QUALITY CONTROLLED ARTIFACT REDUCTION

Martin Richter, Mathias Stöber, and Hartmut Schröder

Circuits and Systems Lab, TU Dortmund University
Otto-Hahn-Str. 4, 44221 Dortmund, Germany
{martin.richter, mathias.stoeber, hartmut.schroeder}@tu-dortmund.de

ABSTRACT

Strongly varying input content to modern flat panel displays makes highly adaptive algorithms mandatory. On the one hand, low quality content is dominated by coding artifacts and therefore strong artifact reduction can increase image quality. On the other hand, too strong processing removes image details in high quality video. These contrary optimization criteria cannot be fulfilled easily by simple processing strategies. This paper proposes a framework utilizing different artifact reduction algorithms while their output is controlled by objective quality measures selecting the most suitable algorithm at run-time. Thus, a dynamic selection of the optimal processing for each input scenario over a wide quality level is possible. Integrating a feedback loop and applying the process again can further enhance the output quality.

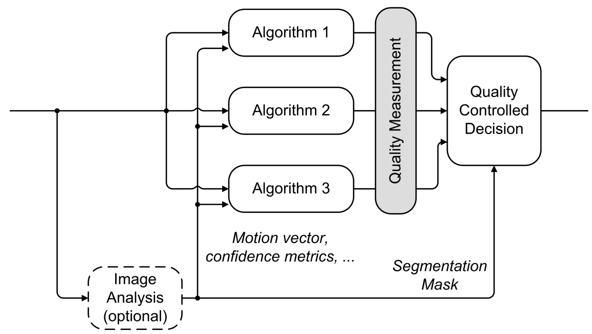

Figure 1: Multiple algorithm blending [7]. The grey block indicates a proposal to integrate image quality control.

1. INTRODUCTION AND RELATED WORK

Over the last few years, the format diversity of input sources to high definition displays has strongly increased. On the one hand, highly compressed video from handheld devices or web portals with dominating artifacts and poor detail level must be handled. On the other hand, high definition content or even 4K resolution video featuring a lot of details and very few artifacts must be processed as well. Without highly adaptive processing, high quality aspects like removing artifacts while preserving sharpness and detail level cannot be fulfilled. In general, static and adaptive filter algorithms [1] ,[2], trainable filters [3], [4], and iterative reconstruction techniques [5], [6], [7] can be distinguished with respect to their adaptivity and flexibility. Whereas static filters cannot even discriminate artifacts and details, adaptive filters carry out an image or alternatively an artifact analysis and switch between several processing modes. Trainable filters are based on an off-line optimization process with the objective to get optimal filter coefficients for image processing tasks like scaling or artifact reduction. But the results of this processing strongly depend on the training material, and several filter coefficient tables may be required for different quality levels. A strategy to increase the adaptivity is multiple algorithm blending which is depicted in Fig. 1 [7]. Several algorithms optimized for certain image features, e.g. edges, textures, or homogenous regions, are applied to the whole image, and then, based on an image analysis discriminating determined features, the best processing is chosen. In the approach already described in [7], the image analysis is only based on information of the input sequence and no output control is carried out.

A further group of processing are iterative reconstruction algorithms with image models to carry out image processing tasks like artifact reduction [5]. These tasks are usually solved by a global on-line optimization process. Thus, suitable processing for a wide range of input qualities can be achieved due to automatically adoption at run-time. In Fig. 2 a generalized flow diagram for this strategy is shown. But the optimization approach can handle only rather simple features due to limitations of the image models. Therefore complex image features, especially present in high resolution sequences, cannot be described well and therefore may be removed during processing. To overcome these limits, system based improvements can be introduced, e.g. by applying content adaptive processing and concentrate the optimization to regions where the internal image model is fulfilled [7].

The contribution of this paper is a framework to determine the potentially best processing for each input quality by introducing image quality measures to control the reached output quality. Fig. 1 and 2 show solutions how these measures can be used to improve the performance. Instead of relying on information about the input sequence only, the output is controlled and the best processing is selected based on this control unit (Fig. 1). The limited mathematical model in iterative optimization techniques (Fig. 2) can be improved by checking the quality after each iteration, stopping the process if an objective quality degradation is reached. As will be depicted later, not only optimal parameter settings for each algorithms but the optimal execution order of several algorithms can be determined at run-time. This paper is organized as follows:

- The proposed framework, its components and the derivation of an optimization function is described in section 2. Detailed validation and theoretical assumptions will be given to guarantee a good output quality.
- Several application scenarios, e.g. switching between algorithms on different levels like pixel or sequence level are presented in section 3.
- An evaluation of this framework compared to state-of-the-art processing is given in section 4.
- The results of this paper are summarized in section 5.

