Trends and Challenges in Next-generation Video Coding

Jian-Liang Lin, Ph.D.
Advanced Technology Development Div.
MediaTek Inc.
OUTLINE

- Introduction to Video Compression
- Overview of Video Coding Standards
- H.264 Video Coding Standard
- Next-generation video coding
OUTLINE

• Introduction to Video Compression
• Overview of Video Coding Standards
• H.264 Video Coding Standard
• Next-generation video coding
Video communications

- Compress still images or moving video before transmission

- Key issues:
  - compression efficiency and image quality
  - computational complexity
  - frame rate
## The need for image compression

- **512 x 512 pixel color image**
  
  \[512 \times 512 \times 24\text{bits} = 786 \text{ Kbytes}\]

- **Video conference QCIF (quarter common intermediate format)**
  
  \[(176 \times 144 + 88 \times 72 + 88 \times 72) \times 8 \times 25 = 7.6 \text{ Mbits/s}\]

- **Digital television**
  
  \[(720 \times 576 + 360 \times 288 + 360 \times 288 ) \times 8 \times 25 = 124 \text{ Mbits/s}\]

- **High definition television - HDTV**
  
  \[(1920 \times 1080) \times 1.5 \times 8 \times 60 = 1493 \text{ Mbits/s}\]
Y u v Formats

Y, u, v Image

PCM 8 bits

\[ Y = 0.30 \, R + 0.59 \, G + 0.11 \, B \]

\[ u = B - Y \]

\[ v = R - Y \]

Y = 0.30 R + 0.59 G + 0.11 B

Y, u, v Image

PCM 8 bits

Y, u, v Image

PCM 8 bits

Y = 0.30 R + 0.59 G + 0.11 B

u = B - Y

v = R - Y

RGB 24 bits/pixel

YUV 4:2:0 (12 bits/pixel)
Essential of Compression

- Remove redundant information
  - **Spatial** redundancy:
    - Neighboring samples have similar values
  - **Frequency** redundancy:
    - Some information can be discarded
  - **Temporal** redundancy
    - Neighboring frames in a video sequence are similar
Remove Temporal redundancy

Reference Frame

Current Frame

Motion Vector

Search Range

Ω
Motion Compensation Approach

- Motion Vectors
  - Inform decoder exactly where in the previous image to get the data
  - Motion vector would be zero for a static background
Remove Spatial and Frequency redundancy

- Remove spatial redundancy: DCT, Wavelet, Matching Pursuit …
- Discard smaller coefficients which are less significant
- Assign shorter codes to more probable symbols

Transform

Quantization

Entropy coding

Compressed bitstream

Motion Estimation

Prediction error

video frames
OUTLINE

- Introduction to Video Compression
- Overview of Video Coding Standards
- H.264 Video Coding Standard
- Next-generation video coding
Video Coding Standards

**ITU-T**
- **VCEG**
  - H.261 (1990)
  - H.263 (1995/96)
  - H.263+ (1997/98)

**ISO/IEC**
- **MPEG**
  - MPEG-1 (1993)
  - MPEG-2 (H.262) (1994/95)
  - MPEG-4 v1 (1998/99)
  - MPEG-4 v2 (1999/00)
  - MPEG-4 v3 (2001)

Timeline:
- 1990
- 1992
- 1994
- 1996
- 1998
- 2000
- 2002
- 2003

Source: MEDIATEK Advanced Technology Development Div.
## Multimedia Communications Standards and Applications

<table>
<thead>
<tr>
<th>Standards</th>
<th>Application</th>
<th>Video Format</th>
<th>Raw Data Rate</th>
<th>Compressed Data Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.320 (H.261)</td>
<td>Video conferencing over ISDN</td>
<td>CIF/QCIF</td>
<td>37 Mbps</td>
<td>&gt;=384 Kbps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9.1 Mbps</td>
<td>&gt;=64 Kbps</td>
</tr>
<tr>
<td>H.323 (H.263)</td>
<td>Video conferencing over Internet</td>
<td>4CIF/CIF/QCIF</td>
<td></td>
<td>&gt;=64 Kbps</td>
</tr>
<tr>
<td>H.324 (H.263)</td>
<td>Video over phone lines/ wireless</td>
<td>QCIF</td>
<td>9.1 Mbps</td>
<td>&gt;=18 Kbps</td>
</tr>
<tr>
<td>MPEG-1</td>
<td>Video distribution on CD/ WWW</td>
<td>CIF</td>
<td>30 Mbps</td>
<td>1.5Mbps</td>
</tr>
<tr>
<td>MPEG-2</td>
<td>Video distribution on DVD/digital TV</td>
<td>CCIR601 4:2:0</td>
<td>128 Mbps</td>
<td>3-10 Mbps</td>
</tr>
<tr>
<td>MPEG-4</td>
<td>Multimedia distribution over Inter/Intra net</td>
<td>QCIF/CIF</td>
<td></td>
<td>28-1024 Kbps</td>
</tr>
<tr>
<td>GA-HDTV</td>
<td>HDTV broadcasting</td>
<td>SMPTE296/295</td>
<td>&lt;=700 Mbps</td>
<td>18--45 Mbps</td>
</tr>
<tr>
<td>MPEG-7</td>
<td>Multimedia databases (content description and retrieval)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Overview of Video Coding Standards (1/2)

