EBook Gratis

APRENDIZAJE video

Free unaffiliated eBook created from
Stack Overflow contributors.

#video

Tabla de contenido

Acerca de 1

Capítulo 1: Empezando con el video 2

Observaciones 2

Examples 2

Comprender los archivos multimedia independientes 2

General 3

Vídeo 3

Audio 5

Resumen de analisis 5

Más 6

Capítulo 2: Entendiendo una presentación de medios 7

Observaciones 7

Examples 7

Comprender los archivos multimedia independientes 7

General 8

Vídeo 8

Audio 10

Resumen de analisis 10

Más 11

Comprender los componentes de las pistas de vídeo 11

Entendiendo las presentaciones de transmisión adaptativa DASH 14

Capítulo 3: Relaciones de aspecto de video 19

Observaciones 19

Examples 19

Relación de aspecto de la pantalla (DAR) 19

Relación de aspecto de la imagen (PAR) 20

Relación de aspecto de la muestra (SAR) 20

Relación de aspecto de píxeles 21

Creditos 22

Acerca de

You can share this PDF with anyone you feel could benefit from it, downloaded the latest version from: video

It is an unofficial and free video ebook created for educational purposes. All the content is extracted from Stack Overflow Documentation, which is written by many hardworking individuals at Stack Overflow. It is neither affiliated with Stack Overflow nor official video.

The content is released under Creative Commons BY-SA, and the list of contributors to each chapter are provided in the credits section at the end of this book. Images may be copyright of their respective owners unless otherwise specified. All trademarks and registered trademarks are the property of their respective company owners.

Use the content presented in this book at your own risk; it is not guaranteed to be correct nor accurate, please send your feedback and corrections to info@zzzprojects.com

Capítulo 1: Empezando con el video

Observaciones

La reproducción de video está en el centro y en una amplia gama de soluciones modernas, con el software y los estándares evolucionando rápidamente. Para comprender este campo, primero debe comprender los múltiples aspectos involucrados en el trabajo con el video:

- La información de color sin procesar capturada en el mundo físico se suele *codificar* utilizando un códec, un algoritmo cuyo propósito es representar estos datos en forma comprimida, a menudo sacrificando algunos detalles visuales en favor de una mayor compresión.
- Para la reproducción, se ejecuta el algoritmo inverso: los datos se *decodifican* para volver a convertirse en información de color en bruto que se puede suministrar a un dispositivo de salida (por ejemplo, un monitor).
- Entre la codificación y la decodificación, los datos comprimidos se *empaquetan* para su almacenamiento, lo que puede implicar la combinación de pistas de diferentes tipos en un solo archivo o la segmentación del contenido en un gran número de pequeños segmentos.
- El video se *entrega* al dispositivo del usuario final mediante una tecnología de entrega, que puede ser tan simple como la descarga de un archivo a través de HTTP o considerablemente más complejo, e incluye retroalimentación en vivo de la infraestructura de red y la adaptación automática de los niveles de calidad.
- El contenido premium generalmente se *encripta* antes del empaquetado y solo se puede reproducir en un reproductor equipado con una tecnología DRM que garantiza la seguridad de descifrado de la clave durante el uso y protege activamente contra la captura de salida.

Si bien la parte visual es obviamente dominante en visibilidad, el audio y el texto también desempeñan un papel clave en las presentaciones de medios, al proporcionar características en varios idiomas que hacen que el contenido sea accesible para una amplia audiencia. En la mayoría de los flujos de trabajo, las pistas de audio y texto se manejan de una manera equivalente a las pistas de video, codificadas, decodificadas, empaquetadas y entregadas en las mismas líneas.

Todos estos aspectos, y más, deben cuidarse en una solución moderna, que garantice una experiencia agradable para los usuarios finales.

Examples

Comprender los archivos multimedia independientes

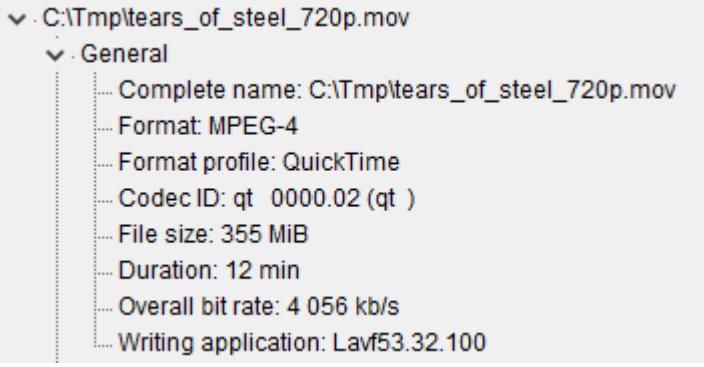
El contenido de muestra utilizado aquí es Tears of Steel, de Blender Foundation. Específicamente, usaremos la descarga titulada "HD 720p (~ 365MB, mov, 2.0)". Este es un archivo único que termina con la extensión "mov" y se reproducirá en casi cualquier reproductor multimedia moderno.

Tenga en cuenta que la página de descarga ofrece subtítulos como descargas de archivos SRT independientes. En este contenido de muestra, no hay subtítulos entregados en el mismo archivo. Por lo tanto, dejamos el análisis de subtítulos fuera del alcance de este ejemplo.

Una forma fácil de analizar varios archivos multimedia es usando la herramienta / biblioteca MediaInfo . Si bien la funcionalidad de análisis que se muestra aquí utiliza la GUI por simplicidad, todas las características también están disponibles a través de la API MediaInfo.

Al abrir este archivo en la GUI de MediaInfo y cambiar a la vista de árbol, verá tres secciones: General, Video y Audio. El primero contiene información básica sobre el archivo, mientras que los dos restantes describen una pista multimedia que se encuentra en este archivo. Examinemos la información más relevante en cada sección de la salida.

General



Los primeros parámetros de interés son *Formato* y *Formato de perfil* . El primero indica que el **formato de empaquetado** es del conjunto de estándares MPEG-4. MPEG-4 define el *formato de archivo de medios base ISO* y el *formato de empaquetado MP4* . Además, Apple ha creado su propia especificación que se deriva de estas, nombradas en MediaInfo como el perfil "QuickTime".

> Nota: Tenga cuidado de no confundir MP4 y MPEG-4; el primero se refiere a un formato de empaquetado específico en el conjunto de estándares internacionales MPEG-4, que también incluye códecs de video y audio. Esto puede llevar a confusión, así que evite usar el término MPEG-4 cuando haga referencia a cualquier otra cosa que no sea el conjunto completo de estándares.

