Deep Video Color Propagation

Simone Meyer, Victor Cornillère, Abdelaziz Djelouah, Christopher Schroers, Markus Gross

Paper ID 30

Abstract. Traditional approaches for color propagation in videos rely on some form of matching between consecutive video frames. Colors are then propagated both spatially and temporally. These methods, however, are computationally expensive and do not take advantage of semantic information of the scene. In this work we propose a deep learning framework for color propagation that combines a local strategy, to propagate colors frame-by-frame ensuring temporal stability, and a global strategy, using semantics for color propagation within a longer range. Our evaluation shows the superiority of our strategy over existing video and image color propagation methods.

1 Introduction

Color propagation is an important problem in video processing and has many applications ranging from color modification for artistic purposes in movies to restoration and colorization of heritage footage. Furthermore, the ability to faithfully propagate colors in videos can have a direct impact on video compression.

Traditional approaches for color propagation rely on optical flow computation to propagate colors in videos either from scribbles or fully colored frames, which is computationally expensive and error prone. Inaccuracies in optical flow can lead to color artifacts which accumulate over time. Recently, deep learning methods have been proposed to take advantage of semantics for color propagation in images [1] and videos [2]. Still, these approaches have some limitations and do not yet achieve satisfactory results on video content.

In this work we propose a framework for color propagation in videos that combines local and global strategies. Given the first frame of a sequence in color, the local strategy warps these colors frame by frame based on the motion. However this local warping becomes less reliable with increasing distance from the reference frame. To account for that we propose a global strategy to transfer colors of the first frame based on semantics, through deep feature matching. These approaches are combined through a fusion and refinement network to synthesizes the final image. The network is trained on video sequences and our evaluation shows the superiority of the proposed method over image and video propagation methods as well as neural style transfer approaches, see Figure 2.

2 Approach

I order to colorize a gray scale image sequence by propagating the given color of the first frame, our proposed approach takes into account two complementary aspects: short range and long range color propagation, see Figure 1.

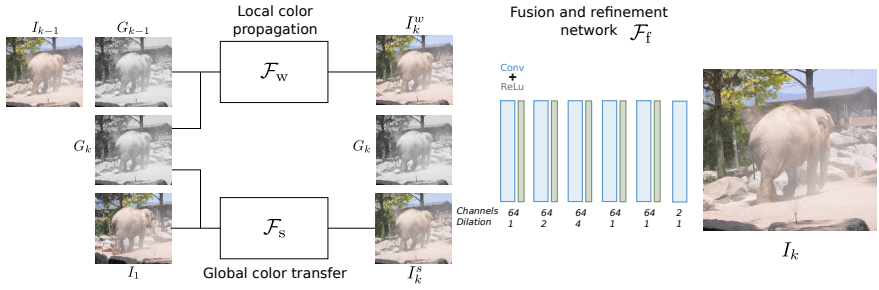


Fig. 1. Overview. To propagate colors in a video we use both short range and long range color propagation. The results of these two steps and the gray scale image are the input to the fusion and refinement network which estimates the final color frame.

The objective of the short range propagation network is to propagate colors on a frame by frame basis. It takes as input two consecutive gray scale frames and estimates a warping function. This warping function is used to transfer the colors of the previous frame to the next one. We choose to use spatially adaptive kernels that account for motion and re-sampling simultaneously [3], but other approaches based on optical flow could be considered as well.

For longer range propagation, simply smoothing warped colors according to the gray scale guide image is not sufficient. Semantic understanding of the scene is needed to transfer color from the first colored frame of the video to the rest of the video sequence. In our case, we find correspondences between pixels of the first frame and the rest of the video. Instead of matching pixel colors directly we incorporate semantical information by matching deep features extracted from the frames. These correspondences are then used in order to sample colors from the first frame. To maintain good quality for the matching, while being computationally efficient, we adopt a two stage coarse-to-fine matching. regions that have similar semantics, whereas the fine matching step considers texture-like statistics that are more effective once a region of interest has been defined. Besides the advantage for long range color propagation, this approach also helps to recover missing colors due to occlusion/dis-occlusion.

To combine the intermediate images of these two parallel stages, we use a convolutional neural network for the fusion and refinement stage. As a result, the final colored image is estimated by taking advantage of information that is present in both intermediate images, i.e. local and global color information.

3 Results

For our evaluation we used various types of videos. This includes videos from DAVIS [4, 5], as well as HD videos from the video compression dataset [6].

Ablation Study. To show the importance of both the local and global strategy, Figure 3 (a) shows an example where color propagation is not possible due to an occluding object, and a global strategy is necessary.

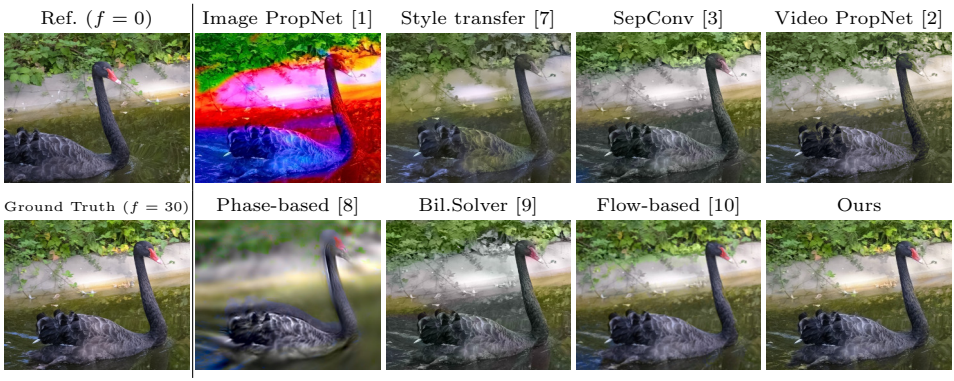


Fig. 2. Color propagation at $f = 30$. Our approach is superior to existing strategies for video color propagation.

