TEMPORALLY CONSISTENT REGION-BASED VIDEO EXPOSURE CORRECTION

Xuan Dong[1], *Lu Yuan*[2], *Weixin Li*[3], *Alan L. Yuille*[3]

Tsinghua University[1], Microsoft Research Asia[2], UC Los Angeles[3]
dongx10@mails.tsinghua.edu.cn, luyuan@microsoft.com, lwx@cs.ucla.edu, yuille@stat.ucla.edu

ABSTRACT

We analyze the problem of temporally consistent video exposure correction. Existing methods usually either fail to evaluate optimal exposure for every region or cannot get temporally consistent correction results. In addition, the contrast is often lost when the detail is not preserved properly during correction. In this paper, we use the block-based energy minimization to evaluate the temporally consistent exposure, which considers 1) the maximization of the visibility of all contents, 2) keeping the relative difference between neighboring regions, and 3) temporally consistent exposure of corresponding contents in different frames. Then, based on Weber contrast definition, we propose a contrast preserving exposure correction method. Experimental results show that our method enables better temporally consistent exposure evaluation and produces contrast preserving outputs.

1. INTRODUCTION

We study the problem of temporally consistent video exposure correction. Exposure is one of the most important factors to evaluate the quality of a video in the applications of video shooting and editing. A good exposure correction method should 1) make all regions well-exposed, 2) keep the exposure temporally consistent for the corresponding object/region over time, and 3) preserve the contrast of pixels after correction. It is challenging due to the large amount of videos and the variations of video contents.

Traditional image/video enhancement methods usually analyze the statistics of the whole image, like [1] and [2]. However, without considering each image region individually, the exposure evaluation is not always optimal for every region. In addition, the contrast of their results is often lost due to the poor preservation of detail in the enhancement algorithms. Region-based exposure correction method like [3] can get all regions well-exposed and compensate the lost contrast. But, directly using it for videos will have significant flickering artifacts due to the lack of consideration for temporal consistency. Some video temporally consistent post-processing algorithms can adjust its results to get temporally consistent output. One kind of methods temporally smooths enhancement parameters for temporal consistency such as [2] [4]. The limitation is when short-term and long-term flickering artifacts exist at the same time, they cannot perfectly remove them due to lack of consideration of correspondence between different frames. Another kind of methods searches correspondence among frames and temporally adjusts the flickering frames using their corresponding flickering-free frames such as [5], [6], and [7]. However, these post-processing algorithms always assume that the image enhancement method is global. As a result, using them to correct the results of the algorithm in [3] will have un-natural outputs because the correction is local.

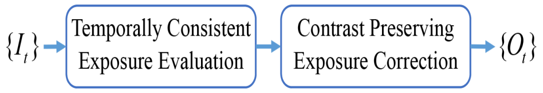

$\{I_t\}$
Temporally Consistent Exposure Evaluation
Contrast Preserving Exposure Correction
$\{O_t\}$

Fig. 1. Pipeline.

We propose a temporally consistent video exposure correction algorithm. As shown in the pipeline (Fig. 1), first, to get temporally consistent exposure evaluation, we use a block-based energy-optimization algorithm to 1) maximize the visibility of contents, 2) keep the relative difference of neighboring blocks, and 3) have temporally consistent evaluation of temporally corresponding blocks. Second, in the exposure correction module, based on the Weber contrast definition [8], we propose a contrast-preserving method to recover the output of each frame. The contributions of the proposed algorithm include: 1) a temporally consistent block-based exposure evaluation algorithm and 2) a contrast-preserving exposure correction algorithm.

We describe the temporally consistent exposure evaluation algorithm in Sec. 2 and the contrast preserving exposure correction algorithm in Sec. 3. Experimental results are shown in Sec. 4 and the paper is concluded in Sec. 5.

2. TEMPORALLY CONSISTENT EXPOSURE EVALUATION

The aim of the temporally consistent exposure evaluation is to evaluate exposure values for each frame so as to 1) maximize the visibility of all contents, 2) keep the relative difference between neighboring regions, and 3) have temporally consistent evaluation of corresponding contents in different frames.

