ADAPTIVE BILATERAL FILTERING USING SALIENCY MAP FOR DEBLOCKING LOW BIT RATE VIDEOS

Ivan Himawan, Andrew J. Zele, Dian Tjondronegoro

Queensland University of Technology
2 George Street, Brisbane, Australia
i.himawan@qut.edu.au, andrew.zele@qut.edu.au, dian@qut.edu.au

ABSTRACT

This paper proposes a novel approach to video deblocking which performs perceptually adaptive bilateral filtering by considering color, intensity, and motion features in a holistic manner. The method is based on bilateral filter which is an effective smoothing filter that preserves edges. The bilateral filter parameters are adaptive and avoid over-blurring of texture regions and at the same time eliminate blocking artefacts in the smooth region and areas of slow motion content. This is achieved by using a saliency map to control the strength of the filter for each individual point in the image based on its perceptual importance. The experimental results demonstrate that the proposed algorithm is effective in deblocking highly compressed video sequences and to avoid over-blurring of edges and textures in salient regions of image.

Index Terms— deblocking, H.264, bilateral filter, post-processing

1. INTRODUCTION

Compressing videos at low bit rates is causing visible coding artefacts and results in low quality videos which degrade viewing experience. If the allocation of bit rates in the background region is too low, the coding impairment can distract the viewer's attention. The widely adopted H.264/AVC format, a state-of-the-art video coding standard, incorporates a mandatory loop filter to reduce blocking artefacts within its encoding loop to ensure a better picture quality. An additional post-filtering out of the coding loop can be employed to improve visual quality of the client's video player.

The coding artefacts, known as the "blocking effect", is caused by quantization error in the compression process that utilize the block-based discrete cosine transform (DCT). In the DCT scheme, the image is divided into 8x8 or 16x16 blocks of pixels and each block is transformed from the spatial domain to the frequency domain. The DCT coefficients from each block are quantized, a lossy process which removes information since it is not possible to determine the exact value of the original fractional number from the rounded integer. Because blocks of pixels are treated independently and the correlation among spatially adjacent blocks is not taken into account, strong edges and discontinuities between adjacent blocks appears as a result of dividing images in blocks prior to coding. The blocking artefacts are mostly visible at low bit rates where quantization errors are greater due to coefficient truncation. The post-filtering can be employed to improve the perceptual quality of the encoded image. Since there is no extra information about the original image, any modification to the images pixels is solely based on the assumptions.

The simplest way to remove the blocking artefacts is to perform low pass filtering, but this will reduce high frequency components and produce blurry image. A more complex procedures such as in [1, 2, 3] involve nonlinear smoothing of pixels. Their methodologies consist of classification step (i.e. edge detection or different methodologies to identify image details and separate them from artificial edges) and filtering step for removing the blocking noise. The drawback of most low pass filtering techniques is that the method to identify image details are not based on perceptual importance.

Aside from pixel domain approaches, a post-processing algorithm can be applied in the DCT domain [4]. A powerful filtering technique such as projections onto convex sets (POCS) needs to satisfy two requirements [5, 6]. First, the closed convex sets which capture the properties of the image free from blocking artefacts (hence close to the original image) need to be defined. Second, the projection operators will be derived onto them so that the decoded image converges to the post-processed image, which is aimed to be free from the blocking artefacts and close to the original image. The POCS methods are not suitable for real-time application since they typically require exhaustive computational power (with many iterations) for the estimates to converge to the desirable solution.

Most of the existing deblocking methods only adapt the filters to local features such as edges and textures to preserve image details, but seldom account for the characteristics of the human visual system (HVS). Few works, such as [7] considers human visual system by integrating the HVS models of the luminance masking, spatial activity masking, and tem-

poral masking in the design of post-filter. By considering the characteristics of the HVS, the strength of the filter can be adjusted in the region where the distortion is more perceivable. For example, since the HVS has difficulties of perceiving spatial details in fast-moving regions, distortions in these regions can be tolerated to some extent and as a result appropriate filter weights should be be applied for frames with different temporal activity.

