# CONTENT-ADAPTIVE ENCODER OPTIMIZATION OF THE H.264/AVC DEBLOCKING FILTER FOR VISUAL QUALITY IMPROVEMENT

*Konstantin Hanke, Peter Hosten, Fabian Jäger*

Institute of Communications Engineering
RWTH Aachen University
52056 Aachen, Germany
{hanke, hosten, jaeger}@ient.rwth-aachen.de

## ABSTRACT

This paper presents a new *self-adapting and content-sensitive optimization* technique for the H.264/AVC in-loop deblocking filter [1], focussing on the *visual* enhancement of the perceived reconstruction quality. Performed frame-wisely at the encoder side, the proposed algorithm first identifies visually important image regions in the currently *decoded and blocky* frame, including natural *and* artificial edge areas, and then optimizes the filter inherent threshold decisions, finding the best trade-off between preserving natural and suppressing artificial edges. As fast optimization criterion, the low complex Edge-PSNR [2] is employed, which has proved a very good congruence to the human visual quality sensation, much better than standard PSNR. In this work, the optimization behavior of all thresholds is analyzed, showing their mostly *good-natured curve shape* for convex optimization, but high *content dependency* of the optimal values. The in-loop coding results for AVC evidence the approach's high capability for visual improvement, whose general design is easily portable to other AVC deblocking based architectures like HEVC.

***Index Terms*—** Advanced video coding (AVC), in-loop deblocking filter, artifact removal, encoder optimization

## 1. INTRODUCTION

One of the main issues of modern video coding is the continuous enhancement of coding efficiency, i.e. the improvement of visual quality at a given bitrate level. Regarding the common activities in standardization of "High Efficiency Video Coding" (HEVC) and the responses to the Call for Proposals [3], trends for a major re-design of already implemented tools are identifiable, as also the case for the deblocking filter, whose current implementation [4] is still very similar to the original AVC design.

In this paper, the AVC deblocking filter [1] and its $\alpha$, $\beta$ activation and $c_0$ clipping thresholds are re-investigated and new capabilities for further quality improvements are identified, which conceptually can be easily transferred to current HEVC. Based on our analysis of the optimization behavior of all thresholds, we found out that the usage of the filter inherent threshold tables is often not the best coding alternative, leading to a noticeable over- or under-deblocking, especially for sequences with complex motion. Based on this observation, we show that the optimal deblocking thresholds do not only depend on the respective compression ratio, but also highly vary with the given video content and its motion complexity, making a side transmission of encoder optimized thresholds suitable and advantageous. Although AVC already has a limited ability of a frame-wise threshold adaptation via sending two offset parameters for $\alpha/c_0$ and $\beta$ within the frame header, the objective of this work is also to enlighten how much the visual quality can be improved by threshold optimization, which implicitly also gives an indication for an optimal choice of these offsets.

In this contribution, we introduce a new frame-wise threshold optimization tool that efficiently adapts all thresholds according to the optimal visual reconstruction quality. The optimization is made on basis of a simplified version of the Edge-PSNR [2], first identifying edge areas in the currently *decoded* frame via standard Canny edge detection [5], then using the found edge areas as region of support for quality measurement. We chose the Edge-PSNR due to three reasons: First, it is *simple*, helping us to keep down complexity during the fairly long-term optimization process, second, it shows a *very good correlation* to the human quality sensation, which has proved to be very sensitive to quality degradation on edges, and third, it *fades out non-edge areas* and helps the optimization process to concentrate on *visually important* image regions. Regarding the last aspect, we test different approaches of *masking* non-edge areas and compare them concerning their error separation and visual improvement capabilities. Last but not least, we present coding results of the proposed optimized deblocking for the AVC in-loop application, using the non-scalable AVC-compliant base-layer embedded in the JSVM-3 [6] software package, evidencing a high visual improvement capability with blocking-free visually optimized reconstructions.

The paper is organized as follows: Section 2 briefly reviews the AVC deblocking filter and its threshold functions. Section 3 presents the new proposed threshold optimization technique and compares different masks for edge detection. Subsequently, Section 4 discusses the optimization behavior of all thresholds and outlines the simulation results, while Section 5 gives final remarks.

