SPATIALLY ADAPTIVE FILTERING FOR REGISTRATION ARTIFACT REMOVAL IN HDR VIDEO

Stephen Mangiat and Jerry Gibson

Electrical and Computer Engineering Department
University of California, Santa Barbara, CA 93106
{smangiat, gibson}@ece.ucsb.edu

ABSTRACT

One method to extend the dynamic range of video captured with inexpensive cameras is to alternate the exposure time between frames and combine the information in adjacent frames using post-processing. This method requires no hardware modification, yet traditionally there is a quality tradeoff. Dynamic range expansion corresponds to an increased number of saturated pixels in individual frames, which along with occlusions contributes to registration artifacts. Therefore, we describe a "High Dynamic Range (HDR) Filter" that can mitigate these artifacts to produce a pleasing HDR video without exact frame registration. This filter builds upon the bilateral filter to smooth frames while maintaining important edges. Additionally, the filter strength locally adapts to corresponding motion vectors. Since regions with poor registration generally correspond to higher motion, smoothing here can reduce artifacts without degrading perceptual quality. Results show a significant improvement for HDR videos with fast local motion within saturated regions.

***Index Terms*—** High Dynamic Range (HDR) Video, Bilateral Filter, Auto-Gain Control

1. INTRODUCTION

High dynamic range (HDR) video aims to accurately record scenes with brightness variations beyond the capabilities of a typical camera sensor. Auto-exposure algorithms attempt to minimize the number of saturated pixels, yet they fail to correctly expose the entire frame. This issue is especially prevalent on handheld devices, which may have video conferencing capabilities. The mobility of the device means that the user is often exposed to extreme lighting conditions. Consequently, poorly lit faces detract from the user experience and contribute to awkward interaction.

Most HDR methods include some way to obtain multiple exposures of a scene, whether using specialized hardware or software. In this way, the bright regions are captured in the shorter exposures while the dark regions are captured in the longer exposures. HDR video introduces many difficulties compared to still imagery due to scene motion, which will appear as ghosting. To eliminate ghosting, [1] used alternating exposures and an "HDR stitching" method utilizing gradient-based optical flow to register adjacent frames. Alternatively, [2] used block-based motion estimation for frame registration and a bilateral filtering step following tone mapping for artifact removal. Both methods were susceptible to registration artifacts when saturation occurred on moving objects in the scene.

The goal here is to outline a new filtering method for registration artifact removal in an HDR video frame created from alternating exposures. Building upon the cross-bilateral and dual-bilateral filters, this "HDR Filter" utilizes edges in the low dynamic range frame, edges in the unfiltered HDR frame, and associated motion vectors to adaptively smooth poorly registered pixels.

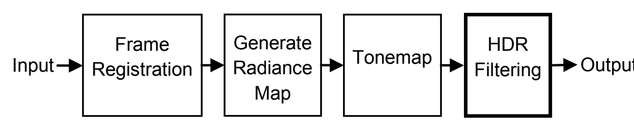


Fig. 1. HDR Video Process

The main steps of our HDR video process are shown in Fig. 1. The input is a sequence of frames captured at alternating short and long exposures using a dual-exposure control algorithm described in Sec. 2. An overview of our frame registration and HDR imaging procedure is provided in Sec. 3. Following an introduction to bilateral filtering in Sec. 4, the HDR filtering technique is defined in Sec. 5. Finally, we present and discuss sample results from a processed output in Sec. 6.

2. REAL-TIME DUAL-EXPOSURE CONTROL

Typical video cameras use a single exposure setting that is adapted according to the statistics of each frame. Since the dynamic range of the scene is usually much larger than that of the camera, this auto-gain control algorithm attempts to minimize the number of saturated pixels. In order to extend this dynamic range, we adapt two exposures (short and long) in real-time, as in [1]. The camera cycles between these two shutter speeds in alternating frames.

Our "dual-exposure" algorithm calculates new shutter speeds every fourth frame. This process is simplified compared to [1], as the camera response function and intensity histograms are unneeded. We also adapt the short and long exposure times independently, without limits on their ratio. Our goal here is to maximize the long exposure and minimize the short exposure, thus maximizing the dynamic range expansion, while maintaining enough non-saturated pixels to adequately register adjacent frames.

At the start of a video capture, both the short and long exposure times are initialized to the same value using standard auto-exposure. This is useful for evaluating the quality of the HDR output with respect to a low-dynamic range video, as shown in Figs. 2 (c) and (d). Once the HDR settings are engaged, the two exposures will separate according to the dynamic range of the scene. In each set of four frames, the first two frames are downsampled and converted to greyscale. For the short exposure, we count the number of underexposed (intensity$<$10) pixels, N_u. Similarly, the number of overexposed (intensity$>$230) pixels, N_o, is determined for the long exposure.