In this paper, an adaptive bilateral filtering method is proposed by considering the perceptual importance based on saliency map for the purpose of removing blocking artefacts in highly compressed videos. Section 2 presents the related works. Section 3 describes the proposed adaptive deblocking filter algorithms. The experimental results are presented and discussed in Section 4 and Section 5 respectively. Finally, conclusion is provided in Section 6.

2. RELATED WORKS

The HVS is particularly sensitive to edges [8]. Edges are used to discriminate objects from the background and establish their shape. Because of this reason, edge detection (i.e. gradient) has been typically used as the first step in the process of scene analysis.

Discontinuities between block boundary pixels are characterized by high-frequency components at the edges of blocks. These artificial edges cause distraction to the viewer because their appearance is perceivable by eyes. One method to eliminate these artificial edges is by performing low-pass filtering in smooth region. This is based on the fact that the variance of DCT quantization errors is higher at regions that contain low-frequency content such as background of an image [9]. On the other hand, signals that do contain high-frequency content such as textured regions of an image, error variance is higher for pixels near the middle than pixels near block boundaries. In low-pass filtering techniques for reducing artefacts, less smoothing is usually applied in the "true" edges of the images to avoid over-blurring the texture details.

In the proposed method, rather than considering only edges, less smoothing is applied to the salient regions which typically constitute the region of interest. The reason for this is that the HVS will direct their attention to any objects that are visually salient in videos and may pay more attention to details in such regions. However, in the regions of fast moving content, the HVS may not have sufficient time to notice spatial details because of the high motion content. Thus, distortions in such regions may be tolerated to some extent. A stronger artefacts reduction should be applied in regions with slow motion contents (i.e. the background).

The proposed technique considers salient regions to give the measure of perceptual importance rather than the use of edges obtained from the local phase coherence analysis to adaptively tuned bilateral filter parameters [10]. Motion, color (using the red-green and yellow-blue color opponent channels), and luminance features are used to construct the saliency map. The fMRI analysis of the human visual cortex shows that strong responses are obtained to chromatic contrast that modulates the red-green and blue-yellow colour opponent mechanisms [11]. Also, the proposed technique considers motion feature to capture different temporal changes within frames. While the adaptive version of the bilateral filter has been proposed in [12], their method to obtain filter parameters to avoid over-blurring texture regions and to reduce block discontinuities has not been based on perceptual importance.

3. PROPOSED DEBLOCKING FILTER ALGORITHM

The proposed method utilizes a spatially adaptive version of the bilateral filter. The bilateral filter adaptively reduce block discontinuities in smooth and "non-salient" regions and avoid over-blurring texture and "salient" regions. Figure 1 shows the framework of post-filtering using adaptive bilateral filtering.

Bilateral filter considers color or gray pixel similarities and geometric closeness of the neighboring pixels. The bilateral filter output at a pixel location $\mathbf{x}$ is calculated as follows:

$$\tilde{I}(\mathbf{x}) = \frac{1}{W} \sum_{\mathbf{y} \in S(\mathbf{x})} G_{\sigma_d}(||\mathbf{y} - \mathbf{x}||) G_{\sigma_r}(|I(\mathbf{y}) - I(\mathbf{x})|) I(\mathbf{y})$$

