MOTION SENSITIVE PRE-PROCESSING FOR VIDEO

Dinei A. Florêncio

Microsoft Research, One Microsoft Way, Redmond, WA 98052

ABSTRACT

Due to the capture process, video signals are generally contaminated by noise. Furthermore, overall video quality can be improved by reducing sharpness in fast moving areas, except in cases where the eye is able to track the motion. In this paper we propose a relatively simple pre-processing method that is able to address both these situations. The proposed pre-processing algorithm is based on selectively removing high frequencies that are not well predicted by the motion compensation. While the decision is made based on temporal tracking, the filtering is done exclusively in the spatial domain, thus avoiding the artifacts produced by other pre-processing methods.

1. INTRODUCTION

Most video signals are contaminated by noise. In fact, noise is essentially intrinsic to image capture devices, including CCD and CMOS sensors. While this noise may be negligible when encoding video at low rates, its importance increases in high quality encoding. Pre-processing video in order to reduce this camera noise may improve the overall quality of the encoded video. Furthermore, even at lower bitrates, the encoding process may benefit from not spending too many bits in areas of the image that have complex motion. A common approach to pre-processing involves a temporal-filter, based on motion-compensated frames [1-3]. Nevertheless, areas that are not perfectly motion compensated will introduce artifacts (or “phantoms”), which will consume bits to encode. In this paper, we propose a new pre-processing method, which reduces the camera noise without introducing these artifacts. Furthermore, the proposed technique also reduces sharpness in non-trackable moving areas, which further helps in improving encoding efficiency.

2. WHAT INFORMATION NEEDS TO BE FILTERED OUT?

While we obviously don’t want to spend any bits encoding the camera noise, it is not clear what other information should be discarded. We first note that camera noise, while spreading over the whole frequency range, is most important at higher frequencies, where it is expensive to encode, and often more intense than the high frequency contents of the desired signal itself. A simplistic solution would be therefore to low-pass each frame. Nevertheless, doing so would also significantly reduce the sharpness of the image, ending up with a lower quality signal. Fortunately, these two sources of high-frequency can be differentiated. The high frequency content of the image is going to be the same from frame to frame, and can therefore be predicted by motion compensation (MC), while the camera noise will be independent from frame to frame, and therefore cannot be predicted by MC. Therefore, similarly to previous approaches, we will use MC to differentiate between these two sources of high frequency energy.

One other important aspect of pre-processing is the need to remove irrelevant detail, even if actually originated from the scene. In particular, detail information in fast moving areas cannot be perceived by humans, unless it can be tracked by the eye. In other words, fine details will be relevant only if they are stationary, or – when moving – if they can be tracked by the eye. Differently from other algorithms, our proposed algorithm will also help remove these irrelevant details from the video sequence. Note again that these details are also concentrated in the high frequencies.

3. SPATIAL VS. TEMPORAL FILTERING

As we mentioned in the previous section, most of the information to be removed consist of high frequencies. These high frequencies could be attenuated by simple spatial or temporal low-pass filtering. Nevertheless, only temporal filtering will attenuate camera noise while at same time preserving the high frequency contents of the actual image. For this reason, previous pre-processing methods generally involve some sort of temporal filtering. The key disadvantage of these temporal filtering methods is the phantoms generated in the image, even when motion compensated filtering is used.

In this paper, we propose a new method, where – in essence – the decision about filtering or not comes from

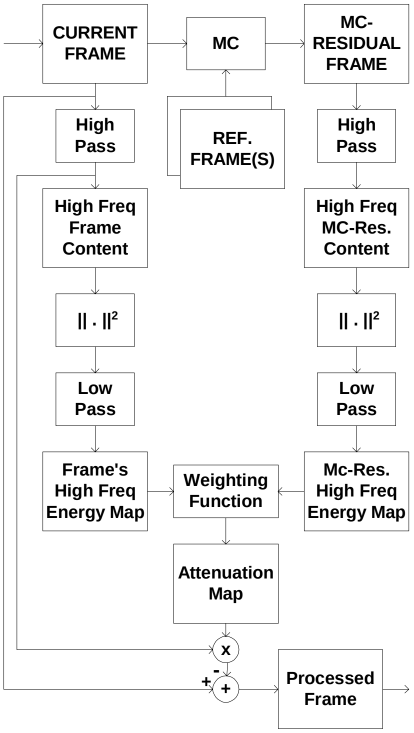

Figure 1 – *Overall diagram of the proposed algorithm***.**

the temporal filtering, but the filtering itself is done in the spatial domain. The proposed method has two advantages: first, it eliminates the artifacts introduced by temporal filtering, and, second, it allow to reduce resolution in areas which have complex motion, helping the encoding process to allocate bits more efficiently to information that can be tracked by the eye.