## 2. DEBLOCKING FILTER DESIGN IN AVC

In order to reduce blocking artifacts in the coding loop, AVC defines an adaptive three-stage deblocking filter [1]. By smoothing all blocky edges around macroblock and subblock boundaries without affecting the image sharpness, it significantly improves the visual quality of the decoded video and, due to its in-loop design, may also positively affect the coding efficiency, offering a better prediction for motion compensation that is closer to the original.

The deblocking filter is adaptive on three stages: On edge, on pixel, and on frame (slice) level. On *edge level*, the filter assigns one *boundary strength* parameter (BS) to each $4 \times 4$ luminance transform block within a macroblock. All vertical and horizontal subblock edges are filtered using BS values from 0 to 4 controlling the employed deblocking filter strength. The underlying BS conditions are hierarchically tested starting with $\mathrm{BS} = 4$ and $\mathrm{BS} = 3$ indicating the strongest filter modes applied only on intra-coded macroblock

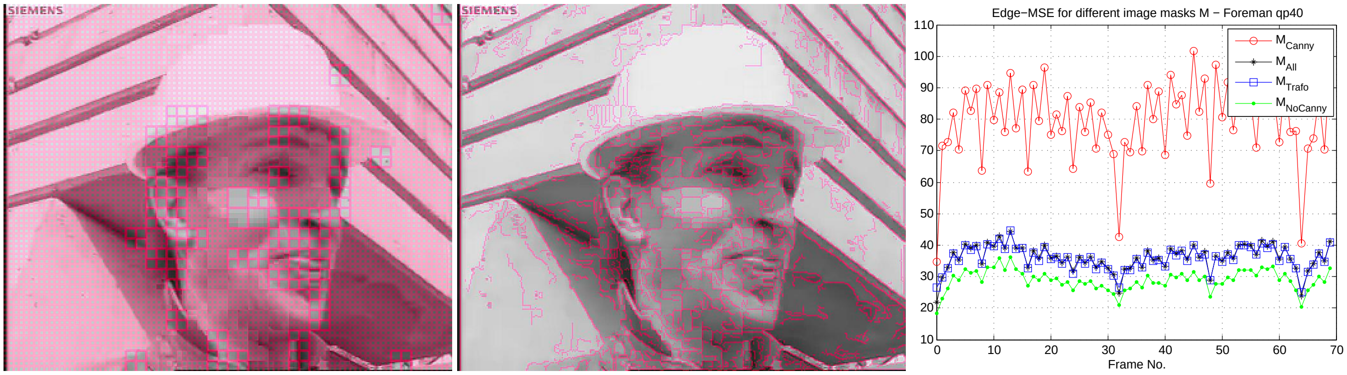


**Fig. 1**. Examples for edge mask $M_{\text{Trafo}}$ (left) and $M_{\text{Canny}}$ (middle), and performance comparison of all masks (right) for *Foreman* QP 40.

and subblock boundaries, respectively. The lower strengths are assigned to blocks having coded residuals (BS = 2) and blocks that are predicted using different reference frames or higher motion vector differences (BS = 1). In all other cases, via BS = 0, the deblocking filter is fully deactivated and no further analysis on sample level is performed. Unmodified, this partitioning is still applied in HEVC [4].

On *pixel level*, two *activation thresholds* $\alpha$ and $\beta$ are defined that make a sample-wise decision on whether to perform filtering across the current boundary or not. If we denote the current regarded samples of one line across the edge inside two adjacent $4 \times 4$ blocks by $p_3, p_2, p_1, p_0 \mid q_0, q_1, q_2, q_3$, the pixels $p_0$ and $q_0$ being directly located near the edge are only filtered if

$$|p_0 - q_0| \quad < \quad \alpha\big(\text{Index}_{\alpha,c_0}\big), \tag{1}$$

while samples within blocks get filtered if one condition of

$$|p_1 - p_0| \quad < \quad \beta\big(\text{Index}_{\beta}\big), \tag{2}$$
$$|q_1 - q_0| \quad < \quad \beta\big(\text{Index}_{\beta}\big), \tag{3}$$

is fulfilled. Samples $p_2$ and $q_2$ are only deblocked for BS = 4.