 ISSN 2076-1465

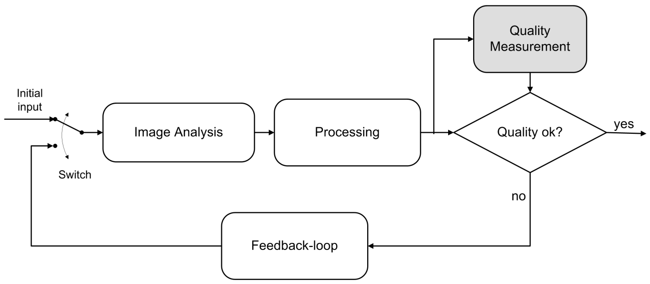

Figure 2: Iterative reconstruction algorithm. The grey block shows how image quality control can be integrated.

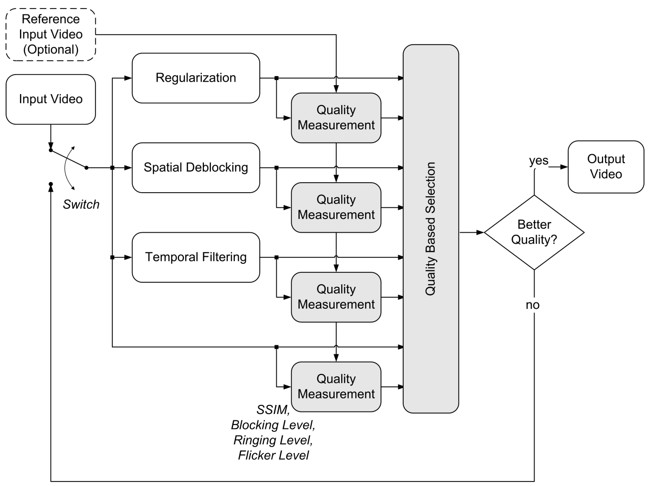

Figure 3: Block diagram of general framework for our system concept.

2. FRAMEWORK FOR QUALITY CONTROLLED PROCESSING

This section deals with a system overview describing each component. Because the quality selection function is a crucial part of the system and its performance, it will be presented in detail. Mathematical optimization functions must fulfil certain mathematical constraints. Thus, a validation of the function based on extensive measurements and theoretical considerations are discussed afterwards.

2.1 System overview

Fig. 1 and 2 already depict potential solutions how to integrate image quality measures into known processing schemes. By combining both approaches, the framework in Fig. 3 can be obtained. Several processing strategies already used for artifact reduction, deblocking [7], regularization [8], [7], and temporal filtering [9] are selected for processing. The main reason for the selection of these three algorithms is that they are suited to reduce different coding artifacts under certain situations. Blocking can be reduced by spatial deblocking in still and moving sequences, temporal filtering can remove it in moving sequences. Ringing can be reduced by regularization and temporal filtering, and flicker can strongly be reduced by temporal filtering. Because each of these methods comprises an internal image analysis, blur due to erroneous processing can for the most part be avoided. The output of each algorithm is fed to an objective quality measurement unit applying an optimization function described in the next section. Based on the results of this optimization function for each processing, the best processing is determined for each input. After this selection process, the result is fed back and the methods are applied again. The implemented framework is flexible, allowing the selection process on a pixel-by-pixel, frame-by-frame, or sequence-by-sequence basis. If no improvements could be achieved by processing in the actual iteration, the final output is obtained.

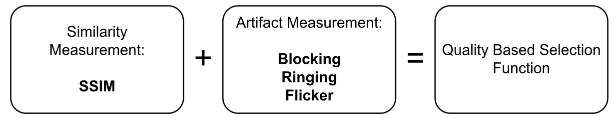

Figure 4: Basic blocks of optimization function.

