MOTION-COMPENSATED TECHNIQUES FOR ENHANCEMENT OF LOW-QUALITY COMPRESSED VIDEOS

Ling Shao, Ihor Kirenko, Andre Leitao, Piotr Mydlowski

Philips Research Europe, High Tech Campus 36, Eindhoven 5656 AE, The Netherlands

ABSTRACT

Algorithms for enhancement of low quality compressed videos are described and evaluated in this paper. Cascaded and combined spatio-temporal filters using hierarchical motion estimation and occlusion area detection are investigated for removal of severe coding artifacts and temporal flickering. Both objective and subjective evaluations prove that temporal filtering can significantly improve the video quality of low bit rate video sequences. The proposed methods could therefore be a differentiating feature for future IPTV design.

Index Terms— Adaptive filters, motion-compensated filtering, compression artifacts removal.

1. INTRODUCTION

The ever-expanding availability of internet videos has prompted the popularity of IPTV. Internet videos are usually highly compressed and often have much lower quality than that of the broadcasting channels. Therefore, different video enhancement algorithms should be designed to cope with this new challenge.

Spatial filtering is commonly used for sharpness enhancement or noise/artifacts reduction in TV signal post-processing. In other words, for video samples, algorithms are based on information available only within the currently processed frame. The process of enhancing quality of the sequence is done relying on local dependencies or structure of a spatial region. Spatial algorithms, like trained filters [1-3], bilateral filters [4-5] or filters using complexity measures [6-7], can solve many problems. However, some artifacts such as flickering, extensive blockiness and noise cannot be removed efficiently using 2D methods. Due to their temporal nature, removal of these highly resistant artifacts needs to take under consideration information not only from the current frame but also from the adjacent ones.

This paper is aimed at the investigation of possible advantages of motion-based algorithms for enahancement of low-quality compressed videos. Section 2 describes the motion-compensated compression artifacts and noise removal solution. The objective and subjective evaluations of the method in comparison to spatial filtering are presented in Sections 3 and 4. Finally, Section 5 concludes our paper.

2. MOTION-COMPENSATED FILTERING

Low bit rate videos which are commonly displayed on IPTV usually contain compression artifacts, such as blocking, ringing and mosquito noise. These artifacts often can be effectively reduced by spatial filtering methods. Still there are artifacts, which are not easily visible on the single frame basis, but they become quite noticeable in a video sequence. One of the most common and annoying examples is flickering. Fig. 1 illustrates the characteristic of temporal flickering. For those kind of artifacts, motion-compensated filtering will be beneficial.

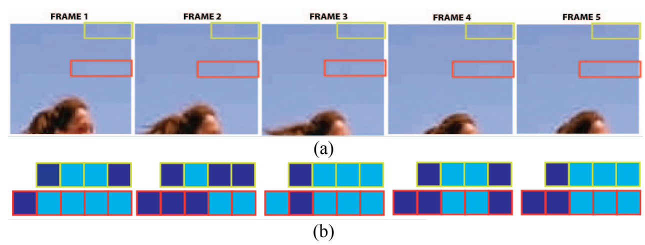

Fig. 1. Illustration of flickering effect. (a) parts of compressed frames; (b) zoomed pixel areas.

Motion-compensated filtering fully utilizes temporal information. It relies on motion estimation to create vectors for precise matching between corresponding pixels of consecutive frames in time. Using the results of this estimation, motion-compensated algorithms retrieve pixels along motion vectors to perform temporal filtering. Low-pass filtering of pixels in the time domain allows reduction of temporal artifacts, which can not be addressed by spatial filters.

Three-dimensional recursive search (3DRS) block-matching algorithm [8] has been implemented in many TV systems for motion estimation. 3DRS takes into account both spatial and temporal coherence and is very computationally efficient. However, for compressed materials the motion estimation tends to be unreliable due to the interference of coding artifacts. To solve this problem, we applied a hierachical motion estimation approach based on 3DRS using resolution down-scaling. Using spatial down-scaling, the coarser motion vectors are obtained from block-matching at a lower spatial resolution and can be successively refined at higher resolutions (Fig. 2).

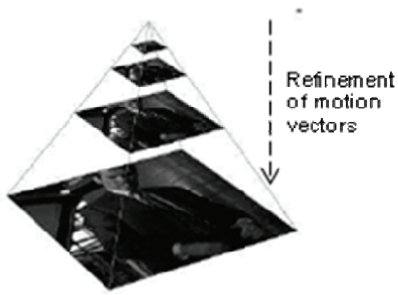

Fig. 2. Different scales in hierarchical motion estimation.