A third threshold, the *clipping threshold* $c_0$, controls the filter's smoothing effect by limiting the adjustment $\Delta$ of the filtered samples to a maximum possible value of

$$|\Delta| \quad \leq \quad c_0\big(\text{Index}_{\alpha,c_0}, \text{BS}\big). \tag{4}$$

It uses the same table index as $\alpha$ but, additionally, depends on the boundary strength, i.e. it increases towards higher BS [1]. After

$$\text{Index}_{\alpha,c_0} \quad = \quad \text{clip}\,\big(\text{QP} + \text{Offset}_{\alpha,c_0}, 0, 51\big), \tag{5}$$
$$\text{Index}_{\beta} \quad = \quad \text{clip}\,\big(\text{QP} + \text{Offset}_{\beta}, 0, 51\big), \tag{6}$$

all three thresholds further depend on the *quantization parameter* (QP), locally derived as average QP of the two edge adjacent blocks, and increase with higher QP / lower bitrates, amplifying the deblocking effect. Like the QP, the indices (5) and (6) are limited to integers from 0 to 51. In HEVC, the activation rules (1) – (3) were changed, removing the $\alpha$ threshold by coupling with $c_0$, jointly called $t_c$ [4].

On the highest *frame level*, the deblocking filter strength can be globally controlled by $\text{Offset}_{\alpha,c_0}$ and $\text{Offset}_{\beta}$, which have to be transmitted as side information in the slice header. Negative values mean less filtering, positive values increase the smoothing effect. In AVC, even these frame-wise adaptive parameters are intended for a global filter adjustment to actually given video characteristics or scene complexity. Our following approach also relies on this frame-wise adjustment, but optimizes all threshold values on their own, non regarding fixed threshold tables or offset constraints.

## 3. THRESHOLD OPTIMIZATION ALGORITHM

For optimization and analysis of the threshold's behavior, at the encoder side a high number of possible threshold triples $(\alpha_i, \beta_i, c_{0i})$ is tested. For this task, we define the *frame-wise objective function*

$$\mathcal{F}(\alpha_i, \beta_i, c_{0i}) = \mathcal{Q}_\alpha\,(\alpha_i) + \mathcal{Q}_\beta\,(\beta_i) + \mathcal{Q}_{c_0}\,(c_{0i}), \tag{7}$$

on which basis the optimization is performed. Herein,

$$\mathcal{Q}_\theta\,(\theta_i) = \mathcal{S}\Big(X', O, M\Big) \quad \text{with} \quad X' := \mathcal{D}\big(X, \theta_i\big) \tag{8}$$

denotes an *abstract quality function* of threshold $\theta \in \{\alpha, \beta, c_0\}$, that employs a *pixel-wise quality measure* $\mathcal{S}(\cdot)$. $\mathcal{S}$ evaluates the quality of the the *deblocked frame* $X'$ compared to the *original frame* $O$, where $X$ is built by standard *AVC deblocking* $\mathcal{D}(\cdot, \theta_i)$ of the currently *decoded blocky frame* $X$ using a threshold value $\theta_i$ instead of the standard AVC value $\theta_{\text{avc}}$. In $\mathcal{S}$, only those pixel locations $(m, n)$ are considered that belong to the given region of support, specified by non-zero entries in the *image mask* $M(m, n) > 0$. Thus, pixels without mark, i.e. $M(m, n) = 0$, have no influence on $\mathcal{S}$.

Following [2], we can now employ the *(edge) mask-based*

$$\textit{Edge-PSNR} = 10\log_{10}\left(\frac{P^2}{\textit{Edge-MSE}}\right) = 10\log_{10}\left(\frac{K_M \cdot P^2}{\textit{Edge-SSE}}\right) \tag{9}$$

as concrete PSNR-based implementation for $\mathcal{S}$ in (8), where $P = 255$ for 8-bit sequences, *Edge-MSE* is the *mean square error* relative to the total number of pixels $K_M$ that are positively marked in $M$, and

$$\textit{Edge-SSE}\,\left(X', O, M\right) \quad = \sum_{\forall M(m,n)>0} \left(X'(m,n) - O\,(m,n)\right)^2 \tag{10}$$

is the corresponding *sum of squared errors* between the $X'$ and $O$, again only applied on positively marked pixels. From (9) and (10), obviously all edge mask-based measures become identical to the respective standard measures, if the mask frame $M$ is completely marked. Since $M$ and thus $K_M$ only determined once per frame, from the optimization point of view, minimizing the *Edge-SSE* is equal to maximizing the *Edge-PSNR*, but has the big advantage of never reaching infinite values. Therefore, for analysis, we employ the *Edge-SSE* to simplify and stabilize the optimization process.