Once N_u is known for the most recent short exposure, we update the short exposure time as follows. If $N_u/N < 20\%$, i.e. fewer than 20% of the total number of pixels (N) are underexposed, then the short exposure time is *decreased* according to a schedule adapted from [1]. On the other hand, if $N_u/N > 30\%$, then the short exposure must be *increased* in a similar manner. Finally, if N_u is between 20% and 30% then no change is made. In this way, the short exposure time is kept as low as possible while maintaining enough non-saturated pixels for registration. The long exposure is updated in the same way using N_o. Figures 2 (a) and (b) show a short and long exposure captured using this dual-exposure algorithm.

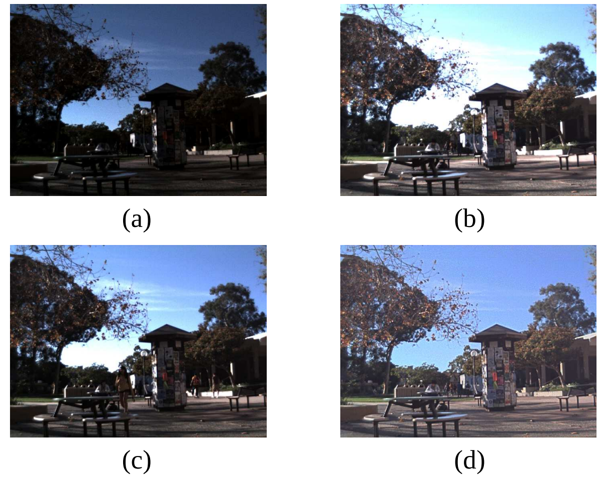


Fig. 2. Dual-Exposure Control: (a) Short Exposure (b) Long Exposure (c) Standard Auto-Exposure: Saturation causes a white sky and shadows obscure details (d) HDR Output: Enhanced colors and local contrast, without saturation (Images best viewed in color)

3. HDR IMAGING

Given a sequence of alternating exposures provided by the dual-exposure algorithm, the task is to utilize neighboring frames to predict a second exposure for each time instant. Ideally, this prediction should represent exactly the same scene as the current frame, though this is impossible due to occlusions and non-overlapping regions. Still, frame registration provides useful results.

Details of our frame registration method are found in [2]. The first step is to boost the short exposures to match the long exposures using the camera response function, which can be calculated using a number of methods such as [3]. Once adjacent frames are at approximately the same global brightness, we perform block-based motion estimation (ME) between the current frame and both the previous and next frames.

The best matches on a block-by-block basis are then chosen to generate a differently exposed prediction for the current frame. Since saturated regions in the current frame are unusable for matching, we replace the predictions for saturated blocks using bi-directional ME calculated directly between the previous and next frames. Additionally, we refine the prediction on a pixel-wise basis by replacing pixels that are too bright to appear underexposed in the current frame or too dark to appear overexposed in the current frame [2].

Following registration, radiances given by the current frame and predicted frame are combined to form a high dynamic range radiance map using the camera response function [3]. The estimated radiance is a weighted average of the two exposures, while saturated pixels are ignored. Since the predicted radiance for a given pixel may be incorrect, [1] weights the predicted radiance less if the radiance disparity between the two exposures is large. However, noise in the current frame also contributes to this disparity. Using only the radiance predicted by the current frame might also lead to flickering artifacts in the output video.

We instead use a simple hat function to weight radiances higher when given by midrange pixel values [3]. The effect of noise and flickering is reduced, though the result is an HDR radiance map that may still be vulnerable to blocking and other artifacts where the predicted radiance is poor. We therefore address these artifacts using filtering, following a local tone mapping step that maps the HDR radiances back into displayable range for low dynamic range media [4].

4. BILATERAL FILTERING

An image filtered by Gaussian convolution is given by [5]

$$\mathrm{GC}[I]_{\mathbf{p}} = \sum_{\mathbf{q}\in\mathcal{S}} G_\sigma(||\mathbf{p}-\mathbf{q}||) I_{\mathbf{q}}, \quad (1)$$

where $\mathcal{S}$ is a neighborhood of pixels about $\mathbf{p}$, $I_{\mathbf{q}}$ is the pixel value at $\mathbf{q}$, and $G_\sigma(x)$ denotes the 2D Gaussian kernel

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (2)$$

Similarly, the bilateral filter replaces each pixel with a weighted average of its neighbors. However, the weight of each neighbor is defined not only by its spatial distance to the center, but by its difference in value. In this way, edges are preserved while smoothing. The bilateral filter is thus defined as

$$\mathrm{BF}[I]_{\mathbf{p}} = \frac{1}{W_{\mathbf{p}}}\sum_{\mathbf{q}\in\mathcal{S}} G_{\sigma_s}(||\mathbf{p}-\mathbf{q}||) G_{\sigma_r}(|I_{\mathbf{p}}-I_{\mathbf{q}}|) I_{\mathbf{q}}, \quad (3)$$

where σ_s is the standard deviation in the spatial domain, σ_r is the standard deviation in the range domain, and $W_{\mathbf{p}}$ is the sum of all weights used for normalization [5].

