) o aproximaciones “avaras” donde por ejemplo, solamente unas pocas opciones de cuantificación son mantenidas para cualquier etapa dada. El problema puede también ser simplificado por medio de modelos de las tasas de distorsión dependientes características ya que no todos los puntos operacionales en el árbol necesitan ser explícitamente computados, o tomando modelos de la tasa y asumiendo que la escala de cuantificación provee una buena estimación de la calidad.

Aplicación de componentes básicos en los algoritmos de codificación de imágenes y video

A continuación, explicaremos brevemente escenarios donde variaciones de las formulaciones genéricas descritas anteriormente han sido halladas muy útiles. Mientras las formulaciones y algoritmos son muy similares, hay diferencias fundamentales que serán descritas mas adelante. Estructuraremos nuestra discusión a través de dos clases de problemas, la restricción de presupuesto y la alocaación restringida en delay (o demora).

Problemas de presupuesto restringido

Ahora veremos unas pocas aplicaciones que involucran optimizaciones en la restricción de presupuesto en el contexto de la codificación de imágenes y video. El problema de la optimización del presupuesto radica en minimizar la distorsión de cuantificación bajo una restricción en la tasa de bits. La formulación más general en el contexto de codificación-transformación involucra la selección de los puntos de operación concernientes a la combinación de la transformación, cuantificación y codificación entropica en pos de realizar la mejor relación de la tasa de distorsión. Dependiendo de la flexibilidad (y complejidad concerniente) del framework uno o todos de los componentes funcionales pueden ser juntamente optimizados. Típicamente, la transformación es mantenida fija y la cuantificación y codificación entropica son juntamente optimizadas. Los modos de cuantificación pueden variar de un simple cuantificador escalar a un vector de cuantificadores bastantes sofisticados, pero abstrayéndonos de estos modos de cuantificación que construyen un conjunto discreto de cuantificadores, un framework diferente puede ser considerado bajo un mismo concepto.

Optimización de la tasa de distorsión basada en la transformación fija

Un buen ejemplo de una aplicación basada en la transformación fija involucra la optimización de las restricciones de sintaxis de la codificación de la imagen y el video como en JPEG, donde la opción del cuantificador (matrices de cuantificación de 8x8 para la imagen) y la opción de la codificación entropica (tabla de Huffman) pueden ser optimizadas sobre una base que determina la tasa de compresión por imagen. El espectro de las aplicaciones generales para optimizar la codificación entropica y los cuantificadores pueden ir de la selección de diferentes cuantificadores a simplemente un cuantificador escalar definido por medio de un factor escalar, como es usualmente hecho por los usuarios de JPEG.

Otra técnica basada en la tasa de distorsión es la codificación de imágenes basadas en wavelets, donde muchos estados del algoritmo de codificación ART realizan altas ganancias en performance por medio del uso de una variedad de sofisticadas técnicas de optimización de la tasa de distorsión. Una lista parcial de algoritmos de codificación que derivan de estas técnicas incluyen el popular framework de wavelets zerotree así como el framework de estimación-cuantificación (EQ), la cuantificación de adaptación hacia atrás y el frameworks de adaptación hacia delante involucrando criterios de clasificación de

subbandas, cuantificación del espacio de frecuencias (SFQ) que optimizan la tasa de distorsión.

Similarmente, técnicas de optimización pueden ser aplicadas con impresionantes ganancias en performance por frameworks de video como MPEG y H263. Hay una cantidad considerable de redundancia espacio-temporal en datos típicos de video y es particularmente importante prestar atención a la dimensión temporal, la cual tiene el volumen de redundancia inherente para una secuencia típica de video.

Las técnicas de optimización de la tasa de distorsión pueden ser aplicadas a la codificación de figuras, donde la relación entre la fidelidad versus la tasa de bits necesitada para representar esa figura puede ser optimizada. Estas técnicas son usadas en el estándar MPEG-4, el cual introduce soporte par objetos de video.

Optimización de la tasa de distorsión basada en la transformación adaptativa.

Mientras los algoritmos de codificación que usan transformación fija pueden ser muy útiles si la clase de señales esta bien situada en algún sentido para la transformación fija, esto puede no ser adecuado para tratar con clases arbitrarias de señales con algunas características desconocidas o que varían en el tiempo. Por ejemplo, para imágenes o segmentos de imágenes que tienen componentes estacionarios de alta frecuencia, la transformación basada en wavelets es una mala elección. Esto motiva a considerar un más poderoso framework adaptativo que puede ser robusto cuando trata con grandes clases de señales de características desconocidas y con variaciones de tiempo o de espacio. En esta aproximación, el objetivo es realizar una transformación de señales adaptativa.