As important aspect of this work, we intensely analyze how different masks $M$ affect the optimization results and, thus, the visual quality. Among many others tested, we here compare the three masks

$M_{\text{All}}$ Mask marked for *all pixel coordinates* $(m, n)$.
$M_{\text{Trafo}}$ Mask marked for *all direct edge pixels* $(p_0, q_0)$.
$M_{\text{Canny}}$ Mask marked for *edge pixels that are detected by the Canny edge detector* [5] in the blocky frame $X$.

While $M_{\text{All}}$ represents the standard full-image PSNR/SSE quality evaluation, $M_{\text{Trafo}}$ considers only pixels that are directly located at the transform boundaries (see Section 2), exemplarily shown in Fig. 1 (left) for varying block sizes from 4×4 to 16×16. This mask is chosen due to its obvious analogy to the $\alpha$ threshold (1), which only regards samples $p_0, q_0$ for the deblocking decision. For $M_{\text{Canny}}$, a Canny edge detection [5] is performed on the *blocky* frame $X$ applying a low and high edge detection threshold of 2 and 32, respectively, and a filter size of 3x3 pixels, which has shown best visual results. For these detection settings, not only natural edges, but also visually annoying blocking artifacts are marked in $M_{\text{Canny}}$, especially visible in the image center of Fig. 1 (middle). This feature is desired here, since both edge kinds should be optimized, finding the best visual trade-off of smoothing artifacts and preserving natural edges.

Fig. 1 (right) compares all masks regarding their properties to detect (edge) areas with a high error energy in $X$, since these areas are visually most important for optimization. As observable, $M_{\text{Canny}}$ achieves the best performance in separating regions with high MSE from the remaining non-detected pixels ($M_{\text{NoCanny}}$), while the separation of $M_{\text{Trafo}}$ is negligible small and very similar to $M_{\text{All}}$, giving us a first hint for the later shown very good visual quality of $M_{\text{Canny}}$.

## 4. THRESHOLD OPTIMIZATION ANALYSIS AND SIMULATION RESULTS

Our tests are performed within the framework of JSVM-3 [6], only using the non-scalable AVC compliant base-layer in standard high profile settings (activated FRExt etc.) For optimization, standard AVC deblocking is performed with optimized threshold parameters, that are transmitted as entropy-coded side information, but only if they achieve a performance gain compared to standard thresholds. For rate-distortion analysis, this bitrate increment is taken into account, but only changes the overall bitrate in a very small manner. The performance is tested on the four CIF sequences *Bus*, *Football*, *Foreman*, and *Mobile* for QP values of 28, 32, 36, 40, 44 and 48.

Concerning important optimization issues, we first regard a single threshold optimization and clarify how different image masks influence the optimization process. Fig. 2 (a) to (c) depict typical curves of the *Edge-SSE*-based $\alpha$ and $c_0$ minimization, respectively. It is conspicuous that despite of the non-linearity of the deblocking filter operations a *good-natured curve shape* can be observed, allowing for the application of fast and low-complex convex optimization techniques. In our tests, however, we still employ full-search optimization, testing all suitable threshold triples in (7), i.e. value ranges of 0-255 for $\alpha$ and $\beta$, and 0-50 for $c_0$, since optima $c_0 > 50$ could never be observed. Comparing Fig. 2 (a) and (b), in extrem cases, the standard SSE optimization using $M_{\text{All}}$ in (a) fails, since $\alpha_{opt} = 0$ fully deactivates the deblocking filter for the current frame, producing a frame with highest PSNR, but with plenty of blocking artifacts. The same applies to $M_{\text{Trafo}}$. The use of the $M_{\text{Canny}}$ mask in (b), however, i.e. the concentration on high MSE edge regions, obviously *stabilizes* the optimization process, which can now find a suitable, *visually pleasing* threshold, implicitly verifying the observations in [2].