4. THE PROPOSED ALGORITHM

The proposed method is based on the idea that trackable parts of the image will produce little motion compensation residual. We therefore compare the high frequency content of the original frame with the high frequency content of the motion compensation residual. High frequencies that are present in the signal, but not in the motion compensated residual will represent trackable content, and should therefore be preserved. All other high frequency should be attenuated, since it will correspond to either camera noise, or non-trackable content. Figure 1 presents a high level diagram of the proposed method. We first produce a motion compensated (MC) frame, and then produce a high pass version of both the original frame and the MC residual. High frequency "energy" content is compared between these two, and a ratio based on these two energy values determines how much to attenuate the high frequency contents of the original frame.

More precisely, given a frame F, we produce a high pass version of the frame, Fhp, given by Fhp = F*hh, where hh is a high pass FIR filter. In our simulations we used a separable 5-tap filter with a smooth transition around .3π, but other filter could be used, tuned to the amount of smoothness desired, or to computational constrains. This high pass frame is than squared and low pass filtered (we used a 7 tap separable LP filter). This produces an estimate of the amount of high frequency in each region of original frame, which we call EF.

A motion compensated frame residual R is also produced. Ideally this motion compensation is based on a similar setup to that to be used by the encoder, in regards to which reference frames to use, search range, block size, and etc. Motion compensation for frames that will be intra-coded (I frames) should use the next P picture as reference. An estimate of the high frequency content in the MC residual ER is then produced by following the same steps used to produce EF.

We the obtain an attenuation map by computing, for each pixel:

$$ATT = \max(0, \min(1, G1.(\mathrm{ER}/(\mathrm{ER}+\mathrm{EF}) \text{ - } \mathrm{B1}))) \tag{1}$$

where G1 is a gain factor and B1 a bias factor. To understand the theoretical values of G1 and B1, let us analyze two extreme cases. First note that if all high frequencies are from the image, and the region is perfectly trackable, the MC residual will be zero. In this case, we do not want to attenuate any of the high frequencies, and therefore the theoretical value for B1 is zero. A higher value may be used to preserve more of the high frequencies, since the MC tends to never exactly match, due to the motion vector precision (e.g., half-pel), or other factors. On the other extreme, if the desired image is completely flat, all high frequencies on both the original frame and on the MC residual will be due to camera noise. If we assume that the MC did not track this camera noise, the energy level in the MC should be around twice the energy level in the original frame. Therefore, setting G1=1.5 will yield ATT=1, and therefore remove all high frequencies from this region of the image, as we would expect. A higher value of G1 will increase noise attenuation, and may also be used to compensate the fact that the MC may track some of the camera noise. Or, a lower value may be used to be conservative and preserve more of the high frequencies. In our experiments, we have set to G1 = 1, and B1 = 0.

The final frame is the obtained by subtracting the (attenuated) high frequency content Fhp from the original frame F, i.e.:

$$OF = F - ATT \cdot Fhp \tag{2}$$

where OF is the output frame, to be provided to the encoder.