Todos los formatos de empaquetado basados en el formato de archivo de medios ISO básico, definidos en la familia de estándares MPEG-4, son muy similares y, a menudo, pueden procesarse con las mismas herramientas, y sus diferencias son en gran medida una cuestión de extensiones personalizadas que a menudo pueden ser seguras. ignorado Por lo tanto, podemos esperar que el video de muestra aquí sea altamente compatible con todos los reproductores de video modernos.

Vídeo

```
Video
    ID: 1
    Format: AVC
    Format/Info: Advanced Video Codec
    Format profile: Main@L3.1
    Format settings, CABAC: No
    Format settings, ReFrames: 2 frames
    Format settings, GOP: M=4, N=18
    Codec ID: avc1
    Codec ID/Info: Advanced Video Coding
    Duration: 12 min
    Bit rate: 4 000 kb/s
    Width: 1 280 pixels
    Height: 534 pixels
    Display aspect ratio: 2.40:1
    Frame rate mode: Constant
    Frame rate: 24.000 FPS
    Color space: YUV
    Chroma subsampling: 4:2:0
    Bit depth: 8 bits
    Scan type: Progressive
    Bits/(Pixel*Frame): 0.244
    Stream size: 338 MiB (95%)
    Writing library: x264 core 122
    Encoding settings: cabac=0 / ref=2 / deblock=1:0:0 / ar
    Language: English
```

El detalle más importante sobre la pista de video es el códec que se utiliza para transformar los datos de color en bruto en una forma comprimida. El nombre del códec lo proporciona el parámetro *Formato* .

AVC también se conoce como H.264 y es el códec de video que hoy en día es el más extendido, compatible con prácticamente todos los dispositivos modernos y plataformas de software. Una pista de video codificada con AVC seguramente se reproducirá en casi cualquier reproductor.

Los códecs a menudo tienen múltiples *perfiles* que permiten que la funcionalidad del códec se divida en niveles, lo que permite la evolución de la tecnología de forma controlada. El parámetro de *perfil de formato* indica que este video usa el perfil principal. Este perfil es relativamente poco común, ya que casi todos los dispositivos modernos son compatibles con el perfil Alto, que ofrece una mayor eficiencia de compresión.

La calidad de la pista de video es a menudo de suma importancia. Aquí vemos los factores críticos expresados por los parámetros *Velocidad de bits* , *Anchura* y *Altura* . Los dos últimos insinúan que se trata de una pista de video de 720p, que se considera una calidad HD de gama baja. La imagen es en realidad más corta verticalmente que el marco estándar de 720p de 1280x720 píxeles.

La velocidad de bits mide la cantidad de datos que la forma comprimida del flujo de video ocupa, en promedio, por cada segundo de reproducción. Este es un parámetro crucial para la optimización, ya que la cantidad de datos entregados es una fuente importante de costos en soluciones de video a gran escala.

Los puntos de datos anteriores sobre la calidad del video son simplemente hechos que obtenemos del análisis: cualquier juicio sobre la idoneidad de estos parámetros es un tema que necesitaría mucho más análisis y se aborda por temas separados en esta categoría de documentación, al igual que muchos otros puntos finos de Trabajando con pistas de video.

Audio

- Audio
  - ID: 2
  - Format: MPEG Audio
  - Format version: Version 1
  - Format profile: Layer 3
  - Mode: Joint stereo
  - Codec ID: .mp3
  - Duration: 12 min
  - Bit rate mode: Constant
  - Bit rate: 192 kb/s
  - Channel(s): 2 channels
  - Channel positions: Front: L R
  - Sampling rate: 44.1 kHz
  - Compression mode: Lossy
  - Stream size: 16.8 MiB (5%)
  - Language: English
  - Default: Yes
  - Alternate group: 1

Una vez más, conocer el códec utilizado para codificar los datos de audio es de importancia crítica. Esto se expresa mediante los parámetros de *perfil Formato* y *Formato* . "MPEG Audio Layer 3" es más comúnmente conocido como MP3 y es un formato de audio universalmente compatible que se puede esperar que se reproduzca en cualquier lugar.

Al igual que con el video, los parámetros de calidad de audio son los segundos puntos de datos más importantes, expresados principalmente por el parámetro *Velocidad de bits* .

Resumen de analisis

El contenido se empaqueta utilizando un formato de empaquetado muy popular, construido sobre el conjunto de estándares MPEG-4. Se codifica utilizando códecs de audio y video universalmente adoptados. A partir de esto, queda claro que el video debe ser de fácil acceso para todos los espectadores: la compatibilidad y la disponibilidad fueron clave para sus autores.

El uso de MP3 muestra la antigüedad del contenido del ejemplo, ya que ya no se considera a la altura de los competidores modernos, en cambio, AAC (Codificación de audio avanzada) es el sostén de la familia en el campo de los códecs de audio.

Lo mismo se puede decir sobre el uso del perfil principal H.264. Es muy raro que se use un perfil H.264 además de Alto, dado que casi todos los decodificadores lo admiten, lo que permite a todos aprovechar la eficiencia mejorada que ofrecen las características de Alto perfil.

Las tasas de bits utilizadas son ligeramente más altas de lo esperado para el entorno actual. Esto puede explicarse por el deseo de los autores de alta calidad o simplemente por las limitaciones de los codificadores que estaban disponibles cuando se creó el contenido.

Más

Otras herramientas útiles para el análisis de archivos multimedia son FFprobe , que forma parte del paquete de software FFmpeg, y las herramientas Bento4 para trabajar con archivos MP4. Ambos también están disponibles en forma de biblioteca. Son capaces de realizar más análisis de bajo nivel que MediaInfo, en situaciones en las que necesita examinar elementos individuales que conforman los archivos multimedia.

Lea Empezando con el video en línea: https://riptutorial.com/es/video/topic/5690/empezando-con-el-video

Capítulo 2: Entendiendo una presentación de medios

Observaciones

La reproducción de video está en el centro y en una amplia gama de soluciones modernas, con el software y los estándares evolucionando rápidamente. Para comprender qué constituye una presentación multimedia, primero debe comprender los múltiples aspectos involucrados en el trabajo con el video:

- La información de color sin procesar capturada en el mundo físico se suele *codificar* utilizando un códec, un algoritmo cuyo propósito es representar estos datos en forma comprimida, a menudo sacrificando algunos detalles visuales en favor de una mayor compresión.
- Para la reproducción, se ejecuta el algoritmo inverso: los datos se *decodifican* para volver a convertirse en información de color en bruto que se puede suministrar a un dispositivo de salida (por ejemplo, un monitor).
- Entre la codificación y la decodificación, los datos comprimidos se *empaquetan* para su almacenamiento, lo que puede implicar la combinación de pistas de diferentes tipos en un solo archivo o la segmentación del contenido en un gran número de pequeños segmentos.
- El video se *entrega* al dispositivo del usuario final mediante una tecnología de entrega, que puede ser tan simple como la descarga de un archivo a través de HTTP o considerablemente más complejo, e incluye retroalimentación en vivo de la infraestructura de red y la adaptación automática de los niveles de calidad.
- El contenido premium generalmente se *encripta* antes del empaquetado y solo se puede reproducir en un reproductor equipado con una tecnología DRM que garantiza la seguridad de descifrado de la clave durante el uso y protege activamente contra la captura de salida.