- **H.261**:  
  - First video coding standard, targeted for video conferencing over ISDN  
  - Uses block-based hybrid coding framework with integer-pel MC

- **H.263**:  
  - Improved quality at lower bit rate, to enable video conferencing/telephony  
  - below 54 kbps (modems or internet access, desktop conferencing)  
  - Half-pel MC and other improvement

- **MPEG-1**:  
  - Video on CD and video on the Internet (good quality at 1.5 mbps)  
  - Half-pel MC and bidirectional MC

- **MPEG-2**:  
  - TV/HDTV/DVD (4-15 mbps)  
  - Extended from MPEG-1, considering interlaced video
Overview of Video Coding Standards (2/2)

- **MPEG-4**
  - To enable object manipulation and scene composition at the decoder
  - Interactive TV/virtual reality
  - Object-based video coding: shape coding
  - Coding of synthetic video and audio: animation

- **MPEG-7**
  - To enable search and browsing of multimedia documents
  - Defines the syntax for describing the structural and conceptual content

- **H.264**
  - Improved coding efficiency (by having more options for optimization)
Basic Structure for Video Standard
MPEG-4 Overview

- Entire scene is decomposed into multiple objects
  - Object segmentation is the most difficult task!
  - But this does not need to be standardized 😊
- Each object is specified by its shape, motion, and texture (color)
  - Shape and texture both changes in time (specified by motion)
Mesh Animation

- An object can be described by an initial mesh and MVs of the nodes in the following frames
- MPEG-4 defines coding of mesh geometry, but not mesh generation
Body and Face Animation

- MPEG-4 defines a default 3-D body model through body definition table (BDP) and face definition table (FDP)
- The body and face can be animated using the body animation parameters (BAP) and face animation parameters (FAP)
OUTLINE

- Introduction to Video Compression
- Overview of Video Coding Standards
- H.264 Video Coding Standard
- Next-generation video coding
Overview of H.264

- H.264 can encode video with approximately 3 times fewer bits than comparable MPEG-2 encoders.
- H.264 is up to twice as efficient as MPEG-4 Part 2 (natural video) encoding.

you can see the difference of encoding via MPEG-2 and H.264 at 100 kpbs.
Performance of Video Standards

![Graph showing the performance of video standards over time.](image)
H.264 Features

- **High compression performance**
  Many advanced compression techniques are adopted.
  - Advanced Intra-Prediction
  - Strong Motion Isolation (4x4, ¼-pel resolution)
  - Multiple Reference Frames
  - Weighted Bi-Prediction
  - Context-adaptive VLC/BAC
  
  **Average bit rate reduction of 50%**

- **Exact match decoding**
  - Integer Transform

- **Improved Perceptual Quality**
  - In-Loop Deblocking Filter

- **Network friendliness**
  - NAL (Network abstraction layer)
  - Enhanced Error Resilience

Very High Complexity!
H.264: Motion Compensation Accuracy

- Macroblock Partitioning

<table>
<thead>
<tr>
<th>MB Types</th>
<th>16x16</th>
<th>16x8</th>
<th>8x16</th>
<th>8x8</th>
</tr>
</thead>
<tbody>
<tr>
<td>8x8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8x4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4x8</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4x4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Motion vector accuracy ¼ pixel

MPEG-2
- 16x16 block size
- Square shape
- ½ pel motion vector
  ➔ Weak Motion Isolation !

MPEG-4
- 8x8 block size
- Square shapes
- ½ pel motion vector
  ➔ Moderate Motion Isolation !!

