# Region-based Temporally Consistent Video Post-processing

Xuan Dong[1], Boyan Bonev[2] Yu Zhu[3] Alan L. Yuille[2]
[1]Tsinghua University. [2]UC Los Angeles. [3]Northwestern Polytechnical Universi

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 [3], color grading [2], 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 [3], and color grading [1] [2]. The prior is based on the observation that in a local region, image enhancement methods tend to 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 SCE prior can be described that within a local region $i$, there will exist an enhancement curve $\alpha_i$ to reconstruct the region $i$. $\alpha_i$'s independent variable is only the intensity of the pixels and the reconstruction results should be very similar with the original enhancement results $E$, 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 $\alpha_i$. Borrowing the concept of reconstruction quality from video coding, we use Peak Signal-

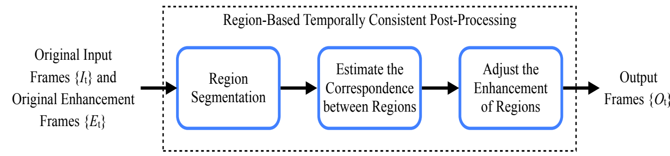


Figure 1: Pipeline of our region-based temporally consistent video post-processing.

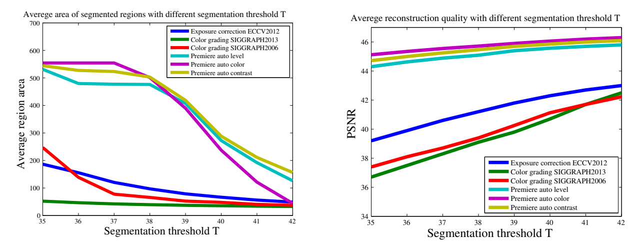


Figure 2: Average square root of the area and the reconstruction quality of segmented regions with different threshold $T$. The enhancement algorithms include Pr auto level, auto color, auto contrast, exposure correction [3], color grading [1] [2].

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.

To verify how good the prior is, we propose a SCE based segmentation method 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 threshold $T$, the average area for each region will decrease and the average reconstruction quality will increase due to the increase of the reconstruction quality requirement. As shown, all the algorithms have high average area (more than $60 \times 60$) for most $T$. 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. This gives very strong evidence for the SCE prior. Based on the SCE prior, a Markov Random Field (MRF) optimization model is proposed to adjust the enhancement of regions of all frames for temporal consistency.

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.

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

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

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

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.