We formulate it as a block-based energy minimization problem and the energy function considers these three factors, i.e., maximizing the visibility of blocks, keeping the relative difference of neighboring blocks, and having temporally consistent evaluation of corresponding blocks. Comparing with global methods, the block-based formulation could get better exposure evaluation for different regions. Comparing with the pixel-based formulation, the block-based formulation has the advantage that regions can better represent the visibility of contents and measure the relative difference of neighboring regions. In addition, we use fixed-size blocks instead of flexible regions like [3] because it is easier for temporal corresponding contents searching. Although some edge blocks may cover two different objects, we optimize the final result by considering all blocks together and the small number of edge blocks will not have big effect on the final results.

2.1. Block-based energy optimization

We use S-curve for the exposure evaluation, i.e., for each frame, we estimate a S-curve as the exposure evaluation curve. The reason to use S-curve is that it can adjust both over- and under- exposed regions to well-exposed levels. In addition, since the S-curve evaluation is global, it can avoid unnatural results of small regions caused by motion estimation outliers and defects of the design of the energy function. Thus, the optimization goal for each frame is the optimal S-curve exposure curve that satisfies the three goals, i.e. maximizing the visibility of all blocks, keeping the relative difference between neighboring blocks, and temporally consistent evaluation of corresponding blocks in different frames. S-curve α is defined as

$$\alpha(I(x):\phi_s,\phi_h) = I(x)+\phi_s\times\alpha_\Delta(I(x))-\phi_h\times\alpha_\Delta(1-I(x)), \tag{1}$$

where $I(x)$ and $\alpha(I(x):\phi_s,\phi_h)$ are the input and output luminance of pixel x (the luminance ranges from 0 to 1). ϕ_s and ϕ_h are the shadow and highlight amounts. The incremental function $\alpha_\Delta(l)$ is defined as $\alpha_\Delta(l) = \kappa_1 l \exp(-\kappa_2 l^{\kappa_3})$, where κ_1, κ_2 and κ_3 are default parameters (κ_1 = 5; κ_2 = 14; κ_3 = 1.6) so that the modified tonal range will fall in $[0, 0.5]$. An example S-curve is shown in Fig. 2.

Because α is determined by ϕ_s and ϕ_h, we optimize ϕ_s and ϕ_h and use the optimal ϕ_s^*, ϕ_h^* to get the optimal S-curve α^*. The optimal ϕ_s^*, ϕ_h^* of each frame are obtained by:

$$\begin{aligned}\{\phi_s^*,\phi_h^*\} &= \arg\min_{\{\phi_s,\phi_h\}} E(\phi_s,\phi_h)\\ = \arg\min_{\{\phi_s,\phi_h\}} &\sum\nolimits_i \Big(E_i + \frac{\lambda_s}{4}\sum_{j\in\Omega(i)} E_{ij} + \frac{\lambda_t}{K}\sum_{a\in A(i)} E_{ia}\Big)\end{aligned} \tag{2}$$

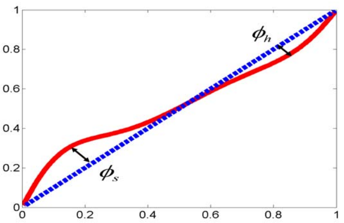


Fig. 2. An example S-curve.

where energy function $E(\phi_s,\phi_h)$ includes data term E_i to maximize the visibility of blocks, smooth term E_{ij} to keep the relative difference of neighboring blocks, and temporal term E_{ia} to have temporally consistent results of corresponding blocks. $\Omega(i)$ is the set of 4-neighboring blocks of block i. $A(i)$ is the set of K-nearest neighboring temporally corresponding blocks of block i in the feature space. The searching method for $A(i)$ in reference frames will be discussed in Sec.2.2. Given a pair of ϕ_s, ϕ_h, the S-curve result I_o of input I is got by $I_o = \alpha(I;\phi_s,\phi_h)$. λ_t and λ_s are set as 1 and 2 in our experiments. This encourages

The data term E_i is defined as

$$E_i = \begin{cases} +\infty, & (z_i-0.5)(z_{i_o}-0.5)<0\\ -(z_{i_o}-z_i), & z_i, z_{i_o} < 0.5\\ -(z_i-z_{i_o}), & z_i, z_{i_o} > 0.5\end{cases} \tag{3}$$

where z_i is the average luminance of block i in I and z_{i_o} is the average luminance of block i in I_o. The data term encourages the optimization results of blocks as close to 0.5 as possible so that the under- and over- exposed regions can be adjusted to well-exposed levels.