H.264
- 4x4 block size
- Arbitrary shapes
- ¼ pel motion vector
  ➔ Strong Motion Isolation !!!
Multiple Reference Frames

- Support for multiple reference pictures. It gives significant compression when motion is periodic in nature.
½- and ¼-Pel Interpolation

- Samples at ½-pel positions (yellow) are interpolated using a 6-tap filter.
- Samples at ¼-pel positions (green) are the average of two neighboring ½-pel pixels.
Intra Prediction

Directional spatial prediction (H.264)
(9 types for 4x4 luma, 4 types for 16x16 luma, 4 types for chroma)

- e.g., Mode 4:
  - diagonal down/right prediction
  - a, f, k, p are predicted by
    \((A + 2Q + 1 + 2) \gg 2\)
4x4 Intra-Prediction

- 9 prediction modes
16x16 Intra-Prediction

- 4 prediction modes

0 (Vertical)

1 (Horizontal)

2 (DC)

3 (Plane)
Comparison to Other Standards

![Graph showing the comparison of quality (PSNR) to bit-rate for different standards: H.264, MPEG-4, H.263, and MPEG-2. The graph is labeled with 'Foreman QCIF (at 10 frames/s).']
OUTLINE

- Introduction to Video Compression
- Overview of Video Coding Standards
- H.264 Video Coding Standard
- Next-generation video coding
Is there a Moore’s law of compression?

Ref: Prof. Dr. Touradj Ebrahimi, VISNET-II Summer School
Improvement of Compression Ratio

- A roughly 50% gain in about every 5 years
**Next-generation video coding**

- **H.265** - Long-term video coding standard beyond H.264
- Under study for potential development by the ITU-T Video Coding Experts Group (VCEG)
- High coding efficiency (50% bit-rate reduction compared with H.264/AVC)
- Loss/error robustness

- **Key Technical Area (KTA)** – developed as the software platform based on JM11
High-performance Video Coding

- Call for Proposals on High-Performance Video Coding (HVC)
- Coordination between VCEG and MPEG
- Target: higher compression capability than the existing AVC standard
- Display resolutions: up to 1080p and Ultra HD
- Final Call for Proposals: 2010/01/22
- Final standard approval: 2012/07
New Technologies adopted in KTA (1/4)

- Adaptive Interpolation Filter (AIF)
  - 2-D non-separable AIF (AD08, AE16)
  - Separable AIF (COM16-C219, AG10)
  - Directional Interpolation Filter (DIF) (AG21, AG22, AH17, AH18)
  - Low Complexity Directional Interpolation Filter (AI12)
  - Enhanced AIF (EAIF) (C464, AI38, AJ30)
  - Switched Interpolation Filter with Offset (AI35, AJ29)
  - High Precision Interpolation (AI33)
  - Single Pass Encoding using Switched Interpolation Filters with Offset (AJ29)
  - Single-Pass Encoding Using Multiple Adaptive Interpolation Filters (AK26)
Filter support for E-AIF and E-DAIF

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>B2</td>
<td>B3</td>
<td>B4</td>
<td>B5</td>
<td>B6</td>
</tr>
<tr>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C4</td>
<td>C5</td>
<td>C6</td>
</tr>
<tr>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td>D5</td>
<td>D6</td>
</tr>
<tr>
<td>E1</td>
<td>E2</td>
<td>E3</td>
<td>E4</td>
<td>E5</td>
<td>E6</td>
</tr>
<tr>
<td>F1</td>
<td>F2</td>
<td>F3</td>
<td>F4</td>
<td>F5</td>
<td>F6</td>
</tr>
</tbody>
</table>
New Technologies adopted in KTA (2/4)

- Motion Compensation
  - Competition-Based MV prediction (AC06)
  - Motion compensated prediction with 1/8-pel (AD09)
  - Extended Block Sizes (COM16-C123)
New Technologies adopted in KTA (3/4)

- Transform and Quantization
  - Mode-dependent directional transform (MDDT) (AF15, AG11, AJ24, AI36)
  - Adaptive Quantization Matrix Selection (AQMS) (AC07, AD06, AF08, AI19)
  - Rate Distortion Optimized Quantization (RDO-Q) (AH21)
  - Prediction error coding in Spatial and Frequency domain (AB06, AD07, AE15)
Prediction error coding in Spatial and Frequency domain
New Technologies adopted in KTA (4/4)

- **In-loop Filter**
  - Block-based Adaptive Loop Filter (BALF) (AI18, AJ13)
  - Quadtree-based adaptive loop filter (COM16-C181, AK22)

- **Other**
  - Internal bit depth increase (IBDI) (AE13, AF07)
Block-based Filter On-Off approach
Thank You!
Yesterday is history
Tomorrow is a mystery
Today is a gift That is why it is called the present