MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
13-27

High Dynamic Range Video through Fusion of Exposure-Controlled Frames

Seung-Jun Youm, Won-Ho Cho, Ki-Sang Hong
Department of Electrical and Computer Engineering
Pohang University of Science and Technology
{sjyoum, ellescho, hongks}@postech.ac.kr

Abstract

In this paper, we present a method for generating high dynamic range (HDR) video through fusion of exposure-controlled frames in a stationary video camera system. Frames captured by commercial video cameras have a limit dynamic range, even if we use auto-exposure control (AEC), and these frames often show saturation areas which may degrade the image quality. To overcome these, in the video sequence alternating short and long exposures, we combine neighboring frames using an image fusion approach, since images under different exposures have different visibility in a scene. When we fuse exposure frames, there are two considerations; controlling exposures and handling moving objects. For the first, we control short and long exposures with the AEC approach. For the other, we adjust fusion masks to reduce ghosting artifacts by moving objects. Our method has the advantage of automatically selecting proper exposure times to make outputs more visible than fixed exposure ones. It also reduces motion blur distortion caused by moving objects in output frames, by using the mask adjustment method which is simpler and faster than frame registration methods. To verify the proposed method, we show experimental results using a real scene.

1 Introduction

The real world scene contains a wide dynamic range of illuminance(radiance) values(up to 1:500,000). The brightness captured by commercial digital cameras or camcorders(using CCD or CMOS sensors), is generally 8bit-levels (1:256). Because these low dynamic range devices cannot span several orders of magnitude in the scene and cannot map all illuminance values to 256 pixel values as shown in Fig.1, the captured images often show over- or under-saturated regions as seen in Fig.2. To solve this problem, many researchers have studied HDR imaging or the HDR video algorithm with multiple exposures.

One of those HDR methods is the radiance map approach. It obtains a full radiance map of the scene using multi-exposure images[1, 2] and then compresses the map for a display using a tone-mapping method[5, 6, 7]. For video sequences, Yourganov[3] alters the shutter speeds and combines two images into one which has a higher dynamic range. However, the moving object problem, as well as an exposure control, is not handled. Kang *et al.*[8] obtains the source video sequence alternating long and short exposure frames, and then generates HDR video frames from neighboring three exposure frames through HDR stitching and temporal tone-mapping. However, the exposure control algorithm is not given concretely and it takes a long computation time to register frames with different exposures in HDR stitching process.

Another HDR method, the image fusion approach [4, 11], is simpler and faster than the radiance map approach. But they use fixed exposures and are not concerned with the moving object problems. Recently, Youm[10] selects short and long exposure times from auto exposure time using a camera response curve and fuses these three exposure images. It also handles moving object problems using moving object regions between images with different exposures. However, an accurate camera curve is needed in order to get proper short and long exposure times and find reasonable moving object regions. Extending this approach to video sequences, this paper addresses the problem of controlling exposures and handling moving objects without a camera response curve.

Our work obtains video sequences alternating short and long exposures and then combines neighboring exposure frames, which is a similar framework to Kang *et al*.[8]. However, we specifically present exposure control algorithm based on AEC setting, which selects proper short and long exposures and combine consecutive two exposure frames using the pyramid based fusion approach.

This paper is organized as follows. Section 2 describes the overall proposed algorithm and we present each stage of our work; exposure control in section 3 and fusion of neighboring frames in section 4. In section 5 we show the experimental results. In section 6, we conclude this paper

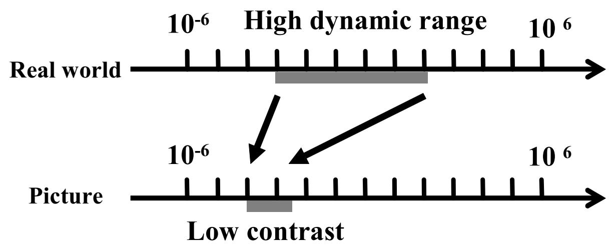


Figure 1. Mapping HDR illuminance of scenes to 8bit value of display devices makes low contrast.