The smooth term E_{ij} is defined as

$$E_{ij} = ((z_{j_o}-z_{i_o})-(z_j-z_i))^2, \tag{4}$$

where z_j and z_{j_o} are the average luminance of block j in I and I_o, respectively. The smooth term is designed to keep the relative difference between neighboring blocks. The combination of data term and smooth term can encourage exposure to get close to 0.5 but will stop at some intensity by the smooth term so that the relative difference between neighboring blocks can be kept.

The temporal term E_{ia} is defined as:

$$E_{ia} = w_{ia}^t|(z_{i_o}-z_i)-(z_{a_o}-z_a)|, \tag{5}$$

where block a is block i's corresponding block, z_a and z_{a_o} are the average luminance of block a in input frame and its S-curve adjustment result, weight value $w_{ia}^t = \exp(-\frac{(z_i-z_a)^2}{0.1^2})$. When there are similar contents in neighboring frames, weight of E_{ia} will increase so as to keep the evaluated exposure values of current block i consistent with block a.

To infer the optimal $\{\phi_s^*\}$, we use the candidate values of ϕ_s from 0 to 1, with intervals of 0.1, and select out the value with minimum energy cost. The optimal $\{\phi_s^*\}$ of all frames are inferred by Loopy Belief Propagation. This is efficient because the solution space is small. The optimal $\{\phi_h^*\}$ is calculated in the same way. ϕ_s and ϕ_h could be inferred one by one because ϕ_s only change the shadow parts and ϕ_h only change the highlight parts of the frame. For simplification, we denote the optimal S-curve α^* as α in Sec. 3.

2.2. Correspondence searching

We aim at finding the set of K-nearest neighboring temporally corresponding blocks $A(i)$ in all frames for each block i. Since the temporal term in the energy function only considers the exposure condition of blocks, and blocks under similar exposure conditions have the same influence on the energy function, we measure the similarity between blocks by their difference of feature values instead of their physical motion relationships. To accelerate the searching process, we organize all blocks of all frames with a traditional kd-tree [9], and blocks are organized according to their feature values which include average luminance of the block itself and its 4-neighboring blocks. For each block i, its set of KNN blocks $A(i)$ is searched in the kd-tree using the method in [9]. For videos with any resolution, we resize them to 400×320 and the block size is fixed to 40×40.

3. CONTRAST PRESERVING EXPOSURE CORRECTION

Directly using the optimized curves α to adjust frames can have well-exposed and temporally consistent results, however, the contrast of the outputs is often lost, especially at the pixels whose luminance is close to 0.5. It is because those pixels and their neighboring pixels tend to be adjusted to the same luminance, i.e., 0.5, and their original contrast between each other is reduced. We observe that whether the frame is well-exposed and the exposure is temporally consistent is mostly determined by the base layer, and whether the contrast is preserved is mostly determined by the detail layer. Thus, first, we separate each input frame into base and detail layer. Second, we use the optimized curves to adjust base layers of the input frames to get base layers output. Third, we make use of Weber contrast definition [8] to recover the contrast preserving output from the base layer output.

3.1. Base layer adjustment

First we separate the input frame into base layer B and detail layer D using the guided filter in [10], i.e., $I \rightarrow B + D$. In detail, we set the radius to 4% of the height of the input frame I and use guided filter to filter the input frame I to get the base layer B. And detail layer D is obtained by $D = I - B$. Second, we use the optimized exposure curve α to adjust the base layer. This could be expressed by

$$B_o(p) = \alpha(B(p)), \tag{6}$$

where $B_o(p)$ is the background value of pixel p in output frame, and $B(p)$ is the background value of pixel p in input frame.

3.2. Contrast preserving output computation

After getting the output base layer, we make use of Weber contrast definition [8] to compute the contrast preserving output frame. Contrast $C(p)$ at pixel p is defined as

$$C(p) = \frac{I(p) - B(p)}{B(p)}, \tag{7}$$

where $I(p)$ and $B(p)$ are the luminance and background values of pixel p. We enforce contrast of output and input frames equal to each other at each pixel, i.e.,

$$\frac{O(p) - B_o(p)}{B_o(p)} = \frac{I(p) - B(p)}{B(p)}, \tag{8}$$

where $O(p)$ is the luminance the output frame at pixel p. Using Eq. (6), we get

$$O(p) = \frac{\alpha(B(p))}{B(p)} I(p). \tag{9}$$

Here, we can get contrast preserving output O from α, I and B using Eq. (9). When $B(p) = 0$, we set $B(p) = 0.0001$.

