An Improved Huber-MAP Post Filtering Method for DCT-based Compressed Video Sequences

Jie Xiang Yang, Hong Ren Wu
School of Electrical and Computer Engineering
Royal Melbourne Institute of Technology
Melbourne, VIC 3001 Australia
s3173885@student.rmit.edu.au, henry.wu@rmit.edu.au

Abstract—Maximum a posteriori (MAP) post processing is widely investigated to reduce digital video or image coding artifacts. Many MAP post-filtering methods have adopted the Huber-Markov random field (HMRF) as a prior probability model. However, a drawback of the Huber-MAP post-filter is that it blurs image structures and details in video sequences when removing coding artifacts, such as blocking and ringing. This paper proposed an improved prior model to preserve edge information in images of video sequences as well as to reduce coding artifacts. Experimental results show the performance improvements of the proposed approach over the Huber-MAP post-filter. The performance of the proposed approach is also comparable to that of the H.264 post-filter, in terms of the peak signal to noise ratio (PSNR) and the visual picture quality.

Keywords—edge preservation; artifact reduction; post processing; Maximum a posterior; Huber-Markov Random Field

I. INTRODUCTION

Currently, most of the widely adopted video compression systems are based on the Discrete Cosine Transform (DCT). With the block-based DCT (BDCT) video coding techniques, such as those used in the H.264, H.263, MPEG-4 and MPEG-2 standards, the amount of data required for a video sequence can be significantly reduced for transmission or storage. This, however, may introduce visible coding artifacts [1], such as blocking and ringing. These artifacts degrade the visual quality of the video and need to be removed and suppressed in order to provide quality video applications.

Over the years, significant efforts have been made to reduce the annoying blocking and ringing artifacts. Generally, there are two approaches to digital video coding artifacts reduction. One approach is to implement a loop-filter in the encoder, such as the ITU-T H.261 or H.264 loop-filter. The other is to implement a post-filter in the decoder. In the post-filtering approach, iterative post-filtering scheme forms a special category and includes the projection onto convex sets (POCS) [2] [3] and the maximum a posterior (MAP) [4] [5] methods. The latter method is mainly derived from Bayesian estimation, which uses a Huber-Markov random field (HMRF) as a prior model for the estimation [4]. The HMRF model has been considered to be more robust than the Gaussian-Markov random field (GMRF) model since the GMRF model is unable to exclude outliers, which eventually causes blurring of the image details with linear low-pass filtering characteristics. In contrast, the HMRF model achieves a compromise between coding artifact reduction and edge preservation thanks to its lighter penalty to the large discontinuities in the images [6].

However, the above HMRF model employed in the existing MAP post-filter is not robust enough to restrict those extra outliers, i.e. the HMRF model still blurs edge features, especially the sharp or high contrast edges. The reason will be discussed in detail shortly. Further work is required to focus on preserving the edge features while removing the coding artifacts.

This paper proposed a modified HMRF prior model to better preserve edges and suppress blocking and ringing artifacts. Experimental results demonstrate significant improvements, in terms of the visual quality and the PSNR, in the proposed filter over the Huber-MAP filter and the H.264 post-filter.

II. THE HUBER-MAP POST FILTER FOR CODING ARTIFACT REDUCTION

A. Maximum a Posterior (MAP) Estimate

MAP is one of the Bayesian Inferences. It aims at the optimal approximation of the original event or the evidence, given the observed event. Hence, it has been applied in the image or video post processing for its statistical nature. Here is the mathematical expression of the MAP estimation [7],

$$\hat{\theta}_{MAP} = \arg \max_{\theta} f_{\theta}(\theta \mid y)$$

$$= \arg \max_{\theta} \left[ f_{y \theta}(y \mid \theta) f_{\theta}(\theta) \right]$$

(1)

where $\Theta$ is the source parameter space of the random process, $\theta$ is the source parameter vector, $Y$ is the observed parameter space, $y$ is the observed parameter vector, $\hat{\theta}$ is the MAP estimate. $f_{\theta}(\theta \mid y)$ is the posterior which means the probability of the evidence $\theta$ with the condition of $y$. $f_{y \theta}(y \mid \theta)$ is the likelihood that means the probability of $y$ under the condition of $\theta$. $f_{\theta}(\theta)$ is the prior meaning the probability of the source parameter $\theta$. Moreover, $f_{\theta}(y)$ is...
supposed to be 1 and abbreviated, since all the probabilities of the observed video exist.