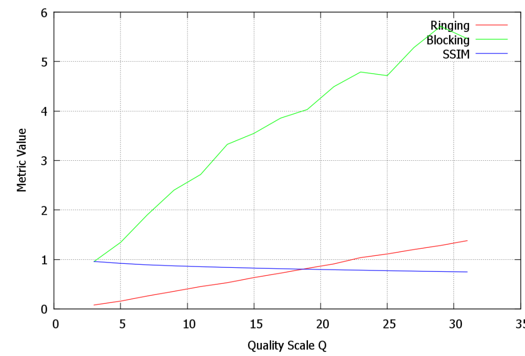

Figure 5: Evaluation of each quality metric for MPEG-2 coded sequences over all quality levels.

2.2 Implementation and validation of optimization function

Crucial for the performance of the artifact reduction framework is the implementation of a suitable optimization or cost function. To guarantee a unique solution, the function must be convex with only one global optimum. Moreover, a good reproduction of the human visual system and its quality perception and in the best case even of a display model must be achieved. One simple quality measure (e.g. PSNR or SSIM) can only measure the overall quality and is not focussed on (local) coding artifacts. Thus a combination of several metrics is necessary. Known from iterative optimization techniques [5], a function with two blocks is chosen as illustrated in Fig. 4: structural similarity index (SSIM, [11]) is used to guarantee similarity between input and output sequence, and specific artifact measures (blocking level [10], flicker level [12], ringing measure) are used as constraint. In optimization techniques, this second part is interpreted as image model. Contrary to (global) optimization, no solving strategy is applied. The output after each iteration is only determined by the processing of each algorithm. The function only evaluates the results of each processing. As already discussed in [11] SSIM correlates much better with the visual perception than e.g. the PSNR and is used therefore.

Several options are possible to build up an optimization function: non-linear combination (maximum, minimum), fuzzy classification rules, Pareto optimality, or simple linear combination like weighted averaging. Because this pa-

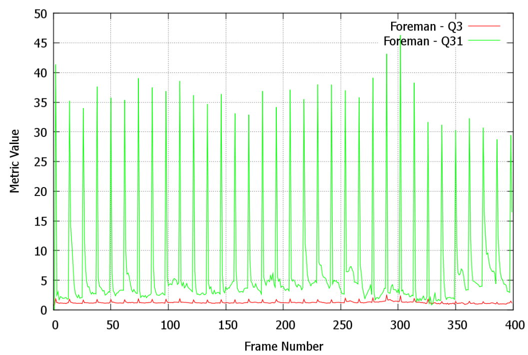

Figure 6: Blocking level measurement for quality levels 3 and 31 of 'foreman' sequence coded with MPEG-2.

per focuses on a general concept, the latter is chosen for demonstration. To create a final optimization function, normalization and/or distortion functions are necessary to fulfil the mathematical prerequisites. After extensive tests the final equation has been obtained:

$$f = SSIM_{norm} - \alpha \times \{BL_{norm} + RL_{norm} + FL_{norm}\} \qquad (1)$$

In this equation a higher value represents a better image quality. SSIM is measured using the processed result and the (given) optimal reference. *BL*, *RL*, and *FL* are blocking, ringing, and flicker level. *norm* represents a normalization to scale each measure to the same dynamic range. α is a weighting parameter controlling the selection process which focus strongly on similarity to the input or on artifact reduction. Setting α to 0.15 reaches a good compromise between both properties. Fig. 5 and 6 depict results of the tests carried out to determine this function. As can be seen in Fig. 5, each of the measures has a different range of values motivating the normalization. As normalization, the output of the measure between coded input and reference (uncompressed raw material) is used in equation (1). Additionally it can be observed that all measures behave nearly linear between each quality step, so no distortion function (e.g. logarithm, exponentional) is necessary. To determine these behaviors, several sequences were coded with MPEG-2 at quality scales from 1 (high quality) to 31 (poor quality). But the output of the quality function can strongly vary locally. Fig. 6 illustrates this for the blocking level of different frames of a MPEG-2 coded sequence. The strong presence of blocking in the intra frame coded pictures is especially visible in the output of this measure which is less visible at higher quality levels. Because each of the chosen measures is computed locally for every pixel and can then be summed up to a measure for a frame or the whole sequence, switching between algorithms is possible on pixel, frame, and sequence level.
Because the optimal input quality is not available in every scenario, e.g. if this framework is applied in a consumer device after decoding, the concept from equation (1) can only be used to develop optimal signal processing chains or for parameter tuning and the like. To overcome these limit, SSIM is measured using the processed sequence and the distorted input. But the distorted input contains image features but also artifacts and therefore the SSIM of 1 is not optimal. A solution to this problem is to introduce a deviation δ and replace the first term in equation (1) by a $|\delta - SSIM|$ term. This term penalizes a SSIM value above and below δ, overcoming the problem that 1 is not optimal on the one hand and preventing blur in case of a too strong deviation from the input on the other hand:

$$f = |\delta - SSIM_{norm}| + \alpha \times \{BL_{norm} + RL_{norm} + FL_{norm}\} \quad (2)$$