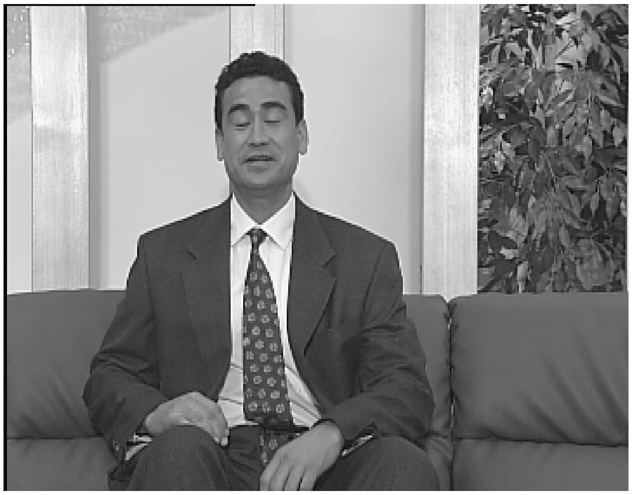

Figure 2 – *Original frame*

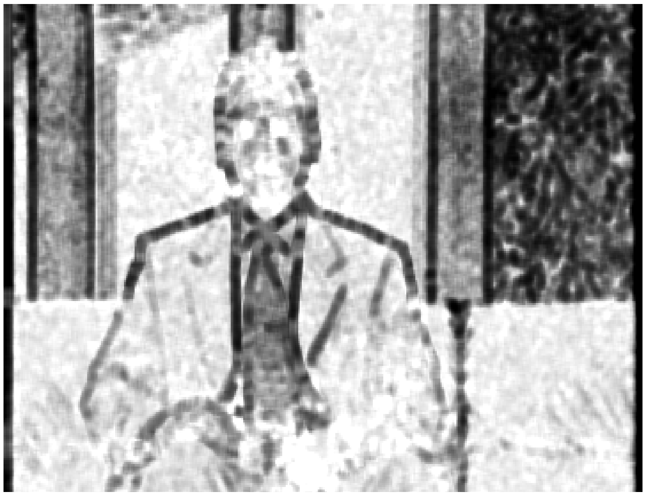

Figure 4 - *Attenuation map (whiter = attenuate more)*

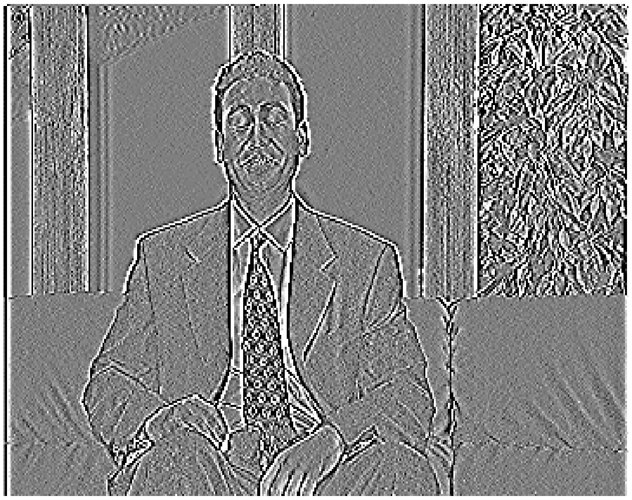

Figure 3 - *High Frequency content of original frame*

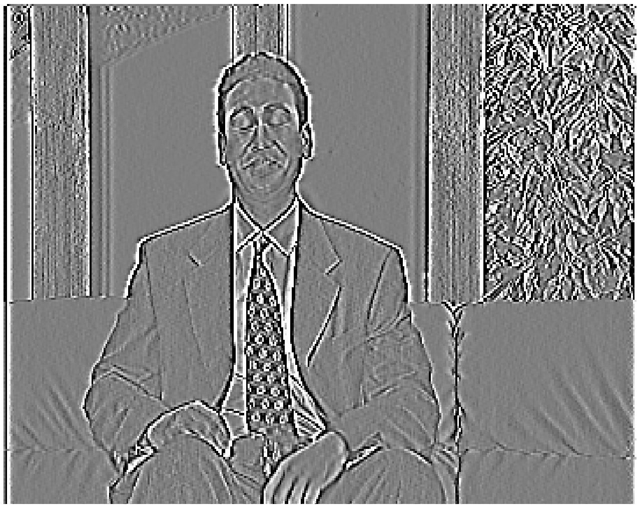

Figure 5 - *High frequency content of processed frame*

5. RESULTS

Figures 2-8 illustrate some of the results obtained by applying the algorithm. Figure 2 is the original frame 50 of the MPEG-4 test sequence SEAN, while Figure 3 represents the high frequency content in that same frame. Figure 4 shows the attenuation map, which highlights regions where the high frequency content was not tracked by the MC, and should therefore be attenuated. Note that all textured regions of the background (e.g. the plants and the textured columns) appear dark in 4, meaning they will have their detail information preserved. The same is true of regions that are well tracked by the MC, like the dark suit contour, or the tie. In contrast, flat regions of the background (where high frequencies are dominated by camera noise) will be low-passed, like the sofa or the wall. Similarly, regions that have complex motion (e.g., face, hands) will have the high frequency detail removed, making it easier for the encoder to allocate bits to the relevant information. This attenuation of the undesired high frequencies can be observed by comparing Figures 3 and 5, which shows the high frequency content of the output (processed) frame. The main effect of the proposed algorithm is to reinforce video elements that can be tracked by the eye, and therefore the difference is mostly noticeable in watching video. Figure 6 compares the coding error in terms of SNR. Note that even though the desired result of the proposed technique is mostly to improve subjective quality by improving quality of trackable content, it even improves the SNR marginally by around 0.2dB. Figures 7 and 8 show the same frame after coding.