Generally, in the video post processing, it seems impossible to reconstruct the video sequence as the same as the original, due to the loss of information during the compression process. Nevertheless, the MAP post-filtering method presents a closer approximation of the original using the received compressed video. Within the MAP post-filtering framework, the video compression process is treated as the likelihood which is not discussed in this paper, while the video source is considered as the prior. In recent years, the HMRF is regarded as a better model than the GMRF for its edge preservation performance [6]. However, such edge protection feature is yet to be more robust.

B. MAP Post Filtering Using Huber Prior Model

If the temporal factor is not considered, each spatial correlated video frame can be treated as a Markov random field (MRF) model. Some non-Gaussian type MRF model has been argued as a good image model for smooth regions and discontinuities. Such a prior model is referred to as the HMRF model [4]. Its mathematical expression is given by,

$$p(\theta) = \frac{1}{G} \exp \left\{ -\lambda \sum_{c} \rho_{\downarrow}(d'_{c}, \theta) \right\}$$  \hspace{1cm} (2)

where $G$ is a normalizing constant, $\lambda$ is a regularization parameter greater than zero, $c$ and $C$, respectively, represents a local group of pixels and the set of all these local groups, $d'_{c}$, is a collection of the linear operators, generally a subtraction between the pixel value which is being estimated and its neighbor pixels. $\rho_{\downarrow}(\cdot)$ is the Huber $\rho$–function [8],

$$\rho_{\downarrow}(u) = \begin{cases} u^2, & |u| \leq T \\ T^2 + 2T |u - T|, & |u| > T \end{cases}$$  \hspace{1cm} (3)

The parameter $T$ is the threshold that reduces the effect when the input ‘$u$’ exceeds this threshold. Fig. 1 shows the Huber $\rho$–function and its first derivation $\psi$–function. The Huber $\psi$–function is also given as,

$$\psi_{\downarrow}(u) = \begin{cases} 2u, & |u| \leq T \\ \text{sgn}(u) \times 2T, & |u| > T \end{cases}$$  \hspace{1cm} (4)

The MAP process can be reduced to the following equation when combining (1), (2) and (3),

$$\hat{\theta} = \arg \min_{\theta} \left\{ \sum_{c} \rho_{\downarrow}(d'_{c}, \theta) \right\}$$  \hspace{1cm} (5)

$\hat{\theta}$ is a simplified estimation using Huber prior model. Since the video source and the compression process are two independent processes, the MAP is equal to maximizing the prior and maximizing the likelihood separately.

Due to the convexity of the Huber $\rho$–function, the optimization of (5) can be solved with a gradient descent method [9], see (6),

$$\theta^{(\omega+1)} = \theta^{(\omega)} - \alpha^{(\omega)} g(\theta^{(\omega)})$$  \hspace{1cm} (6)

where $\theta^{(\omega)}$ denotes the estimate in the $\omega$ iteration, $\alpha^{(\omega)}$ is a gradient step, and $g(\theta)$ is the gradient function,

$$g(\theta) = \lambda \sum_{c} d'_{c} \psi_{\downarrow}(d'_{c}, \theta)$$  \hspace{1cm} (7)

In (6), the initial condition $\theta^{(0)}$ uses the received/observed pixel data. When the difference between $\theta^{(\omega+1)}$ and $\theta^{(\omega)}$ is no larger than a pre-set threshold value, the iterative process terminates. In (7), the Huber $\psi$–function is considered as the core in the estimating process. An image frame is supposed to be locally smooth and, hence, the $\psi$–function is usually very small in smooth regions. When the blocking or ringing artifact is encountered, the $\psi$–function will be relatively large, and then the region is smoothed with such a large quantity. In other words, sharp edges will be blurred by their smooth region neighbors using the Huber prior model.