(1)

where $G_{\sigma_d}(||\mathbf{y}-\mathbf{x}||) = \exp^{-\frac{||\mathbf{y}-\mathbf{x}||^2}{2\sigma_d^2}}$ is a geometric closeness function and $G_{\sigma_r}(|(I(\mathbf{y})-I(\mathbf{x})|) = \exp^{-\frac{|I(\mathbf{y})-I(\mathbf{x})|^2}{2\sigma_r^2}}$ is a photometric similarity function. $S(\mathbf{x})$ is a spatial neighborhood of $\mathbf{x}$, and $W = \sum_{\mathbf{y} \in S(\mathbf{x})} G_{\sigma_d}(||\mathbf{y}-\mathbf{x}||) G_{\sigma_r}(|I(\mathbf{y})-I(\mathbf{x})|)$ is the normalization constant. There are two parameters which control the behavior of bilateral filter. The geometric spread σ_d determines the desired amount of low-pass filtering in the spatial domain. A greater value of σ_d results in more blur effect because more pixels in the neighborhood are weighted and combined together. The σ_r controls the weights in the intensity domain. Pixels with values closer to each other are weighted and combined together according to the value of σ_r. Typically, the value of σ_r is chosen to be larger than the discontinuity amount for the discontinuity to be eliminated.

3.1. Obtaining Guidance Map

The Fourier phase spectrum of an image contains crucial information about its geometry such as contours. Using the reconstruction of the phase spectrum from the quaternion Fourier Transform of an image, the underlying contours of an image can be detected. The reconstruction of the phase spectrum of an image indicates the position of "objects" if there is less periodicity or less homogeneity in vertical or horizontal orientation. This information has been exploited in [13] to detect salient objects in the image. The advantages of using
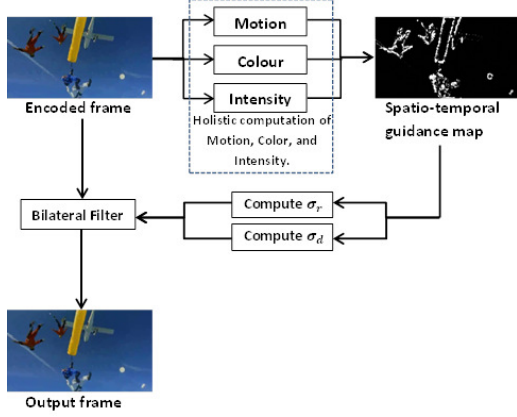
Fig. 1. The framework of post-filtering using adaptive bilateral filtering.

spectral analysis model to obtain saliency map are simple to implement and fast, although other techniques to determine saliency map may also be used.

The guidance map for bilateral filtering is obtained by creating a quaternion image which consist of four features. Using the reconstruction of the phase spectrum, the visually salient and highly textured region of the image characterized by high-frequency components can be located in the image. In summary, the proposed deblocking for video are driven by two high-level principles, as follows:

- Smoothing smooth non-salient regions and in areas with slow motion content because distortions in such regions will be causing distraction to the viewer.

- Avoid blurring the regions of the image that are characterized by high frequency components such as regions with rich textures and perceptually salient features.

A video with a T number of frames is processed frame by frame as an image $I(x, y, t)$, where x and y are the location of each pixel and $t = 1, 2, ..., T$. The RGB color frame (consist of $r(t)$, $g(t)$, and $b(t)$ which are red, green, and blue channel of $I(x, y, t)$) is decomposed into luminance $L(t) = (r(t) + g(t) + b(t))/3$ and two color opponent-component channels that are [13][1] $RG(t) = R(t) - G(t)$ and $BY(t) = B(t) - Y(t)$. The $R(t) = r(t) - \frac{(g(t)+b(t))}{2}$, $G(t) = g(t) - \frac{(r(t)+b(t))}{2}$, $B(t) = b(t) - \frac{(r(t)+g(t))}{2}$, and $Y(t) = \frac{(r(t)+g(t))}{2} - \frac{|r(t)-g(t)|}{2} - b(t)$. The motion channel are calculated by $M(t) = |L(t) - L(t - \tau)|$ where τ is the latency coefficient and set to $\tau = 3$.

