Region-based Temporally Consistent Video Post-processing

Xuan Dong
Tsinghua University
dongx10@mails.tsinghua.edu.cn

Boyan Bonev
UC Los Angeles
bonev@ucla.edu

Yu Zhu
Northwestern Polytechnical University
zhuyu1986@mail.nwpu.edu.cn

Alan L. Yuille
UC Los Angeles
yuille@stat.ucla.edu

Abstract

We study the problem of temporally consistent video post-processing. Previous post-processing algorithms usually either fail to keep high fidelity or fail to keep temporal consistency of output videos. In this paper, we observe experimentally that many image/video enhancement algorithms enforce a spatially consistent prior on the enhancement. More precisely, within a local region, the enhancement is consistent, i.e., pixels with the same RGB values will get the same enhancement values. Using this prior, we segment each frame into several regions and temporally-spatially adjust the enhancement of regions of different frames, taking into account fidelity, temporal consistency and spatial consistency. User study, objective measurement and visual quality comparisons are conducted. The experimental results demonstrate that our output videos can keep high fidelity and temporal consistency at the same time.

1. Introduction

The consumption of videos is increasing dramatically in video streaming and surveillance systems. This results in mass demands for video enhancement of exposure , color, contrast, etc. In computer vision, there exist many image enhancement algorithms such as exposure correction [22], color grading [4], etc. Their enhancement effects are very impressive, and they are used in many video applications and systems such as video editing softwares like Adobe Premiere (Pr), mobile phone apps like Instagram, etc.

However, there are usually significant flickering artifacts when performing video enhancement, or image enhancement methods frame by frame for videos, due to lack of built-in temporal consistency. To remove these artifacts is non-trival because they have a profound effect on the visual quality. In addition, in practical systems, we usually only have access to the input videos and the original enhancement videos (with flickering artifacts), and do not know or cannot have access to the enhancement algorithms. For example, 1) the enhancement algorithms of industrial softwares are not known to the public, like Pr and Instagram. 2) For embedded/hard-ware enhancement algorithms, the device may not provide interfaces to revise the algorithms for temporal consistency. 3) In practical development of a software or an application for video editing, several enhancement algorithms may be required which are all different. So designing a temporally consistent method for each separate algorithm will be time-consuming. In such cases, it is desirable to do temporally consistent enhancement as post-processing, by simply analyzing the input videos and original enhancement videos.

In this paper, we study the problem of temporally consistent video post-processing when the original input and enhancement videos are available. The goal is to keep both temporal consistency and fidelity of the output videos. 1) Temporal consistency means that for the same objects in different frames, the enhancement should be consistent. 2) Fidelity means that the final results should have similar effects as the original enhancement videos. In the other words, the output frames should be similar with the original enhancement results at non-flickering frames, and the objects in flickering frames should be adjusted referred to the corresponding objects in non-flickering frames. The challenges of the problem include: 1) the original enhancement methods are unknown and cannot be revised at all, 2) motions of videos are complicated, 3) the method should be able to remove flickering artifacts caused by different enhancement methods.

We discover experimentally the spatially consistent enhancement (SCE) prior which is valid for many leading image enhancement methods, including Pr auto color, auto level, auto contrast, exposure correction [22], and color grading [2] [4]. The prior is based on the observation that in a local region, image enhancement methods tend to

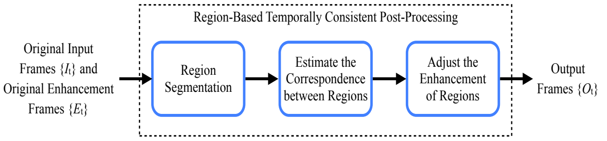

Region-Based Temporally Consistent Post-Processing
Original Input Frames $\{I_t\}$ and Original Enhancement Frames $\{E_t\}$
Region Segmentation
Estimate the Correspondence between Regions
Adjust the Enhancement of Regions
Output Frames $\{O_t\}$

Figure 1. Pipeline of our region-based temporally consistent video post-processing.

keep the enhancement values consistent for pixels with the same rgb values. Based on this prior, we propose a region-based temporally spatially consistent adjustment method. The pipeline is shown in Fig. 1. The inputs include the original input frames and the original enhancement frames. 1) Based on the prior, each frame is segmented into several regions. 2) Corresponding regions between different frames are estimated. 3) a Markov Random Field (MRF) optimization model is used to adjust the enhancement of regions of all frames.

The advantage of the proposed algorithm includes that: 1) it can post-process flickering results of any enhancement method as long as it enforces the SCE prior, 2) our results can keep high fidelity and 3) temporal consistency. The contributions in this paper are as follows: 1) The SCE prior is discovered and experimentally proved. 2) A region-based temporally spatially enhancement algorithm is proposed to post-process videos with flickering artifacts.

Experimental results show that our proposed algorithm performs better than the frame-wise enhancement algorithms including Pr auto color, auto level, auto contrast, exposure correction [22], color grading [2] [4] and the temporally consistent video adjustment algorithms including [4] [6] [8] [11] in user study, objective and visual quality comparisons.

2. Related Works

For energy function based image enhancement methods, in [16] [12], a temporal term is added into the original energy function for temporal consistency. In [15], properly designed filters are used to substitute for the energy minimization process, so as to accelerate optimization driven methods. [5] extends the 2-D image filter to a 3-D temporal-spatial filter. The limitation of these methods is that they have to revise the original enhancement algorithm, so they cannot be directly used for this paper's problem.

