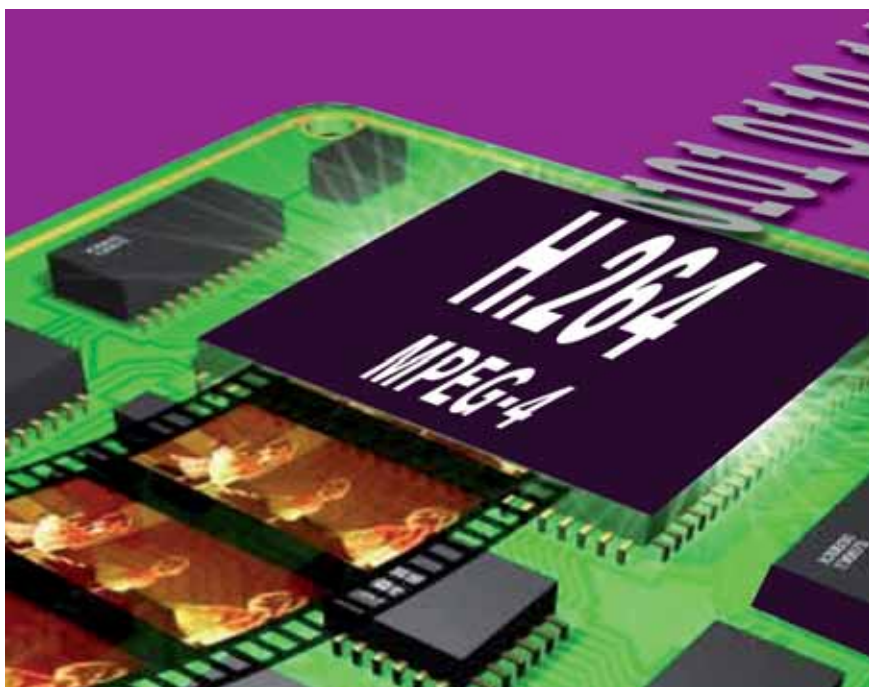


Estándar de compresión de video H.264



Con capacidad de reducir hasta un 80% el tamaño de un archivo de video digital, la compresión H.264 es un estándar abierto que está comenzando a ganar el mercado tanto del video en red como en aplicaciones comerciales, como TV de alta definición, videojuegos y hasta transmisiones satelitales. Explicamos aquí de qué se trata y como funciona básicamente esta tecnología.

Introducción

H.264 es un estándar abierto con licencia compatible con las técnicas más eficientes de compresión de video disponibles. Sin comprometer la calidad de la imagen, un codificador H.264 puede reducir el tamaño de un archivo de video digital en más de un 80% si se compara con Motion JPEG y hasta un 50% comparado con el estándar MPEG-4 Parte 2. Esto significa que se requiere menos ancho de banda y espacio de almacenamiento para los archivos de video. O, visto de otra manera, se puede lograr mayor calidad de imagen de video para una frecuencia de bits determinada.

Se trata de un estándar definido de manera conjunta por organizaciones de normalización de los sectores de telecomunicaciones e industrias de TI y ya se ha utilizado en nuevos aparatos electrónicos como teléfonos móviles y reproductores de video digital.

En el sector de la videovigilancia, H.264 encontrará su mayor utilidad en aplicaciones donde se necesiten velocidades y resoluciones altas, como en la vigilancia de autopistas, aeropuertos y casinos, lugares donde por regla general se usa una velocidad de 30/25 (NTSC/PAL) imágenes por segundo. Es aquí donde las ventajas económicas de

un ancho de banda y un almacenamiento reducidos se harán sentir de forma más clara.

La compresión H.264 puede acelerar también la adopción de cámaras megapixel, ya que con esta tecnología pueden reducirse los archivos de gran tamaño y las frecuencias de bits sin que la calidad de la imagen se vea afectada. El estándar también tiene sus exigencias: aunque H.264 permite ahorrar en costos de ancho de banda y almacenamiento, también necesita cámaras de red y estaciones de control de mejor rendimiento.

Desarrollo de H.264

H.264 es el resultado de un proyecto conjunto entre el grupo de expertos de codificación de video de ITU-T y el grupo de expertos de imágenes en movimiento de ISO/IEC (MPEG). ITU-T es el sector que coordina los estándares de telecomunicaciones en nombre de la Unión Internacional de Telecomunicaciones.

