HDR Video Synthesis Using Superpixel-Based Motion Estimation

Tu Van Vo, Chul Lee
Pukyong National University

요 약

We propose a novel high dynamic range (HDR) video synthesis algorithm using alternatively exposed low dynamic range (LDR) videos. We first develop a superpixel-based illumination invariant correspondence estimation algorithm. Then, we propose a reliability weight to further improve the quality of the synthesized HDR frame. Experimental results show that the proposed algorithm provides high-quality HDR frames compared to conventional algorithms.

1. 서론

High dynamic range (HDR) imaging is becoming more popular and getting more interests from both academia and industry owing to its faithful representation of the scene. Since most of the researches on HDR imaging has focused on high-quality synthesis of still images, HDR video generation receives relatively less attention. Although a few specialized HDR camera systems have been developed to capture HDR videos, they are too expensive to be used in practical applications. Instead, software-based approaches have been proposed to be employed in consumer devices.

A common approach to synthesize an HDR video is to capture multiple low dynamic range (LDR) frames taken with different exposure times and then merge them using alignment techniques. The main challenge of this approach is that it requires accurate motion estimation between consecutive frames to obtain high-quality HDR frames. Kang et al. [1] employed an optical flow-based solution to estimate small motions. However, it may fail to estimate accurate correspondences when complex motions exist. Kalantari et al. [2] employed a patch-based correspondence estimation scheme, but its computational complexity is too high to be used in practical applications.

In this work, we propose a robust HDR video synthesis algorithm using a superpixel-based illumination invariant motion estimation technique to address the aforementioned motion problems between adjacent frames. We further reduce artifacts caused by inaccurate correspondence estimation in the synthesized frame by developing a weighted averaging algorithm. Experimental results show that the proposed algorithm provides higher synthesis qualities than the conventional algorithms [2], [3].

2. 제안하는 기법

To acquire an HDR video using a conventional video camera, we first capture input frames with alternating exposure times for each frame. Suppose that we are given three consecutive frames $I_{t-1}$, $I_t$, and $I_{t+1}$. Since each frame has limited dynamic range, they may contain poorly-exposed regions, i.e., under- or over-exposed. We first define those regions in the reference frame $I_t$ using a threshold value $L_h$.

Specifically, a pixel $I_t(x)$ at location $x$ is defined as well-exposed, if $L_h < I_t(x) < 255 - L_h$.

Since an HDR frame is synthesized by merging alternatively exposed LDR frames, correspondences between the frames should be accurately estimated. We first estimate correspondences between frames $I_{t-1}$ and $I_{t+1}$, then merge them to synthesize an HDR frame as done in [4]. However, pixels in a poorly-exposed region contain no information, resulting in failure in correspondence estimation. Therefore, in a poorly-exposed region, instead of estimating the motion field from $I_{t-1}$ to $I_t$ directly, we estimate the motion field from $I_{t-1}$ to $I_{t+1}$ and halve it. In this work, we propose a superpixel-based motion estimation scheme, assuming that, if a pair of pixels in $I_{t-1}$ and $I_{t+1}$ match, their neighboring pixels are likely to match as well. Specifically, for a pixel at $x$, its neighboring pixels are defined in the intersection of

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2016R1C1B2010319).
superpixel [5] $S_k$ and the support window $P_k$. Let $D^{t-1}_y$ and $D^{t+1}_y$ denote the DASC descriptors [6] of the pixels $y$ in $I_{t-1}$ and $I_{t+1}$, respectively. Then, we define two costs $C_{t-1} = \frac{1}{\sigma^2} \sum_{y \in \Omega} (D^{t-1}_y - \mu)^2$ and $C_{t+1} = \frac{1}{\omega} \sum_{y \in \Omega} \sum_{v \in \Omega} (D^{t+1}_y - \mu)^2$, where the weight of $\omega$ and the normalizing parameter $\omega$ are given by $\omega = \exp(-\frac{1}{\sigma^2})$ with $\sigma = 8$ and $\omega = \sum_{y \in \Omega} (D^{t+1}_y - \mu)^2$, respectively. The optimal motion vector $v$ is obtained by minimizing the overall cost

$$C(v) = \|C_{t-1} - C_{t+1}\|^2.$$  \hfill (1)

The motion vector $v$ obtained by minimizing the cost in (1) may be inaccurate in some regions, resulting in artifacts in the synthesized HDR frames. To improve the accuracy of the correspondence, we quantify the reliability of the motion vector using the structural similarity between the reference frame $I^t$ and the warped frame $I^{t-1}$, which is warped from $I^{t-1}$ to $I^{t}$ using the estimated motion vector. Specifically, the reliability is defined as

$$r^{t-1}(x) = \text{SSIM}(I^t(x), I^{t-1}(x))$$  \hfill (2)

where SSIM($\cdot$,$\cdot$) is the structural similarity index between two blocks centered at $x$. If $r^{t-1}(x)$ is small, the corresponding motion vector is unreliable so as to assign a small weight to $I^{t-1}(x)$ in the HDR synthesis. As $r^{t-1}(x)$ increases, $I^{t-1}(x)$ contributes more to the final HDR frame. Therefore, the reliability map effectively reduces artifacts in the synthesized HDR frame caused by inaccurate correspondence estimation.

Then, the final weight $a_t(x)$ is obtained by multiplying the reliability term and luminance term $I_t(x)$ [7]

$$a_t(x) = r_t(x) \times I_t(x).$$  \hfill (3)

Finally, we synthesize the HDR frame by weighted averaging the irradiance maps in each frame

$$h_t(x) = \sum_{t=1}^{T} a_t(x) \times f(l_t(x))$$  \hfill (4)

where $h_t(x)$ is the HDR irradiance value at location $x$ and $f(.)$ denotes the camera response function [7].

3. 실험 결과

We evaluate the performance of the proposed HDR video synthesis algorithm on the alternatively exposed LDR dataset in Fig. 1. The parameter $I_0$ is fixed to 12, and the size of the block $P_k$ is $7 \times 7$. The number of superpixels in a frame is 5000.

Fig. 1 compares the synthesized results of the 31th frame of the Student sequence obtained by Kalantari et al.’s algorithm [2], Li et al.’s algorithm [3], and the proposed algorithm. In Figs. 1(b) and (c), Kalantari et al.’s algorithm and Li et al.’s algorithm provide artifacts, e.g., missing object structures, due to inaccurate correspondence estimation in poorly-exposed regions. However, we see that the proposed algorithm in Fig. 1(d) provides the highest-quality result by estimating accurate correspondences using the superpixel-based motion estimation.

4. 결론

We proposed a robust algorithm to synthesize HDR videos from alternatively exposed LDR videos. We first estimated the correspondences between two frames taken with different exposure times via a superpixel-based illumination invariant motion estimation. We also developed a reliability weight to further improve the synthesis performance by reducing artifacts. Experimental results on real datasets show that the proposed algorithm can estimate accurate motions between frames and obtain high-quality HDR videos.

5. 참고문헌