C. Improved Huber Prior Model

To avoid blurring the edges, another threshold, LT is introduced to the Huber functions to separate the true edge from the artifacts, as shown in Fig. 2. The modified Huber $\rho$– and $\psi$–functions are shown as,

$$\rho_{\downarrow,LT}(u) = \begin{cases} u^2, & |u| \leq T \\ T^2 + 2T |u - LT|, & T < |u| \leq (LT - T) \\ -(|u| - LT)^2 + 2T(LT - T), & (LT - T) < |u| \leq LT/2T(LT - T), & |u| > LT \end{cases}$$  \hspace{1cm} (8)

$$\psi_{\downarrow,LT}(u) = \begin{cases} 2u, & |u| \leq T \\ \text{sgn}(u) \times 2T, & T < |u| \leq (LT - T) \\ \text{sgn}(u) \times (-2)(|u| - LT), & (LT - T) < |u| \leq LT/0, & |u| > LT \end{cases}$$  \hspace{1cm} (9)
Here, the small threshold T has the same definition as that in the Huber functions, but the difference is the additional threshold LT that eliminates the impact of the true edges. When the difference between two pixel values is larger than LT, it will be treated as an edge. Therefore, its effect to the gradient function is limited to zero. There is a smooth transition from the expansion of the Huber curve to the top bound in the modified ρ−function, so that it can be continuously derivable. Moreover, it is allowable that the edge pixel values fluctuate around the threshold LT. That is, the artifacts within the range (-LT, LT) are suppressed, while the other information outside this range is preserved. The prior probability, the iterative estimate and the gradient can be rewritten as follows,

$$p(\theta) = \frac{1}{G} \exp \left\{ -\lambda \sum_{c \in C} p_{T,L}(d_1 \theta) \right\}$$  \hspace{1cm} (10)$$

$$\theta^{(n+1)} = \theta^{(n)} - \alpha^{(n)} g(\theta^{(n)})$$  \hspace{1cm} (11)$$

$$g(\theta) = \lambda \sum_{c \in C} d_1 \psi_{T,L}(d_1 \theta)$$  \hspace{1cm} (12)$$

From this point of view, a modified HMRF prior is proposed to model the video source. Both the smoothness and the sharpness of the image are well presented with the utilization of such a prior model.

III. EXPERIMENTAL RESULTS

In this experiment, four full D1 resolution (720x480 pixels) mono-chrome video sequences are used for comparing the above post-filters. Each sequence contains at least 172 frames or at most 260 frames, which makes the results more reliable.

Figs. 3-7 demonstrate a comparison of the 3D Animation video sequences processed with different post-processing methods. These clips are compressed using JM12.3 codec (H.264 standard codec) with the Quantization Parameter (QP) 46. While Figs. 8-12 demonstrate another comparison of the Mobile & Calendar video sequences. The QP of these encoded clips is 28. Once these video clips are decoded, they are separately processed by the H.264 post-filter, the Huber-MAP filter and the proposed filter, in comparison with the reconstructed video without post-filtering.

As seen in Figs. 6 and 7, in the 3D Animation clips, most of the blocking and ringing artifacts appearing in Fig. 4 are suppressed by the Huber-MAP filter and the proposed filter, and the black lines in the proposed filtered image look darker than those in the Huber-MAP filtered image; while the ringing artifact remains in the H.264 post-filtered image, (see Fig. 5). On the other hand, in Mobile & Calendar clips, the contours of leaves look blurry in Fig. 11, comparing with Fig. 12, where these contours look as sharp as those in Fig. 9 and 10. The original frames are also presented in Figs. 3 and 8.