Si bien la parte visual es obviamente dominante en visibilidad, el audio y el texto también desempeñan un papel clave en las presentaciones de medios, al proporcionar características en varios idiomas que hacen que el contenido sea accesible para una amplia audiencia. En la mayoría de los flujos de trabajo, las pistas de audio y texto se manejan de una manera equivalente a las pistas de video, codificadas, decodificadas, empaquetadas y entregadas en las mismas líneas.

Todos estos aspectos, y más, afectan la composición, el formato y el uso de una presentación de medios. Deben entenderse como un todo para hacer un uso efectivo de las tecnologías de los medios.

Examples

Comprender los archivos multimedia independientes

El contenido de muestra utilizado aquí es Tears of Steel , de Blender Foundation.

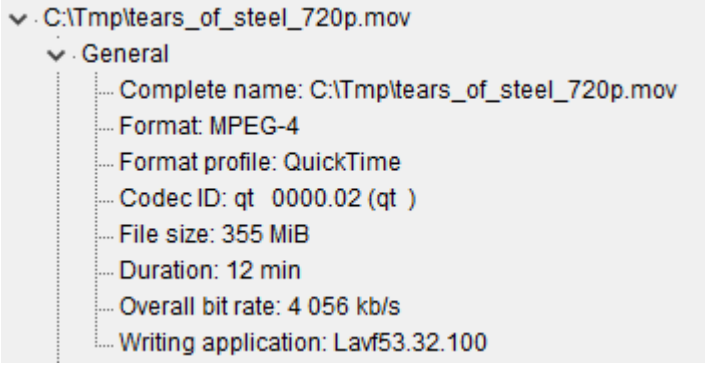
Específicamente, usaremos la descarga titulada "HD 720p (~ 365MB, mov, 2.0)". Este es un archivo único que termina con la extensión "mov" y se reproducirá en casi cualquier reproductor multimedia moderno.

Tenga en cuenta que la página de descarga ofrece subtítulos como descargas de archivos SRT independientes. En este contenido de muestra, no hay subtítulos entregados en el mismo archivo. Por lo tanto, dejamos el análisis de subtítulos fuera del alcance de este ejemplo.

Una forma fácil de analizar varios archivos multimedia es usando la herramienta / biblioteca MediaInfo . Si bien la funcionalidad de análisis que se muestra aquí utiliza la GUI por simplicidad, todas las características también están disponibles a través de la API MediaInfo.

Al abrir este archivo en la GUI de MediaInfo y cambiar a la vista de árbol, verá tres secciones: General, Video y Audio. El primero contiene información básica sobre el archivo, mientras que los dos restantes describen una pista multimedia que se encuentra en este archivo. Examinemos la información más relevante en cada sección de la salida.

General

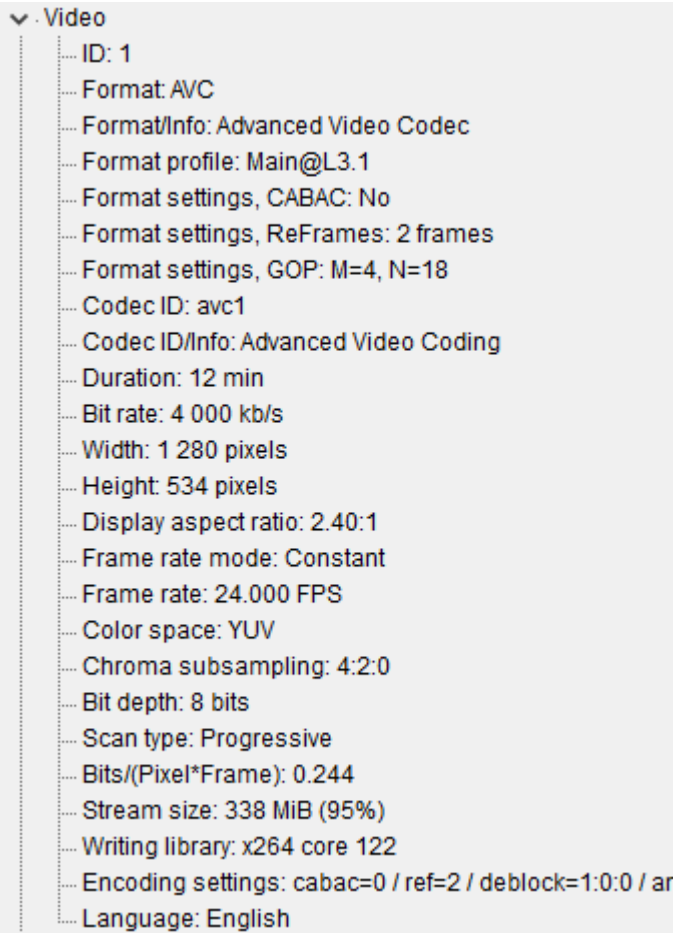


Los primeros parámetros de interés son *Formato* y *Formato de perfil* . El primero indica que el **formato de empaquetado** es del conjunto de estándares MPEG-4. MPEG-4 define el *formato de archivo de medios base ISO* y el *formato de empaquetado MP4* . Además, Apple ha creado su propia especificación que se deriva de estas, nombradas en MediaInfo como el perfil "QuickTime".

> Nota: Tenga cuidado de no confundir MP4 y MPEG-4; el primero se refiere a un formato de empaquetado específico en el conjunto de estándares internacionales MPEG-4, que también incluye códecs de video y audio. Esto puede llevar a confusión, así que evite usar el término MPEG-4 cuando haga referencia a cualquier otra cosa que no sea el conjunto completo de estándares.

Todos los formatos de empaquetado basados en el formato de archivo de medios ISO básico, definidos en la familia de estándares MPEG-4, son muy similares y, a menudo, pueden procesarse con las mismas herramientas, y sus diferencias son en gran medida una cuestión de extensiones personalizadas que a menudo pueden ser seguras. ignorado Por lo tanto, podemos esperar que el video de muestra aquí sea altamente compatible con todos los reproductores de video modernos.

Vídeo



El detalle más importante sobre la pista de video es el códec que se utiliza para transformar los datos de color en bruto en una forma comprimida. El nombre del códec lo proporciona el parámetro *Formato* .

AVC también se conoce como H.264 y es el códec de video que hoy en día es el más extendido, compatible con prácticamente todos los dispositivos modernos y plataformas de software. Una pista de video codificada con AVC seguramente se reproducirá en casi cualquier reproductor.