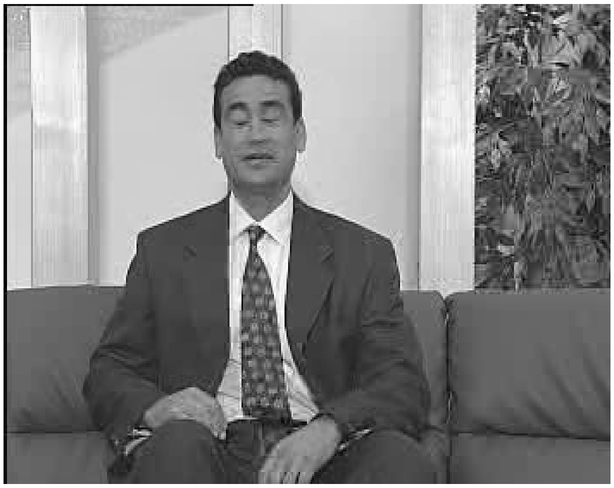

Figure 7 – *Encoded Frame without pre-processing.*

6. COMPUTATIONAL COMPLEXITY AND OTHER CONSIDERATIONS

The most computationally intensive step of the proposed algorithm is, of course, the motion estimation. Nevertheless, this part of the processing requirements could be completely eliminated by integrating the algorithm with the encoding process. This is only possible because the algorithm does not require use of subsequent frames, as in previous pre-processing algorithms based on motion compensation. The real computational requirement of the algorithm is therefore mostly due to the filtering process. Separable filter help reduce the computational complexity, which is around 80 ops/pixel.

Another issue of note is the handling of interlaced video. In this case, the motion compensation can be handled in the same way the encoder operates, but the temporal filtering has to be made within a field, to avoid producing motion blur.

7. CONCLUSIONS

We have presented an algorithm that uses temporal information to selectively attenuate high frequency content in the image. This helps improve coding efficiency, by removing detail information whenever it cannot be tracked by the eyes. By avoiding the temporal filtering common to other algorithms, we have avoided the artifacts produced by the regions where complex motion is present. The basic operation of the algorithm is simple, and the Motion compensation reflects the same choices to be used for the actual coding, allowing the algorithm to share the same motion vector search, except for intra-coded frames, where motion vectors may need to be computed (or re-used).

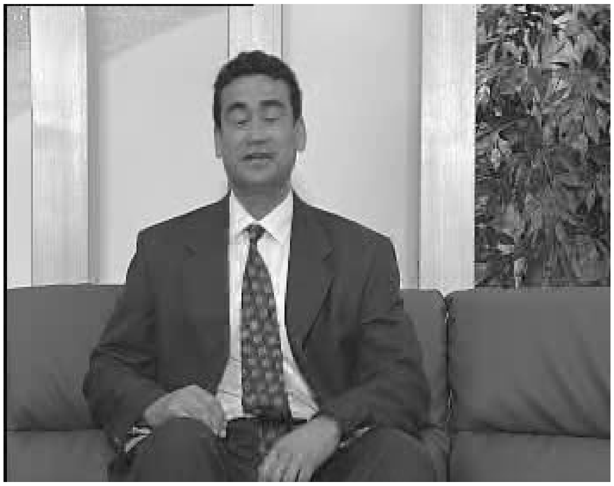

Figure 8 – *Encoded frame using proposed pre-processing.*

REFERENCES

[1] F. Dekeyser, P. Bouthemy, and P. Perez, "Spatial-temporal Wiener filtering of image sequences using a parametric model," *Proc. ICIP'00*, pp. 1586-89, 2000.

[2] P. R. Giaconni, G. A. James, S. Minelly, and A. Curley, "Motion-compensated multi-channel noise reduction of colour film sequences," *SPIE Journal of Electronic Imaging*, vol. 8-3, pp. 246-254, July 1999.

[3] K. J. Boo and N. K. Bose, "A motion-compensated spatial-temporal filter for image sequences with signal dependent noise," *IEEE Trans. Cir. Syst. For Video Tech.*, vol. 8-3, pp. 287-298, June 1998.

[4] A. Kundu, "Motion estimation by image content matching and application to video processing," *Proc. ICASSP'96*, vol.IV, pp. 1902-5, 1996.

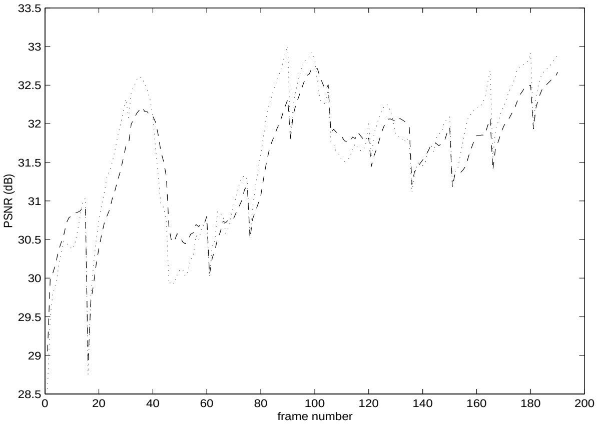

Figure 6 – *PSNR results (dotted = with pre-processing; dashed = standard encoding).*