Fast Hybrid Simulation for Accurate Decoded Video Quality Assessment on MPSoC Platforms with Resource Constraints

Deepak Gangadharan and Roger Zimmermann
Department of Computer Science, National University of Singapore
and
Samarjit Chakraborty
Institute for Real-Time Computer Systems, TU Munich
Motivation (Typical MPSoC architecture for video decoders)

- Mobile devices with video decoders have resource (frequency, buffer etc) constraints
- Resources can be determined based on the output quality requirement
- Necessity for an efficient framework to explore the trade-off between resources and display quality
Motivation (Conventional method - Design space exploration of resources for no loss in quality)

- System should satisfy the “no quality loss” constraint for all test video clips
- A model of the system running a decoder application is used in the system simulator
- All video clips simulated to obtain decoder task workload values
- High simulation time!!
Motivation (Reducing simulation time)

- Hybrid simulation based task workload estimation
- Simulation-based task workload estimation
- Model-based task workload estimation
Organization

- Related Work
- Overview of hybrid simulation-based quality assessment framework
- MPEG2/MPEG4 workload models
- Frame discard strategy
- Experimental results
## Related Work

- Very little work on studying the trade-offs between scarce system resources (frequency and buffer) and decoded video quality important for mobile devices.

- Yanhong et al. [1] investigate trade-offs between quality of MPEG-4 decoded video and processor frequency:
  - Use expensive simplescalar simulations to find task workload values.

- Koumaras et al. [2] propose an end-to-end video quality prediction framework with packet losses in a network transmission scenario:
  - Cannot be adapted to mobile devices because PSNR value estimation is not accurate.

---


Overview of our framework

- **Task Differentiation** + **Hybrid Simulation**
  - Training set clips for task differentiation and workload model derivation
  - Test clips for (simulation of light tasks only + workload model equipped bitstream analysis)

- **Workload model + Bitstream analysis**

- **MPSoC platform details** (number of processing elements)
- **System resource values** (Frequency + Buffer values)

- **Quality Assessment**
  - **PSNR calculator**
  - **quality_{out}**
  - **Frame discard strategy**
  - **(exec_{task1}, exec_{task2}, exec_{task3})**
Step 1 - Task Differentiation

- Tasks in a decoder application are differentiated based on a **lightness** property of the tasks (using training set video clips)

- For a task $t_k$, where $1 \leq k \leq N_T$, **lightness** property is determined by

  $$T_k \cdot F_{th} \leq T_{tot}$$

- Light tasks and other tasks which cannot be accurately modeled are simulated

- Accurate task workload models are derived for the remaining tasks
MPEG-2/MPEG-4 decoder task differentiation

MPEG-2 decoder ($F_{th} = 0.2$)

Light and Simulated
- VLD

Heavy and Modeled
- IDCT
- MC

MPEG-4 decoder ($F_{th} = 0.3$)

Light and Simulated
- VLD
- AC-DC prediction

Heavy and Modeled
- IDCT
- MC
Step 2 - Hybrid Simulation

- Workload of heavy tasks (IDCT+MC for both MPEG-2 and MPEG-4 decoder) from Step 1 are estimated by extracting certain easily obtainable parameters from the video stream.

- Accurate workload models are used for IDCT and MC (accuracy measured in terms of a frame drop deviation condition).

- Workload of light tasks from Step 1 are estimated by simulation as deriving accurate workload model is difficult.
Frame Drop Deviation (FDD)

\[
b_{\text{sim}}(t) = r_{f_1} + B_{\text{bot sim}} \quad \text{and} \quad b_{\text{model}}(t) = s_{f_2} + B_{\text{bot model}}, \quad t \\
\]

\( r_{f_1} \) – r no. of MBs of \( f_1 \)-th frame at the top of the buffer and 0 \% r \% FSIZE

\( s_{f_2} \) has the same interpretation as \( r_{f_1} \)

• According to FDD, everytime a buffer overflow occurs, \( \text{FDD} = 0 \Rightarrow |f_1 - f_2| = 0 \Rightarrow f_1 = f_2 \)
Workload model for heavy tasks in MPEG-2/MPEG-4

- Video clips from training set used to derive the workload models
- Workload models were developed for portable ISA (PISA).
- VLD task workload **accuracy** is difficult to achieve by using a model in both MPEG-2 and MPEG-4. AC-DC prediction workload accuracy is also difficult to attain in MPEG-4.
MPEG-2 MC workload model

• MC workload depends on 6 parameters of the MPEG-2 frame namely
  1) Y component’s x-dimension is HALF-PIXEL
  2) Y component’s y-dimension is HALF-PIXEL
  3 & 4) U or V component’s x and y-dimension is HALF-PIXEL
  5) forward or backward motion compensation is required
  6) motion compensation window is 16x8 or 16x16

• Depending on the frame type, the MC routine is called with different parameters

• Look-up table with 64 processor workload values is used to model the MC workload
  Workload_{MC} = LUT(x), where x is a 6-bit value
MPEG-2 IDCT workload model

- IDCT workload depends on the number and position of non-zero IDCT coefficients in the 8x8 block structure of MB.