Los códecs a menudo tienen múltiples *perfiles* que permiten que la funcionalidad del códec se divida en niveles, lo que permite la evolución de la tecnología de forma controlada. El parámetro de *perfil de formato* indica que este video usa el perfil principal. Este perfil es relativamente poco común, ya que casi todos los dispositivos modernos son compatibles con el perfil Alto, que ofrece una mayor eficiencia de compresión.

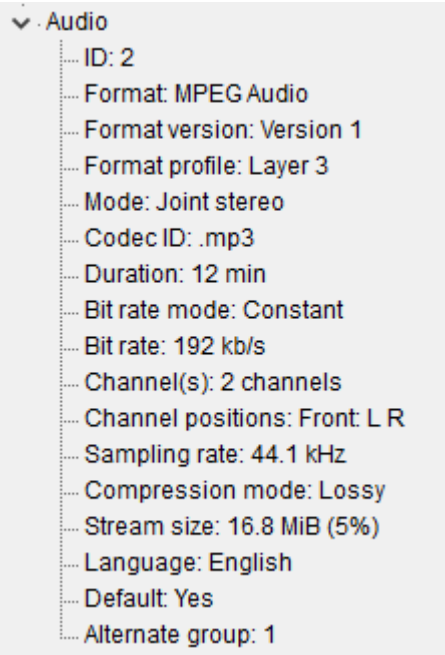
La calidad de la pista de video es a menudo de suma importancia. Aquí vemos los factores críticos expresados por los parámetros *Velocidad de bits* , *Anchura* y *Altura* . Los dos últimos insinúan que se trata de una pista de video de 720p, que se considera una calidad HD de gama baja. La imagen es en realidad más corta verticalmente que el marco estándar de 720p de 1280x720 píxeles.

La velocidad de bits mide la cantidad de datos que la forma comprimida del flujo de video ocupa,

en promedio, por cada segundo de reproducción. Este es un parámetro crucial para la optimización, ya que la cantidad de datos entregados es una fuente importante de costos en soluciones de video a gran escala.

Los puntos de datos anteriores sobre la calidad del video son simplemente hechos que obtenemos del análisis: cualquier juicio sobre la idoneidad de estos parámetros es un tema que necesitaría mucho más análisis y se aborda por temas separados en esta categoría de documentación, al igual que muchos otros puntos finos de Trabajando con pistas de video.

Audio



Una vez más, conocer el códec utilizado para codificar los datos de audio es de importancia crítica. Esto se expresa mediante los parámetros de *perfil Formato* y *Formato* . "MPEG Audio Layer 3" es más comúnmente conocido como MP3 y es un formato de audio universalmente compatible que se puede esperar que se reproduzca en cualquier lugar.

Al igual que con el video, los parámetros de calidad de audio son los segundos puntos de datos más importantes, expresados principalmente por el parámetro *Velocidad de bits* .

Resumen de analisis

El contenido se empaqueta utilizando un formato de empaquetado muy popular, construido sobre el conjunto de estándares MPEG-4. Se codifica utilizando códecs de audio y video universalmente adoptados. A partir de esto, queda claro que el video debe ser de fácil acceso para todos los espectadores: la compatibilidad y la disponibilidad fueron clave para sus autores.

El uso de MP3 muestra la antigüedad del contenido del ejemplo, ya que ya no se considera a la altura de los competidores modernos, en cambio, AAC (Codificación de audio avanzada) es el sostén de la familia en el campo de los códecs de audio.

Lo mismo se puede decir sobre el uso del perfil principal H.264. Es muy raro que se use un perfil H.264 además de Alto, dado que casi todos los decodificadores lo admiten, lo que permite a todos aprovechar la eficiencia mejorada que ofrecen las características de Alto perfil.

Las tasas de bits utilizadas son ligeramente más altas de lo esperado para el entorno actual. Esto puede explicarse por el deseo de los autores de alta calidad o simplemente por las limitaciones de los codificadores que estaban disponibles cuando se creó el contenido.

Más

Otras herramientas útiles para el análisis de archivos multimedia son FFprobe , que forma parte del paquete de software FFmpeg, y las herramientas Bento4 para trabajar con archivos MP4. Ambos también están disponibles en forma de biblioteca. Son capaces de realizar más análisis de bajo nivel que MediaInfo, en situaciones en las que necesita examinar elementos individuales que conforman los archivos multimedia.

Comprender los componentes de las pistas de vídeo.

Este ejemplo explorará cómo ver el diseño de una pista de video y cómo extraer las imágenes individuales que contiene.

El contenido de muestra utilizado aquí es Tears of Steel , de Blender Foundation. Específicamente, usaremos la descarga titulada "HD 720p (~ 365MB, mov, 2.0)". Este es un archivo único que termina con la extensión "mov" y se reproducirá en casi cualquier reproductor multimedia moderno.

Usaremos las herramientas mp4info y mp4dump de la suite Bento4 para el diseño de la pista y el análisis de la estructura, y FFmpeg para extraer las imágenes individuales que conforman la pista de video.

La película de muestra utiliza el formato de empaquetado "QuickTime" (MOV), que se basa en el formato de archivo de medios ISO básico, un estándar internacional que subyace en todos los formatos de empaquetado de la familia de formatos de archivo MP4. Esto lo hace altamente compatible con la mayoría de las herramientas disponibles y permite un análisis fácil.

Primero examinemos la estructura general del archivo. Todos los archivos de medios basados en el formato de archivo de medios ISO se estructuran como una jerarquía de cuadros, un mini sistema de archivos de tipo. Use la utilidad mp4dump para extraer la estructura de la caja, ejecutando el siguiente comando:

```
mp4dump tears_of_steel_720p.mov
```

La salida será similar a la siguiente:

```
[ftyp] size=8+12
  major_brand = qt
  minor_version = 200
```

```
  compatible_brand = qt
[wide] size=8+0
[mdat] size=8+371579623
[moov] size=8+598972
  [mvhd] size=12+96
    timescale = 1000
    duration = 734167
    duration(ms) = 734167
  [trak] size=8+244250
    [tkhd] size=12+80, flags=f
      enabled = 1
      id = 1
      duration = 734167
      width = 1280.000000
      height = 534.000000
...
```

Esto representa la estructura interna del archivo. Por ejemplo, aquí ve un cuadro de *moov* que tiene un encabezado de 8 bytes y 598972 bytes de contenido. Este cuadro es un contenedor para varios cuadros de metadatos que describen el contenido del archivo. Para obtener más información sobre el significado de las distintas cajas y sus propiedades, consulte ISO / IEC 14496-12 .