[4] [18] [14] [13] [10] [9] [3] first enhance videos frame by frame. Then, based on the characteristics of the known enhancement algorithms, they propose different methods to detect and remove the flickering frames. The limitation of them is that they assume the enhancement is a global transformation. So, they cannot keep fidelity of the output videos for local enhancement algorithms. [21] do not use the global enhancement assumption. They first estimate correspondence between frames, and then temporally filtering matched pixels. For unmatched pixels, a reflectance completion algorithm is proposed to blend those pixels with neighboring matched pixels. The limitation is that the reflectance completion algorithm is specifically designed for the problem of separating images into shading and reflectance layers, and cannot be directly used for other enhancement algorithms.

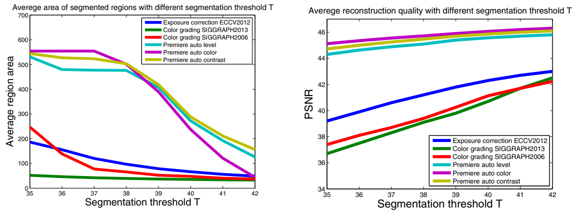


Figure 2. Average square root of the area and the reconstruction quality of segmented regions with different threshold T. The enhancement algorithms include Pr auto level, auto color, auto contrast, exposure correction [22], color grading [2] [4].

There are some temporally consistent post-processing methods for unknown original enhancement algorithms, including [8],[19],[6],[11]. All of them only use the original enhancement videos with flickering artifacts for temporally consistent enhancement. But we propose to make use of both original input and enhancement videos. They first propose different algorithms to find sparse correspondence between frames. For matched pixels, the pixel values are temporally filtered. For unmatched pixels, global transformation is used according to matched pixels. Because they are designed for global enhancement methods, their results will have low fidelity and cannot keep temporal consistency perfectly if the enhancement methods are local. In addition, in [6], the errors of enhancement are accumulated over time, and the key frames selection is not adaptive. In [11], the enhancement curve is estimated by a smooth piecewise-quadratic spline with 7 knots at (0,0.2,0.4,0.6,0.8,1). When the estimation of the 7 knots has some errors, the spline will enlarge the errors to the whole dynamic range.

The comparisons of [8],[4],[6],[11] with our proposed algorithm are shown in Sec. 5.

3. Spatially Consistent Enhancement Prior

In this section, first, we mathematically describe the SCE prior, i.e., within a local region, the enhancement of pixels is consistent. Second, since the regions that enforce prior vary a lot in size, shape, and location for different images and different enhancement algorithms, we propose a region segmentation method to get the regions with the SCE property. Third, the prior is experimentally verified by the segmentation results.

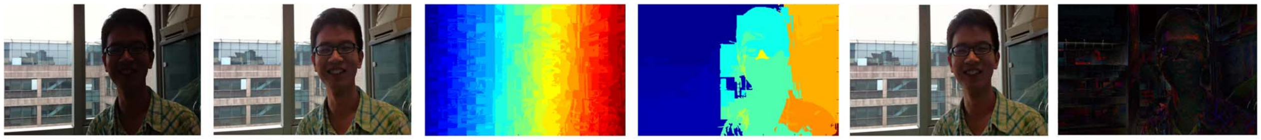


Figure 3. Example of region based reconstruction. Left to right: input map I, original enhancement result E using exposure correction [22], super pixels segmentation result [1], regions merging result, reconstruction result R, and absolute difference between E and R (enlarged in 5 times).

3.1. Description of the SCE prior

For an original enhancement method F to enhance an input image I, if a segmented region i is given, the original enhancement result can be written as $E_i(x) = F(x, I(x)), x \in i$, where E_i is the original enhancement result of region i. x is the pixel belonging to region i. The SCE prior is based on the observation that at local regions within the same object/scene, many image enhancement methods tend to keep the enhancement values the same or very similar for pixels with the same rgb values. In other words, within a local region i, there will exist an enhancement curve α_i to reconstruct the region i. α_i's independent variable is only the intensity of the pixels and the reconstruction results should be very similar with the original enhancement results, i.e., $E_i \approx \alpha_i(I(x)), x \in i$. We define

$$R_i(x) = \alpha_i(I(x)), x \in i, \tag{1}$$

where R_i is the reconstruction result of region i using α_i. Borrowing the concept of reconstruction quality from video coding, we use Peak Signal-to-Noise Ratio (PSNR) to measure the similarity between R_i and E_i , i.e.,

$$RQ_i = PSNR(R_i, E_i), \tag{2}$$

where RQ_i is the reconstruction quality of region i. According to the prior, the reconstruction quality RQ_i should be very high for good enhancement.

3.2. Region segmentation

Since the regions that enforce the SCE prior vary a lot in size, shape, and location for different images and different enhancement algorithms, we propose a region segmentation method to find these regions.

First, for a given region i, we verify whether the region enforces the SCE prior. To do so, we use standard histogram matching [20] to estimate the curve α_i for region i. In histogram matching, the histograms H_I and H_E of I and E_i within region i are computed. Then, using H_I and H_E the cumulative, distribution functions C_I and C_E are computed. Next, for each gray level $G_1 \in [0, 255]$, we find the gray level G_2, for which $C_I(G_1) = C_E(G_2)$, and this is the result of histogram matching function: $M(G_1) = G_2$. RGB channels are computed respectively to form the reconstruction function α_i. If the enhancement is consistent within the region i, histogram matching can get the correct estimation of the truth enhancement curve. Thus, using Eq. (1) and Eq. (2), we can get high RQ_i. Here, we set a threshold T. If $RQ_i > T$, the region is seen as enforcing the SCE prior.