The four features are represented by a quaternion image $q(t)$ which has four channels that is $q(t) = M(t) + RG(t)\mu_1 + BY(t)\mu_2 + L(t)\mu_3$, where $\mu_i, i = 1, 2, 3$, satisfies $\mu_i^2 = -1$, $\mu_1 \perp \mu_2, \mu_2 \perp \mu_3, \mu_1 \perp \mu_3, \mu_3 = \mu_1\mu_2$. The

---
[1]A variant of many opponent-color models.

$q(t)$ can be further represented in *symplectic* form, $q(t) = q_1(t) + q_2(t)\mu_2$, where $q_1(t) = M(t) + RG(t)\mu_1$, $q_2(t) = BY(t) + L(t)\mu_1$.

The frequency domain representation $Q(t)$ of quaternion image $q(t)$ is the computed by means of Quaternion Fourier Transform (QFT). In polar form, Q (t is dropped for clarity) can be represented as $Q = ||Q|| \exp^{\mu\Phi}$, where Φ is the phase spectrum of Q and μ is a unit pure quaternion. The image's phase spectrum is obtained by setting the $||Q||$ to a unity (i.e. $||Q|| = 1$). The spatio-temporal guidance map is obtained by calculating g that is the inverse Fourier transform of Q.

The saliency map can be constructed in different sizes to simulate the various view distances between the observer and the scene. The coarser resolution is obtained from a smaller size resembling the observer looking at the scene from a long distance such that the global features are emphasized whereas the fine details of features are omitted. Conversely, constructing the saliency map at larger sizes will display the fine details of local features. To highlight both local and global features, three maps which are from 1, 1/4, and 1/8 of the original frame size are combined with weights are empirically set to be 0.6, 0.3, and 0.1 respectively [14].

The contrast of the spatio-temporal guidance map g is enhanced by using the contrast-stretching transformation function $\tilde{g} = \frac{1}{1+(m/g)^E}$ to make the salient region more apparent in comparison to the non-salient region. Figure 2 shows results from the transformed spatio-temporal guidance map $\tilde{g}$ with $m = 0.05$ and $E = 5.0$, normalized to values between 0 and 1. A comparison with phase congruency image (matlab code is available at http://www.csse.uwa.edu.au/~pk/research/matlabfns) is also shown.

The shape of the filter is independently tuned for each individual point in the image based on its perceptual importance. The σ_r and the σ_d must be sufficiently large to remove blockiness in smooth regions. For regions with perceptually significance details, σ_r and σ_d are set to be sufficiently small to avoid blurring details. The σ_r is calculated for a pixel location $\mathbf{x}$,

$$\sigma_r(\tilde{g}(\mathbf{x})) = max(n_r, k_r(\sigma_{r_{max}} - \tilde{g}(\mathbf{x}))) \qquad (2)$$

where $\sigma_{r_{max}}$ is the maximum values of $\tilde{g}(\mathbf{x})$ and n_r is a minimum level of filtering and k_r is a constant parameter controlling the mapping from $\tilde{g}(\mathbf{x})$ to $\sigma_r(\tilde{g}(\mathbf{x}))$.

To calculate σ_d which is the level of spatial smoothing, the spatio-temporal guidance map is smoothed using a low pass filter to remove high spatial frequency noise. This results in a map of $\tilde{o}(\mathbf{x})$ and σ_d is calculated as,

$$\sigma_d(\tilde{o}(\mathbf{x})) = max(n_d, k_d(\sigma_{d_{max}} - \tilde{o}(\mathbf{x}))) \qquad (3)$$

