Motion-Aware KNN Laplacian for Video Matting

Dingzeyu Li[1], Qifeng Chen[2], Chi-Keung Tang[3]

[1]Columbia University, [2]Stanford University, [3]The Hong Kong University of Science and Technology

Motivation

The fundamental problem to solve in video matting is to produce spatio-temporally coherent clusters of moving foreground pixels. There are several difficulties:

- Intensive Preprocessing

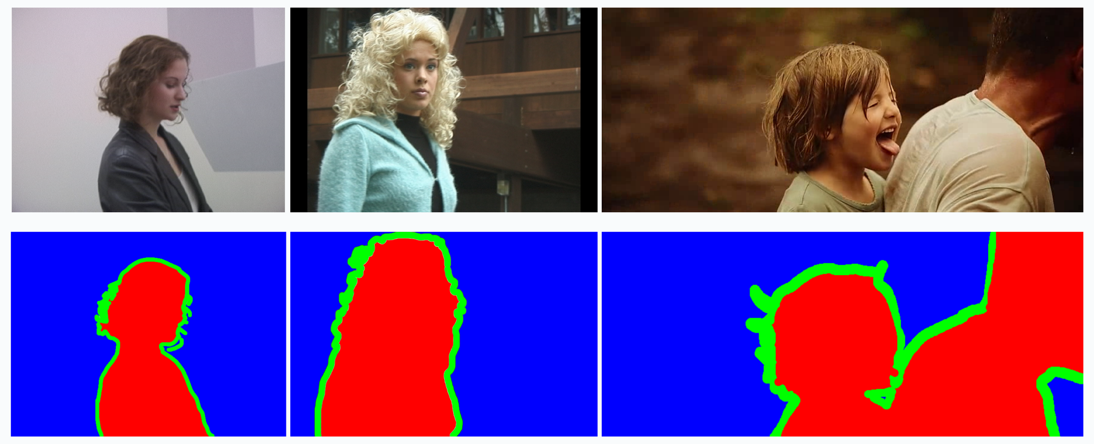

Figure 1 : Traditional methods require every input frame with trimap mark-ups, which could be labor intensive.

- Temporal Incoherence: Per-frame based algorithms produce jittering artifacts, relying on postprocessing or user interaction to mitigate the undesired effects.

Our Work

- Two-Frame Laplacian & Nonlocal Neighbor Selection

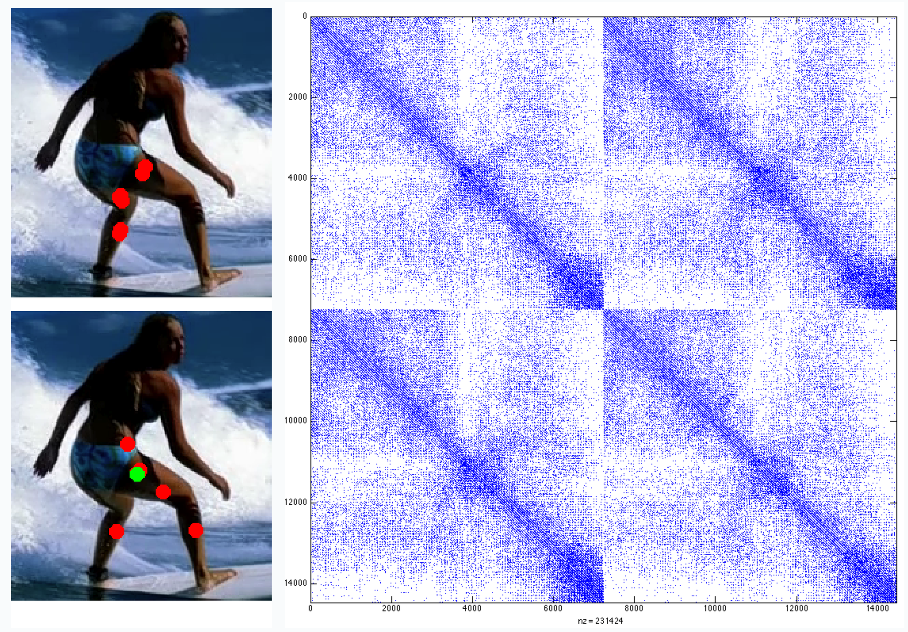

Figure 2 : Left shows K nearest neighbors (red) of the selected point (green); right shows a typical sparse nonlocal two-frame affinity matrix A in KNN video matting.

- Motion-Aware Feature Vector

The feature vector $X(i)$ at a pixel i in frame t is:

$$X_t(i) = (\underbrace{\lambda_s(x,y)}_{\text{Spatial}} \quad \underbrace{\lambda_f(u_f,v_f,u_b,v_b)}_{\text{Optical Flow}} \quad \underbrace{P(i,\lambda_p)}_{\text{Patch}})_t \qquad (1)$$

The objective function defined on alpha value x :

$$g(x) = x^T L x + \lambda \left[\sum_{i\in \mathbf{m}_b} x_i^2 + \sum_{i\in \mathbf{m}_f} (1-x_i)^2 \right] \qquad (2)$$

The optimal solution is

$$x = (L+\lambda D)^{-1}(\lambda \mathbf{m}_f). \qquad (3)$$

Analysis

In feature vector (1), λ_s controls the amount of spatial coherence, λ_f the influence of optical flow, and λ_p the size of an image patch. K is the number of nearest neighbors for nonlocal matching.