Fig. 2 (d) and (e) depict typical distributions of found $\alpha$ and $\beta$ optima over all frames having *one and the same* macroblock QP. For one given QP, the standard AVC filter only assigns *one fixed* threshold value. As shown, the optima are highly spread, particularly for $\alpha$. The same holds for $c_0$ in all BS variations. The observed high content dependency of $\alpha$, $\beta$ and $c_0$ indicates the high potential of threshold customization, suggesting that an optima transmission to the decoder, directly or via offsets, may be advantageous for the overall coding performance, especially for motion complex scenes in *Football* and *Foreman*, where the *Edge-SSE* is mostly improved.

| BD-Rate (%) | *Bus* | *Football* | *Foreman* | *Mobile* |
|---|---|---|---|---|
| qp48-36 | -0.1711 | -0.8095 | -0.6432 | -0.2901 |
| qp44-32 | -0.3321 | -0.6229 | -0.4977 | -0.3595 |
| qp40-28 | -0.3849 | -0.7282 | -0.3279 | -0.2077 |

**Table 1**. Bjøntegaard delta [7] rate savings of the proposed optimized deblocking over standard AVC for different QP ranges.

For these complex scenes, Fig. 3 presents visual results of in-loop coding tests for QP 40, depicting standard AVC results on the left side, and results for $M_{\text{All}}$ standard SSE minimization in the middle. As derivable, the optimization using $M_{\text{All}}$ is already better than standard AVC. The best visual performance, however, is achieved by optimizing $\alpha$, $\beta$ and $c_0$ based on the Canny edge detection mask $M_{\text{Canny}}$. This results holds for all sequences, but the best visual gain is observed for *Football* and *Foreman*, as already mentioned.

Regarded from a bitrate saving's point of view, however, the overall performance gain of our approach over AVC is very small, up to a maximum of $0.8\%$ as stated in Tab. 1, measured on basis of Bjøntegaard's Delta evaluation method [7]. This marginal rate gain relies on the fact, that all appearing savings from enhanced prediction are mostly compensated by the additional threshold transmission rate. However, the actual *visual* quality gain, as shown in Fig. 3, cannot be sufficiently reflected by this nominal rate decrement.

## 5. CONCLUSION AND OUTLOOK

In this work, we have investigated new capabilities of visual quality improvement by optimizing the deblocking filter thresholds at the encoder side. The main outcomes of our work are: All thresholds 1.) show an almost good-natured curve shape, and 2.) a high content dependency in their optima. 3.) Excluding low-distorted non-edge regions from quality measurement employing the Canny-based Edge-PSNR [2] significantly aids the optimization process in finding a visually satisfying optimum. The next step in further enhancing our approach is a precise prediction of the required side information bitrate in order to employ a frame-wise rate-distortion optimized mode switching, achieving an even more improved visual quality.

## 6. REFERENCES

[1] P. List, A. Joch, J. Lainema, G. Bjøntegaard, and M. Karczewicz, "Adaptive deblocking filter," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.13, no.7, pp. 614–619, July2003.

[2] C. Lee, S. Cho, J. Choe, T. Jeong, W. Ahn, and E. Lee, "Objective video quality assessment," *SPIE Optical Engineering*, vol. 45, no. 1, 017004, Jan. 2006.