where $\sigma_{d_{max}}$ is the maximum values of $\tilde{o}(\mathbf{x})$ and n_d is a minimum level of filtering and k_d is a constant parameter controlling the mapping from $\tilde{o}(\mathbf{x})$ to $\sigma_d(\tilde{o}(\mathbf{x}))$.
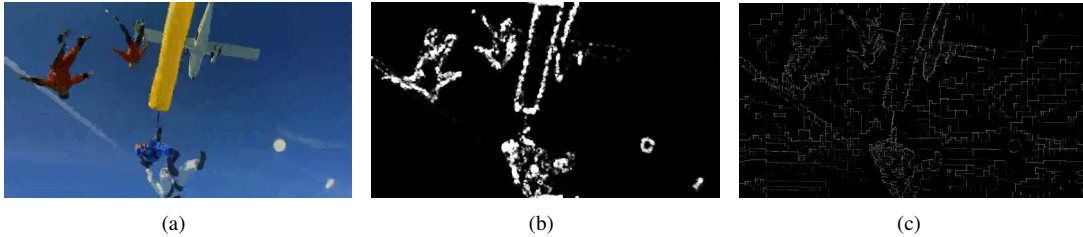
Fig. 2. (a) original image, (b) spatio-temporal guidance map, (c) phase congruency image which capture corners and edges.

The parameters k_r and k_d are analogous to the range filtering value σ_r and the geometric spread σ_d which control the behaviour of bilateral filter respectively. Given a spatio-temporal guidance map, the final value of k_r and k_d (obtained from Equation 2 and 3 respectively) will be perceptually-tuned to the changes in motion, color (using the red-green and yellow-blue color opponent channels), and luminance of the image.

4. EXPERIMENTS

The peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index[2] were used to quantify the performance of the proposed method. Because the blocking artefacts results in the abrupt changes of pixels across boundaries, a specific type of metric is required to measure such distortions. In addition to PSNR and SSIM, the $MSDS_t$ [15] (calculated on 16x16 block) was used to measure changes of intensity gradient of the pixels close to boundary of two blocks. For coarsely quantized blocks, a difference in the intensity gradient across the block boundary is expected to be larger than blocks which are not coarsely quantized.

Four video clips (labeled here as video AR, CR, BW, and DP and available at http://thediemproject.wordpress.com) with a total number of 50 frames and original bitrate of approximately 1400kbps were used as test videos. To evaluate the performance of deblocking filter for low quality videos, the test videos were encoded with low target bit rates of 150, 200, 300kbps. The parameters k_r and k_d for the proposed adaptive bilateral filter in this experiments were chosen empirically to have value of 5 and 3 respectively. The same value of parameters which are the σ_r and σ_d for the fixed bilateral filter were chosen to have also value of 5 and 3 respectively. The parameters n_r and n_d for the minimum level of filtering were set to 0.05.

Table 1 shows PSNR, SSIM, and $MSDS_t$ results from the encoded videos when the in-loop deblocking filter is disabled, using DCT-based deblocking method [4] (it is chosen to represent DCT domain based approaches), fixed bilateral filter, and proposed adaptive bilateral filter respectively. In the last row of Table 1, PSNR, SSIM, and $MSDS_t$ results

[2]available at http://nma.web.nitech.ac.jp/fukushima/opencv/ssim/ssim-e.html. SSIM is computed in RGB color channels and average the values.

are shown for videos processed with proposed techniques after the default in-loop deblocking filter is enabled.

The bilateral filter is popular because of simplicity and fast algorithms and naive implementation of the bilateral filter can be extremely slow. To speed up the computation of bilateral filter, the separable bilateral filter which approximate the 2D bilateral filter by two 1D bilateral filters were chosen [16] due to its ease of implementation (achieving about 4x speed up in the computation). The results from separable kernel method is presented in Table 2. Apart from separable filter kernel, other fast bilateral filter implementation exist such as the use of square box function for the spatial weight [17] and bilateral grid [18]. While these methods may run faster than the separable kernel method, the color image may need to be processed separately for each channel or in the case of bilateral grid, it requires 5D convolution, which adds an additional computation time to run. The issue of other fast implementation of adaptive bilateral filter will be left for further works.

For a frame at resolution WxH with an image dimension d, the total time complexity of the proposed adaptive bilateral filtering is $O(WHlog_2(WH) + zmdWH)$ where z is the total number of required real addition/multiplication operations [13], m is the size of filtering kernel for separable bilateral filter.