In this case, a lower value represents a better image quality. Further improvements are possible by weighting each term depending on a global image analysis dividing the input sequence into edge, flat and textured regions, consider motion perception or display models. With this information the perception of the human visual system can better be modelled, because artifacts are masked in specific regions (e.g. textures) and are strongly visible in flat regions. Moreover, the erroneous detections of artifacts can be prevented by this procedure. E.g. ringing can only be present at edges surrounded by homogenous regions. For normalization, either fixed values or a specific processing, e.g. described in [9], can be used as reference material.

2.3 Theoretical considerations

For a validation of this framework concept, a theoretical proof is essential to underlie the detailed measurements. In contrast to off-line optimization techniques, not every possible parameter set is tested to obtain the output solution. In our case, we propose the processing by several known algorithms, temporal deblocking [9], spatial deblocking [10], and spatial regularization [7] thus allowing only deterministic step sizes determined by these algorithms. Because each algorithm is subjectively tuned by its parameters and internal and adaptive image analysis, completely describing the overall framework mathematically is very challenging. The advantage of the additional quality measurement is that after every step each processing is evaluated and the best one is chosen. The iterative processing can be formulated mathematically:

$$I_{n+1} = f(A_1(I_n), A_2(I_n), ..., A_m(I_n), I_n) \qquad (3)$$

A_1 to A_m denote the algorithms used for processing of input image I_n at iteration step n and the output image I_{n+1} is a combination of every processed output by each algorithm and the input image. Fig. 7 (a) depicts this selection process for every pixel position. Convergence is reached when no algorithm leads to an improvement and the image from the previous iteration is bypassed. This convergence is guaranteed due to the assumption (based on extensive measurements and the formulation of equations (1)-(3)) that f is convex and thus only improvements are possible otherwise the unprocessed input is selected. In certain scenarios a fixed iteration number is sensible to guarantee real-time processing or to reduce hardware costs.

3. APPLICATION SCENARIOS

The framework can be used for several applications and lead to improved solutions for artifact reduction: as a real-time scenario in consumer electronics selecting the optimal algorithm execution order for each sequence at run-time, as a controlled off-line framework for optimization of the best parameter settings, or as improved objective evaluation function. The 'granularity' of this framework can differently be tuned to every scenario. The selection process can be

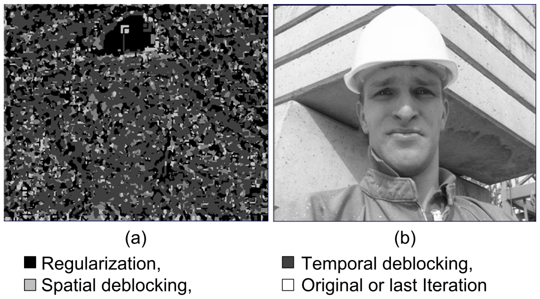

Figure 7: Algorithm selection process on pixel level.

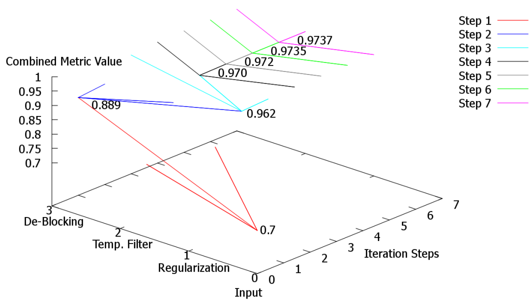

Figure 8: Algorithm selection in iterative quality based framework for 'akiyo' sequence.

Figure 9: 'foreman' sequence coded with Motion JPEG using quality scale 3 (a)-(c) and quality scale 19 (d)-(e). Input (a), (d), STABLE [9] (b), (e), and proposed framework with reference input (c), (f).

applied on pixel level, frame level or sequence level. Using pixel level and optimization function (1) is sensible to determine the potentially best processing for every input situtation providing a 'benchmark' for other algorithms and frameworks. Fig. 7 depicts a map illustrating the selection process on pixel level. As can be seen, at certain positions different algorithms are preferred over the other.