Las muestras de medios reales en sí mismas (las imágenes comprimidas y las formas de onda de audio) se almacenan en el cuadro *mdat* , cuyo contenido es opaco para la utilidad mp4dump.

Para excluir datos irrelevantes y simplificar el flujo de trabajo de análisis (este ejemplo se centra en la pista de video), ahora eliminamos la pista de audio de nuestra película de muestra. Ejecuta el siguiente comando:

```
ffmpeg -i tears_of_steel_720p.mov -an -vcodec copy video_track.mov
```

> Tenga en cuenta que el paso anterior también eliminará varios elementos de extensión personalizados del video de entrada, empaquetando la esencia del contenido visual en un nuevo archivo contenedor y descartando cualquier otra cosa. Si hace esto en un escenario de producción, ¡asegúrese de que realmente puede descartar todos los demás elementos del archivo de entrada!

Las pistas de vídeo codificadas son una secuencia de imágenes. Con el códec H.264 utilizado aquí, y todos los demás códecs modernos utilizados comúnmente, las imágenes pueden ser de varios tipos diferentes:

- Imágenes I: estas son imágenes independientes, que se pueden decodificar únicamente utilizando los datos contenidos en la imagen.
- Imágenes P: toman otra imagen como línea de base y aplican una transformación a esa imagen (por ejemplo, "mueva estos píxeles particulares a la derecha en 5 píxeles").
- Las imágenes B, similares a las imágenes P, pero bidireccionales, también pueden hacer referencia a imágenes del futuro y definir transformaciones como "estos píxeles particulares que serán completamente visibles en 5 imágenes ahora son visibles en un 10%".

La combinación exacta de tipos de imágenes puede elegirse libremente por el flujo de trabajo de

codificación, lo que crea muchas oportunidades de optimización, aunque ciertos casos de uso pueden restringir la flexibilidad disponible, por ejemplo, al requerir que esté presente un fotograma I a intervalos de exactamente 2 segundos.

Ejecute el siguiente comando para ver la estructura de la imagen de la pista de video:

```
mp4info --show-layout video_track.mov
```

Además de presentar una forma legible por humanos de los metadatos del archivo en general, verá una impresión detallada del diseño de la imagen de la pista de video.

```
...
  00000959 [V] (1)  size=  7615, offset=15483377, dts=491008 (39958 ms)
  00000960 [V] (1)* size=104133, offset=15490992, dts=491520 (40000 ms)
  00000961 [V] (1)  size= 16168, offset=15595125, dts=492032 (40042 ms)
  00000962 [V] (1)  size=  4029, offset=15611293, dts=492544 (40083 ms)
  00000963 [V] (1)  size= 24615, offset=15615322, dts=493056 (40125 ms)
  00000964 [V] (1)  size=  4674, offset=15639937, dts=493568 (40167 ms)
  00000965 [V] (1)  size= 18451, offset=15644611, dts=494080 (40208 ms)
  00000966 [V] (1)  size= 95800, offset=15663062, dts=494592 (40250 ms)
  00000967 [V] (1)  size= 30271, offset=15758862, dts=495104 (40292 ms)
  00000968 [V] (1)  size= 10997, offset=15789133, dts=495616 (40333 ms)
  00000969 [V] (1)  size= 28458, offset=15800130, dts=496128 (40375 ms)
  00000970 [V] (1)  size=  9593, offset=15828588, dts=496640 (40417 ms)
  00000971 [V] (1)  size= 24548, offset=15838181, dts=497152 (40458 ms)
  00000972 [V] (1)  size=  6853, offset=15862729, dts=497664 (40500 ms)
  00000973 [V] (1)  size= 27698, offset=15869582, dts=498176 (40542 ms)
  00000974 [V] (1)  size=  7565, offset=15897280, dts=498688 (40583 ms)
  00000975 [V] (1)  size= 24682, offset=15904845, dts=499200 (40625 ms)
  00000976 [V] (1)  size=  5535, offset=15929527, dts=499712 (40667 ms)
  00000977 [V] (1)  size= 38360, offset=15935062, dts=500224 (40708 ms)
  00000978 [V] (1)* size= 82466, offset=15973422, dts=500736 (40750 ms)
  00000979 [V] (1)  size= 13388, offset=16055888, dts=501248 (40792 ms)
  00000980 [V] (1)  size=  2315, offset=16069276, dts=501760 (40833 ms)
  00000981 [V] (1)  size= 21983, offset=16071591, dts=502272 (40875 ms)
  00000982 [V] (1)  size=  3384, offset=16093574, dts=502784 (40917 ms)
  00000983 [V] (1)  size= 22225, offset=16096958, dts=503296 (40958 ms)
...
```

Cada fila en esta impresión es una imagen contenida en la pista de video. Aquellos marcados con un asterisco como `(1)*` son I-imágenes. Puede ver cómo son las de mayor tamaño, mientras que las otras permiten una mayor compresión al hacer referencia a las imágenes existentes y solo describen las diferencias.

> La lista también contiene el desplazamiento de los datos de la imagen en el archivo de video y la marca de tiempo de decodificación de la imagen, lo que permite una mayor correlación y análisis. Tenga en cuenta que el orden de decodificación / sincronización de las imágenes no es necesariamente el mismo que el orden de presentación / sincronización. Si existen imágenes B en el video, solo se pueden decodificar *después de* las imágenes a las que hacen referencia, incluso si se presentan *antes que* las imágenes referenciadas.

Después de haber aprendido algo sobre la estructura de la pista de video, ejecute el siguiente

comando para extraer las 30 imágenes que comienzan en la marca de 40 segundos como archivos PNG:

```
ffmpeg -i video_track.mov -ss 00:00:40 -vframes 30 picture%04d.png
```

> Las imágenes extraídas se descodificarán completamente, como aparecerían en un reproductor de video; no es posible (sin herramientas extremadamente especializadas) obtener una representación visual de los datos sin procesar en fotogramas P o fotogramas B.

Observe cómo la séptima imagen generada es un cambio completo de escena en el video. Puede correlacionar esto fácilmente con la salida de mp4info anterior: la séptima imagen que comienza desde la marca de los 40 segundos (número 00000966) es mucho más grande en tamaño que las cercanas. Los cambios de escena son difíciles de codificar, ya que actualizan la imagen completa y contienen muchos datos nuevos. Si al codificador no se le otorga suficiente indulgencia para optimizar los cambios de escena (es decir, no se le permite generar una imagen grande), la salida visual será de baja calidad o "bloqueada" hasta la siguiente imagen I. Al examinar la asignación de ancho de banda (bytes) a varias imágenes, puede obtener información sobre los artefactos visuales que pueden aparecer repentinamente en un video.

Entendiendo las presentaciones de transmisión adaptativa DASH

DASH es la tecnología de transmisión adaptativa más ampliamente implementada en soluciones modernas, utilizada para entregar video en una amplia variedad de escenarios. La mejor manera de entender las presentaciones DASH es observar la actividad de la red que tiene lugar durante la reproducción.