La idea detrás del framework de transformación adaptativa es reemplazar la transformación fija con una gran librería de transformaciones que pueden adicionalmente ser inspeccionadas en forma eficiente. La librería de transformaciones puede ser ampliamente general y puede incluir, por ejemplo, la familia de segmentaciones espaciales quadtree o tamaños variables de bloques de la DTC (por ejemplo de 4x4, 8x8, 16x 16). Similarmente, uno puede tener la descomposición estándar wavelet y modificar varios de los parámetros de estas (filtros, la estructura de árbol, incluyendo el número de niveles de la descomposición) para una imagen en particular o para las partes de una imagen. Estas técnicas han sido desarrolladas para proveer una ganancia substancial sobre las técnicas de descomposición adaptativas.

Problemas de aloación de restricción de delay

Esta clase de problemas, también referidos en la formulación 4, es típicamente encontrada en transmisiones de video bajo restricciones de delay. La visión más popular sobre el tema es referido como un problema de control de buffer pero, las restricciones sobre el delay son más generales. Las técnicas de tasa de distorsión han sido aplicadas para controlar la tasa bajo condiciones de transmisión CBR.

El mecanismo tradicional de feedback directo usado en los algoritmos de control de buffer, reemplaza la escala de cuantificación que es controlada por el llenado del buffer por

un mecanismo de feedback que controla el valor del multiplicador de Lagrange a ser usado en la selección de los puntos operacionales óptimos.

El rol de la tasa de distorsión en la codificación conjunta de la fuente y el canal

Nos hemos enfocado (con la excepción de los párrafos previos) en los métodos de tasa de distorsión para la codificación de la fuente cuando tratamos con fuentes que son imágenes o video. Ahora nos dirigimos brevemente al problema de la aplicabilidad de dichos métodos para un problema más grande de transmisión de imagen y video, específicamente en el contexto de la codificación conjunta de la fuente y el canal. En la sección, “Técnicas de tasa de distorsión para codificación conjunta de la fuente y el canal” se expone la esencia del problema. Aquí, daremos una mirada a algunas de las aplicaciones dirigidas a estas técnicas y proveeremos algunos puntos de actividades recientes en este campo.

Background y formulación del problema

Con la explosión en aplicaciones que involucran a la comunicación de imágenes y video, y como estas proporcionaron un “Boom” en aplicaciones dirigidas al multimedia en Internet, el problema de la comunicación de imágenes ha asumido recientemente mucho interés e importancia, porque los datos visuales representan por ahora el mayor porcentaje del tráfico multimedial. Una pregunta que surge naturalmente es: ¿Por qué necesitamos re-inventar la comunicación de datos por la actual explosión multimedial? Hay muchas razones para volver a examinar los paradigmas y sistemas existentes. Primeramente es que los diseños de los actuales links de comunicación son erróneos para las fuentes de imágenes y video, también se debe considerar las características inherentes en la codificación y transmisión de imágenes/video así como (i) altas variaciones en el tiempo y características del canal, (ii) alta tolerancia de la fuente a la pérdida del canal y (iii) desigual importancia de los bits transmitidos. Algunos atributos importantes de las fuentes de imágenes/video son resumidos a continuación:

- La métrica de performance es la calidad visual entregada (por ejemplo el error cuadrático medio o más correctamente, la distorsión perceptiva) de la fuente esto se debe a que la cuantificación de la fuente y la distorsión del canal están bajo restricciones de recursos como lo es el ancho de banda y la energía de transmisión. Esto contrasta con los criterios de performance usados comúnmente, como por ejemplo la tasa de errores en los bits que son comunicados a través de una red de comunicación, que es apropiada para la transmisión de datos tradicional.
- Los errores desiguales de un stream de bits de un video típico (por ejemplo, los vectores de movimiento representados en los bits o los detalles en las texturas de las áreas de los frames) enfatizan el deseo de tener una aproximación en capas para la codificación de la fuente y el canal, y así tener sistemas con carácter multiresolución en vez del framework convencional de “resolución simple” de transmisiones digitales.

- Debido a los requerimientos en los delays en las aplicaciones de video, existe la necesidad de incluir restricciones de buffer (estrategias eficientes de control de tasa).

En la sección, “Técnicas de tasa de distorsión para codificación conjunta de la fuente y el canal”, nosotros estábamos motivados por la necesidad de la codificación conjunta de la fuente y el canal debido a las deficiencias prácticas del principio de separación, así como la inaplicabilidad teórica para un número de escenarios de comunicación multiusuario de interés como el broadcast y el multicast. Así la potencial ganancia en performance puede ser alcanzada si cerramos la interacción entre las funciones de codificación de la fuente y el canal. El entendimiento de la superioridad de que una aproximación conjunta para la codificación de la fuente y el canal, ha iniciado recientemente numerosas actividades de investigación en esta área.