- To adhere to FDD condition, certain shortcut conditions of fast IDCT implementations are also taken into consideration.

- The IDCT workload model proposed is:
  \[
  \text{Workload}_i = \text{basis}_i + \text{®} \cdot n_{\text{sig}}_i - \bar{\£} \cdot \text{scnt}_i, \quad ; i
  \]
  \[
  \text{Workload}_{IDCT} = \sum_i \text{Workload}_i, \quad 1 \cdot i \cdot 6
  \]

- From training set simulations:
  \[
  \text{basis}_i = 1965/1852/708/595
  \]
  \[
  \text{®} = 113 \text{ and } \bar{\£} = 143
  \]
Frame discard strategy

- Multimedia frames are required to be discarded when the resources are insufficient to process them.

- Frame discard strategy – Simple mechanism of dropping entire frame if one incoming MB cannot be accommodated in the buffer.

- Dropped frames are replaced by previous accepted frames.
Frame discard strategy (...contd)

- **Rule 1**: An I-frame drop is followed by drop of an entire group of pictures (GOP). No. of drops will be a minimum of GOP length

- **Rule 2**: A P-frame drop is followed by drop of the subsequent frames in the GOP. No. of drops will depend on the position of P-frame in the GOP

- **Rule 3**: A B-frame drop does not require any other frame drop

- **Rule 4**: A frame is dropped if any of its MBs causes overflow or it is dependent on a previously dropped frame
PSNR calculation

- A difference based PSNR (quality metric) is calculated as the original video frames are not available.

- Noise – Mean square error (MSE) between the actual pixel values of dropped frame and pixel values of last accepted frame in display order.

- PSNR value calculated using

\[
\text{psnr} = 10 \times \log_{10} \left( \frac{255 \times 255}{(\text{MSE}_r + \text{MSE}_g + \text{MSE}_b)} \right) 
\]

\[
(\text{MSE}_r)_n = \sum_{w=0}^{W-1} \sum_{h=0}^{H-1} \left( r_d(h, w, n) - r_c(h, w, n) \right)^2 
\]

\[
\text{MSE}_r = \sum_{n=0}^{N_{\text{drop}}-1} (\text{MSE}_r)_n
\]
Experimental Results
(Speed of Hybrid Simulation for MPEG-2)

Simulation Time (in mins)

100b_080  time_080  mult_080  susi_080  tens_080

~ 5 times faster

~ 8 times faster

VLD+MC+IDCT

Only VLD
Experimental Results
(Speed of Hybrid Simulation for MPEG-4)

~ 3 times faster (short length clips)

~ 3 times faster (long duration clips)

Simulation Time (in mins)

VLD + AC-DC + MC + IDCT
VLD + AC-DC

mobile  waterfall  paris  bridge
Experimental Results
(Quality Estimation Accuracy)

- PSNR values obtained using model-based workload values were verified using simulation-based workload values.

- PSNR values computed for 1500 resource combinations of \( f_{PE1}, f_{PE2}, B_1 \) and \( B_2 \).

- PSNR values found for the basemap as well as newmap task mapping configurations.

- basemap – 98% accuracy for motion video clips. Only § 0.3% deviation in PSNR values for the remaining 2%. 100% accuracy for the still videos.

- newmap – 100% accuracy in PSNR estimation for all clips.
Experimental Results
(PSNR vs $f_{PE1}$ trade-off)

tens_080

![Graph showing PSNR vs PE1 frequency for tens_080.

v700_080

![Graph showing PSNR vs PE1 frequency for v700_080.]

PSNR (in dB) vs PE1 frequency ($f_{PE1}$ in MHz)
Experimental Results
(PSNR vs $f_{PE2}$ trade-off)

PSNR (in dB)

PE2 frequency ($f_{PE2}$ in MHz)

basemap
newmap

v700_080

tens_080
Experimental Results
(PSNR vs $B_1$ trade-off)

**tens_080**

- **basemap**
- **newmap**

**v700_080**

- **basemap**
- **newmap**

Buffer 1 Size ($B_1$ in MBs)
Experimental Results
(PSNR vs $B_2$ trade-off)

PSNR (in dB)

Buffer 2 Size ($B_2$ in MBs)
Concluding Remarks

• Central focus:
  To devise a hybrid simulation strategy in order to enable the system designers to rapidly arrive at quantitative estimates of quality degradations for video clips in the context of resource constraints

• Framework also very useful to understand the non-trivial influences of system resources on the quantitative quality degradations

• Can be used to quickly estimate optimal points in resource space for desired quality

• Influence of other components in the framework such as frame discard strategy on the quality estimate can be studied rapidly