Second, we propose to merge neighboring regions to see how large the region can be with $RQ_i > T$. To begin with, we segment the input images into a set of super pixels using the SLIC algorithm [1] because of its simplicity and speed. Then, we try to merge each pair of neighboring super pixels. For neighboring super pixels i and j, we estimate the enhancement curve of their merged region $i \cup j$ using histogram matching, reconstruct the enhancement result using Eq.(1), and compute the reconstruction quality using Eq. (2). If $RQ_{i \cup j} > T$, these two super pixels will be merged together. Otherwise, they will not be merged. The merging is done iteratively until none of the neighboring regions can be merged together.

An example is shown in Fig. 3. The super pixels segmentation result is obtained by segmenting the original input image using SLIC [1]. The segmentation figure shows that most of the regions cover a large area and the reconstruction result looks very similar to the original enhancement result. Even if we enlarge the difference between the original enhancement result and the reconstruction result in 5 times, the difference is still small.

3.3. Verification of the SCE prior

We collect a set including 3000 input and original enhancement result pairs from image and video results. The images are from published paper's experimental results and search engines. The videos are from the most popular movies. The enhancement methods we test include Pr autocolor, auto-level, auto-contrast, exposure correction [22], color grading [2] and [4]. The images/videos are resized to 640 × 480.

To verify how good the prior is, we use the proposed segmentation method with different T to segment the images. Then, we compute the average area and reconstruction quality for all regions of the images. Figure 2 shows the result. We can see that with the increase of T, the average area for each region will decrease and the average reconstruction quality will increase due to the increase of the reconstruction quality requirement. The average area and PSNR of Pr enhancement algorithms is larger than the other three algorithms when T is small. The reason is that Pr algorithms are more like global adjustment method while the other three algorithms are relatively more local. Although the average area varies, all the algorithms have high average area (more than 60×60) for most T. This gives very strong evidence for the SCE prior. At the same time, all of the algorithms can get reconstruction quality higher than 37 db. This is a relatively high reconstruction quality and the visible loss is not big.

4. Region-based Temporally Spatially Consistent Adjustment

We segment each frame into several regions and estimate the original enhancement curves of the regions using the SCE prior, estimate correspondence relationships between regions in different frames, optimize the original enhancement curves of regions to get temporally consistent enhancement curves of regions, and reconstruct temporally consistent frames using the optimized curves. The motivation of our algorithm are that 1) according to the SCE prior, each segmented region can be reconstructed with an enhancement curve, and the reconstruction can keep high fidelity. In addition, 2) correspondence relationships of regions in different frames are easier to be found because they only need correspondence of sparse pixels instead of dense correspondence. This can reduce the requirement of motion estimation accuracy.

4.1. Region segmentation

For each frame, we propose to segment each frame into different regions with the principle that 1) there should exist an enhancement curve for each region to reconstruct the region with high reconstruction quality, 2) the regions area should be as large as possible so as to reduce the requirement of motion estimation requirement.

We use the segmentation method in Sec. 3 because it has been verified to be good for different enhancement algorithms. The set of segmented regions for all frames are denoted as Ω_R, and $\Omega_R = \{i^t : t = 1..M, i = 1..N(t)\}$, where i^t is the region i of frame t, M is the total frame number, $N(t)$ is the total region number in frame t. After getting Ω_R, the enhancement curve of each region is estimated using histogram matching.

4.2. Estimating correspondence between regions

We first compute dense correspondence of pixels between neighboring frames using SIFT Flow [17] because of its accuracy. Then, we link the corresponding pixels over different frames to get the motion of a scene point over time, and call it a motion path. Any two pixels along the same motion path are seen as corresponding pixels. Although SIFT Flow is designed for dense correspondence estimation, due to incorrect estimations and occlusions, the motion estimation results of some pixels are outliers. To keep the accuracy of estimated motion paths, we propose to measure the confidence of the pixels along the motion paths. It is measured by the distance in SIFT field. If a pixel has the confidence value larger than a threshold T_{SIFT}, it is detected as outliers and the motion path will stop at this pixel. As a result, the correspondence pixels between two frames become sparse. To avoid cumulative errors during linking motion vectors frame by frame, for each frame, as shown in Fig. 4, we estimate its correspondence with not only the neighboring frames, but also the frames whose intervals are k. When two frames have large time intervals, the linking can skip $k-1$ frames.

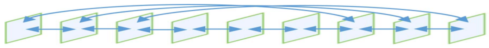


Figure 4. Optical flow estimation among all frames.

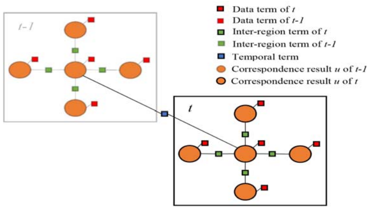


Figure 5. Temporal-spatial belief propagation. The proposed objective function takes into account the temporal term, spatial term, and data term .

