EE 5359

Low Complexity H.264 encoder for mobile applications

Thejaswini Purushotham

Student I.D.: 1000-616 811

Date: February 18, 2010
Objective

The objective of the project is to implement a low-complexity encoder for mobile applications using machine learning algorithm.

Motivation

H.264 is currently one of the most widely accepted video coding standards in the industry. It enables high quality video at very low bitrates. H.264 is a block-oriented motion-compensation based video codec developed by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Pictures Expert Group (MPEG). H.264 has a highly complex encoder which leads to a good performance in terms of bit-rate. The high-computational complexity of H.264 and real-time requirements of video systems represent the main challenge to overcome the development of efficient encoder solutions. The computational complexity in H.264 comes from the motion estimation and mode decision techniques implemented in the encoder. Hence by reducing the computation complexity of the motion estimation and mode decision techniques, it will be possible to confide with the real-time requirements of the video systems.

Details

Overview of H.264:

H.264 [1] is a standard for video compression, and is equivalent to MPEG-4 Part 10, or MPEG-4 AVC (for advanced video coding) (Fig. 2 and 3). As of 2008, it is the latest block-oriented motion-compensation-based video standard developed by the ITU-T Video coding experts group (VCEG) together with the ISO/IEC moving picture experts group (MPEG), and it was the product of a partnership effort known as the joint video team (JVT). The ITU-T H.264 standard and the ISO/IEC MPEG-4 part 10 standard (formally, ISO/IEC 14496-10) are jointly maintained so that they have identical technical content.
The standardization of the first version of H.264/AVC was completed in May 2003. The JVT then developed extensions to the original standard that are known as the fidelity range extensions (FRExt).
These extensions enable higher quality video coding by supporting increased sample bit depth precision and higher-resolution color information, including sampling structures known as YUV 4:2:2 and YUV 4:4:4. Several other features are also included in the fidelity range extensions, such as adaptive switching between 4x4 and 8x8 integer transforms, encoder-specified perceptual-based quantization weighting matrices, efficient inter-picture lossless coding, and support of additional color spaces. The design work on the fidelity range extensions was completed in July 2004, and the drafting work on them was completed in September 2004.

Scalable video coding (SVC) [4] as specified in Annex G of H.264/AVC allows the construction of bitstreams that contain sub-bitstreams that conform to H.264/AVC. For temporal bitstream scalability, i.e., the presence of a sub-bitstream with a smaller temporal sampling rate than the bitstream, complete access units are removed from the bitstream when deriving the sub-bitstream. In this case, high-level syntax and inter prediction reference pictures in the bitstream are constructed accordingly. For spatial and quality bitstream scalabilities, i.e. the presence of a sub-bitstream with lower spatial resolution or quality than the bitstream, network abstraction layer (NAL) units are removed from the bitstream when deriving the sub-bitstream. In this case, inter-layer prediction, i.e., the prediction of the higher spatial resolution or quality signal by data of the lower spatial resolution or quality signal, is typically used for efficient coding. The scalable video coding extension was completed in November 2007[4].

Some of the features adopted in H.264 for enhancement of prediction, improved coding efficiency and robustness to data errors/losses are listed as follows.

Features for enhancement of prediction are as follows.

- Directional spatial prediction for intra coding

- Variable block-size motion compensation with small block size (Fig. 4)

![Fig. 4: Various block sizes in H.264 for motion estimation/compensation](image.png)
- Quarter-sample-accurate motion compensation
- Motion vectors over picture boundaries
- Multiple reference picture motion compensation
- Decoupling of referencing order from display order
- Decoupling of picture representation methods from picture referencing capability
- Weighted prediction
- Improved “skipped” and “direct” motion inference
- In-the-loop deblocking filtering

Features for improved coding efficiency are as follows.

- Small block-size transform
- Exact-match inverse transform (Fig. 5)

\[
T_{4x4} = \begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & 1 & -1 & -2 \\
1 & -1 & -1 & 1 \\
1 & -2 & 2 & -1
\end{bmatrix}
\]

\[
T_{8x8} = \begin{bmatrix}
8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \\
12 & 10 & 6 & 3 & -3 & -6 & -10 & -12 \\
8 & 4 & -4 & -8 & -8 & -4 & 4 & 8 \\
10 & -3 & -12 & -6 & 6 & 12 & 3 & -10 \\
8 & -8 & -8 & 8 & 8 & -8 & -8 & 8 \\
6 & -12 & 3 & 10 & -10 & -3 & 12 & -6 \\
4 & -8 & 8 & -4 & -4 & 8 & -8 & 4 \\
3 & -6 & 10 & -12 & 12 & -10 & 6 & -3
\end{bmatrix} \times \frac{1}{8}
\]

Fig. 5: Forward 4x4 and 8x8 integer transforms [3]

- Short word-length transform
- Hierarchical block transform
- Arithmetic entropy coding
- Context-adaptive entropy coding
Features for robustness to data errors/losses are as follows.

- Parameter set structure
- NAL unit syntax structure
- Flexible slice size
- Flexible macroblock ordering (FMO)
- Arbitrary slice ordering (ASO)
- Redundant slices (RS)
- Data partitioning
- SP/SI synchronization/switching pictures

Profiles in H.264

H.264 standard defines numerous profiles, as listed below.