H.264 es el nombre usado por ITU-T, mientras que ISO/IEC utiliza el nombre MPEG-4 Parte 10/AVC, ya que lo presenta como parte de su suite MPEG-4. La suite MPEG-4 incluye, por ejemplo, MPEG-4 Parte 2, que es un estándar que ya ha sido usado por codificado-

res de video basados en IP y cámaras de red.

Diseñado para solucionar una serie de debilidades de estándares de compresión de video anteriores, H.264 satisface con éxito su objetivo de admitir:

- Implementaciones que ofrezcan una reducción de la frecuencia de bits del 50%, a partir de una calidad de video fija y comparada con otros estándares de video
- Robustez frente a errores, de forma que se toleren los errores de transmisión a través de varias redes
- Capacidades de baja latencia y mejor calidad para latencias mayores
- Especificación de sintaxis directa que simplifique las implementaciones
- Decodificación de coincidencia exacta, que define cuántos cálculos numéricos debe realizar un codificador y un decodificador para evitar que se acumulen errores

H.264 tiene también la flexibilidad suficiente como para admitir una amplia gama de aplicaciones con diferentes requisitos de frecuencia de bits. Por ejemplo, en aplicaciones de video de entretenimiento -que incluye retransmisiones, satélite, cable y DVD- H.264 podrá ofrecer un rendimiento de entre 1 y 10 Mbit/s con una alta latencia,

Continúa en página 144

Viene de página 140

mientras que en servicios de telecomunicaciones puede ofrecer frecuencias de bits inferiores a 1 Mbit/s con baja latencia.

Cómo funciona la compresión de video

La compresión de video implica reducir y eliminar datos redundantes del video para que el archivo de video digital pueda enviarse y almacenarse de manera eficiente.

En este proceso se aplica un algoritmo al video original para crear un archivo comprimido y ya listo para ser transmitido o guardado. Para reproducir el archivo comprimido, se aplica el algoritmo inverso y se crea un video que incluye prácticamente el mismo contenido que el video original. El tiempo que se tarda en comprimir, enviar, descomprimir y mostrar un archivo es lo que se denomina latencia. Cuanto más avanzado es el algoritmo de compresión, mayor es la latencia a igual potencia de procesamiento.

El par de algoritmos que funcionan conjuntamente se denomina códec de video (codificador/decodificador). Los códecs de video que usan estándares diferentes no suelen ser compatibles entre sí. Es decir que el contenido de video comprimido con un estándar no se puede descomprimir con otro estándar diferente.

Por ejemplo, un decodificador MPEG-4 Parte 2 no funcionará con un codificador H.264. Esto ocurre simplemente porque un algoritmo no puede decodificar correctamente los datos de salida del otro algoritmo, pero es posible usar muchos algoritmos diferentes en el mismo software o hardware, con objeto de poder comprimir varios formatos.

Los diferentes estándares de compresión utilizan métodos diferentes para reducir los datos y, en consecuencia, los resultados en cuanto a frecuencia de bits y latencia son diferentes.

Los resultados de los codificadores que utilizan el mismo estándar de compresión pueden variar también debido a que el diseñador de un codificador haya decidido usar diferentes conjuntos de las herramientas definidas por un estándar.

Siempre que los datos de salida de un codificador se ajusten al formato de un estándar, pueden realizarse implementaciones diferentes. Esto es una ventaja, ya que cada implementación puede tener sus propios objetivos y presupuestos. Los codificadores profesionales que no funcionan en tiempo real y que se utilizan para el control de medios ópticos deberían tener la opción de poder ofrecer un mejor video codificado que un codificador de hardware en tiempo real para videoconferencias integrado en un dispositivo portátil. Un estándar determinado, en consecuencia, no puede garantizar la frecuencia de bits o la calidad. Además, el rendimiento de un estándar no se puede comparar adecuadamente con otros estándares, ni incluso con otras implementaciones del mismo estándar, sin definir primero como se implementa.

Un decodificador, a diferencia de un codificador, debe implementar todas las partes necesarias de un estándar para descodificar una transmisión de bits compatible. Esto se debe a que un estándar especifica exactamente la forma en la que el algoritmo de descompresión debe restaurar cada bit de un archivo comprimido.

El gráfico siguiente compara la frecuencia de bits, partiendo de la misma calidad de imagen, entre los siguientes estándares de video: Motion JPEG, MPEG-4 Parte 2 (sin compensación de movimiento), MPEG-4 Parte 2 (con com-

pensación de movimiento) .264 (perfil de base).