In Fig. 8 a selection process for the 'akiyo' sequence on sequence level is depicted. As can be seen, at first a spatial deblocking is selected because major parts in this sequence are still (background) and blocking is the most annoying degradation in this sequence. Afterwards, spatial regularization is applied further improving image quality, mainly by reducing ringing artifacts at edges. Then, temporal filtering is carried out several times reducing flicker and remaining artifacts until no improvements regarding the optimization function (1) are possible. The number of iterations strongly depends on the input material and varied between 3 and 15 in our experiments. For this sequence, the optimization framework described in this paper validates the proposed combination of deblocking and regularization introduced in [7]. The strongest improvements are achieved by spatial deblocking in the first iteration. After several iteration steps, only minor objective improvements are possible. Therefore, computational effort can be saved by stopping before convergence of the framework is reached. For other sequences and quality levels, different algorithm execution orders will be the result, e.g. with less temporal filter stages because there is less temporal correlation between consecutive frames. Other tasks like objective parameter optimization can directly be applied to this framework by replacing the different algorithms (regularization, deblocking, and temporal filtering) by one and the same method tuned with different parameters.

4. EVALUATION AND RESULTS

In this section objective and subjective results of the framework are presented. As already described, extensive tests with different sequences (e.g. akiyo, foreman and football) coded with different coding standards (MPEG-2, MPEG-4, Motion JPEG) at quality scales 1 to 31 were carried out. Moreover, the reachable performance compared to other actual approaches is described, mainly spatio-temporal image content adaptive coding artifact reduction (STABLE) [9] and texture preserving spatial regularization with deblocking as pre-processing (TextPres) [7]. As depicted in literature, these methods are comparable or sometimes even better than other state-of-the-art coding artifact reduction algorithm. Fig. 9 shows results of different algorithms for a high quality (quality scale 3) and low quality (quality scale 19) sequence emphasizing the high flexibility of the proposed framework. The results of our optimization framework measure the image quality against the perfect (unprocessed) reference. The other methods do not have this advantage and this fact must be considered during evaluation of the superior performance at both quality levels regarding artifact reduction but even detail re-establishment. Up to a certain degree, an upper anchor for benchmarking can be accomplished by the novel framework. In comparison to this, the spatio-temporal image content adaptive artifact reduction, comprising spatial deblocking, regularization and temporal multi-frame (motion compensated) filtering strongly increases the performance compared to the input sequence, but certain details cannot be re-established because the optimal (input) case is not known and must be estimated by internal assumptions within made by from image models. Whereas for the low bit-rate case image quality strongly improves due to artifact removal, in the high quality input some image details are removed degrading the quality. These subjective observations can be confirmed by objective measures (SSIM, see Table I). The SSIM measure is lower for the sequences processed by the spatio temporal

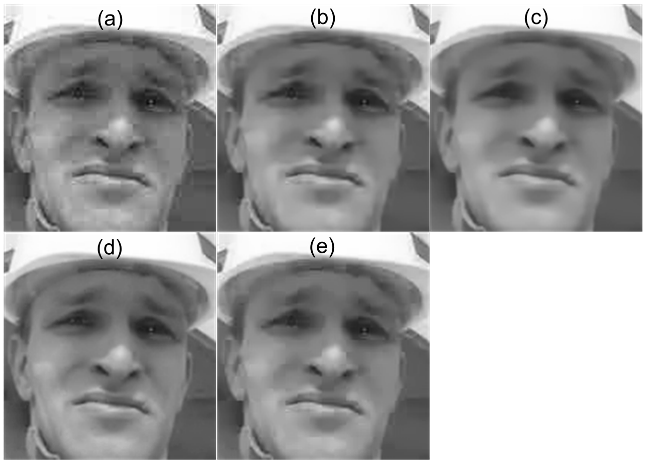

Figure 10: Motion JPEG coded sequence 'foreman' using quality scale 19 (a), processed by TextPres [7](b), STABLE [9] (c), and proposed framework with (d) and without (e) ideal reference input.