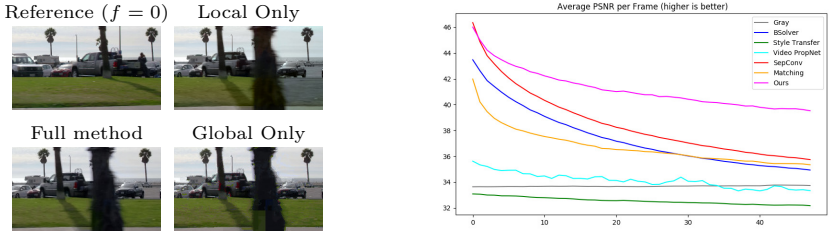
Comparisons. To show the advantage of our approach, we run compare to a large range of methods including color propagation in images [1, 9] and video [2, 8, 10] as well as photo-realistic style transfer [7], see Figure 2.

Photo-realistic style transfer methods [7] propagate colors of a reference image to replicate the global look but struggle to transfer the exact colors.

Given a partially colored image, propagating the colors to the entire image can be achieved using the bilateral space [9] or deep learning [1]. To extend these methods to video, we compute optical flow between consecutive frames [11] and use it to warp the current color image. These methods achieve satisfactory color propagation on the first few frames but the quality quickly degrades.

Relying on optical flow to propagate colors in a video is the most common approach such as Xie *et al.* [10]. However, their costly method is limiting as processing 30 HD frames requires several hours. We achieve similar or better quality in one minute. Instead of optical flow, spatially adaptive kernel can be used to account for the motion [3]. This corresponds to the local only baseline. Phase-based representation can also be used for edit propagation in videos [8]. This original approach to color propagation is however limited by the difficulty in propagating high frequencies. Recently, video propagation networks [2] were proposed to propagate information forward through a video. But by relying on standard bilateral features (i.e. colors, position, time) colors can be mixed and propagated from incorrect regions, which leads to the global impression of washed out colors. Furthermore, their performance vary largely depending on the sequence leading to a reduced numerical performance, see Figure 3(b).

Quantitative evaluation. Our test set consists of 69 videos which span a large range of scenarios with videos containing various amounts of motions, occlusions/dis-occlusion, change of background and object appearing/disappearing. In Figure 3(b) we plot the temporal behavior of the different methods, as error evolution over time averaged for all sequences. On the first frames, our results are almost indistinguishable from a local strategy but we quickly see the benefit of the global strategy. Our approach consistently outperforms related approaches for every frame and is able to propagate colors within a much larger time frame.



(a) Ablation study (b) Average error per frame as PSNR

Fig. 3. (a) Using local color propagation only preserve details but is sensitive to occlusion/dis-occlusion. Using only global color transfer does not preserve details and is not temporally stable. (b) The average error per frame shows the temporal stability of our method and its ability to maintain a higher quality over a longer period.

4 Conclusions

In this work we have presented a new approach for color propagation in videos. Thanks to the combination of a local strategy, that consists of a frame by frame image warping, and a global strategy, based on feature matching and color transfer, we have augmented the temporal extent to which colors can be propagated. Our extended comparative results show that the proposed approach outperforms recent methods in image and video color propagation as well as style transfer.

References

1. Zhang, R., Zhu, J.Y., Isola, P., Geng, X., Lin, A.S., Yu, T., Efros, A.A.: Real-time user-guided image colorization with learned deep priors. arXiv:1705.02999 (2017)
2. Jampani, V., Gadde, R., Gehler, P.V.: Video propagation networks. In: CVPR. (2017)
3. Niklaus, S., Mai, L., Liu, F.: Video frame interpolation via adaptive separable convolution. In: CVPR. (2017)
4. Perazzi, F., Pont-Tuset, J., McWilliams, B., Van Gool, L., Gross, M., Sorkine-Hornung, A.: A benchmark dataset and evaluation methodology for video object segmentation. In: CVPR. (2016)
5. Pont-Tuset, J., Perazzi, F., Caelles, S., Arbeláez, P., Sorkine-Hornung, A., Van Gool, L.: The 2017 davis challenge on video object segmentation. arXiv:1704.00675 (2017)
6. Wang, H., Katsavounidis, I., Zhou, J., Park, J., Lei, S., Zhou, X., Pun, M.O., Jin, X., Wang, R., Wang, X., et al.: Videoset: A large-scale compressed video quality dataset based on jnd measurement. JVCI (2017)
7. Li, Y., Liu, M.Y., Li, X., Yang, M.H., Kautz, J.: A closed-form solution to photo-realistic image stylization. arXiv:1802.06474 (2018)
8. Meyer, S., Sorkine-Hornung, A., Gross, M.: Phase-based modification transfer for video. In: ECCV. (2016)
9. Barron, J.T., Poole, B.: The fast bilateral solver. In: ECCV. (2016)
10. Xia, S., Liu, J., Fang, Y., Yang, W., Guo, Z.: Robust and automatic video colorization via multiframe reordering refinement. In: ICIP. (2016)
11. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime tv-l 1 optical flow. In: Joint Pattern Recognition Symposium. (2007)