4. EXPERIMENTAL RESULTS

A series of experiments are conducted on a Windows PC (Intel Core 2 Duo T6500 at 2.0 GHz with 3GB of RAM). The implementation language is Matlab. No GPU or multi-core acceleration is used in our experiments. The resolution of videos is 640 × 480. The processing time of major steps is shown in Table 1. There are 100 videos in our experiments. We compare the proposed method with the image exposure correction [3] (named ECCV12), and the video virtual exposure method [2] (named SIG05). Since the temporally consistent video post-processing algorithms in [5] [7] can remove flickering artifacts, we use them to post-process the results of ECCV12 for temporal consistency (named ECCV12+ICCP13 and ECCV12+SIG13). When implementing ECCV12+ICCP13, the input videos are first adjusted by ECCV12, and then the results of ECCV12 are corrected by ICCP13. When implementing ECCV12+SIG13, first, the frame-wise exposure evaluation S-curves are estimated by the method in [3], second, the SIG13 method is used to temporally smooth the S-curves, last, the exposure correction in [3] is performed with the smoothed S-curves. Please see the input videos and output videos of different algorithms in the supplementary materials.

Steps	Time/ms
Feature calculation	2.3
Building of kd-tree and correspondence searching	1.6
Energy minimization	2.6
Base/detail separation	43.3

Table 1. The average time cost of major steps of our algorithm for each frame. Input frame resolution is 640 × 480.

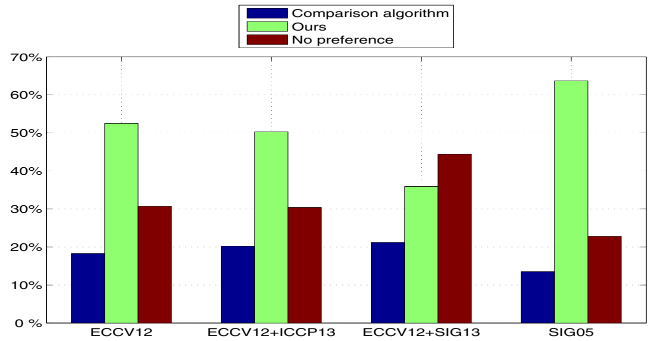
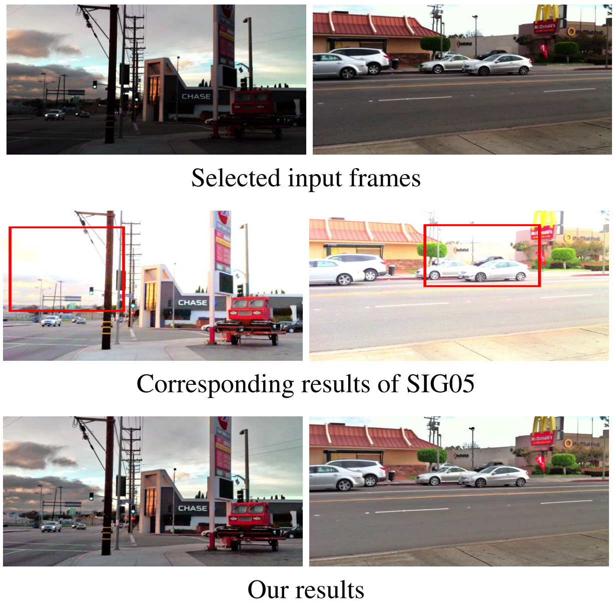


Fig. 3. User study. We compare the proposed method with ECCV12 [3], ECCV12+ICCP13 [5], ECCV12+SIG13 [7], and SIG05 [2]. Each color bar shows the average percentage of favored video.

User study is conducted in our experiments. The comparison algorithms include the ECCV12, ECCV12+ICCP13, ECCV12+SIG13, and SIG05. We invited 10 volunteers (6 males and 4 females) to perform pairwise comparison between our result and results of ECCV12, ECCV12+ICCP13, ECCV12+SIG13, SIG05. For each pairwise comparison, the subject has three options: better, or worse, or no preference. Subjects are allowed to view each video clip pair back and forth for the comparison. To avoid the subjective bias, the order of pairs, and the video order within each pair are randomized and unknown to each subject. This user study is conducted in the same settings (room, light, and monitor). The user study results are summarized in Fig. 3. Each color bar is the averaged percentage of the favored video over all subjects. From results, we can see that in the comparison with different algorithms, the participants overwhelmingly select our result over the ECCV12 (52.2% vs. 18.9%), ECCV12+ICCP13 (50.4% vs. 20.2%), ECCV12+SIG13 (30.8% vs. 21.5%), SIG05 (63.8% vs. 13.9%).