Figure 2. Pictures with saturated regions

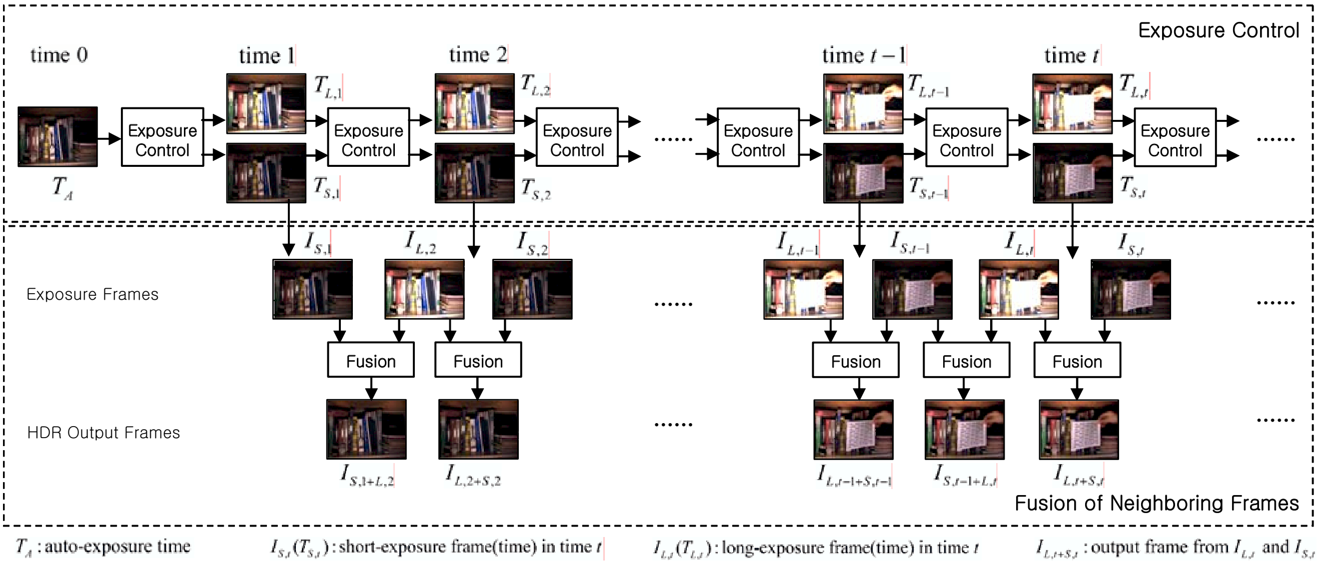


Figure 3. Overview of the proposed method

2 The Proposed HDR Video Algorithm

Our algorithm has the overall flow shown in Fig.3 and consists of two parts; exposure control during capturing frames and fusion of neighboring ones handling moving objects. Each stage has an overview as follows.

Exposure control

In this stage, we select the proper short and long exposure times through exposure control to get long and short exposure frames from a video camera. In the case of conventional video cameras, the automatic exposure (AE) system has a sampling area or the region of interest (ROI) to calculate proper exposures. Similarly, we find ROI masks for bright and dark regions of the scene, from the current long exposure frame. Using these masks, we apply the *selecting exposure* method[9] to control short and long exposure times in each frame. According to these exposure times, we obtain the long and short exposure video frames.

Fusion of neighboring frames

This stage has inputs of short and long exposure frames obtained from previous stage and outputs HDR frames by fusing neighboring exposure frames. There are two purposes we want to achieve in this process. The first is to extend the dynamic range by combining short and long exposure frames. They give us good information about the bright or dark regions of the scene and it is important to seamlessly fuse those areas obtained from these frames. We use the pyramid based approach to combine temporally adjacent frames by using the *saturation region mask* which is the mask of highlighted regions in long exposure frames. The second is to handle moving objects to remove distortions in HDR output frames. To reduce ghosting artifacts by moving objects, we adjust the *saturation region mask* by using the *change region mask* that is obtained by the change detection between previous and current long (short) exposure frames.

3 Exposure Control

Our exposure control method is similar to the AEC system. We find sampling areas or ROI masks for dark (bright) regions in a scene and select a long (short) exposure time that shows the visibility of those regions well.