5. DISCUSSION

PSNR and SSIM results from the proposed technique show improvement over H.264/AVC coded images with in-loop deblocking filter disabled. It can be seen that across the tested bit rates, greater improvement is more obvious at low bit rates. When the in-loop filter is enabled, there is small improvement of SSIM values over coded images with in-loop filter enabled but not for the PSNR values. In particular, the reduction in $MSDS_t$ values is due to less blockiness after the videos are processed using the proposed technique. The DCT-based deblocking [4] shows an increase of PSNR values only at 150kbps, and an increase of SSIM values only at 150 and 200kbps. However, the $MSDS_t$ values reduce a little which indicate that sharpness along block edges still be preserved. The separable kernel method of bilateral filter shows improvement in terms of PSNR and SSIM over coded images without in-loop filter enabled, albeit less improvement than the brute

Table 1. *The PSNR, SSIM, and $MSDS_t$ (lower values is better for $MSDS_t$) results of H.264/AVC encoded videos using default in-loop deblocking filter, DCT-based deblocking, standard bilateral filtering, and the proposed adaptive bilateral filtering for various bit rates. The results are presented as an average from four test videos.*

	150 kbps			200 kbps			300 kbps		
	PSNR	SSIM	$MSDS_t$	PSNR	SSIM	$MSDS_t$	PSNR	SSIM	$MSDS_t$
no deblocking	28.56	0.784	199.3	30.09	0.826	197.3	31.98	0.871	186.3
in-loop deblocking enabled	29.01	0.802	37.4	30.58	0.841	46.1	32.53	0.885	56.2
DCT-based [4]	28.64	0.793	172.8	30.09	0.830	176.1	31.85	0.870	170.9
Fixed Bilateral	28.66	0.802	1.1	29.79	0.834	1.1	30.86	0.866	1.2
Proposed Bilateral	**28.80**	**0.801**	**117.8**	**30.30**	**0.839**	**129.7**	**32.15**	**0.879**	**134.9**
Proposed Bilateral (after in-loop)	**29.02**	**0.807**	**25.2**	**30.54**	**0.844**	**32.4**	**32.44**	**0.886**	**42.9**

Table 2. *The PSNR, SSIM, and $MSDS_t$ (lower values is better for $MSDS_t$) results of the proposed adaptive bilateral filtering using separable kernel implementation for various bit rates. The results are presented as an average from four test videos.*

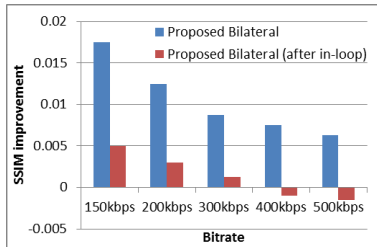|	150 kbps			200 kbps			300 kbps		
	PSNR	SSIM	$MSDS_t$	PSNR	SSIM	$MSDS_t$	PSNR	SSIM	$MSDS_t$
Proposed Bilateral	**28.70**	**0.794**	**149.5**	**30.23**	**0.833**	**155.5**	**32.12**	**0.876**	**154.2**
Proposed Bilateral (after in-loop)	**29.02**	**0.805**	**29.3**	**30.58**	**0.843**	**37.0**	**32.51**	**0.886**	**47.6**

Fig. 3. The tendency of performance versus bitrates from the proposed adaptive bilateral filter.

force implementation.

Figure 3.a. shows a frame (in-loop filter is disabled) from video AR. Figure 3.b. and Figure 3.c. show same post-processed frames using fixed bilateral filter and proposed adaptive bilateral filter respectively. The frame and the corresponding post-processed frame for video CR are shown in Figure 3.d., 3.e., and 3.f respectively. It can be seen that the blocking artefacts are eliminated but the texture and edge details in salient regions are not blurred using the proposed method (e.g. see airplane and parachuting people in video AR and see phone cable and window frames in video CR). A significant reduction of blockiness which is reflected from a very low $MSDS_t$ value, is obtained using bilateral filter with fixed parameters at the cost of over-blurring the texture and edge details. The separable kernel method for improving the speed of bilateral filter may still avoid over-blurring the texture and edge details, but the overall quality improvements are not as large as brute force version for the tested videos.