Este ejemplo utiliza Fiddler para capturar y analizar el tráfico de la red del navegador, aunque cualquier herramienta similar también será suficiente. Usaremos el reproductor de código abierto dash.js para la reproducción de video.

Para nuestro contenido de demostración, utilizaremos los vectores de prueba DASH de Axinom , específicamente la variante de 1080p de un solo período del vector de prueba "Clear".

Con su captura de red en ejecución, abra el reproductor de muestra de compilación nocturna dash.js en cualquier navegador moderno e ingrese la URL http://media.axprod.net/TestVectors/v6-Clear/Manifest_1080p.mpd en el cuadro de texto. Presione Cargar para iniciar la reproducción. Observará cómo se descargan los siguientes archivos:

```
http://media.axprod.net/TestVectors/v7-Clear/Manifest_1080p.mpd
http://media.axprod.net/TestVectors/v7-Clear/2/init.mp4
http://media.axprod.net/TestVectors/v7-Clear/15/init.mp4
http://media.axprod.net/TestVectors/v7-Clear/18/init.mp4
http://media.axprod.net/TestVectors/v7-Clear/18/0001.m4s
http://media.axprod.net/TestVectors/v7-Clear/2/0001.m4s
http://media.axprod.net/TestVectors/v7-Clear/15/0001.m4s
http://media.axprod.net/TestVectors/v7-Clear/15/0002.m4s
http://media.axprod.net/TestVectors/v7-Clear/5/init.mp4
http://media.axprod.net/TestVectors/v7-Clear/5/0002.m4s
http://media.axprod.net/TestVectors/v7-Clear/18/0002.m4s
http://media.axprod.net/TestVectors/v7-Clear/5/0003.m4s
...
```

El primer archivo es el manifiesto de presentación, un documento XML cuyo formato se define en ISO / IEC 23009-1 . Esto describe la presentación DASH a una profundidad suficiente para que el jugador entienda cómo reproducirla.

Si mira dentro del manifiesto, verá varios elementos de `AdaptationSet` , cada uno de los cuales describe una única adaptación del contenido. Por ejemplo, hay un conjunto de adaptación para la pista de video, tres conjuntos de adaptación para tres idiomas de audio y cinco conjuntos de adaptación para cinco idiomas de subtítulos.

Los conjuntos de adaptación interior son elementos de `Representation` . Para el conjunto de adaptación de video, hay varios de estos: cada representación contiene el mismo contenido visual codificado con un nivel de calidad diferente. Cada conjunto de adaptación de audio y texto solo tiene una representación.

```
AdaptationSet [ segmentAlignment=true group=1 maxWidth=1920 maxHeight=1080 maxFrameRate=24 par=16:9 la
    <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main" xmlns="urn:mpeg:dash:schema:mpd:2011" />
    <SegmentTemplate timescale="1200000" media="$RepresentationID$/$Number%04d$.m4s" startNumber="1" d
    <Representation id="8" mimeType="video/mp4" codecs="hev1.2.4.L63.90" width="512" height="288" frameRat
    <Representation id="9" mimeType="video/mp4" codecs="hev1.2.4.L63.90" width="640" height="360" frameRat
    <Representation id="10" mimeType="video/mp4" codecs="hev1.2.4.L90.90" width="852" height="480" frameRa
    <Representation id="11" mimeType="video/mp4" codecs="hev1.2.4.L93.90" width="1280" height="720" frameR
    <Representation id="12" mimeType="video/mp4" codecs="hev1.2.4.L120.90" width="1920" height="1080" fram
AdaptationSet [ segmentAlignment=true group=2 lang=en ]
    <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main" xmlns="urn:mpeg:dash:schema:mpd:2011" />
    <SegmentTemplate timescale="24000" media="$RepresentationID$/$Number%04d$.m4s" startNumber="1" dura
    Representation [ id=15 mimeType=audio/mp4 codecs=mp4a.40.29 audioSamplingRate=48000 startWithSAP=1 b
        <AudioChannelConfiguration schemeIdUri="urn:mpeg:dash:23003:3:audio_channel_configuration:2011" valu
AdaptationSet [ segmentAlignment=true group=2 lang=en-AU ]
    <SegmentTemplate timescale="24000" media="$RepresentationID$/$Number%04d$.m4s" startNumber="1" dura
    Representation [ id=16 mimeType=audio/mp4 codecs=mp4a.40.29 audioSamplingRate=48000 startWithSAP=1 b
        <AudioChannelConfiguration schemeIdUri="urn:mpeg:dash:23003:3:audio_channel_configuration:2011" valu
AdaptationSet [ segmentAlignment=true group=2 lang=et-ET ]
    <SegmentTemplate timescale="24000" media="$RepresentationID$/$Number%04d$.m4s" startNumber="1" dura
    Representation [ id=17 mimeType=audio/mp4 codecs=mp4a.40.29 audioSamplingRate=48000 startWithSAP=1 b
        <AudioChannelConfiguration schemeIdUri="urn:mpeg:dash:23003:3:audio_channel_configuration:2011" valu
AdaptationSet [ segmentAlignment=true group=3 lang=en ]
    <Role schemeIdUri="urn:mpeg:dash:role:2011" value="main" xmlns="urn:mpeg:dash:schema:mpd:2011" />
    <Role schemeIdUri="urn:mpeg:dash:role:2011" value="subtitle" xmlns="urn:mpeg:dash:schema:mpd:2011" />
    <SegmentTemplate timescale="1000" media="$RepresentationID$/$Number%04d$.m4s" startNumber="1" durat
    <Representation id="18" mimeType="application/mp4" codecs="wvtt" startWithSAP="1" bandwidth="428" xmlns
```

Para realizar la reproducción, un jugador deberá decidir qué ajustes de adaptación presentar al espectador. Puede tomar esta decisión basándose en la lógica empresarial personalizada o incorporada que desee (p. Ej., Orden de preferencia de idioma). Los conjuntos de adaptación que el autor del contenido considera primario tienen un elemento de `Role` en el manifiesto que declara la función "principal".

Además, el jugador tendrá que decidir qué representación presentar al espectador (si un conjunto de adaptación ofrece representaciones múltiples). La mayoría de los jugadores comienzan de forma conservadora y aplican un algoritmo heurístico que intentará presentar el nivel de calidad máximo que la conexión de red del espectador puede mantener.

El jugador es libre de cambiar el conjunto activo de representaciones y / o conjuntos de adaptación en cualquier momento, ya sea en respuesta a la acción del usuario (seleccionando un

idioma diferente) o la lógica automatizada (las heurísticas de ancho de banda resultan en un cambio de nivel de calidad).