Because the coding artifacts are highly suppressed in the down-scaled images, the motion vectors tend to be smoother, which is also propagated to higher scales. As a result, the obtained motion field is robust to spatially local high-frequency artefacts or noise. Additional to the increase of noise-robustness of the estimation, this approach has also memory size and bandwidth advantages.

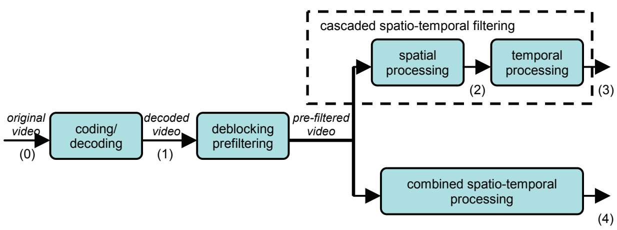

Fig. 3. Flowchart of the spatio-temporal processing.

Two motion-compensated approaches, cascaded and combined, are designed for reduction of coding artifacts and flickering effects. Fig. 3 depicts the processing procedures of the two methods. For both methods, the pre-filtering of block-edges, described in [9], is first applied on the decompressed sequence to reduce the blocking artifacts. Pre-filtering is a necessary and critical part for both the cascaded and combined methods, because the severe blocking artifacts will significantly degrade the reliability of motion estimation if they are not reduced by the pre-filter.

For the cascaded approach, trained filters [1], which rely on a classification-based least squares optimization, is used for spatial filtering. In the cascaded approach spatial and temporal filtering are executed independently. The temporal filtering is based on a motion-compensated Infinite Impulse Response (IIR) filtering. The IIR filter is applied to the current pixel and the corresponding output pixel of the previous frame along the motion vectors:

$$F_i^o = \alpha * F_i + (1-\alpha) * F_{i-1}^o \ , \qquad (1)$$

where F_i^o - output pixel of the current frame, F_i - input pixels of the current frame, F_{i-1}^o - output pixel of the previously filtered frame, α - parameter, which defines the strenght of temporal IIR filtering. The parameter α is defined for every filtered pixels depending on several outputs of the motion estimation algorithm, e.g. occlusion probability, consistency of the motion field :

$$\alpha = (occl[i,j] + incon[i,j]) \qquad (2)$$

where the $occl[i,j]$ parameter defines the probability of occlusion for the pixel with position $[i,j]$, and $incon[i,j]$ is the parameter which estimates the inconsistency of the motion vector field for the same pixel.

We assume that the value of α should be between 0.5 and 1.0. A high causes a stronger contribution of the currnet frame to the resulted filtered frame; thus, the value of α should be increased in the cases when the contribution of the previous frame is not desirable. This might happen, for example, after a scene change, locally in occlusion areas, or when the motion estimation produces unreliable motion vectors.

For the combined solution, a spatio-temporal bilateral filter is implemented. The strength of blurring and the size of the spatial apperture depend on both the motion strength and the local spatial complexity of the region , as definded in the following equation.

$$STF[I]_{\mathbf{p}} = \frac{1}{W_{\mathbf{p}}} \sum_{\mathbf{q}\in S} G_{\sigma_s}\left(\|\mathbf{p}-\mathbf{q}\|\right) G_{\sigma_r}\left(|I_{\mathbf{p}} - I_{\mathbf{q}}|\right) I_{\mathbf{q}} \qquad (3)$$

where I and $STF[I]$ are the input and output of filtering, G_{σ_s} and G_{σ_r} are Gaussian functions for space and range respectively, S includes pixels in neighbourhoods of both the current pixel P in the current frame and the corresponding pixel in the previous frame according to motion estimation. σ_s and σ_r are both dependent on compelxity of the region and the motion strength.

Intuitively, more smoothing is applied, if the amount of motion analysed during the motion estimation is large and if the spatial region is less complex. This is explained by the fact that a fast motion results in a motion blur, thus more smoothing will not destroy perceived sharpness. Moreover, artifacts usually are more visible in less complex spatial regions, therefore stronger filtering can be applied to flat regions without a risk of destroying image details.

Only spatial filtering is applied for areas, which have a high probability of being occluded, or in other words, when the parameter $occl[i,j]$ is larger than a pre-defined threshold. The temporal part of the cascaded filtering is skipped also in cases of scene change.

3. OBJECTIVE EVALUATION

The quality of the investigated algorithms was evaluated using objective measures. Eight different sequences as illustrated in Fig. 4 are used in the evaluation.

Fig. 4. Snapshots of 8 test sequences.