After the estimation of sparse pixels correspondence between frames, we estimate corresponding regions in different frames. If two regions in different frames have corresponding pixels, they are marked as corresponding regions. Otherwise, they are marked as un-corresponding regions.

$$\chi(i^{t_1}, j^{t_2}) = \begin{cases} 1, & \text{if } i^{t_1}, j^{t_2} \text{ have corresponding pixels} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

χ defines whether the region i of frame t_1 and the region j of frame t_2 are corresponding regions. In addition, for each region, we could find the set of its corresponding regions in all frames, i.e., $C({i_1}^{t_1}) = \{i^t : \chi({i_1}^{t_1}, i^t) = 1, i^t \in \Omega_R\}$, where ${i_1}^{t_1}$ is the region i_1 of frame t_1, $C({i_1}^{t_1})$ defines the set of its corresponding regions in different frames. Ω_R is the set of all of the regions in all frames.

4.3. Region-based temporally spatially optimization

We propose a region-based temporal spatial optimization method to adjust the enhancement curves of regions. Our goal is that 1) for regions in non-flickering frames, the adjusted enhancement curves should be the original enhancement curves α. And 2) for regions in flickering frames, the adjusted enhancement curves should be one of the curves of the regions in non-flickering frames. To achieve the goal, for each region, we let it pick one enhancement curve from its corresponding regions and itself, i.e., the solution space for each region is the curves of its corresponding regions and itself, i.e., $u(i^t) \in C(i^t)$. No matter whether the region belongs to flickering or non-flickering frames, the desired curve is within the solution space. The number of corresponding regions for each region depends on the video contents. In our experiments, the corresponding regions for each region is about several hundred on average. The solution space is not big and it could help avoid unnatural enhancement curves. We define the solution space for all regions of all frames as U, and $U = \{u(i^t) : i^t \in \Omega_R, u(i^t) \in C(i^t)\}$, where $u(i^t)$ is the picked corresponding region of i^t. The adjustment of enhancement curves of regions is modeled as a MRF problem, as shown in Fig. 5. The nodes of the MRF are the regions of different frames. The optimization problem is to select one of the corresponding regions for each region so that the enhancement curve of the selected region can help get the minimum energy costs under the MRF constraints. After getting the optimized corresponding regions relationships $u^*(i^t)$ for all regions in all frames, **we get the optimized enhancement curve $\alpha^*_{i^t}$ of any region i^t as** $\alpha_{u(i^t)}$, and use them to reconstruct each frame using Eq. (1). The optimal solution U^* of the MRF is obtained by $U^* = \arg\min_U E(U)$, and the objective function E is defined as:

$$E(U) = \sum_{i^t \in \Omega_R} [E_{data}(u(i^t)) + \lambda_1 E_{temporal}(u(i^t)) + \lambda_2 E_{spatial}(u(i^t))], \tag{4}$$

where the variable $u(i^t)$ is the picked corresponding region of i^t, the data term E_{data} aims at keeping fidelity of regions in non-flickering frames, the temporal term $E_{temporal}$ aims at keeping temporal consistency of regions in flickering regions. Although the neighboring regions have different enhancement curves, we propose the spatial term $E_{spatial}$ to keep the difference of their enhancement consistent, so as to avoid spatial inconsistent enhancement. λ_1 and λ_2 are the weights of the temporal and spatial terms, respectively. In detail,

$$E_{data}(u(i^t)) = ||\alpha_{u(i^t)} - \alpha_{i^t}||_2^2, \tag{5}$$

where α_{i^t} is the original enhancement curve for region i^t. By making the optimized curves as similar as the original enhancement curves, the data term can keep the optimal enhancement as similar as the original enhancement for non-flickering frames so as to keep high fidelity.

$$E_{temporal}(u(i^t)) = \sum_{{i_1}^{t_1} \in C(i^t)} w^{temporal}_{i^t,{i_1}^{t_1}} ||\alpha_{u(i^t)} - \alpha_{{i_1}^{t_1}}||_2^2, \tag{6}$$

where ${i_1}^{t_1}$ is the corresponding regions of region i^t. $w^{temporal}_{i^t,{i_1}^{t_1}}$ is the temporal weight between region i^t and ${i_1}^{t_1}$. And $w^{temporal}_{i^t,{i_1}^{t_1}} = \frac{CP(i^t,{i_1}^{t_1})}{\sum_{i'^{t'} \in C(i^t)} CP(i^t,i'^{t'})}$, where $CP(i^t, {i_1}^{t_1})$ is the number of corresponding pixels between region i^t and ${i_1}^{t_1}$. In the temporal term, we make each pair of corresponding regions similar with each other to keep temporally consistent enhancement. This can achieve temporally consistent adjustment of regions curves in flickering frames.

$$E_{spatial}(u(i^t)) = \sum_{j^t \in \Omega_N(i^t)} w^{spatial}_{i^t,j^t} ||\beta_{u(i^t)} - \beta_{u(j^t)}||_2^2, \tag{7}$$

where region j^t is the neighboring region of region i^t, $\Omega_N(i^t)$ is the set of neighboring regions of region i^t, and $\beta_{(i^t)}$ is the fitted global curve of frame t, and the spatial weight $w^{spatial}_{i^t,j^t} = \frac{RA(j^t)}{\sum_{{j_1}^t \in \Omega_N(i^t)} RA({j_1}^t)}$, where $RA(j^t)$ is the area of the region j^t. For each frame, we fit a global curve β to measure the enhancement. For neighboring regions, we keep the frames curves of selected regions as similar as possible so as to keep the enhancement difference of neighboring regions consistent, even if the frame curve itself may have a low reconstruction quality.