El elemento `SegmentTemplate` define la estructura de URL que el jugador puede usar para acceder a las diferentes representaciones. Un factor clave de las presentaciones DASH es que el contenido se divide en pequeños segmentos de unos pocos segundos cada uno (4 segundos en el caso de nuestra película de muestra), que se descargan de forma independiente. Cada representación también tiene un segmento de inicialización, llamado "init.mp4" para esta película de muestra, que contiene la configuración del decodificador específica de la representación y, por lo tanto, debe cargarse antes de que se pueda procesar cualquier otro segmento de esa representación.

> El comportamiento descrito aquí es preciso para el perfil DASH Live, que es la variante más utilizada de DASH. También existen otros perfiles con comportamientos ligeramente diferentes, no cubiertos aquí. ¡Preste atención al atributo "perfil" en el elemento raíz del manifiesto DASH para asegurarse de que esta descripción se aplique a sus videos!

Al examinar la lista de URL obtenidas de la captura de tráfico de la red y compararla con la información proporcionada por el manifiesto, concluirá que el jugador realizó las siguientes acciones después de descargar el manifiesto:

1. Descargue los segmentos de inicialización para las representaciones 2 (video 360p), 15 (audio en inglés) y 18 (subtítulos en inglés).
2. Descargue los primeros segmentos de las tres representaciones anteriores (0001.m4s).
3. Descarga el segundo segmento de la representación de audio.
4. A partir del segundo segmento de video, ¡cambie a la transmisión de video de 1080p! Esto se indica mediante la descarga del segmento de inicialización y el segundo segmento de representación 5 (video 1080p).
5. Continuar descargando más segmentos de las representaciones activas.

Al observar la actividad de la red, resulta fácil observar las decisiones que toma en funcionamiento un reproductor de transmisión adaptable DASH. Dicho reproductor es simplemente un mecanismo que descarga segmentos de varias pistas y los proporciona consecutivamente a un motor de reproducción de medios, cambiando las pistas según corresponda.

Los vectores de prueba DASH de Axinom también contienen archivos de almacenamiento que le permiten descargar la presentación completa para el análisis a nivel de sistema de archivos. Encontrará que los archivos en el disco son exactamente como están en el nivel de la red. Esto significa que las presentaciones DASH pueden ser servidas por servidores HTTP arbitrarios, sin la necesidad de ninguna lógica personalizada del lado del servidor.

Un aspecto del DASH de perfil en vivo que complica el análisis es que las muestras de medios se distribuyen en una gran cantidad de segmentos. La mayoría de las herramientas de análisis de medios no pueden procesar segmentos individuales, operando solo en pistas completas. A menudo puede superar esta limitación simplemente concatenando los segmentos de una sola representación, comenzando con el segmento de inicialización. Por ejemplo, en Windows puede

usar el siguiente comando:

```
copy init.mp4 /b + 0001.mp4 /b + 0002.mp4 /b + 0003.mp4 /b track.mp4
```

Esto creará un archivo track.mp4 que contiene los primeros tres segmentos de medios de una representación. Si bien no es idéntica en estructura a un archivo MP4 independiente, la mayoría de las herramientas (como mp4info y FFmpeg) aún pueden analizar dicho archivo sin una pérdida significativa de funcionalidad.

Lea Entendiendo una presentación de medios en línea:
https://riptutorial.com/es/video/topic/5879/entendiendo-una-presentacion-de-medios

Capítulo 3: Relaciones de aspecto de video

Observaciones

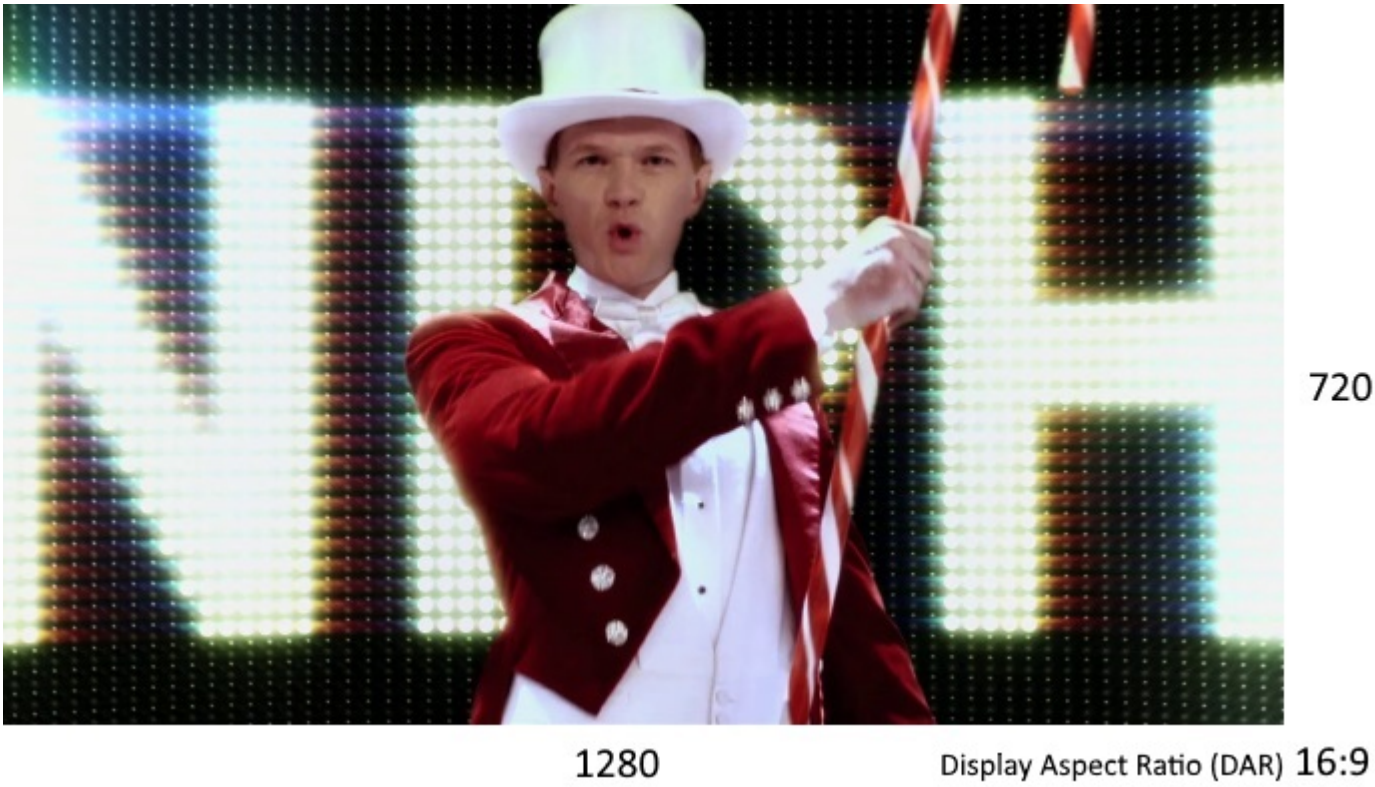
Las relaciones de aspecto a menudo se expresan como una relación ancho: altura que a menudo, pero no siempre, se simplifica y, a veces, también como un entero entero de coma flotante.