The *cross*-bilateral filter is a variant allowing color information to be smoothed according to the edges of a second image [5]. This is especially useful for reducing registration artifacts in an HDR frame, as the HDR color information can be smoothed according to the edges in the low dynamic range (LDR) input frame [2].

Still, it is undesirable to remove all high frequency content, as simply filtering the entire image might produce. To counteract this in [2], the filtered pixels were only used in the output frame when the difference between the unfiltered and filtered versions was very large (according to perceptual color distance and structural similarity metrics). Additionally, the cross-bilateral filter is only useful where there is edge information in the input frame, so it cannot be used in saturated regions since it will then simply act as a spatial Gaussian and excessive smoothing will be perceived as flickering. Consequently, [2] did not filter pixels in saturated regions, and passes noticeable artifacts for fast moving objects containing saturated pixels.

5. HDR FILTERING

The HDR filter addresses the limitations of the bilateral filtering method in [2] by filtering both non-saturated and saturated regions. This is accomplished by observing that faster moving objects may be smoothed more without significant degradation of perceptual quality. Furthermore, it is no longer necessary to apply similarity metrics between the unfiltered and filtered frames, as the final HDR output is simply the output of the filter.

To start, we first note that the filter should only smooth regions with registration artifacts. Though there is no ground truth knowledge of where these artifacts are located, the filter strength can locally adapt to the *likelihood* of artifacts. This adaptation is an adjustment of filter standard deviation in the range domain, σ_r. For instance, if the likelihood of artifacts is higher, then σ_r is increased for a smoother output.

To measure the likelihood of artifacts, we use the largest motion vector length in the neighborhood of each pixel. The faster an object moves, the more likely there will be registration errors. Furthermore, blurring appears more natural for fast moving objects. Therefore we define a spatially adaptive standard deviation

$$\sigma_r(\mathbf{p}) = \alpha \max_{\mathbf{q}\in\mathcal{S}}(||\mathrm{MV}_{\mathbf{q}}||), \quad (4)$$

where $\mathcal{S}$ is a neighborhood around pixel $\mathbf{p}$, and α is a constant used to adjust the amount of smoothing. Plugging $\sigma_r(\mathbf{p})$ into Eq. (3) yields a bilateral filter whose strength adapts to the likelihood of registration errors. However, the adjustment of σ_r still has no effect over saturated regions in the input frame since there is no edge information.

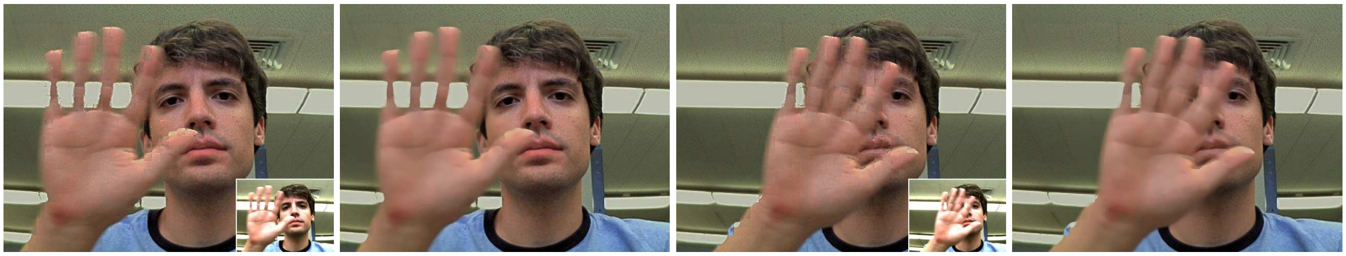
To make the filter useful in both saturated and non-saturated regions, we also utilize the edges of the unfiltered HDR frame in saturated regions. Here, the filter acts as a *dual*-bilateral filter [5], since the weight of each neighboring pixel now depends on the edges of two images (the LDR input frame and unfiltered HDR frame). We now define σ_I as the adaptive standard deviation in the range domain for edges in the LDR input frame (I), and similarly σ_J for the edges in the unfiltered HDR frame (J). The final HDR filter is defined as