Un codificador H.264 creó hasta un 50% menos bits por segundo para una secuencia de video de muestra que un codificador MPEG-4 con compensación de movimiento. El codificador H.264 fue al menos tres veces más eficiente que un codificador MPEG-4 sin compensación de movimiento y al menos seis veces más eficiente que Motion JPEG.

Perfiles y niveles de H.264

El grupo implicado en la definición de H.264 se centró en la creación de una solución sencilla y elegante, limitando al mínimo las opciones y funciones. Un aspecto importante de este estándar es que, al igual que otros, proporciona sus capacidades en forma de perfiles (conjuntos de características algorítmicas) y niveles (clases de rendimiento) que admiten producciones populares y formatos comunes.

H.264 tiene siete perfiles, cada uno de los cuales va destinado a una clase concreta de aplicaciones. Cada perfil define que conjunto de características puede usar el codificador y limita la complejidad de implementación del decodificador.

Lo más probable es que las cámaras de red y los codificadores de video usen un perfil denominado "base", que está destinado, sobre todo, a aplicaciones con recursos informáticos limitados. El perfil base es el más adecuado dado el rendimiento disponible para un codificador en tiempo real integrado a un producto de video de red. Asimismo, el perfil permite una baja latencia, algo importante en la videovigilancia y especialmente importante para permitir el control de movimiento PTZ en tiempo real en cámaras de estas características. H.264 tiene 11 niveles o grados de capacidad para limitar los requisitos de rendimiento, ancho de banda y memoria. Cada nivel define la frecuencia de bits y la frecuencia de codificación en macrobloques por segundo para resoluciones que van de QCIF a HDTV y más allá. Cuanto mayor es la resolución, mayor es el nivel requerido.

En qué consisten los fotogramas

Según cual sea el perfil de H.264, el codificador puede utilizar diferentes tipos de fotogramas (imágenes estáticas): fotogramas I, fotogramas P y fotogramas B.

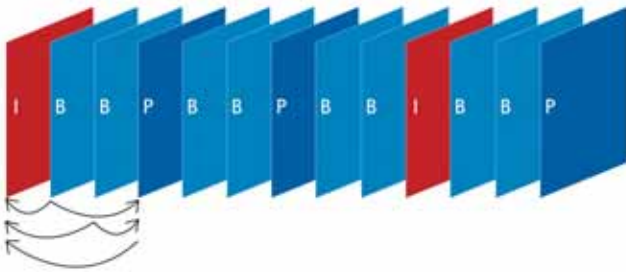
Un fotograma I, o intrafotograma, es una imagen autó-

Viene de página 144

noma que se puede codificar de forma independiente sin hacer referencia a otras imágenes. La primera imagen de una secuencia de video es siempre un fotograma I. Los fotogramas I sirven como puntos de inicio en nuevas visualizaciones o como puntos de resincronización si la transmisión de bits resulta dañada. Los fotogramas I se pueden utilizar para implementar funciones de avance o retroceso rápido o de acceso aleatorio. Un codificador insertará automáticamente fotogramas I a intervalos regulares o a petición de nuevos clientes que puedan incorporarse a la visualización de una transmisión. La desventaja de este tipo de fotogramas es que consumen muchos más bits, pero por otro lado no generan demasiados artefactos (defectos).

Un fotograma P (de interfotograma Predictivo), hace referencia a partes de fotogramas I o P anteriores para codificar el fotograma. Los fotogramas P suelen requerir menos bits que los fotogramas I, pero con la desventaja de ser muy sensibles a la transmisión de errores, debido a la compleja dependencia con fotogramas P o I anteriores.

Un fotograma B, o interfotograma Bipredictivo, es un fotograma que hace referencia tanto a fotogramas anteriores como posteriores.



Secuencia típica con fotogramas I-, B y P. Un fotograma P sólo puede hacer referencia a fotogramas I o P anteriores, mientras que un fotograma B puede hacerlo a fotogramas I o P anteriores o posteriores.

Cuando un decodificador de video restaura un video decodificando la transmisión de bits fotograma a fotograma, la decodificación debe comenzar siempre por un fotograma I. Los fotogramas P y B, en caso de usarse, deben decodificarse junto a los fotogramas de referencia.