Table 1: SSIM results for different algorithms for Motion JPEG coded sequences. (P.w.r. = proposed framework with reference, P.n.r. = proposed framework without reference)

	Akiyo Q3	Akiyo Q19	Foreman Q3	Foreman Q19
Coded inp.	0.9812	0.9293	0.9632	0.8565
TextPres [7]	0.9743	0.9434	0.9474	0.8731
STABLE [9]	0.9748	0.9454	0.9481	0.8822
P.w.r.	0.9841	0.9605	0.9725	0.9157
P.n.r.	0.9735	0.9445	0.9443	0.8765

image content adaptive processing (STABLE) than for the unprocessed but compressed input at quality scale 3. For the lower quality, strong improvements are even objectively possible. For our proposed framework with the given reference (P.w.r.) subjective and objective improvements are possible for low and high quality validating the optimization framework and its underlying mathematical function.

Fig. 10 compares the results of the proposed optimization framework for two input cases. In (d) the optimal reference sequence is available (d) and in (e) it is not. As can be seen, even if the reference is not available (e) a strong improvement can be achieved compared to the unprocessed input (a), but remaining blocking is clearly visible. For benchmarking, results of the texture preserving regularization (TextPres) (b) and spatio-temporal image content adaptive processing (STABLE) (c) are shown, too. If the reference is not available, the subjectively and for the low bit-rate case optimized spatio-temporal processing (STABLE) outperforms the framework presented in this paper. These results are confirmed by the SSIM measurements in Table I which is remarkable because SSIM is one measure used in the optimization process of the framework. The limits of this framework can be explained by the fixed parameters δ and α determining the weighting between artifact measurement and similarity to the given distorted output, with a predermined deviation by δ from the input content. If these parameters are chosen adaptively, higher performance will be possible.

5. CONCLUSION

In this paper a highly adaptive coding artifact reduction framework based on objective quality selection of a suitable algorithm for every input scenario is presented. The flexibility of this framework allows automatic tuning to input sequences with completely different quality. This is extremely important for actual displays, to process both low bit-rate content and HD well. Using this framework and its quality selection function, optimal parameter sets or alternatively the optimal execution order of several algorithms can automatically be determined. Depending on processing time and costs, an iterative feedback loop can be integrated into the framework with either a fixed number of iterations or computation until convergence. Subjective and objective results show a high image quality for this framework comparable or in many cases superior to conventional methods without quality measurement. Further applications of this framework can be an improved objective evaluation for artifact reduction algorithms, because similarity to the original input on the one hand and the degree of distortions on the other hand are measured.

REFERENCES

[1] G. de Haan, Video Processing for Multimedia Systems, University Press Eindhoven, 2000.

[2] H. Schröder, H. Blume One- and Multidimensional Signal Processing: Algorithms and Applications in Video Processing, WileyBlackwell, 2000.

[3] T. Kondo et al., Picture conversion apparatus and learning method, Sony Corporation, US Patent 6323905, 2001.

[4] L. Shao, H. Zhang and G. de Haan, An Overview and Performance Evaluation of Classification-Based Least Squares Trained Filters, IEEE Transactions on Image Processing, Oct. Vol. 10, 2008, pp. 1772-1782.

[5] G. Aubert and P. Kornprobst, Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations, Springer 2006.

[6] T. F. Chan and J. Shen, Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods, SIAM, 2005.

[7] C. Dolar, M. Richter, H. Schröder, Total Variation Regularization Filtering for Video Signal Processing IEEE Int. Symposium on Consumer Electronics, Kyoto, 2009.

[8] L. I. Rudin, S. Osher, and E. Fatemi,. Nonlinear total variation based noise removal algorithms, Physica D: Nonlinear Phenomena, vol. 60, no. 1-4, 1992.

[9] M. Richter, C. Dolar, and H. Schröder, Coding Artifact Reduction by Temporal Filtering, IEEE Int. Symposium on Consumer Electronics, Kyoto, 2009.

[10] P. Piastowksi, Image processing to reduce blocking artifacts, US Patent Application US20060274959 2006

[11] B. Wang, S. Sheikh, Image Quality Assessment: From Error Visibility to Structural Similarity, IEEE Trans. on Image Proc., Vol .13, No. 4, 2004

[12] D. Vo and T. Nguyen, Directional motion-compensated spatio-temporal fuzzy filtering for quality enhancement of compressed video sequences, IEEE Int. Conf. on Image Processing (ICIP), San Diego, 2008