MPEG-2 is applied for compression in our experiments. We first employ PSNR to evaluate the fidelity of the filtered sequences in comparison to the original video (0). The following sequences have been analyzed (Fig. 3):

(1) : non-processed degraded sequence (MPEG-2 coded with 500 Mbps bit rate);

(2) : degraded sequence processed with pre-filter and spatial trained filter;

(3) : first pre-filtered, then processed with trained filter and finally with the IIR temporal filter;

(4) : first pre-filtered, then processed with combined spatio-temporal filter;

Table 1. PSNR results.

sequence	core	corv	girl	h	niro	out	st	star
(0) to (1)	31,32	30,41	30,60	28,37	26,39	29,70	28,59	32,36
(0) to (2)	32,41	29,59	29,95	26,63	26,36	28,06	28,01	31,33
(0) to (3)	32,37	29,56	30,09	26,65	26,43	27,39	28,14	31,49
(0) to (4)	31,85	30,73	31,13	28,54	26,80	29,19	29,23	32,93

From Table 1, we can see that the combined spatio-temporal algorithm in general produces slightly better results. The spatial filter based on trained filter and the cascaded spatio-temporal method give similar performance. Unfortunetely, PSNR can not always serve as a valid estimation of artefact reduction efficiency. Therefore, we used other metrics as well.

To assess the de-blocking capability of the algorithm, non-reference Block-Edge Impairment Metric (BIM) [10] is used.

Table 2. BIM results.

sequence	core	corv	girl	h	niro	out	st	star
(1)	4,49	4,62	4,54	3,23	3,30	5,93	4,64	4,84
(2)	1,61	1,45	1,31	1,34	1,48	1,89	1,76	1,52
(3)	1,47	1,34	1,28	1,30	1,35	1,80	1,58	1,37
(4)	1,74	1,56	1,57	1,54	1,60	2,93	1,99	1,70
(0)	1,14	1,26	1,06	1,06	1,28	1,50	1,06	1,16

Accroding to Table 2 the cascaded spatio-temporal filter provides BIM results closest to uncompressed video. Spatial filtering using trained filter gives the second best results, while the combined method turns out to be not very effective for blocking artifact reduction according to BIM.

Unfortunately, only objective estimation cannot be always used for reliable judgement of the visual quality. It is especially noticeable in case of degradation involving not only spatial but also the temporal domain.

4. SUBJECTIVE EVALUATION

To make the algorithm comparison complete, a subjective evaluation is introduced in this section. The perception test is carried out in a typical lab environment. All sequences are displayed on a LCD panel in pairs, requiring viewers to select one, which has better visual quality. To allow the differentiation of opinions, viewers not only had to decide which video sample is better but also how big the difference is using following weights:

[2] :: right sequence is surely better
[1] :: right sequence is a bit better
[-1] :: left sequence is a bit better
[-2] :: left sequence is surely better

Pairs of the sequences were shown twice in quasi-random order to test the consistency of the viewers. The participating group consisted of experts in video processing and people with no background knowledge in this field. All of them were asked to evaluate the given video sequences in respect to blockiness, sharpness and especially temporal flickering. Figures 5-10 show the results of subjective estimations of different algorithms. Figure 5 shows that most participants thought the spatially processed sequences look slightly better than the non-processed ones. This is not surprising, because temporal artifacts (especially flickering) are still significantly visible after spatial filtering.

Conclusion can be made from Figures 5-10 that both cascaded and combined spatio-temporal filtering can significantly improve the video quality, since most viewers considered the spatio-temporally processed sequences are "surely better" than both the non-processed and the spatially filtered ones. The difference between the cascaded method and the combined filter is subtle (Fig. 10), but in average, the cascaded algorithm performs slightly better.

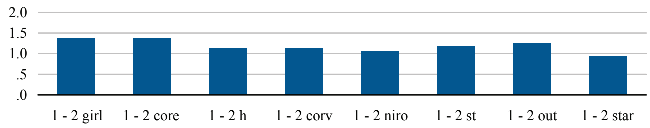

Fig. 5. Evaluation of non-processed (1) and spatially filtered (2) sequences.

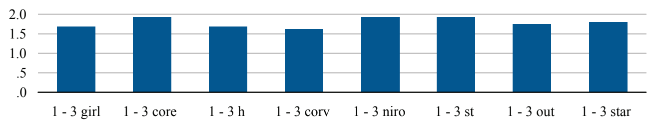

Fig. 6. Evaluation of non-processed (1) sequence and result of cascaded filtering (3).