En el perfil base de H.264 sólo se utilizan fotogramas I y P. Este perfil es el ideal para cámaras de red y codificadores de video, ya que la latencia se reduce gracias a la ausencia de fotogramas B.

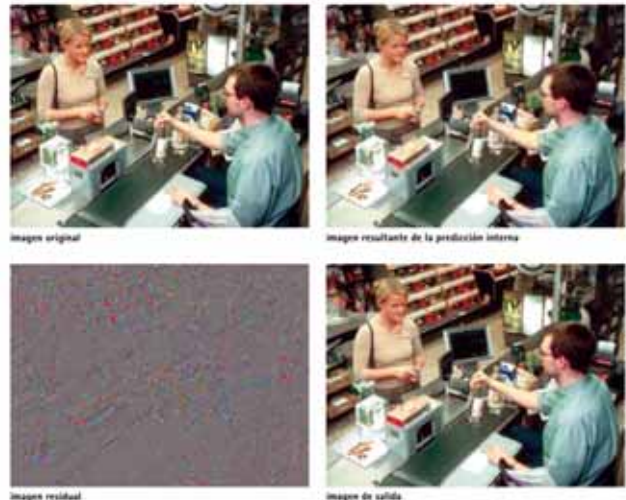
Eficiencia de H.264

H.264 lleva la tecnología de compresión de video a un nuevo nivel. Con H.264 se introduce un nuevo y avanzado esquema de predicción interna para codificar fotogramas I. Este esquema reduce en gran medida el tamaño en bits de un fotograma I y conserva una gran calidad, permitiendo la predicción sucesiva de bloques de píxeles más pequeños dentro de un macrobloque de un fotograma. Esto se realiza buscando los píxeles coincidentes entre los píxeles anteriormente codificados que rodean a un nuevo bloque de 4x4 píxeles que se va a codificar internamente. Mediante la reutilización de los valores de píxel que ya han sido codificados, se puede reducir drásticamente el tamaño en bits. Esta nueva predicción interna es una parte fun-

damental de la tecnología H.264 y ha demostrado ser muy eficiente. A modo de comparación, si se utilizaran exclusivamente fotogramas I en una transmisión H.264, el tamaño del archivo sería mucho menor que si se utilizara una transmisión Motion JPEG, que sólo utiliza fotogramas I.



Ilustraciones de algunos de los modos de predicción interna que pueden tener lugar en la codificación de 4x4 píxeles dentro de uno de los 16 bloques que forman un macrobloque. Cada uno de los 16 bloques de un macrobloque se pueden codificar usando métodos distintos.



Las imágenes de arriba ilustran la eficiencia del esquema de predicción interna de H.264, mediante el cual la imagen resultante de la predicción interna puede enviarse "gratis". Sólo es necesario codificar el contenido residual y los modos de predicción interna para crear la imagen de salida.

La compensación de movimiento basada en bloques (usada para codificar los fotogramas P y B) también ha sido mejorada en el estándar H.264. Un codificador puede elegir entre buscar bloques coincidentes (con una precisión inferior a un píxel) en unas pocas o en muchas áreas de uno o varios fotogramas de referencia. El tamaño y la forma del bloque también se puede ajustar para mejorar las coincidencias encontradas. En áreas donde no se encuentran bloques coincidentes en un fotograma de referencia, se utilizan macrobloques codificados internamente. El alto grado de flexibilidad de la compensación de movimiento basada en bloques es útil en la vigilancia de escenas abigarradas, manteniendo la calidad en aplicaciones exigentes. La compensación de movimiento es el aspecto más exigente de un codificador de video, y las diferentes formas y grados en el que puede implementarlo el codificador H.264 puede afectar a la eficiencia de la compresión del video.

Continúa en página 152

Viene de página 148

Con H.264, los típicos defectos pixelados (que pueden verse en video muy comprimido usando los estándares Motion JPEG y MPEG en vez de H.264) se pueden reducir usando un filtro de eliminación de bloques en bucle. Este filtro suaviza los bordes de los bloques con intensidad adaptativa para ofrecer un video descomprimido casi perfecto.



Los artefactos (defectos) pixelados en la imagen muy comprimida de la izquierda se reducen al aplicar un filtro de eliminación de bloques, como se muestra en la imagen de la derecha.