Table 1 presents the peak signal to noise ratio (PSNR) of the above post-filtered sequences, and the simulation parameters, i.e., the QP, the parameter λ, and the two thresholds T and LT of the proposed filter. The Huber-MAP filter also shares the same parameter λ and the lower threshold T with the proposed filter. The result shows that, the proposed filter achieved PSNR improvements from the Huber-MAP post-filter, in sequences Mobile & Calendar with QP 28 and QP 32, and Color Characters with QP 28, about 0.4 dB, 0.7 dB, and 0.4 dB gains respectively. That is because more high contrast details of the sequences are preserved by the proposed filter. While compared with the H.264 post-filter, the PSNR performance of the proposed filter is comparable. On the other hand, in sequences 3D Animation with QP 36 and QP 46, and Suzie with QP 36, the PSNR performance of the proposed filter is just comparable with the Huber-MAP post-filter and the H.264 post-filter. That is because the video compression process discards many picture details and introduces significant artifact, the proposed filter is just similar with the Huber-MAP filter, however, it still has about 0.5 dB PSNR gain over the unfiltered sequence.

IV. CONCLUSION AND FUTURE WORK

This paper introduced modified Huber ρ− and ψ−functions to better differentiate edge features from the digital coding artifacts in compressed digital video. A performance improvement in terms of visual quality and the PSNR of the video sequences has been achieved using statistical methods in the video post-processing. With the proposed model, video details are well preserved while the blocking and ringing artifacts are visibly reduced. Further work is required to preserve fine image details which have the same order of magnitudes as those of the coding artifacts. Temporal fluctuation noise is another issue that needs to be addressed in the future. In addition, achieving a practical application of the proposed iterative post-filter requires a reduction of the computational complexity of the filter implementation.

ACKNOWLEDGMENT

The first author is a recipient of an RMIT postgraduate scholarship. He acknowledges the technical assistance and supports by L. Mei, Dr. D. Tan, Dr. D. Wu, and S. Chai at RMIT.

REFERENCES

TABLE I. THE PSNR PERFORMANCES OF MONO-CHROME VIDEOS USING DIFFERENT POST-FILTERS

<table>
<thead>
<tr>
<th>Sequence Name</th>
<th>Frames</th>
<th>Quantization Parameter</th>
<th>$\lambda$</th>
<th>$T$</th>
<th>LT</th>
<th>PSNR of Mono Chrome Video (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unfiltered</td>
</tr>
<tr>
<td>Mobile &amp; Calendar</td>
<td>260</td>
<td>28</td>
<td>0.0020</td>
<td>3</td>
<td>16</td>
<td>35.74</td>
</tr>
<tr>
<td>Mobile &amp; Calendar</td>
<td>260</td>
<td>32</td>
<td>0.0020</td>
<td>3</td>
<td>16</td>
<td>32.46</td>
</tr>
<tr>
<td>3D Animation</td>
<td>260</td>
<td>36</td>
<td>0.0040</td>
<td>3</td>
<td>26</td>
<td>35.47</td>
</tr>
<tr>
<td>3D Animation</td>
<td>260</td>
<td>46</td>
<td>0.0065</td>
<td>3</td>
<td>56</td>
<td>29.71</td>
</tr>
<tr>
<td>Color Characters</td>
<td>196</td>
<td>28</td>
<td>0.0025</td>
<td>3</td>
<td>12</td>
<td>39.39</td>
</tr>
<tr>
<td>Suzie</td>
<td>172</td>
<td>36</td>
<td>0.0025</td>
<td>3</td>
<td>16</td>
<td>35.61</td>
</tr>
</tbody>
</table>


Figure 3. Original Frame in The 3D Animation Sequence
Figure 4. Unfiltered Frame in The 3D Animation Sequence
Figure 5. H.264 Post-filtered Frame in The 3D Animation Sequence

Figure 6. Huber-MAP Filtered Frame in The 3D Animation Sequence

Figure 7. Proposed Filtered Frame in The 3D Animation Sequence

Figure 8. Original Frame in The Mobile & Calendar Sequence
Figure 9. Unfiltered Frame in The Mobile & Calendar Sequence

Figure 10. H.264 Post-filtered Frame in The Mobile & Calendar Sequence

Figure 11. Huber-MAP Filtered Frame in The Mobile & Calendar Sequence

Figure 12. Proposed Filtered Frame in The Mobile & Calendar Sequence