In Fig. 4, results of the SIG05 method [2] and our method are compared. We can notice that the results of SIG05 are over-exposed at the regions marked in red. It is because the enhancement parameters of the algorithm need setting manually and are not adaptive to different videos. As a result, the exposure level is not evaluated properly for the video, leading to the over-exposure. On the contrary, our results are well-exposed since the exposure evaluation of our method is adaptive to different videos. We compare with ECCV12 in Fig. 5. Directly using ECCV12 to adjust the frames one by one leads to significant flickering artifacts, such as the regions marked in red. The reason is that the segmentation results of the same contents in different frames are sometimes quite different, which leads to un-consistent exposure evaluation. However, our results can keep temporal consistency well because the corresponding blocks over time are considered when evaluate the exposure of each frame.


Selected input frames

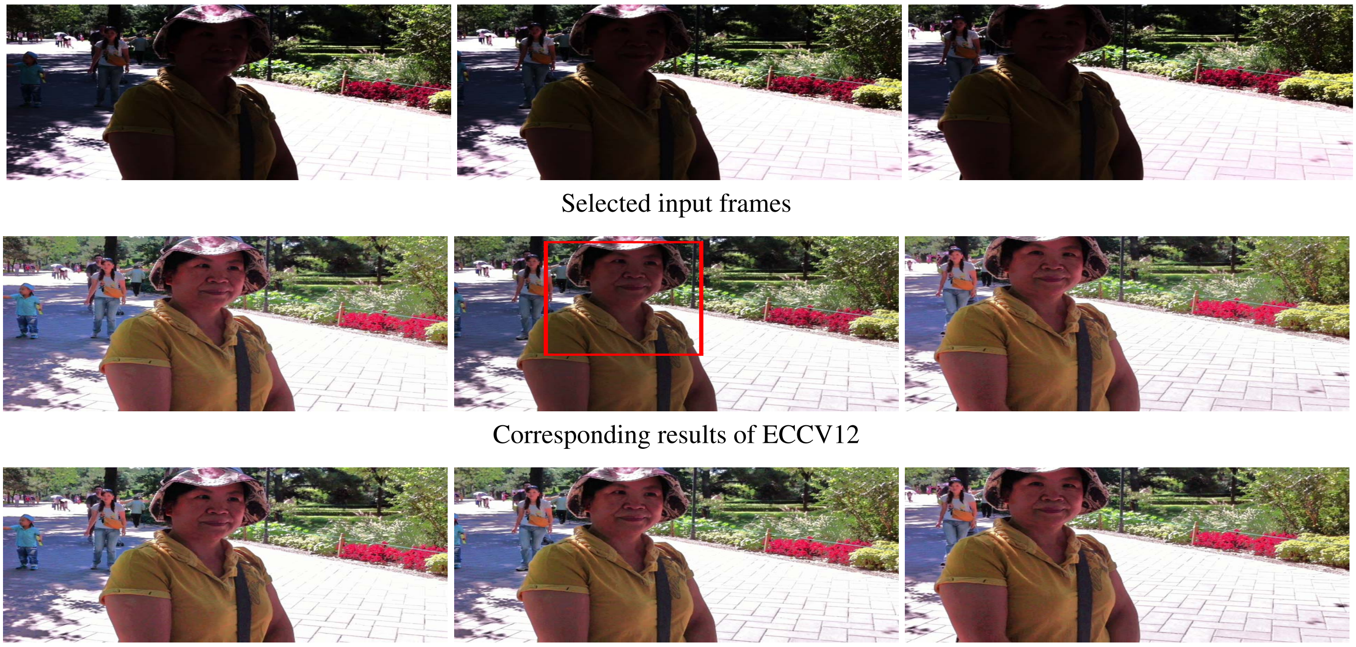
Corresponding results of SIG05

Our results

Fig. 4. Comparisons with SIG05 in [2]. The regions marked in red show un-proper exposure evaluation.