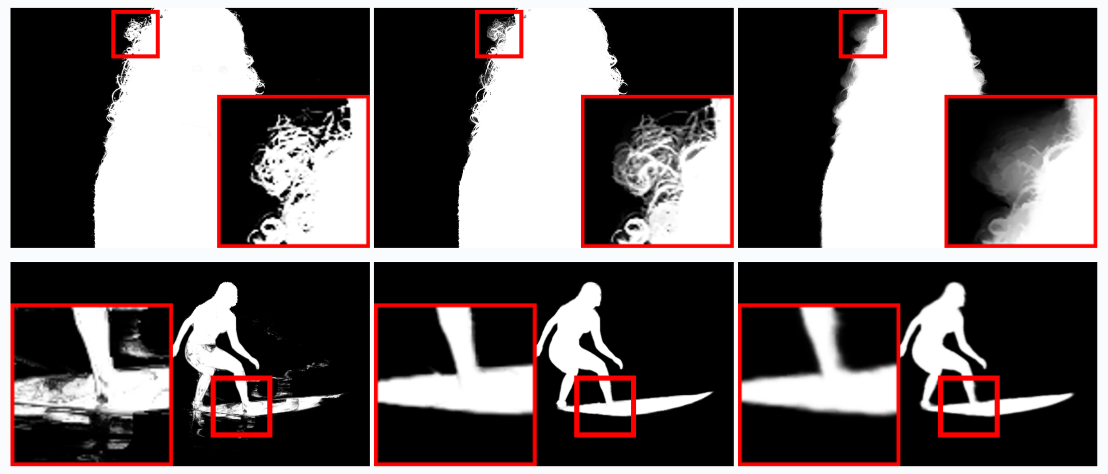

Figure 3 : Analysis on parameter λ_s.

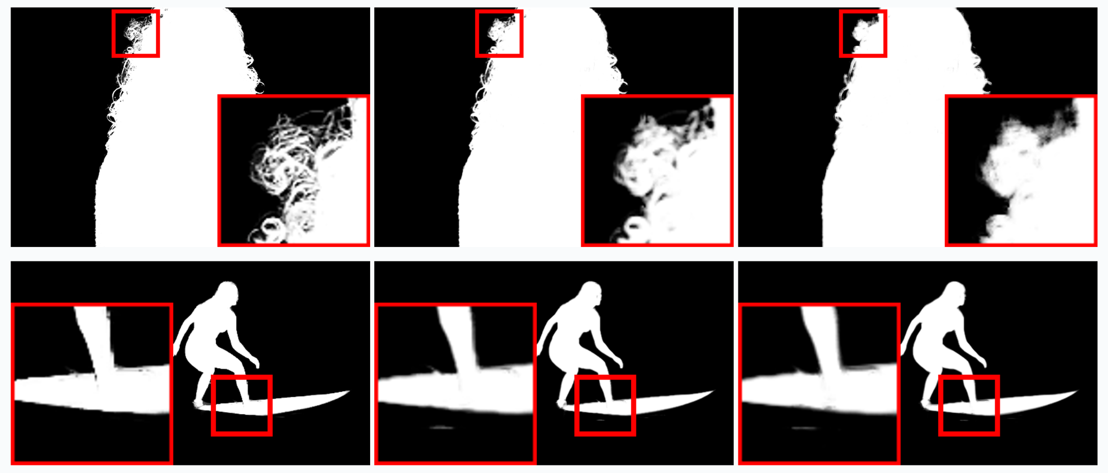

Figure 4 : Analysis on parameter λ_p.

Figure 5 : Analysis on parameter λ_f.