The proposed method is evaluated for bitrates up to 500kbps to see the tendecy of performance versus bitrate. The SSIM improvements are calculated for deblocking videos coded both with in-loop deblocking filter disabled and when it is enabled. From Figure 3, the performance of the proposed method is only effective for low bit rate videos. When the in-loop deblocking filter is enabled, the proposed method does not improve the SSIM index for videos coded above 300kbps.

6. CONCLUSION

This paper presents a novel approach to video deblocking using adaptive bilateral filtering based on a saliency detection model. The model considers color, intensity, and temporal changes between frames to give the measure of perceptual significance of image region. By adapting the parameters of bilateral filter based on saliency map, each pixel in the image can be adaptively tuned for improved perceptual quality. Results show that the proposed algorithm improves the objective quality of highly compressed video sequences (i.e. H.264/AVC format with in-loop filtering is disabled). Moreover, over- blurring of edges and textures in salient regions of image are avoided. Ongoing research is investigating the applications of the proposed method for video denoising and other perceptual filtering operations.
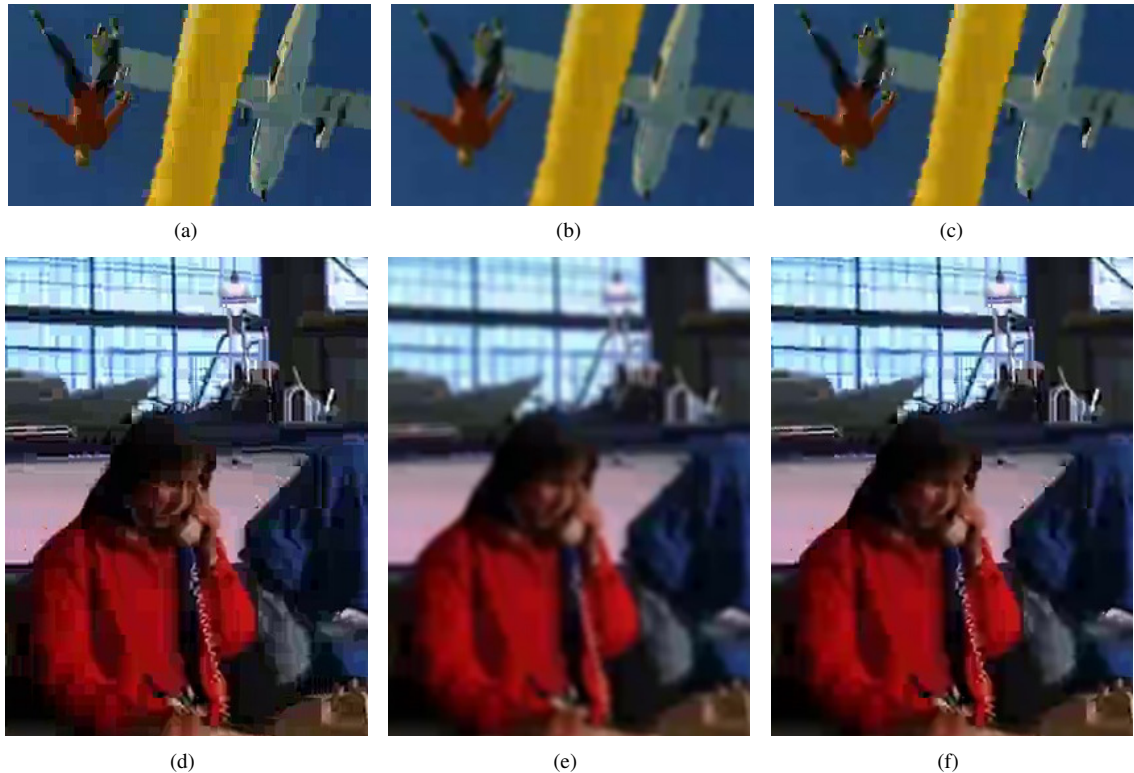
Acknowledgements