Todas las siguientes relaciones de aspecto tienen el mismo valor expresado de diferentes maneras:

- 1280: 720
- 16: 9
- 1.777777777777777777777777777778

Examples

Relación de aspecto de la pantalla (DAR)



Esta es una captura de pantalla de un video en reproducción. Verá un video normal de 16: 9 como lo esperaría ver en cualquier solución de video moderna. Esto, la relación de aspecto que ve el espectador, es lo que se denomina *relación de aspecto de visualización* o DAR.

De los parámetros ilustrados, vemos que DAR = 1280: 720 = 16: 9 = 1.777777777777777777777777777778.

Relación de aspecto de la imagen (PAR)

Internamente, todos los videos son solo una serie de fotos. Echemos un vistazo a una de esas imágenes.

Eso se ve extraño, ¿verdad? En efecto. Las imágenes que componen un video pueden tener una relación de aspecto diferente de la DAR, la mayoría de las veces por razones algorítmicas (por ejemplo, solo los tamaños que son múltiplos de 16 pueden comprimirse con el algoritmo elegido). Esto se llama *relación de aspecto de la imagen* o PAR.

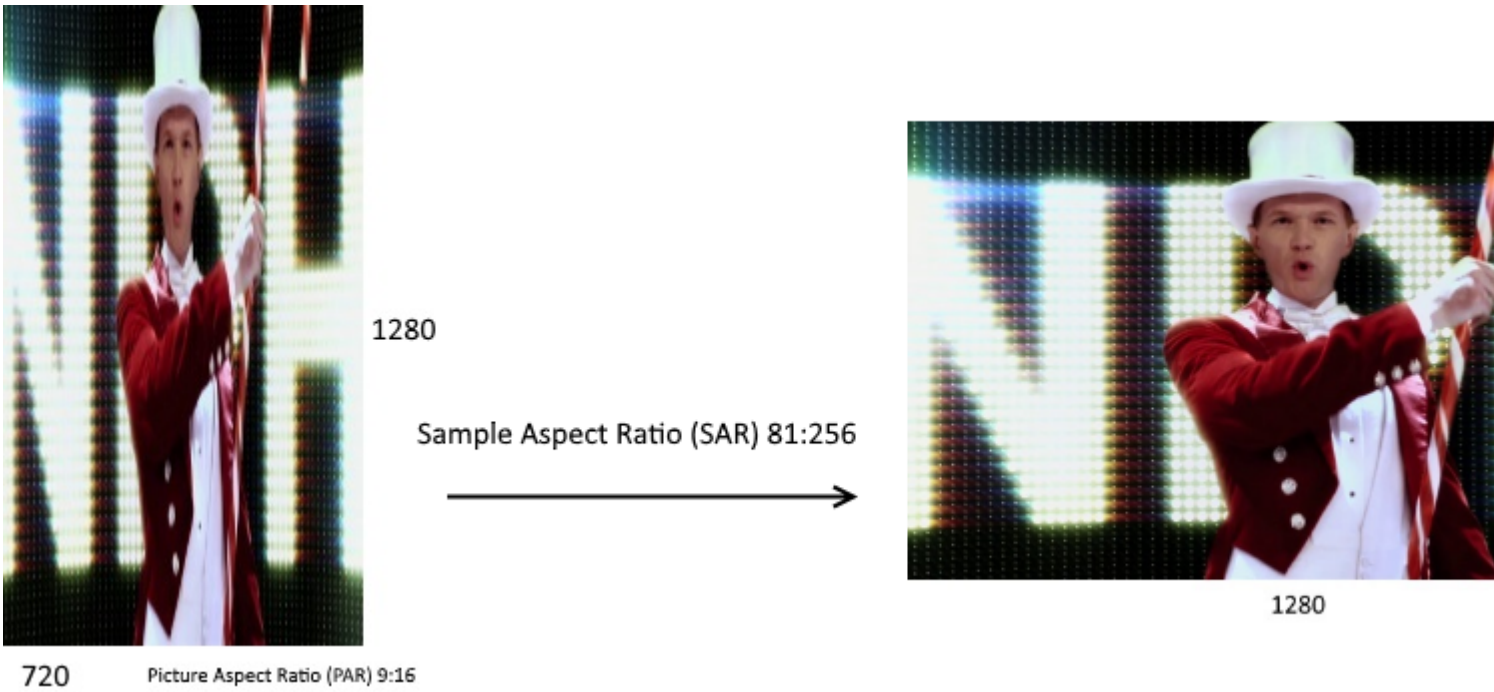
En este ejemplo, tenemos las dimensiones de la imagen exactamente invertidas entre la forma mostrada (16: 9) y la imagen real, por lo que el PAR es 9:16. Normalmente, las diferencias son más pequeñas, pero este ejemplo es exagerado para mayor claridad.

De los parámetros ilustrados, vemos que PAR = 720: 1280 = 9:16 = 0.5625

Relación de aspecto de la muestra (SAR)

Como lo indica el ejemplo de la *relación de aspecto de* la *imagen* , los videos son series de imágenes que no necesariamente tienen la misma relación de aspecto que el resultado final que se mostrará al usuario.

Entonces, ¿cómo se obtiene de esas imágenes estiradas a la salida que se muestra normalmente? Necesitas un factor de estiramiento! Este factor de estiramiento se aplica a la imagen para llevarla a la relación de aspecto correcta para su visualización. Esta es la *relación de aspecto de* la *muestra* o SAR.



El factor de estiramiento a menudo se expresa como una proporción de dos enteros. Lo calculas como SAR = PAR / DAR.

De los parámetros ilustrados, vemos que SAR = 9:16 / 16: 9 = (9/16) / (16/9) = 81/256 = 3.160493827160493827160493827160 5

Relación de aspecto de píxeles

Este es otro nombre para la *relación de aspecto de la muestra* y debe evitarse, ya que las siglas naturales (PAR) entran en conflicto con la *relación de aspecto de la imagen* .

Lea Relaciones de aspecto de video en línea:
https://riptutorial.com/es/video/topic/5713/relaciones-de-aspecto-de-video

Creditos

S. No	Capítulos	Contributors
1	Empezando con el video	Community, Sander
2	Entendiendo una presentación de medios	Sander
3	Relaciones de aspecto de video	Sander