Overview de aplicaciones y metodologías

Los problemas de transmisión de imágenes y video anteriormente formulados vienen en varios “sabores”. De particular interés son las comunicaciones de imágenes y video sobre redes heterogéneas como Internet y ATM, así como el video en WIRELESS, también en los escenarios como el broadcast y multicast. Para aplicaciones sobre redes hay restricciones impuestas por la red relacionadas al porcentaje de la tasa y los picos máximos en la misma. La capacidad disponible del canal puede fluctuar debido a la congestión de la red.

En los escenarios de codificación conjunta de la fuente y el canal han sido históricamente dirigidos por 2 ideologías de alto nivel. Estas pueden ser clasificadas, muy rudimentariamente en métodos de transmisión “digitales” versus “analógicos”.

La clase digital esta basada en la alocaión óptima de bits entre la fuente de codificación digital y el canal. Los bits de la codificación de la fuente corresponden a una compresión digital y un stream de codificación entropica. Los bits de la codificación del canal corresponden a la información de paridad de un código de corrección de error digital. Una aproximación popular basada en la tasa de distorsión consiste en minimizar la distorsión esperada de la fuente de fin-a-fin (debido a la cuantificación de la fuente y los defectos en el canal de transmisión).

Debe ser notado que esta ideología “digital”, mientras permite altos índices de compresión de la fuente debido a la codificación entropica, puede también llevaremos a incrementar el riesgo de propagación de errores. La solución más popular es insertar periódicamente capacidades de resincronizacion usando empaquetamiento. La resultante resincronizacion y sobrecarga de empaquetamiento que son necesarias para tratar con este tipo de errores afecta la eficiencia del sistema en su totalidad. Luego el problema se convierte en uno de optimizar este balance.

Un ejemplo de un área donde las ideas de codificación conjunta de la fuente y el canal han tenido un gran impacto es en la comunicación sobre redes heterogéneas. En particular, el caso del multicast en un ambiente heterogéneo esta bien situado para la codificación de multiresolución de la fuente y el canal. La idea es muy simple: dar a cada usuario la mejor calidad posible para desplegarla en una infraestructura flexible de la tasa de bits objetivo. Más precisamente una transmisión multicast puede ser conceptualizada como estar en un árbol de multiresolucion, el cual es un conjunto de árboles que acarrean varias resoluciones. Cada usuario entonces alcanza muchos niveles del árbol multiresolucion y estos dan las capacidades de acceso. Mientras que

actualmente la mayoría de los links de comunicación son cableados, esta claro que los componentes mobile están tomando más y más importancia. Cada uno de los esquemas debe ser ubicado para cada ambiente en particular, posiblemente con fuentes entre los componentes cableados y WIRELESS.

Resumen

En esta monografía hemos provisto un overview de las técnicas de optimización de la tasa de distorsión como las usadas en aplicaciones prácticas de imágenes y video. Hemos comenzado por establecer un enlace entre estas técnicas y la teoría desarrollada de la tasa de distorsión en el campo de la teoría de la información. La codificación de imágenes y video basados en estándares pueden beneficiar las técnicas de optimización y permitieron al codificador optimizar la selección de los parámetros de codificación, mientras se preservaba la compatibilidad con el decodificador. Cuando entonces definimos un problema general de aloación de recursos y dimos dos ejemplos concretos: Aloación de presupuesto restringido y aloación de demora (delay) restringida. Explicamos en detalle la optimización Lagrangian y técnicas de programación dinámica, la cual se ha convertido en una herramienta esencial para resolver estos problemas de aloación. Esto nos permitió dar un overview de las aplicaciones donde la optimización de la tasa de distorsión ha probado ser muy útil. De hecho, describimos como dichas técnicas pueden también ser halladas muy útiles en los frameworks de codificación conjunta de la fuente y el canal.

Bibliografía

- [1] A. Ortega and K. Ramchandran, "*Rate-Distortion Methods for Image and Video Compression*", IEEE Processing Magazine, 1998.
- [2] G.J. Sullivan and T. Wiegand, "*Rate-Distortion Optimization*" IEEE Signal Processing Magazine, vol. 15, no. 6, Nov. 1998.
- [3] M. Kalman, P. Ramanathan, and B. Girod, "*Rate-Distortion Optimized Streaming with Multiple Deadlines*," Proc. IEEE International Conference on Image Processing, ICIP-200.
- [4] M. Vetterli and J. Kovacevic, Wavelets and Subband Coding. Prentice Hall, 1995.
- [5] B. Bellman, *Dynamic Programming*. Prentice Hall, 1987.
- [6] Zhihai He and Sanjit K. Mitra, "*A Unified Rate-Distortion Analysis Framework for Transform Coding*", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, vol. 11, no. 12, Dec 2001.