Conclusiones

H.264 es un gran paso adelante en la tecnología de compresión de video, ofrece técnicas más eficientes en la compresión gracias a sus mejores capacidades de predicción y a su capacidad de recuperación ante errores. Este estándar ofrece nuevas posibilidades a la hora de crear mejores codificadores de video que permiten transmisiones de video de mayor calidad y a mayores resoluciones manteniendo las mismas frecuencias de bits (en comparación a estándares anteriores), o, a la inversa, la misma calidad de video con frecuencias de bits inferiores. H.264 es el primer proyecto conjunto entre ITU, ISO y IEC para lograr un estándar de compresión de video común e internacional. Gracias a su flexibilidad, H.264 ya ha sido aplicado en distintas áreas como el DVD de alta definición (como Blu-ray), la difusión de video digital, incluyendo la TV de alta definición, el almacenamiento de video en línea (como YouTube), la telefonía móvil de tercera generación y en software como QuickTime, Flash y el sistema operativo de los ordenadores Apple, MacOS X así como en consolas de videojuegos como PlayStation 3. Con el apoyo de tantas industrias y aplicaciones para consumidores y profesionales, se espera que H.264 sustituya al resto de estándares y métodos de compresión utilizados en la actualidad.

A medida que el formato H.264 se popularice en cámaras de red, codificadores de video y software de gestión, los diseñadores e integradores de sistemas deberán asegurarse de que los productos y proveedores que elijan usen este nuevo estándar abierto. Y hasta el momento, los productos de video en red compatibles con H.264 y Motion JPEG son ideales para lograr la mayor flexibilidad e integración posibles. ■

El material técnico que se publica en este informe fue proporcionado por Axis Communications a Revista Negocios de Seguridad®. Prohibida su reproducción (parcial o total) sin el expreso consentimiento del autor o este medio.

Métodos básicos de reducción de datos

Existe toda una variedad de métodos que pueden utilizarse para reducir los datos de video, tanto dentro de un fotograma de imagen como entre una serie de fotogramas.

Dentro del fotograma de la imagen, los datos pueden reducirse, simplemente, eliminando la información que no es necesaria y que afecta a la resolución de la imagen.

En una serie de fotogramas, los datos de video pueden reducirse con métodos como la codificación diferencial, utilizada en la mayoría de estándares de compresión de video, H.264 incluido. En la codificación diferencial, un fotograma se compara con otro fotograma de referencia (como un fotograma I o P anterior) y sólo se codifican los píxeles que han cambiado respecto al fotograma de referencia. De esta forma, se reduce el número de valores de píxeles que hay que codificar y enviar.

Con el formato Motion JPEG, se codifican y envían las tres imágenes de arriba como imágenes separadas y únicas, (fotogramas I), in dependencias entre sí.



Con la codificación diferencial (la utilizada en la mayor parte de estándares de compresión de video, H.264 incluido) sólo se codifica entera la primera imagen (fotograma I).



En las dos imágenes siguientes (fotogramas P), existen referencias a la primera imagen en lo que se refiere a elementos estáticos, como la casa, mientras que sólo se codifican las partes móviles, como el hombre que corre, usando vectores de movimiento y reduciendo así la cantidad de información que se envía y almacena.

El grado de codificación se puede reducir aún más si la codificación y decodificación de las diferencias se basa en bloques de píxeles (macrobloques) en vez de en píxeles individuales. La consecuencia es que se comparan áreas más grandes y sólo se codifican los bloques que son significativamente distintos. Asimismo, se reduce la sobrecarga asociada a la localización de las áreas a cambiar.

La codificación diferencial, de cualquier manera, no reducirá los datos de forma significativa si hay mucho movimiento en el video. En este caso, pueden utilizarse técnicas como la compensación de movimiento basada en bloques. La compensación de movimiento basada en bloques tiene en cuenta que gran parte de un fotograma nuevo está ya incluido en el fotograma anterior, aunque quizás en un lugar diferente del mismo. Esta técnica divide un fotograma en una serie de macrobloques. Se puede componer o "predecir" un nuevo fotograma (por ejemplo, un fotograma P) bloque a bloque, buscando un bloque que coincida en un fotograma de referencia. Si se encuentra una coincidencia, el codificador simplemente codifica la posición en la que se ha encontrado el bloque que coincide del fotograma de referencia. La codificación del vector de movimiento, como se denomina, precisa de menos bits que si hubiera de codificarse el contenido real de un bloque.

Ilustración de la compensación de movimiento basada en bloques.