We use Loopy Belief Propagation [7] to estimate U^* for the MRF. After getting the optimized corresponding regions relationships $u^*(i^t)$ for all regions in all frames, we can get the optimized enhancement curve $\alpha^*_{i^t}$ of any region i^t as $\alpha_{u(i^t)}$, and use them to reconstruct each frame using Eq. (1).

5. Experimental Results

There are 6 original enhancement algorithms including Pr auto color, auto level, auto contrast, exposure correction [22], color grading [2] [4]. In the original enhancement of color grading [4], we only use the image enhancement part in [4] to produce the original enhancement results. Besides the original enhancement results of the 6 algorithms, we also compare with the results of the temporally consistent enhancement algorithms including the methods in [6], [4], [8], and [11]. Here, we use both the image and video enhancement parts in [4] to produce the results of the method. For the algorithm in [6], since the key frame selection method is flexible, in our experiment, we have two choices: 1) single key frame, where the first frame of the video is chosen as the key frame, and 2) multiple key frames, where the first frame

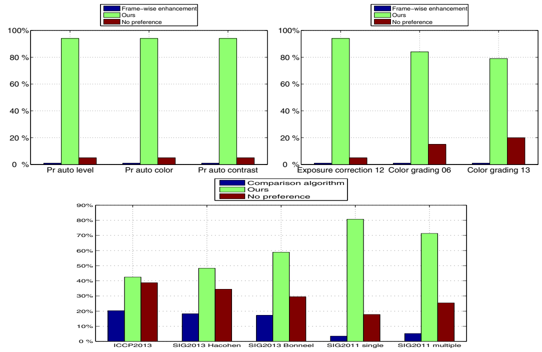
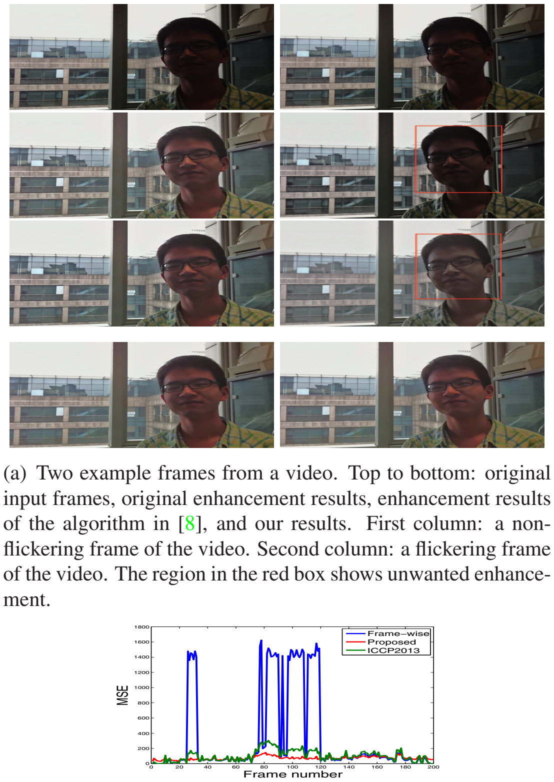


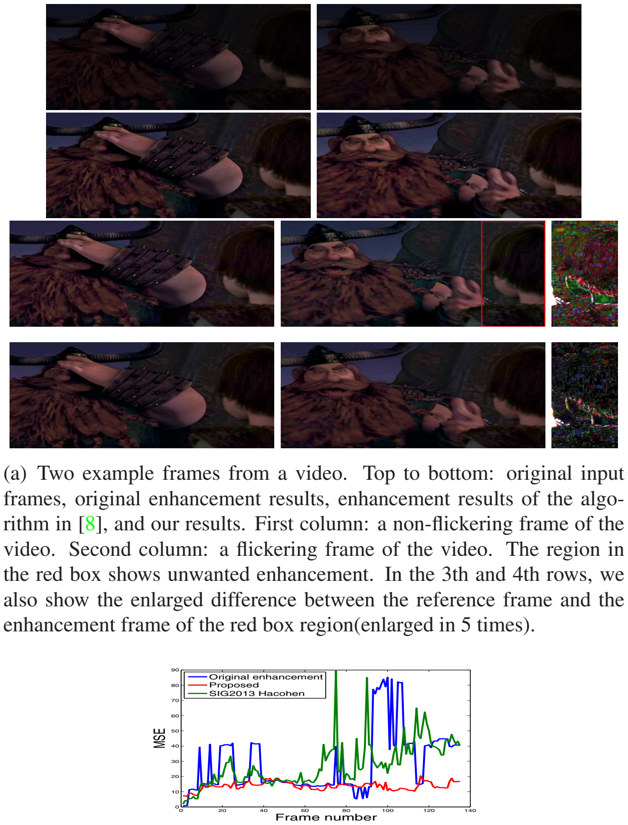
Figure 6. User study results. Pairwise comparison of ours against original enhancement, i.e., Pr auto color, auto level, auto contrast, exposure correction [22], color grading [2], color grading [4], and video post-processing algorithms, i.e., [6],[4],[8],[11]. Each color bar shows the average percentage of favored video.



(a) Two example frames from a video. Top to bottom: original input frames, original enhancement results, enhancement results of the algorithm in [8], and our results. First column: a non-flickering frame of the video. Second column: a flickering frame of the video. The region in the red box shows unwanted enhancement.

(b) Objective results of the same video.

Figure 7. Comparison with the algorithm in [8] in exposure correction enhancement [22]. The video is in the supplementary material.