Fig. 4. Frames from Video AR (top) and Video CR (bottom). (a) and (d) are original images (in-loop deblocking filter is disabled at 200kbps), (b) and (e) are results using standard bilateral filter, (c) and (f) are results using adaptive bilateral filter.

7. REFERENCES

[1] J. Chou, M. Crouse, and K. Ramchandran, "A simple algorithm for removing blocking artifacts in block-transform coded images," *IEEE Signal Process. Letters*, vol. 5, no. 2, pp. 33–35, 1998.

[2] Y.L. Lee, H.C. Kim, and H.W. Park, "Blocking effect reduction of JPEG images by signal adaptive filtering," *IEEE Trans. on Image Process.*, vol. 7, no. 2, pp. 229–234, 1998.

[3] A.S. Al-Fohoum and A.M. Reza, "Combined edge crispiness and statistical differencing for deblocking JPEG compressed images," *IEEE Trans. on Image Process.*, vol. 10, no. 9, pp. 1288–1298, 2001.

[4] Y. Luo and R.K. Ward, "Removing the blocking artifacts of block-based DCT compressed images," *IEEE Trans. on Image Process.*, vol. 12, pp. 838–842, 2003.

[5] Y. Yang, N.P. Galatsanos, and A.K. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed images," *IEEE Trans. on Image Process.*, vol. 4, no. 7, pp. 896–908, 1995.

[6] H. Paek, Rin-Chul Kim, and Sang-Uk Lee, "On the POCS-based postprocessing technique to reduce the blocking artifacts in transform coded images," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 8, no. 3, pp. 358–367, 1998.

[7] S. Wan, M. Mrak, and E. Izquierdo, "Perceptually adaptive joint deringing-deblocking filtering for scalable video coding," in *Proc. of the Int. Conf. on Mobile Multimedia Comm.*, 2006, pp. 1–5.

[8] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *Journal of Physiology*, vol. 195, pp. 215–243, 1968.

[9] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 15, no. 1, pp. 27–38, 2005.

[10] A. Wong, "Adaptive bilateral filtering of image signals using local phase characteristics," *Signal Processing*, vol. 88, no. 6, pp. 1615 – 1619, 2008.

[11] S. Engel, X. Zhang, and B. Wandell, "Colour tuning in human visual cortex measured with functional magnetic resonance imaging," *Nature*, vol. 388, pp. 68–71, 1997.

[12] M. Zhang and B. K. Gunturk, "Compression artifact reduction with adaptive bilateral filtering," in *SPIE Electronic Imaging Conf.*, 2009, vol. 7257.

[13] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. on Image Process.*, vol. 19, pp. 185–198, 2010.

[14] W. Kim, C. Jung, and C. Kim, "Spatiotemporal saliency detection and its applications in static and dynamic scenes," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 21, no. 4, pp. 446 –456, 2011.

[15] G.A. Triantafyllidis, D. Tzovaras, and M.G. Strintzis, "Blocking artifact detection and reduction in compressed data," *IEEE Trans. on Circuits and Syst. for Video Technol.*, vol. 12, no. 10, pp. 877–890, 2002.

[16] T. Q. Pham and L. J. van Vliet, "Separable bilateral filtering for fast video preprocessing," in *IEEE International Conf. on Multimedia&Expo*, 2005, pp. 1–4.

[17] B. Weiss, "Fast median and bilateral filtering," in *ACM SIGGRAPH*, 2006, pp. 519–526.

[18] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," *Int. Journal of Computer Vision*, vol. 81, pp. 24–52, 2009.