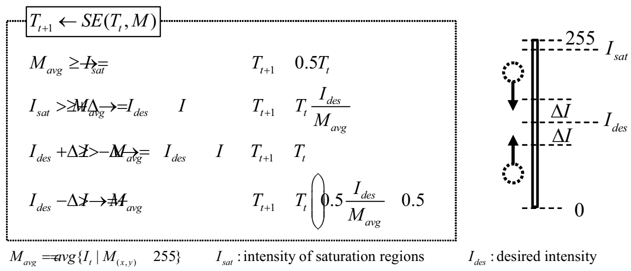


Figure 4. *Selecting exposure* algorithm (*left*) and the concept of converging the average brightness of a mask into the desired intensity (*right*)

In this paper, we use the *selecting exposure* algorithm $SE(\cdot)$, as seen in Fig.4, that finds the next exposure time T_{t+1} from both the current exposure time T_t and the ROI mask M. Using this algorithm, we make an average value of the mask M_{avg} converge into the desired intensity I_{des} ($\pm\Delta I$). For our experiments, we use $I_{sat} = 240$, $\Delta I = 5$ and $I_{des} = 127$. In each time, we obtain exposure frames as follows.

When time t = 0, the long and short exposure times $T_{L,1}$, $T_{S,1}$ are computed by

$$T_{L,1} \leftarrow SE(T_A, M_0),\ T_{S,1} \leftarrow SE(T_A, \overline{M}_0), \tag{1}$$

where $M_{0(x,y)} = \begin{cases} 255 & \text{if } I_{A(x,y)} < 127 \\ 0 & o.w. \end{cases}$, $\overline{M}_{0(x,y)} = |255 - M_{0(x,y)}|$.

According to these exposure times, we get the long and short exposure frames $I_{L,1}$, $I_{S,1}$.

When time t = 1, 2, 3, … , the long and short exposure times $T_{L,t+1}$, $T_{S,t+1}$ are computed by

$$T_{L,t+1} \leftarrow SE(T_{L,t}, M_t),\ T_{S,t+1} \leftarrow SE(T_{S,t}, \overline{M}_t), \tag{2}$$

where $M_{t(x,y)} = \begin{cases} 255 & \text{if } I_{L,t(x,y)} < 240 \\ 0 & o.w. \end{cases}$, $\overline{M}_{t(x,y)} = |255 - M_{t(x,y)}|$.
According to these exposure times, we get the long and short exposure frames $I_{L,t+1}$, $I_{S,t+1}$.

Figure 5. Comparison of the fixed exposure case (*upper row*) with the controlled one (*bottom row*) under an illuminance change

Our exposure control gets proper exposure frames under an illuminance change. In Fig.5, we show an example of short exposure frames and compare the controlled exposure case with the fixed one that is 1/2 of an auto exposure time T_A. When the time is t, we turned on an electric light, in front of the scene.

4 Fusion of Neighboring Frames

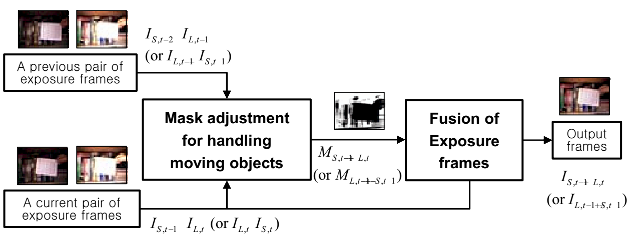


Figure 6. Flow of the frame fusion process

We get HDR output frames combining temporally adjacent frames by the pyramid based fusion method[12] with a fusion mask. To obtain this fusion mask, we need the mask adjustment process which produces saturation and change region masks of moving objects. Fig.6 shows the flow of the frame fusion process which consists of two parts; mask adjustment for handling moving objects and fusion of exposure frames. Each part is described in sections 4.1 and 4.2 for the case where the input frames are $I_{S,t-2}$, $I_{L,t-1}$, $I_{S,t-1}$ and $I_{L,t}$.

4.1 Mask adjustment for handling moving objects

We compute the saturation region mask $M_{sat,t}$ from long exposure frame $I_{L,t}$ using the curve, as shown in Fig.7. This saturation region mask $M_{sat,t}$ is needed for extending the dynamic range of output frames. On the curve in Fig.7 (left), I_{cut} is the cut-off intensity which is the threshold value for the mask $M_{sat,t}$. Experimentally, we have found that I_{cut} =127 gives good results.