$$\mathrm{HDR}[I,J]_{\mathbf{p}} = \frac{1}{W_{\mathbf{p}}}\sum_{\mathbf{q}\in\mathcal{S}} G_{\sigma_s}(\cdot) G_{\sigma_I}(\cdot) G_{\sigma_J}(\cdot) J_{\mathbf{q}}, \quad (5)$$

where

$$G_{\sigma_J}(\cdot) = \begin{cases} 1, \text{ if } 10 < I_{\mathbf{p}} < 230 \\ G_{\sigma_J(\mathbf{p})}(|J_{\mathbf{p}} - J_{\mathbf{q}}|), \text{ elsewhere.} \end{cases} \quad (6)$$

Thus, if the current pixel is not saturated in the input frame, only edges in the input frame are preserved when filtering colors in the HDR frame. Alternatively, if the current pixel is saturated in the current frame, edges in the HDR frame may also be preserved. Still, not all edges in the HDR frame should be preserved in the saturated regions, as then no artifacts would be removed. This is controlled by adjusting the α factor for $\sigma_J(\mathbf{p})$, as shown in Eq. (4). This method is effective in removing artifacts within regions of contiguous color, while preserving very strong edges between differently colored regions.



(a) Unfiltered (b) Filtered (c) Unfiltered (d) Filtered

Fig. 3. HDR Filtering Results: Pixels saturate across the waving hand in the input frames (shown as insets), leading to artifacts (notably on the thumb and between fingers) in the unfiltered HDR frames. The proposed HDR filter smooths artifacts across the entire hand and face, even where there is pixel saturation and no edge information in the current frame.

6. RESULTS

To test the performance of the HDR filter, we processed several HDR sequences of indoor and outdoor scenes[1]. Sample frames for a video exhibiting fast motion are shown in Fig. 3. In the images shown, α in Eq. (4) for σ_I is set to 0.15 and α for σ_J is set to 1. This means that edges in the current frame are more important, while only stronger edges in the unfiltered HDR frame are preserved within saturated regions.

In this video sequence, a hand waves across the screen, generating fast local motion and motion blur. The low dynamic range input frames shown in the insets of Fig. 3 (a) and (c) also show that much of the hand is saturated, leading to significant registration artifacts in the unfiltered HDR frame. The methods described in [1] and [2] cannot remove these types of artifacts since there is no valid edge information available in the input frame.

As shown in Fig. 3 (b) and (d), the HDR filter is able to smooth the blocky regions on the user's hand because the underlying motion vectors are large. Furthermore, the erroneous pixels surrounding the fingers are removed. Pixels neighboring moving objects may be assigned a zero motion vector, even if they belong to the moving object, due to the use of block-based motion vectors. This is accounted for here by using the maximum of all neighboring motion vectors, as seen in Eq. (4).

Low motion regions such as the user's face exhibit very little smoothing, so there is no perceptual degradation or flickering introduced. However, the quality depends on the accuracy of the motion vectors for stationary regions, so it is important to first check all blocks for zero motion. If this check is not made and stationary objects are falsely assigned non-zero motion vectors (perhaps due to a repeating pattern), then flickering might appear. Future work might strengthen the filter near the edges created by block boundaries. The direct implementation is also computationally complex, so future work might also investigate a fast approximation of the HDR filter, similar to the fast bilateral filter.

[1] For videos, please visit http://vivonets.ece.ucsb.edu/HDR.html

7. CONCLUSIONS

We have outlined an effective dual-exposure algorithm, HDR post-processing system, and a new filtering method for the removal of registration artifacts from HDR video. This HDR filter acts as both a cross-bilateral and dual-bilateral filter with adaptive filter strength according to underlying motion vectors, so regions of fast motion are heavily smoothed. This is advantageous in that artifacts are more likely to occur within faster motion regions, where smoothing is less noticeable perceptually. Results show a significant improvement for moving objects that are saturated in the low dynamic range input frame. Future work may address complexity reduction techniques.

8. REFERENCES

[1] S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," in *ACM SIGGRAPH*, New York, NY, USA, 2003, pp. 319–325.

[2] S. Mangiat and J. Gibson, "High dynamic range video with ghost removal," in *SPIE Optical Engineering & Applications*, 2010.

[3] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *SIGGRAPH '97*, New York, NY, USA, 1997, pp. 369–378, ACM Press/Addison-Wesley Publishing Co.

[4] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. on Graphics*, vol. 21, no. 3, pp. 267–276, 2002.

[5] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–73, 2009.