(a) Two example frames from a video. Top to bottom: original input frames, original enhancement results, enhancement results of the algorithm in [8], and our results. First column: a non-flickering frame of the video. Second column: a flickering frame of the video. The region in the red box shows unwanted enhancement. In the 3th and 4th rows, we also show the enlarged difference between the reference frame and the enhancement frame of the red box region(enlarged in 5 times).

(b) Objective results of the same video.

Figure 8. Comparison with the algorithm in [11] in Pr auto level.

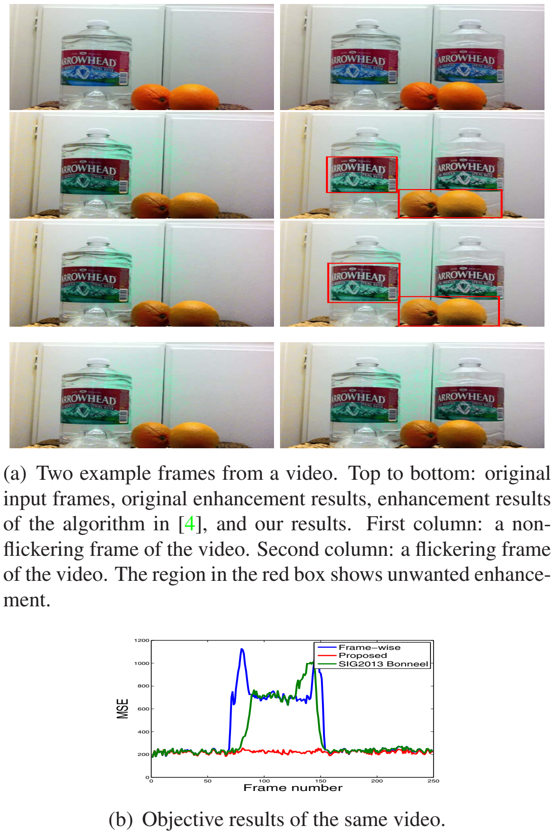
of every 30 frames is chosen as the key frame. There are 100 input videos for each original enhancement algorithm, i.e., 600 videos in total. The input videos are from movie clips and everyday life videos taken by our friends. Please see the enhancement videos of different enhancement algorithms and our algorithm in the supplementary material.

5.1. User study

We invited 12 volunteers (7 males and 5 females) to perform pairwise comparison between our result and the original enhancement result as well as the result of the temporally consistent enhancement algorithms. For each pairwise comparison, the subject had three options: better, or worse, or no preference. Subjects were 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 were randomized and unknown to each subject. This user study was conducted in the same settings (room, light, and monitor). The user study results are summarized in Fig. 6. Each color bar is the averaged percentage of the favored video over all 12 subjects. From the results, we can see that they prefer our results to the results of the original enhancement and the temporally consistent enhancement algorithms in [6], [4], [8], and [11].

5.2. Visual quality and objective comparisons

In order to objectively measure the performance of different temporally consistent algorithms, we select a small



(a) Two example frames from a video. Top to bottom: original input frames, original enhancement results, enhancement results of the algorithm in [4], and our results. First column: a non-flickering frame of the video. Second column: a flickering frame of the video. The region in the red box shows unwanted enhancement.

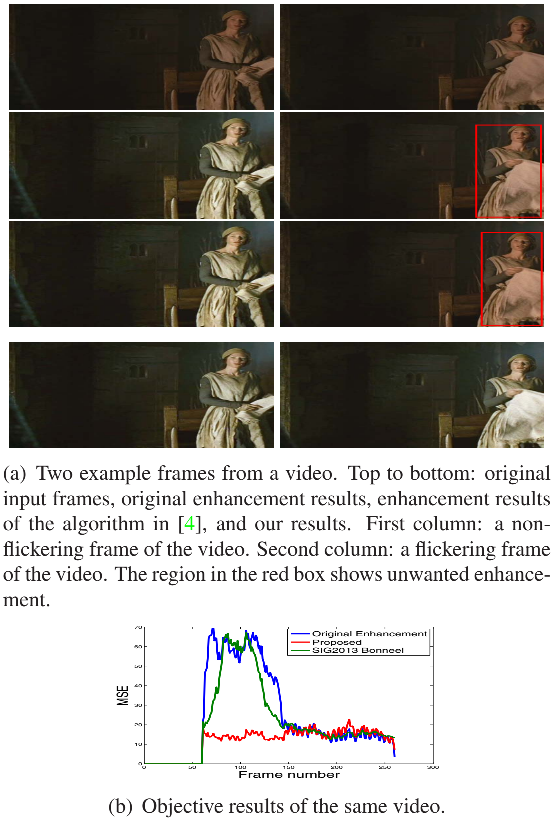
(b) Objective results of the same video.

Figure 9. Comparison with the algorithm in [4] in color grading enhancement [4].