We compute the change region mask $M_{S,t-1}(M_{L,t})$ which is the difference map between the previous frame $I_{S,t-2}(I_{L,t-1})$ and the current frame $I_{S,t-1}(I_{L,t})$, using Eq.(3). These change region masks $M_{S,t-1}$ and $M_{L,t}$, as shown in Fig.8, are used for removing ghosting artifacts in output frames.

$$M_{S,t-1(x,y)} = \begin{cases} 0 & \text{if } |I_{S,t-2(x,y)} - I_{S,t-1(x,y)}| > 20 \\ 255 & o.w. \end{cases}$$
$$M_{L,t(x,y)} = \begin{cases} 0 & \text{if } |I_{L,t-1(x,y)} - I_{L,t(x,y)}| > 20 \\ 255 & o.w. \end{cases} \quad (3)$$

From Eq.(4), we compute the fusion mask $M_{S,t-1+L,t}$ that is used to fuse the neighboring frames $I_{S,t-1}$ and $I_{L,t}$.

$$M_{S,t-1+L,t(x,y)} = \begin{cases} 0 & \text{if } M_{S,t-1(x,y)} = 0 \text{ or } M_{L,t(x,y)} = 0 \\ M_{sat,t(x,y)} & o.w. \end{cases} \quad (4)$$

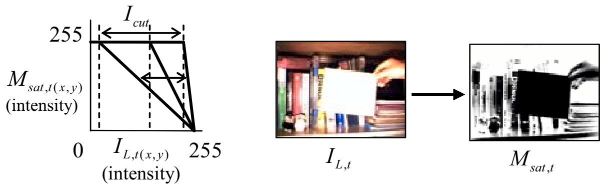


Figure 7. Curve for the saturation region mask (*left*) and obtaining the saturation region mask $M_{sat,t}$ from long exposure frame $I_{L,t}$ (*right*)

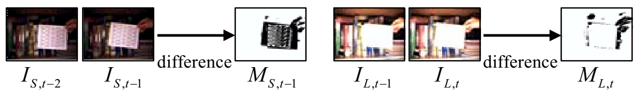
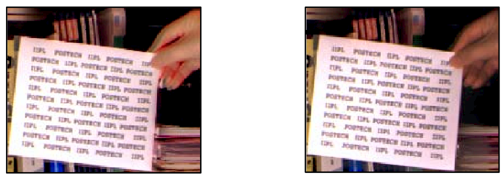


Figure 8. Obtaining the change region masks $M_{S,t-1}$ and $M_{L,t}$ from previous and current pairs of exposure frames, ($I_{S,t-2}$, $I_{S,t-1}$) and ($I_{L,t-1}$, $I_{L,t}$)

In the above process, the distortions appear in output frames if the change region masks are not used. Fig.9(a) shows the artifacts around moving regions in an output frame using only the saturation mask $M_{sat,t}$. Fig.9(b) is the case of using the fusion mask $M_{S,t-1+L,t}$ and shows that ghosting artifacts are reduced in the regions of the paper and hand.



(a) (b)

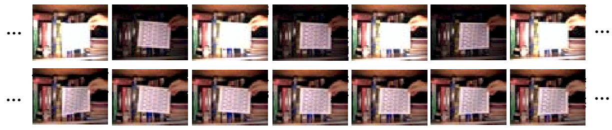
Figure 9. Comparison of output frames (a) without moving object handling mask and (b) with moving object handling mask

4.2 Fusion of exposure frames

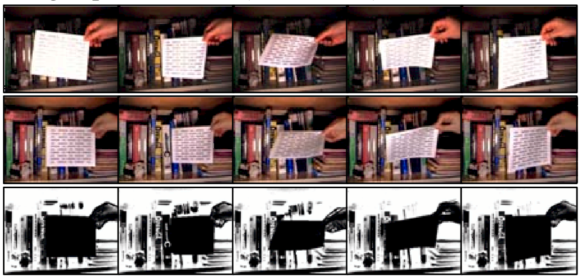
We obtain the output frame $I_{S,t-1+L,t}$ using the pyramid based image fusion method[12] with the mask $M_{S,t-1+L,t}$. The short exposure frame $I_{S,t-1}$ has a high effect on the output frame $I_{S,t-1+L,t}$, if pixel values of the mask $M_{S,t-1+L,t}$ are close to zero. In Eq.(4), we converted pixel values of the fusion mask $M_{S,t-1+L,t}$ into zero, if the difference between exposure frames exists in each pixel. It means that we want to show the moving object region of the short exposure frame in the output frame $I_{S,t-1+L,t}$.

5 Experimental Results

In this section, we show results for two different dynamic scenes: a bright sheet of paper in Fig.10 and a drive along a street in Fig.11. We also compared our results with auto exposure frames in the similar situation. For our experiments, we used a Dragonfly camera from Pt. Grey Research that has SDK to control shutter times or gains. The processing time for each HDR video frame (640×480 pixels) is about 2.5 seconds on a 1.53GHz AMD Athlon XP 1800+ machine.

