A Software-based High-Quality MPEG-2 Encoder Employing Scene Change Detection and Adaptive Quantization

Dirk Farin[1] & Niels Mache[2] & Peter H.N. de With[3]
[1]Univ. Mannheim, Germany, [2]Univ. Stuttgart, Germany,
[3]CMG/Univ. Technol. Eindhoven, Netherlands

Abstract

This paper presents a software-only MPEG encoder implementation which uses adaptive quantization and scene-change detection to enhance the picture quality. Scene-change detection is also used to assign picture coding types allowing easy, lossless cutting at scene-change positions in a later editing process. Although our current implementation concentrates on MPEG-2, all concepts can also be directly applied to MPEG-4 encoders.

Introduction

In the field of MPEG encoding, real-time hardware encoders have become common. However, when real-time processing is not a prerequisite, like in the field of DVD authoring, more flexible software encoders offer the opportunity to increase the picture quality by spending more time and memory on sophisticated image content analysis prior to final encoding.

The combination of quantization and rate-control is an important design consideration in MPEG encoders. As both operate by modifying the macroblock-quantizers step-size, the design of these blocks strongly depends on each other. The selection of optimal quantizers and picture coding-types depending on the actual scene contents is a difficult problem, and dynamic-programming techniques have been proposed to obtain an optimal solution [4]. However, these techniques are far too computationally complex for practical coding applications. Furthermore, techniques based on rate-distortion-characteristics require an objective measure for image quality. However, as we are interested in optimizing the subjectively perceived image quality for which no exact measure is available, alternative techniques should be applied.

Adaptive Quantization and Rate-Control

Adaptive quantization is applied at the macroblock level to reduce the amount of quantization noise in areas where it is most visible to the Human Visual System (HVS). The additional bits which are needed to provide the increased accuracy are obtained by reducing image quality in areas with fine, high-contrast texture (high-activity areas). The HVS is less sensitive to additional noise in these areas and cannot perceive the quality reduction.

Our implementation exploits three properties of the HVS: luminance masking, temporal masking, and activity masking. Luminance masking uses the effect that the human eye adapts to the mean image luminance and thus at the same time is less sensitive to noise at the ends of its dynamic range. The temporal masking effect allows to reduce the image quality after a sudden change of image contents. The HVS tends to integrate pictures over time and needs adaptation time for the situation of a sudden change. We will discuss quantization in presence of the temporal masking in the section *scene change detection* below. Finally, activity masking refers to the effect that noise is less noticeable in areas with high-frequency texture. We observed that MPEG artifacts are particularly visible in blocks that contain both an area with fine texture and an area with low activity. The quantization of the high-frequency coefficients of the fine texture area causes the MPEG-typical ringing effect in the low activity area adjacent to the textured area. It is desired to use more bits to code the blocks which contain both textured and flat areas. Given an image $f(x, y)$, we begin by partitioning each macroblock into 4×4 sub-blocks of size 4×4 pixels. For each of these sub-blocks, the sub-block activity *subact* is calculated as:

$$ subact_{x,y} = \sum_{\substack{0 \leq i \leq 3 \\ 0 \leq j \leq 3}} \left| \frac{\partial f}{\partial x}(x+i, y+j) \right| + \left| \frac{\partial f}{\partial y}(x+i, y+j) \right|. $$

Based on these values, we calculate the overall macroblock busyness (bsy) as the sum of all sub-block activities in the macroblock:

$$ bsy_{x,y} = \sum_{\substack{0 \leq i \leq 3 \\ 0 \leq j \leq 3}} subact_{x+4i, y+4j}. $$

Larger values of bsy correspond to more high-frequency texture. Additionally, we calculate a measure for the risk of ringing (rng) by summing all absolute differences of neighbouring sub-block activities in all 8×8 pixel-blocks in the macroblock (see Figure 1). Adaptive quantization is carried out by adding $qnoise = \alpha \cdot bsy - \beta \cdot rng$ to the MQUANT value, where α and β are appropriate, empirically determined constants.
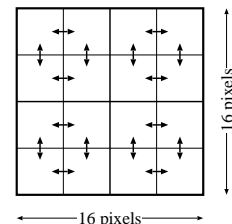

Fig. 1. Calculation of the risk of ringing (rng) in a macroblock. The absolute difference of sub-block activities in the sub-blocks indicated by double-arrows are added.

As the overhead for switching quantization values can be significant for low bit-rates, a post-processing stage applies a modified horizontal median filter on the quantization values

to remove small variations. The MQUANT value is replaced by the median value if the difference between both is less than a threshold value. Moreover, scenes may occur without any high-frequency textures, where the quality could be decreased to reserve bits for the adaptive quantization process and additional quality improvements. If this is detected, adaptive quantization is switched off and the whole image is quantized with a constant quantization value.
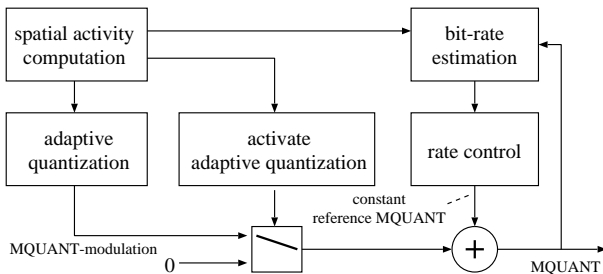

Fig. 2. Adaptive quantization and rate-control flow-graph.