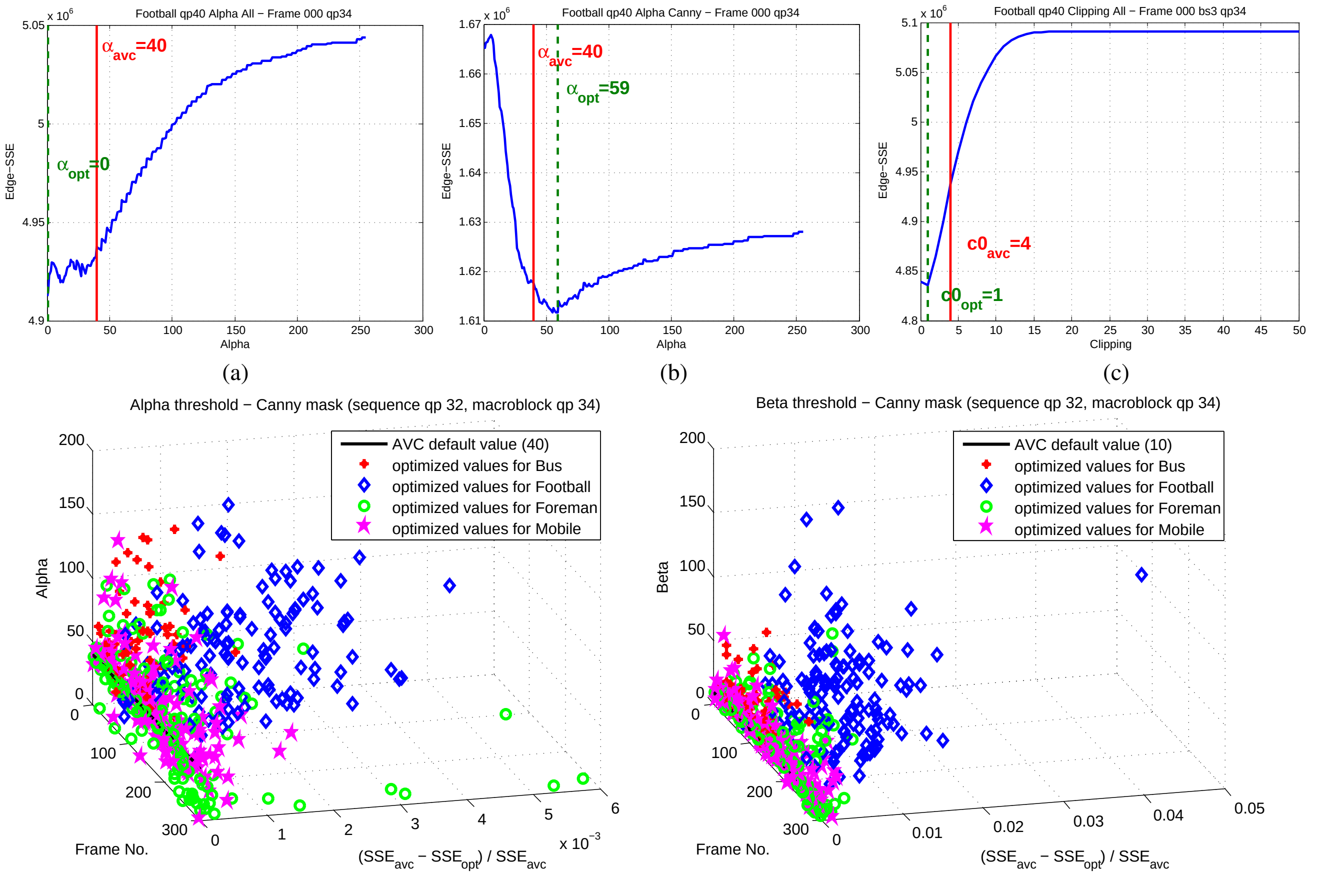
[3] "Table of Proposal Design Elements for High Efficiency Video Coding (HEVC)," Doc. JCTVC-A203, Joint Collaborative Team on Video Coding (JCT-VC), 1st Meeting, Dresden, Apr. 2010.

[4] T. Wiegand, W.-J. Han, B. Bross, J.-R. Ohm, and G. J. Sullivan, "Working Draft 1 of High-Efficiency Video Coding (HEVC)," Doc. JCTVC-C403, Joint Collaborative Team on Video Coding (JCT-VC), 3rd Meeting, Guangzhou, China, Oct. 2010.

[5] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[6] J. Reichel, H. Schwarz, and M. Wien, "Joint scalable video model JSVM-3," Doc. JVT-P202, Joint Video Team (JVT), 16th Meeting, Poznan, Poland, July 2005.