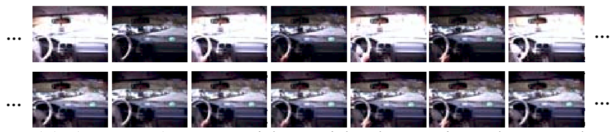


(a) (*top row*) Input video with alternating short and long exposure times, (*bottom row*) HDR video

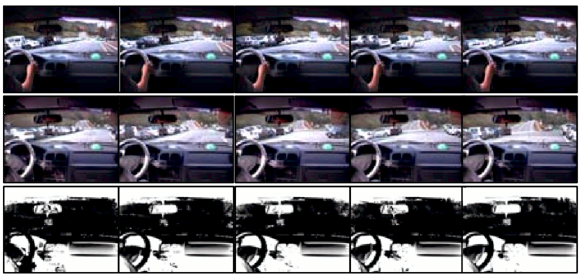


(b) (*top row*) Auto exposure frames, (*middle row*) Our result frames and (*bottom row*) their corresponding fusion masks.

Figure 10. Scene of a bright sheet of paper



(a) (*top row*) Input video with alternating short and long exposure times, (*bottom row*) HDR video



(b) (*top row*) Auto exposure frames, (*middle row*) Our result frames and (*bottom row*) their corresponding fusion masks

Figure 11. Driving scene

Fig.10(a) and 11(a) show the input video with alternating short and long exposure times(*top row*) and the HDR output frames(*bottom row*). Fig.10(b) and 11(b) compare normal frames by AEC setting(*top row*) with HDR frames generated using our method(*middle row*), and also show the fusion masks(*bottom row*) corresponding to our results.

In Fig.10, the scene of a bright sheet of paper, our results are more visible in both the text on a sheet of paper and background than the auto exposure version. In Fig.11, the driving scene, our HDR output frames show good details in the inside and outside of the car. However, our results sometimes show ghosting artifacts, because of large motions.

6 Conclusions

In this paper, we proposed an algorithm to generate HDR video in a stationary video camera system. Our algorithm takes the simple strategy of automatically controlling exposure times and effectively combines bright and dark areas in short and long exposure frames, respectively. We also handle moving objects and reduce ghost artifacts by adjusting fusion masks which is faster and simpler than registering exposure frames.

Our future work is to make a robust system for rapid motion and illuminance change, and also reduce the computational time to make a real-time system.

References

[1] P.E.Debevec and J.Malik, “Recovering High Dynamic Range Radiance Maps from Photographs”, *SIGGRAPH*, pp.369-378, 1997

[2] T.Mitsunaga and S.K.Nayar, “Radiometric Self Calibration”, *CVPR*, 1999

[3] G.Yourganov., "Acquiring HDR Video at Video Rate" Dept. of Computer Science, *Tech. report*, York University, May, 2001

[4] L.Bogoni, M.Hansen and P.Burt, “Pattern-Selective Color Image Fusion”, *Pattern Recognition Society*, pp. 1515-1526, 2001

[5] F.Durand and J.Dorsey, “Fast Bilateral Filtering for the Display of High Dynamic Range Images”, *SIGGRAPH*, 2002

[6] R.Fattal, D.Lischinski and M.Werman, “Gradient Domain High Dynamic Range Compression”, *SIGGRAPH*, 2002

[7] E.Reinhard, M.Stark, P.Shirley and J.Ferwerda, “Photographic Tone Reproduction for Digital Images”, *SIGGRAPH*, 2002

[8] S.B.Kang, M.Uyttendaele, S.Winder and R.Szeliski, “High Dynamic Range Video”, *SIGGRAPH*, 2003

[9] S.K.Nayar and V. Branzoi, “Adaptive Dynamic Range Imaging: Optical Control of Pixel Exposures over Space and Time”, *ICCV*, 2003

[10] S.J.Youm, W.H.Cho and K.S.Hong, “High Dynamic Range Imaging with Multi-Exposures and Handling Moving Objects”, *KSPC*, 2004

[11] W.H.Cho and K.S.Hong, “Extending Dynamic Range of Two Color Images under Different Exposures”, *ICPR*, 2004

[12] P.J.Burt and E.H.Adelson, “A Multi-resolution Spline with Application to Image Mosaics”, *ACM Transactions on Graphics*, pp. 217-236, 1983