For rate-control, the quantization is globally modified on a full picture basis by adding a constant reference value of MQUANT. This approach guarantees an equally distributed quality in the whole image. The reference quantization value is determined by a binary search through the valid range of MQUANTs until the estimated number of bits for the frame matches the predetermined frame-size. The estimation is based on a parametric model of estimating the number of bits per macroblock b depending on $qnoise$:

$$b = \left(\frac{qnoise}{\gamma}\right)^{\lambda} + \epsilon.$$

The parameters γ, λ, and ϵ are determined for each value of MQUANT by a number of training-sequences and adapted during the coding process. This estimation is simple to compute and because the estimation is based on the computations of the adaptive quantization stage, the computational complexity is low.
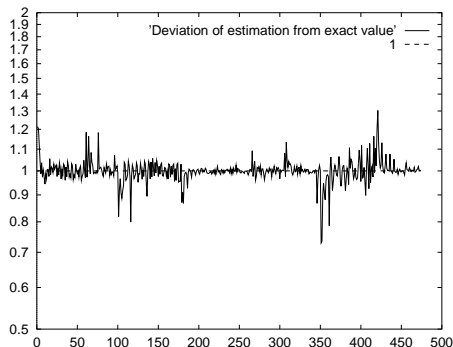

Fig. 3. Rate-control performance. Deviation from estimated bit-rate versus frame number. The deviation is calculated as the ratio of the estimated number of bits to the actual number of bits per frame.

Figure 3 depicts the deviation of the estimated frame-size from the correct frame-size after coding a sample sequence. In general, the estimation matches closely the actual number of bits. After a scene-change, the estimation adapts to the new image statistics after a small number of frames.

SCENE-CHANGE DETECTION

Adjusting the structure of GOPs to the image contents can enhance picture quality by ensuring good reference pictures for motion estimation. By providing excellently positioned points for cutting, editing in the compressed domain without the need for recompression is enabled. Our implementation uses a computationally efficient way of detecting scene-changes, which is based solely on detecting sudden changes in mean picture brightness. The control algorithm further favours equally sized GOPs and avoids to generate very long or short GOPs.
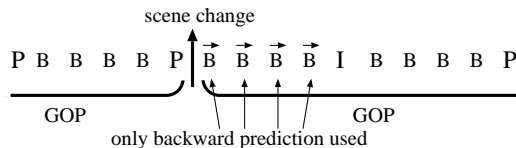

Fig. 4. Optimal structure of the GOPs at a scene-change.

If a scene-change is detected, the GOP structure is adapted to the structure shown in Figure 4. The number of bits to code the initial I-frame is increased, which also enhances the quality of all other frames in the GOP, as the I-frame is the initial reference image. The additional bits are obtained by a strong reduction of the quality of the first B-frames after the scene-change. As the temporal masking-effect of the HVS prevents that low image-quality can be perceived for approximately 100ms after the scene-change, the B-frames can effectively be coded with backward motion-vectors only.

RESULTS

Due to the absence of appropriate metrics for quality perception, objective results for adaptive quantization cannot be given. However, experiments have shown that our system gives a well visible improvement of the perceived image quality, compared to standard coders. In the case of disabled adaptive quantization, the constant quantization value ensures a homogeneous image quality.

To measure the effects of scene-change detection, we disabled rate-control to create VBR streams. With enabled scene-change detection and increased MQUANT around the scene-change, a bit-rate reduction of about 7% was observed (with a scene-change every two seconds). Figure 3 shows that our rate-control mechanism works well and can adapt itself to fastly changing scene contents. We currently work on extending the encoder to include MPEG-4 encoding capabilities.

REFERENCES

[1] D. Farin, N. Mache, and P. H. N. de With. SAMPEG, a scene adaptive, parallel MPEG-2 software encoder. *SPIE Visual Communications and Image Processing*, January 2001.

[2] Mark R. Pickering and John F. Arnold. A perceptually efficient VBR rate control algorithm. *IEEE Transactions on Image Processing*, 3(5):527–531, September 1994.

[3] A. Puri and R. Aravind. Adaptive perceptual quantization for video compression. *SPIE Visual Communications and Image Processing*, 1605:297–300, 1992.

[4] Kannan Ramchandran, Antonio Ortega, and Martion Vetterli. Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders. *IEEE Transactions on Image Processing*, 3(5):533–545, September 1994.