[7] G. Bjøntegaard, "Calculation of average PSNR differences between RD curves," Doc. VCEG-M33, ITU-T VCEG, 13th Meeting, Austin, TX, USA, Apr. 2001.



(a) (b) (c)

(d) (e)

**Fig. 2**. Threshold comparison: Single $\alpha$ optimization 0-255 using (a) $M_{\text{All}}$ mask resulting in $\alpha_{opt} = 0$ and (b) $M_{\text{Canny}}$ mask with suitable and visually much better $\alpha_{opt} = 59$. (c) Typical example for $c_0$ optimization, here BS=3 and suitable value range 0-50. (d)+(e) Typical plot for high content-dependent optima fluctuations (d) for $\alpha$ and (e) smaller for $\beta$. Particularly for *Football* and *Foreman*, the Edge-SSE is significantly improved compared to standard AVC (here, the SSE-gain is plotted in relation to AVC and only *one* macroblock QP is regarded).

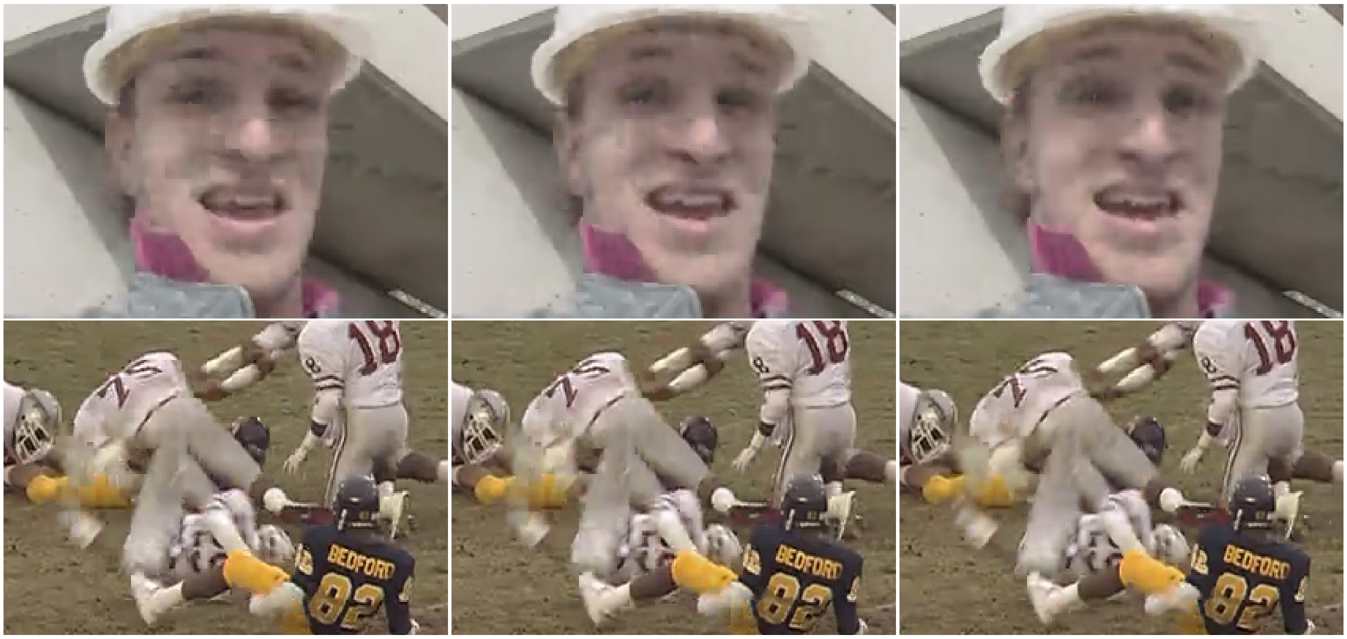


**Fig. 3**. Visual quality comparison of standard AVC encoding (left), $(\alpha, \beta, c_0)$ optimization applying $M_{\text{All}}$ (middle), and $(\alpha, \beta, c_0)$ optimization applying $M_{\text{Canny}}$ (right). All results are given for *Foreman* QP40, frame 141, and *Football*, frame 227, where details were enlarged.