	[8]	[11]	[4]	[6] single	[6] multiple
$\frac{\mu}{\mu_{our}}$	1.14	1.23	1.53	7.7	3.5
$\frac{\sigma^2}{{\sigma^2}_{our}}$	3.8	7.43	38.8	800	400

Table 1. The average ratios of the mean of MSE, i.e., $\frac{\mu}{\mu_{our}}$, and the variance of MSE, i.e., $\frac{\sigma^2}{{\sigma^2}_{our}}$, between the comparison methods and our method of all videos. The comparison methods include the temporally consistent enhancement algorithms in [8], [11], [4], and [6] with single key frame and multiple key frames.

data set of 30 sequences. Within each selected sequence, there are some objects which exist in most of the frames. We select one non-flickering frame as the reference frame, and in the reference frame we manually mark out those objects that exist in most frames. Then we align each frame to the reference frame and compute the difference between the reference frame and the aligned frame within the marked region, quantified by mean squared error (MSE). The mean of MSE can indicate the performance for fidelity and the variance of MSE can indicate the performance for temporal consistency. For each video, the ratio of the mean and variance of MSE between the comparison method and our method is computed. The average ratios of the mean of MSE, i.e., $\frac{\mu}{\mu_{our}}$, and the variance of MSE, i.e., $\frac{\sigma^2}{{\sigma^2}_{our}}$, of all videos are shown in Table 1. As shown, the average ratios of mean and variance of MSE between the comparison algorithm and our method are always higher than 1, which indicate our results have better fidelity and temporal consistency. The videos and more comparisons are shown in the supplementary material.



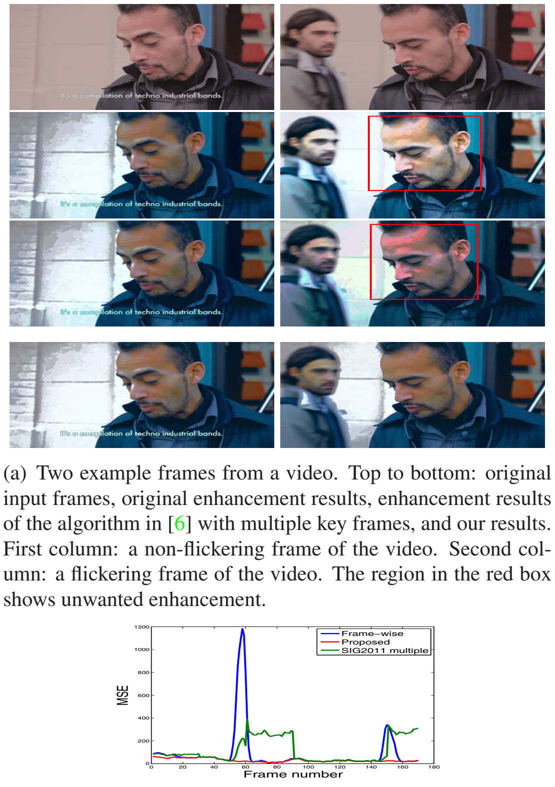
(a) Two example frames from a video. Top to bottom: original input frames, original enhancement results, enhancement results of the algorithm in [4], and our results. First column: a non-flickering frame of the video. Second column: a flickering frame of the video. The region in the red box shows unwanted enhancement.

(b) Objective results of the same video.

Figure 10. Comparison with the algorithm in [4] in Pr auto color enhancement.

Fig. 7 shows the comparison of our algorithm and the algorithm in [8]. The big discontinuities of the original enhancement in the objective result indicate that there are many flickering artifacts. The algorithm in [8] does not remove them perfectly, because they estimate a global adjustment for the original result, but the exposure correction [22] is a local algorithm. Our method performs well because the reconstruction is region-based and does not require the original enhancement algorithm to be global.

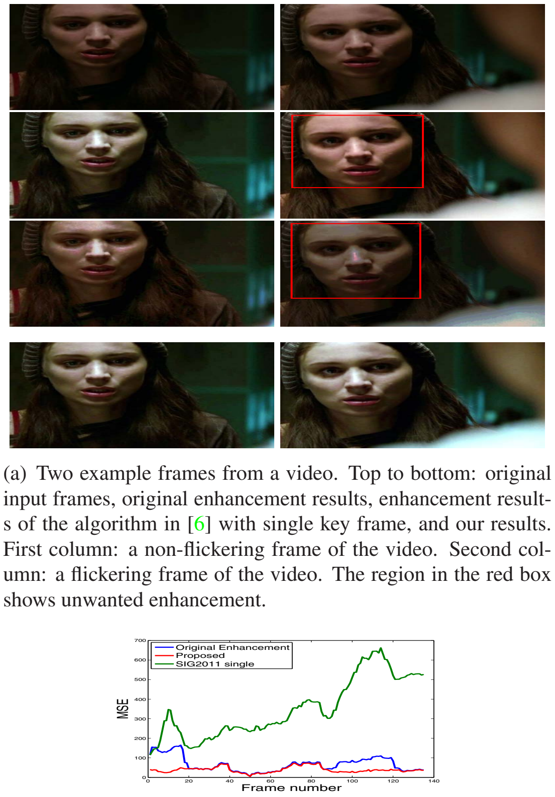
Fig. 8 shows the comparison of our algorithm and the algorithm in [11]. The algorithm in [11] does not remove flickering artifacts perfectly, since their adjustment curve is a spline with 7 knots and when the 7 knots has some errors, the errors will be enlarged to the whole dynamic range. Although the difference is not very large, for videos, the difference can be easily noticed due to temporal changes of the



(a) Two example frames from a video. Top to bottom: original input frames, original enhancement results, enhancement results of the algorithm in [6] with multiple key frames, and our results. First column: a non-flickering frame of the video. Second column: a flickering frame of the video. The region in the red box shows unwanted enhancement.

(b) Objective results of the same video.

Figure 11. Comparison with the algorithm in [6] with multiple key frames method in color grading enhancement [2].



