Eurographics Workshop on Intelligent Cinematography and Editing (2015)
W. Bares, M. Christie, and R. Ronfard (Editors)

Supplemental Material for Key-frame Based Spatiotemporal Scribble Propagation

Pelin Doğan[1], Tunç Ozan Aydın[2], Nikolce Stefanoski[2], Aljoscha Smolic[2]

[1]ETH Zürich, [2]Disney Research Zürich

This is the supplemental material for the work entitled *Key-frame Based Spatiotemporal Scribble Propagation*.

1. GUI

The GUI that allows the user to load key-frames of a video interactively, draw scribbles and conveniently run our propagation method is shown in Figure 1.

Figure 1: *Our user interface for scribble drawing.*

2. Details on Spatial Propagation

2.1. Permeability-Guided Filtering (PGF)

The PGF approach of Aydin et al. [ASC*14] involves iterative application of a 1-dimensional filter along image rows and image columns in a separable way. First an image $J^{(1)}$ is computed by filtering all rows, then image $J^{(1)}$ is filtered along columns and image $J^{(2)}$ is obtained. The method continues alternating between filtering rows and columns until the filtering result converges. In practice, 10 iterations are sufficient for filtering natural images.

A sequence $I_1,\ldots,I_N$ representing values of a color channel of an image row or column is filtered as according to

$$J_i^{(k+1)} = \frac{I_i + \sum_{j=1, j\neq i}^{N} \pi_{ij} J_j^{(k)}}{\sum_{j=1}^{N} \pi_{ij}}, \quad J_i^{(0)} := I_i \tag{1}$$

where I_i is the original image, $J_i^{(k)}$ is the filtering result after k iterations, and $\pi_{ij} \in [0,1]$ is a permeability weight indicating the level of *permeability* between pixels i and j. A permeability of $\pi_{ij} \approx 1$ is interpreted as high permeability. The permeability between two non-neighboring pixels i and j is computed as the product of permeabilities between neighboring pixels that lie on the interval between these pixels, i.e.

$$\pi_{ij} := \begin{cases} 1 & : \quad i = j \\ \prod_{r=i}^{j-1} \pi_{r,r+1} & : \quad i < j \\ \prod_{r=j}^{i-1} \pi_{r,r+1} & : \quad i > j \end{cases}. \tag{2}$$

Thus, a small permeability between neighboring pixels along the way from pixel i to pixel j leads to a low overall permeability π_{ij} and a low diffusion of pixel $J_j^{(k)}$ into $J_i^{(k+1)}$ (cp. Eq. 1). This property is exploited for reducing or even stopping diffusion between certain pixels.

The permeability between neighboring pixels $\pi_{r,r+1}$ is computed based on a variant of the Lorenzian stopping function according to

$$\pi_{r,r+1} := \left(1 + \left|\frac{\bar{I}_r - \bar{I}_{r+1}}{\sigma}\right|^{\alpha}\right)^{-1} \tag{3}$$

where $\bar{I}_r$ indicates the gray value of I, α is a shape parameter, and σ is a scaling parameter. Note that intensity differences between neighboring pixels greater than σ are mapped to permeability weights $\pi_{r,r+1}$ that are smaller than 0.5. Hence, the permeability weights are computed in such a way that a strong spatial diffusion is only achieved if partial derivatives have a magnitude smaller than σ.

2.2. Effect of the Normalization Term

The effect of the normalization term is illustrated in Figure 2.

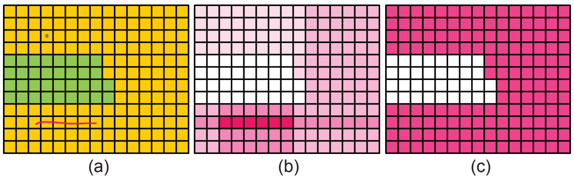

Figure 2: *Illustration of propagation through normalized filtering. For a given image and input scribble (a), the energy of the permeability filtering result generally decreases with increasing distance from the input scribble (b). The normalization term ensures that the scribble colors remain constant throughout the propagated region (c).*

2.3. Metrics for Hybrid Scribble Propagation

Entropy metric

In information theory, entropy is a quantitative measure of uncertainty for a random variable. The term usually refers to *Shannon entropy* which can be described as the average unpredictability of a random variable [Iha93].

If we treat color values of the pixels as random variables and the color histogram within a local region as the probability distribution, we can obtain an entropy measure reflecting the randomness of the color values. Thus, for each pixel of the image, we can compute a *local entropy* by considering its neighborhood as follows:

$$E_{local,p} = \sum n \log_2 n_p, \tag{4}$$

where n contains the local histogram counts in the neighborhood of the pixel p.