- Constrained baseline profile
- Baseline profile
- Main profile
- Extended profile
- High profile
- High 10 profile
- High 4:2:2 profile
- High 4:4:4 predictive profile
- High stereo profile
- High 10 intra profile
- High 4:2:2 intra profile
- High 4:4:4 intra profile
- CAVLC 4:4:4 intra profile
- Scalable baseline profile
- Scalable high profile
- Scalable high intra profile

Table 1 and Table 2 outlines the features of the various profiles in H.264. Fig. 6 gives a graphical comparison of the profiles in H.264.
### Table 1: Features in baseline, main and extended profile [3]

<table>
<thead>
<tr>
<th>Coding Tools</th>
<th>Baseline</th>
<th>Main</th>
<th>Extended</th>
</tr>
</thead>
<tbody>
<tr>
<td>I and P Slices</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CAVLC</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CABAC</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B Slices</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Interlaced Coding (PicAFF, MBAFF)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Enh. Error Resil. (PMO, ASO, RS)</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Further Enh. Error Resil (DP)</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>SP and SI Slices</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

### Table 2: Features in high profile [3]

<table>
<thead>
<tr>
<th>Coding Tools</th>
<th>High</th>
<th>High 10</th>
<th>High 4:2:2</th>
<th>High 4:4:4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Profile Tools</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4:2:0 Chroma Format</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>8 Bit Sample Bit Depth</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>8x8 vs. 4x4 Transform Adaptivity</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Quantization Scaling Matrices</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Separate Cb and Cr QP control</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Monochrome video format</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>9 and 10 Bit Sample Bit Depth</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4:2:2 Chroma Format</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>11 and 12 Bit Sample Bit Depth</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4:4:4 Chroma Format</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Residual Color Transform</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Predictive Lossless Coding</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Machine learning usually refers to the changes in systems that perform tasks associated with the artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction etc. The “changes” might be either enhancements to already performing systems or ab initio synthesis of new systems. It is possible that hidden among large piles of data are important relationships and correlations. Machine learning methods can often be used to extract the relationships (data mining) [11]. The idea is to approximate a function or an unknown value using the statistics of the known data. Statistical methods for dealing with these problems can be considered instances of machine learning because the decision and the estimation rules depend on a corpus of samples drawn from the problem environment.

CONCLUSION

Fig. 6: Comparison of H.264 baseline, main, extended and high profiles [2]

Fig. 7: Building decision trees
Machine learning has been widely used in various image and video processing applications. Content based image and video retrieval (CBIR), content understanding, and more recently video mining are few of the applications where machine learning has been traditionally applied. Statistics like simple mean and variance are used for classifying the MBs. These seemingly simple metrics give very good performance in determining MB mode and prediction mode of MBs [4]. A hierarchy of decision trees based on the relative mean metrics is developed to compute MB modes quickly. These decision trees can be implemented as if-else statements to decide the MB modes. This is computationally less extensive compared to the FS algorithm traditionally used in the JM software for mode decision.

REFERENCES:


REFERENCE BOOKS:


REFERENCE WEBSITES:
ACRONYMS:

ASO          Arbitrary slice ordering
AVC          Advanced Video Coding
B MB         Bi-predicted MB
DCT          Discrete Cosine Transform
DSP          Digital Signal Processing
DVD          Digital Versatile Disc
FMO          Flexible macroblock ordering
FRExt        Fidelity Range Extensions
FS           Full Search
GOP          Group Of Pictures
I MB         Intra Predicted MB
IEC          International Electrotechnical Commission
ISO          International Organization for Standardization
ITU-T        International Telecommunication Union – Transmission sector
JVT          Joint Video Team
P MB         Inter Predicted MB
IDCT         Inverse Discrete Cosine Transform
IQ           Inverse Quantizer
MB           Macroblock
MBAFF        Macroblock level Adaptive Frame/Field
PicAFF       Picture level Adaptive Frame/Field
ME           Motion Estimation
MC           Motion Compensation
MV           Motion Vector
MPEG         Moving Picture Experts Group
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak –to – peak Signal to Noise Ratio</td>
</tr>
<tr>
<td>Q</td>
<td>Quantizer</td>
</tr>
<tr>
<td>R-D</td>
<td>Rate – Distortion</td>
</tr>
<tr>
<td>RS</td>
<td>Redundant slice</td>
</tr>
<tr>
<td>SP/SI</td>
<td>Switched P / Switched I</td>
</tr>
<tr>
<td>SMPTE</td>
<td>Society of Motion Picture and Television Engineers</td>
</tr>
<tr>
<td>SSIM</td>
<td>Structural Similarity Index Measure</td>
</tr>
<tr>
<td>SVC</td>
<td>Scalable Video Coding</td>
</tr>
<tr>
<td>VCEG</td>
<td>Video Coding Experts Group</td>
</tr>
<tr>
<td>VLC</td>
<td>Variable Length Coding</td>
</tr>
<tr>
<td>VLD</td>
<td>Variable Length Decoder</td>
</tr>
<tr>
<td>YUV</td>
<td>Y- Luminance and UV- Chrominance</td>
</tr>
</tbody>
</table>