The temporally consistent post-processing algorithms [5] [7] can be used to repair the flickering artifacts caused by ECCV12. However, as shown in Fig. 6, ECCV12+ICCP13 cannot remove the flickering artifacts perfectly because the algorithm estimates a global correction function to repair frames with flickering artifacts. Since the exposure correction method is local, ICCP13 will produce un-natural enhancement results at flickering frames. ECCV12+SIG13 can remove short-term flickering artifacts, but, as shown in Fig. 6 (d), when the flickering artifacts remain for a long time period, the method will select the flickering frame as the key frame thus fail to remove them. Our algorithm evaluates temporally consistent exposure by considering corresponding frames over time, thus, the flickering artifacts can be removed perfectly.

We also compare with ECCV12 about contrast preservation in Fig. 7. Although ECCV12 adds some amount of details to the S-curve result to preserve contrast, the maximum amount of compensated details is less than two times of the details of the input frame. Thus, when the exposure correction values are large, it cannot perfectly preserve the contrast. On the contrary, our result could preserve the contrast well by



Selected input frames

Corresponding results of ECCV12

Our results

Fig. 5. Comparisons with ECCV12 in [3]. The region marked in red shows a flickering artifact.

enforcing the contrast of output equal to contrast of input.

5. CONCLUSION

We have presented a temporally consistent video exposure correction. First, a block-based energy minimization method is proposed to evaluate temporally consistent exposure, so that 1) the visibility of all contents is maximized, 2) the relative difference between neighboring regions is kept, and 3) temporally consistent exposure of corresponding contents in different frames are evaluated. Second, a contrast preserving exposure correction method is proposed based on Weber contrast definition [8]. Experimental results show that our method can have better temporally consistent exposure evaluation and produce contrast preserving correction results.

6. ACKNOWLEDGE

The work is supported by grant NSF award CCF-1317376.

7. REFERENCES

[1] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, and J. B. Zimmerman, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, 1987.

[2] E.P. Bennett and L. McMillan, "Video enhancement using per-pixel virtual exposures," *ACM SIGGRAPH*, 2005.

[3] L. Yuan and J. Sun, "Automatic exposure correction of consumer photographs," *ECCV*, 2012.

[4] C. Kiser, E. Reinhard, M. Tocci, and N. Tocci, "Real time automated tone mapping system for hdr video," *ICIP*, 2012.

[5] M. Grundmann, C. McClanahan, S.B. Kang, and I. Essa, "Post-processing approach for radiometric self-calibration of video," *ICCP*, 2013.