L-moments metric

L-moments are expectations of certain linear combinations of order statistics and can be defined for any random variable whose mean exists. The main advantage of L-moments over conventional moments is that L-moments suffer less from the effects of sampling variability since they are the linear functions of the data. Thus, L-moments are more robust to outliers in the data and allow more secure inferences to be made from small samples about an underlying probability distribution [Hos90].

3. Details on Temporal Propagation

The illustration of temporal propagation is shown in Figure 3.

4. Additional Results

For another comparison, we ran the method of Lang et al. [LWA*12] on the video sequence from Figure 5 in the main paper using the same set of scribbles with which we generated our method's result. The keyframes were 30 frames apart for both methods. Figure 4 shows that in their results the scribble colors are not always confined within object boundaries. This is especially apparent near the fish with the yellow scribble, but also noticeable near the fish with the magenta scribble (similar artifacts have recently been reported by Ye et al. [YGl*14]). On the other hand, the green scribble does not propagate throughout the whole fish leaving holes as a result. In comparison, our method's result does not suffer from such problems in this example.

It is worth noting that while the video sequence in Figure 4 has complex texture, the motion of the fish is rather slow and linear.

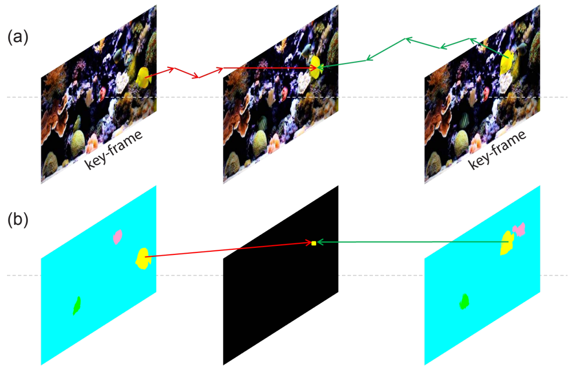

Figure 3: *Temporal propagation of scribble colors from key frames. For each pixel of a given frame, we first determine the corresponding pixels of the nearest two key-frames by following the motion paths (a). Next, we propagate the scribble colors from these corresponding points of the key-frames to the current frame.*

References

[ASC*14] AYDIN T. O., STEFANOSKI N., CROCI S., GROSS M., SMOLIC A.: Temporally coherent local tone mapping of hdr video. In *ACM TOG* (2014). 1

[Hos90] HOSKING J. R. M.: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological) 52*, 1 (1990), pp. 105–124. 2

[Iha93] IHARA S.: *Information theory for continuous systems*, vol. 2. World Scientific, 1993. 2

[LWA*12] LANG M., WANG O., AYDIN T., SMOLIC A., GROSS M. H.: Practical temporal consistency for image-based graphics applications. *ACM TOG 31*, 4 (2012). 2

[YGl*14] YE G., GARCES E., LIU Y., DAI Q., GUTIERREZ D.: Intrinsic video and applications. *ACM TOG 33*, 4 (2014). 2

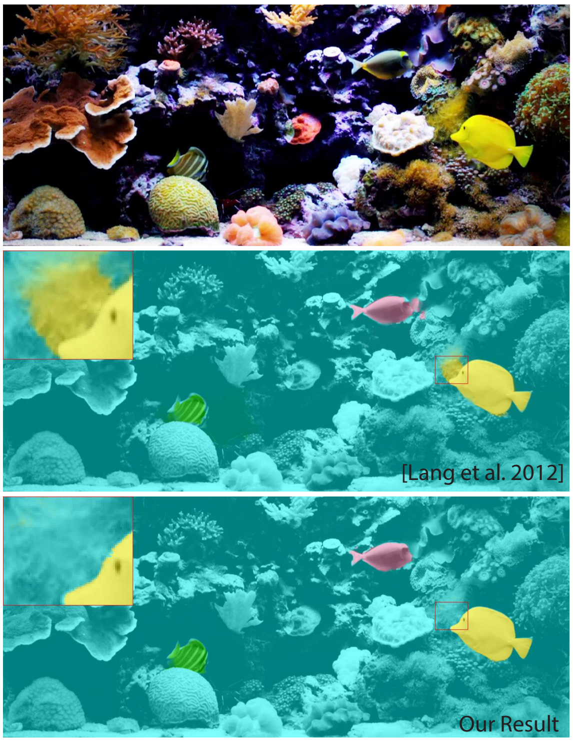

Figure 4: *A comparison of our result with Lang et al. [2012]. Note the leaked colors near the boundaries of the yellow and magenta fish. Key-frames are set 30 frames apart.*