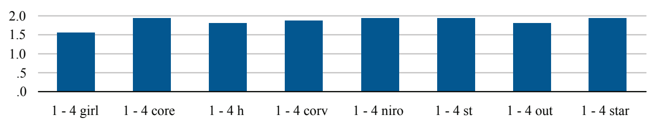

Fig. 7. Evaluation of non-processed (1) sequence and result of combined (4) filtering.

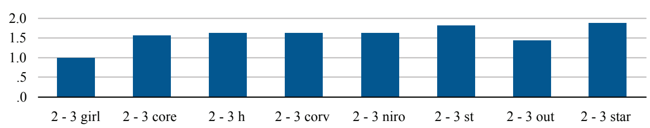

Fig. 8. Evaluation of results of spatial (2) and cascaded (3) filtering.

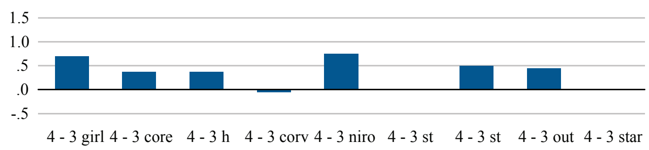

Fig. 9. Evaluation of results of cascaded (3) and combined (4) filtering.

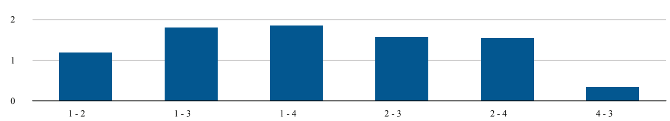

Fig. 10. Summary of subjective evaluation results.

Temporal filtering is effective mostly for low bit rate videos. Sequences compressed at high bit rates do not suffer from blocking artifacts, propagated into the temporal domain. Applying algorithms using temporal filtering on high quality videos might cause blurring or halo effects.

5. CONCLUSION

In this paper algorithms for enhancement of low quality compressed sequences are explained. The algorithms exploit the motion-compensated temporal filtering combined with adaptive spatial filtering. Two approaches are investigated, namely the cascaded and combined solutions.

According to the results of objective and subjective evaluation, spatio-temporal filtering brings a huge benefit in case of sequences suffering from severe coding artifacts (e.g. flickering). Only spatial filtering cannot remove severe coding artifacts, visible as flickering. After application of recursive temporal filtering along estimated motion vectors, the local temporal visual consistency can be recovered and strong blockiness can be removed.

Subjective results of combined and cascaded algorithms are similar, with slight outperformance of the combined algorithm. Objectively, cascaded solution provides lower values of BIM, due to more efficient removal of blockiness, but in most cases leads to lower values of PSNR compared to combined algorithms due to larger blurring.

REFERENCES

[1] 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, Vol. 17(10), pp. 1772-1782, Oct. 2008.

[2] L. Shao, J. Wang, I. Kirenko and G. de Haan, "Quality Adaptive Trained Filters for Compression Artifacts Removal", In Proceedings of the 33rd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, USA, March-April 2008.

[3] L. Shao, H. Hu and G. de Haan, "Coding Artifacts Robust Resolution Up-conversion", In Proceedings of the 14th IEEE International Conference on Image Processing (ICIP), San Antonio, Texas, USA, September 2007.

[4] B. Zhang and J. P. Allebach, "Adaptive bilateral filter for sharpness enhancement and noise removal", IEEE Trans. Image Processing, Vol. 17(5), pp. 664-678, May 2008.

[5] I. Kirenko and L. Shao, "Adaptive Repair of Compressed Video Signals Using Local Objective Metrics of Blocking Artifacts", In Proceedings of the 14th IEEE International Conference on Image Processing (ICIP), San Antonio, Texas, USA, September 2007.

[6] L. Shao and I. Kirenko, "Coding artifact reduction based on local entropy analysis", IEEE Trans. Consumer Electronics, Vol. 53(2), pp. 691-696, May 2007.

[7] O. A. Ojo and T. G. Kwaaitaal-Spassova, "An algorithm for integrated noise reduction and sharpness enhancement", IEEE Trans. Consumer Electronics, Vol. 46(3), pp. 474-480, August 2000.

[8] G. de Haan and P. Biezen, "An efficient true-motion estimator using candidate vectors from a parametric motion model", IEEE Trans. Circuits and Systems for Video Technology, Vol. 8(1), pp. 85-91, Feb 1998.

[9] I. Kirenko, L. Shao, R. Muijs, "Enhancement of Compressed Video Signals Using a Local Blockiness Metric", In Proceedings of the 33rd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, USA, March-April 2008.

[10] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding", IEEE Signal Processing Letters, Vol.4(11), pp. 317-320, Nov. 1997.