Experiments

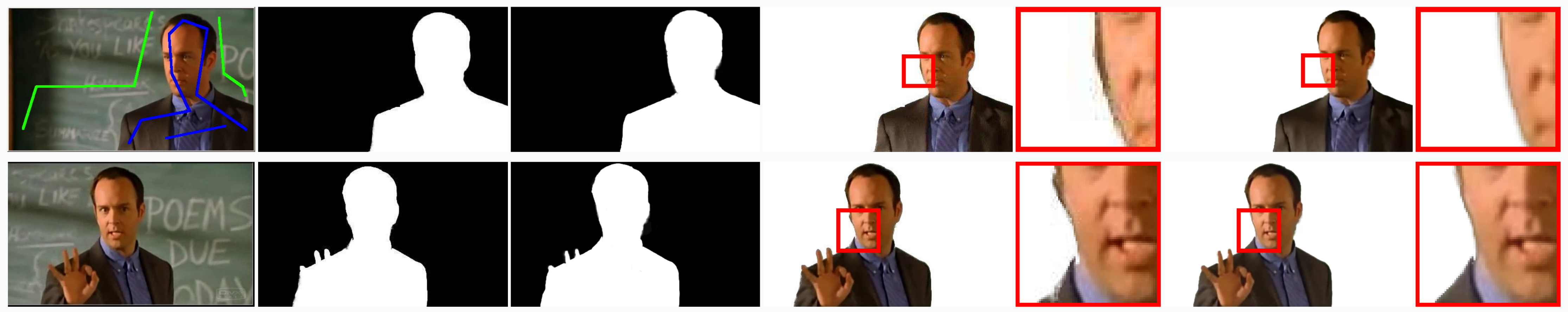

Figure 6 : Comparison with geodesic matting [1] on *talk* using sparse strokes. Only strokes on the first frame are given and all the αs are computed using our closed-form solution. While the α results look similar, αI shows our method extracts a better foreground.

Figure 7 : Comparison with video snapcut on *walk*. Our results (bottom) are robust to stark illumination changes given only a *single* input trimap. The shading on the walking man is constantly changing. In video snapcut (top), the user needs to supply quite a number of additional strokes to achieve a comparable segmentation, for example, by carefully drawn control points on Frame 15 and 51 as well as blue strokes on the intermediate frames.

More examples are available in the paper and the accompanying video.

Quantitative Comparison

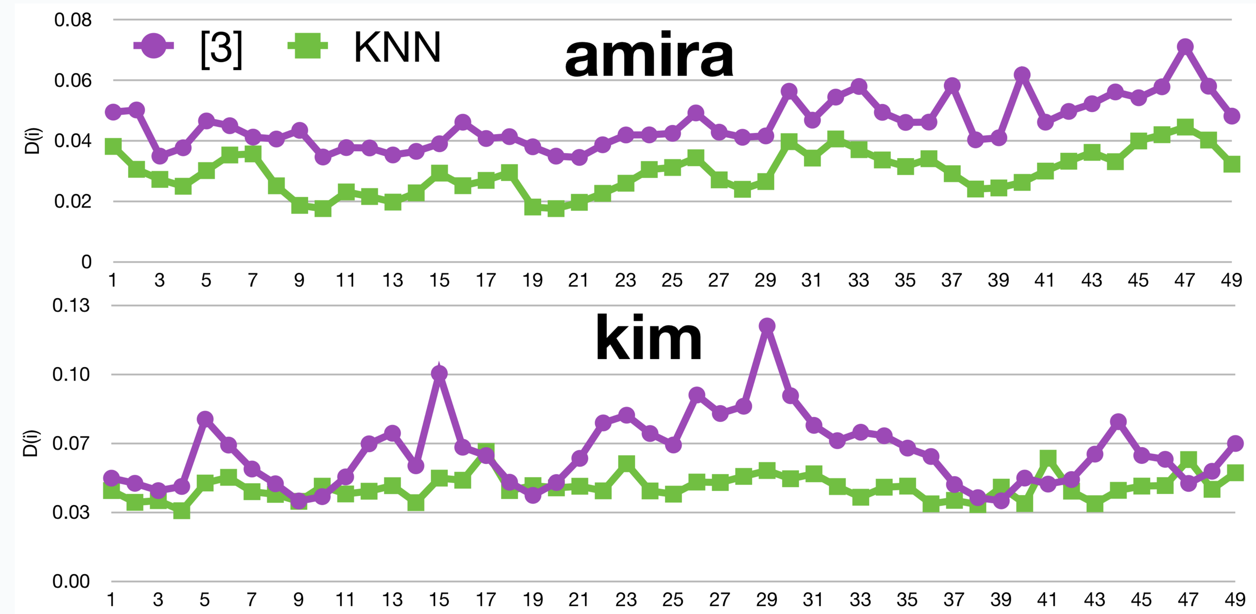

Figure 8 : Quantitative comparison with nonlocal video matting [3].

Conclusions

Our contributions:

- Apply ***nonlocal principle*** to video matting.
- Embed ***motion information*** directly into feature vector.
- Propose a ***two-frame*** affinity matrix.
- Achieve competitive results with ***sparse*** user inputs.

Limitations and future work:

- Enable the keyframes to propagate in both directions.
- Improve the robustness by looking into useful features.

Acknowledgements

Input and output of [1] are courtesy of X. Bai. Input and output of [3] are courtesy of Y.-W. Tai. The research [4] was supported by the Hong Kong Research Grant Council under grant number 619313.

Reference

[1] X. Bai and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. In ICCV, pages 1–8, 2007.
[2] Q. Chen, D. Li, and C.-K. Tang. KNN matting. In CVPR, pages 869–876, 2012.
[3] I. Choi, M. Lee, and Y.-W. Tai. Video matting using multi-frame nonlocal matting laplacian. In ECCV, pages 540–553, 2012.
[4] D. Li, Q. Chen, and C.-K. Tang. Motion-Aware KNN Laplacian for Video Matting. In ICCV, 2013.

Links

Project Website:
`http://tinyurl.com/knnvideomatting`
or
Author's Website:
`http://www.cs.columbia.edu/~dli/`

dli@cs.columbia.edu