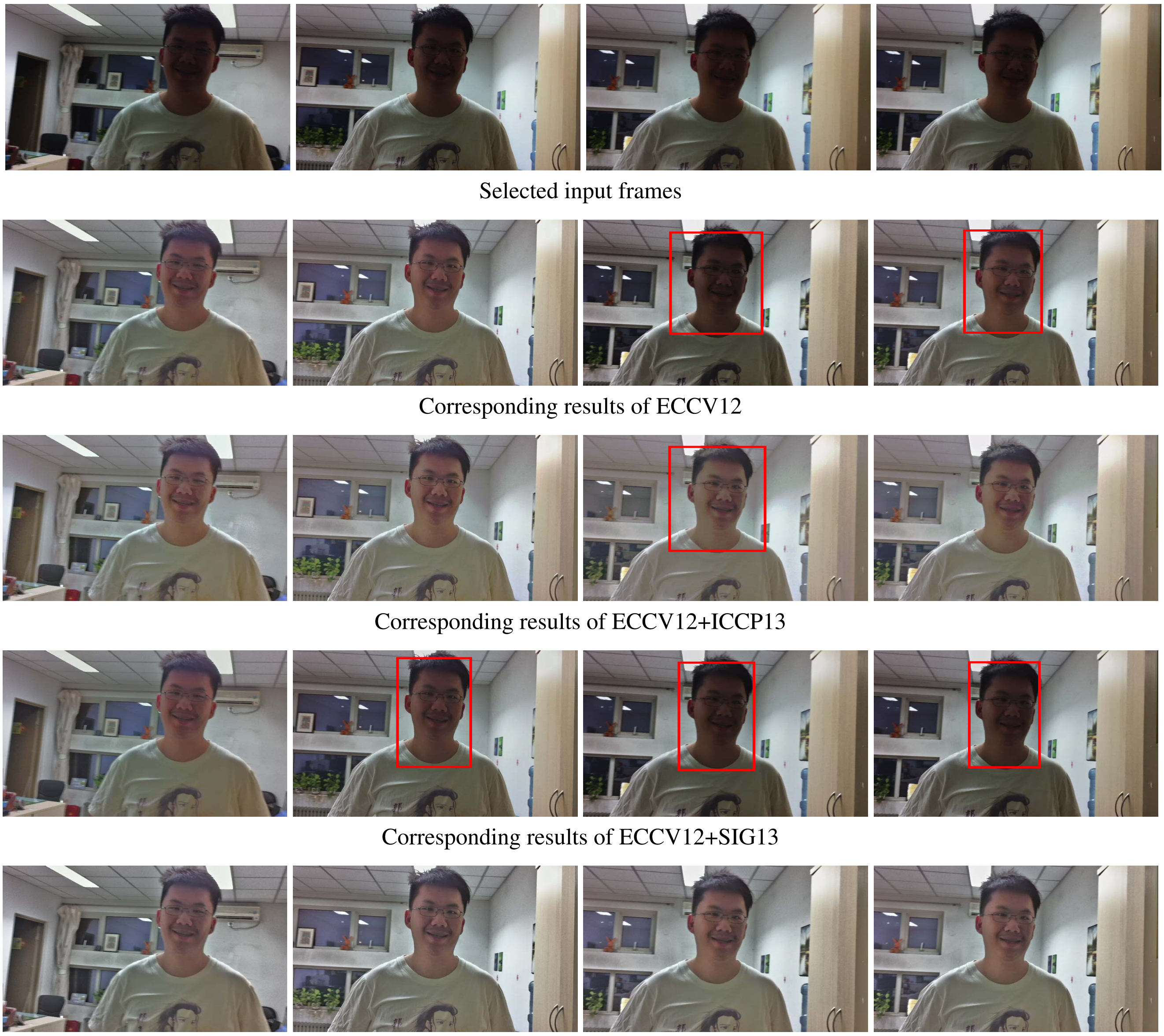
[6] Y. Hacohen, E. Shechtman, D.B. Goldman, and D. Lischinsky, "Optimizing color consistency in photo collections," *ACM SIGGRAPH*, 2013.

[7] N. Bonneel, K. Sunkavalli, S. Paris, and H. Pfister, "Example-based video color grading," *ACM SIGGRAPH*, 2013.

[8] P. Whittle, "The psychophysics of contrast brightness," *Lawrence Erlbaum Associates, Inc*, 1994.

[9] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *ACM Communications*, 1975.

[10] K. He, J. Sun, and X. Tang, "Guided image filtering," *ECCV*, 2010.



Selected input frames

Corresponding results of ECCV12

Corresponding results of ECCV12+ICCP13

Corresponding results of ECCV12+SIG13

Our results

Fig. 6. Comparisons with ECCV12 in [3], ECCV12+ICCP13 in [5], ECCV12+SIG13 in [7]. The regions marked in red show flickering artifacts and un-proper correction results.

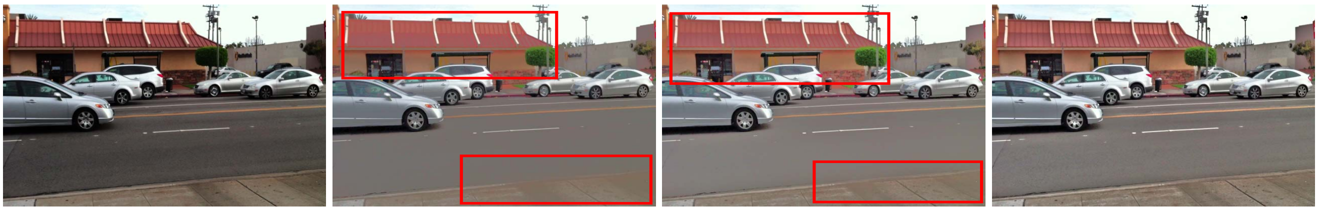


Fig. 7. Comparisons of contrast preservation with direct S-curve adjustment and ECCV12 in [3]. From left to right: input frame, results of direct S-curve adjustment, ECCV12, and our algorithm. The regions marked in red have contrast lost after correction.