(a) Two example frames from a video. Top to bottom: original input frames, original enhancement results, enhancement results of the algorithm in [6] with single key frame, and our results. First column: a non-flickering frame of the video. Second column: a flickering frame of the video. The region in the red box shows unwanted enhancement.

(b) Objective results of the same video.

Figure 12. Comparison with the algorithm in [6] with single key frame method in Pr auto contrast enhancement.

same objects in very short time.

The comparison with the algorithm in [4] is shown in Fig. 9 and 10. Their algorithm fails to remove the long-term flickering artifacts due to their assumption that in the flickering periods the changes of the original enhancement curves are always big. Our method can perform well in these cases because the correspondence between different frames is used for temporal consistency.

We compare with the algorithm in [6] with single key frame and multiple key frames in Fig. 12 and 11. Results of the algorithm in [6] with single key frame have big accumulated errors since each frame only consider the correspondence with its previous frame and errors of one frame will affect all the following frames. The algorithm in [6] with multiple key frames can reduce the accumulated errors. But how to select good key frames adaptively is not well solved and the simple method in our experiments will sometimes choose a flickering frame as the key frame.

6. Conclusions

In this paper, the SCE prior is discovered and experimentally verified that the enhancement of many leading algorithms is consistent in a local region. And a region-based post-processing algorithm for temporal consistency is proposed, by taking into account fidelity, temporal consistency and spatial consistency. User study, objective and visual quality comparisons demonstrate that we can keep both the fidelity and temporal consistency of the output videos.

7. Acknowledge

The work is supported by ONR N00014-12-1-0883 and NIH 5R01EY022247-03.

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012. 3

[2] S. Bae, S. Paris, and F. Durand. Two-scale tone management for photographic look. *ACM Trans. on Graph.*, 25(3):637–645, 2006. 1, 2, 3, 5, 6, 8

[3] R. Boitard, K. Bouatouch, R. Cozot, D. Thoreau, and A. Gruson. Temporal coherency for video tone mapping. *Proc. SPIE 8499, Applications of Digital Image Processing.* 2

[4] N. Bonneel, K. Sunkavalli, S. Paris, and H. Pfister. Example-based video color grading. *ACM Trans. on Graph.*, 32(4):1–11, 2013. 1, 2, 3, 5, 6, 7, 8

[5] Y. Chang, S. Saito, and M. Nakajima. Example-based color transformation of image and video using basic color categories. *IEEE Transactions on Image Processing*, 16(2):329–336, 2007. 2

[6] Z. Farbman and D. Lischinski. Tonal stabilization of video. *ACM Trans. on Graph.*, 30(4):1–9, 2011. 2, 5, 6, 7, 8

[7] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *CVPR*, 16(2):261–268, 2004. 5

[8] M. Grundmann, C. McClanahan, S. Kang, and I. Essa. Post-processing approach for radiometric self-calibration of video. *Int. Conf. Computational Photography*. 2, 5, 6, 7

[9] B. Guthier, S. Kopf, M. Eble, and W. Effelsberg. Flicker reduction in tone mapped high dynamic range video., 2011. Proceedings of the IS&T/SPIE Electronic Imaging (EI) on Color Imaging XVI: Displaying, Processing, Hardcopy, and Applications. 2

[10] Y. Hacohen, E. Shechtman, D. Goldman, and D. Lischinsky. Non-rigid dense correspondence with applications for image enhancement. *ACM Trans. Graph.*, 30. 2

[11] Y. Hacohen, E. Shechtman, D. Goldman, and D. Lischinsky. Optimizing color consistency in photo collections. *ACM SIGGRAPH*. 2, 5, 6, 7

[12] N. K. Kalantari, E. Shechtman, C. Barnes, S. Darabi, D. B. Goldman, and P. Sen. Patch-based high dynamic range video. *ACM Trans. on Graph.*, 32(6):1–8, 2013. 2

[13] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Trans. on Graph.*, 22(3):319–325, 2003. 2

[14] C. Kiser, E. Reinhard, M. Tocci, and N. Tocci. Real time automated tone mapping system for hdr video. *IEEE International Conference on Image Processing*, pages 2749–2752, 2012. 2

[15] M. Lang, O. Wang, T. Aydin, A. Smolic, and M. Gross. Practical temporal consistency for image-based graphics applications. *ACM Trans. on Graph.*, 31(4):1–8, 2012. 2

[16] C. Lee and C. Kim. Gradient domain tone mapping of high dynamic range videos. *IEEE International Conference on Image Processing*, 3:461–464, 2007. 2

[17] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. Freeman. Sift flow: Dense correspondence across difference scenes. *Proceedings of the European Conference on Computer Vision*, 3:28–42, 2008. 4

[18] R. Mantiuk, S. Daly, and L. Kerofsky. Display adaptive tone mapping. *ACM Trans. on Graph.*, 27(3):1–10, 2008. 2

[19] T. Oskam, A. Hornung, R. W. Sumner, and M. Gross. Fast and stable color balancing for images and augmented reality. *International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pages 49–56, 2012. 2

[20] D. Shapira, S. Avidan, and Y. Hel-Or. Multiple histogram matching. *IEEE International Conference on Image Processing*. 3

[21] G. Ye, E. Garces, Y. Liu, Q. Dai, and D. Gutierrez. Intrinsic video and applications. *ACM Trans. on Graph.* 2

[22] L. Yuan and J. Sun. Automatic exposure correction of consumer photographs. *Proceedings of the 12th European Conference on Computer Vision*, pages 771–785, 2012. 1, 2, 3, 5, 6, 7