Temporally Coherent Video Matting

Sun-Young Lee[*] Jong-Chul Yoon [†] In-Kwon Lee[‡]
Dept. of Computer Science, Yonsei University, Korea

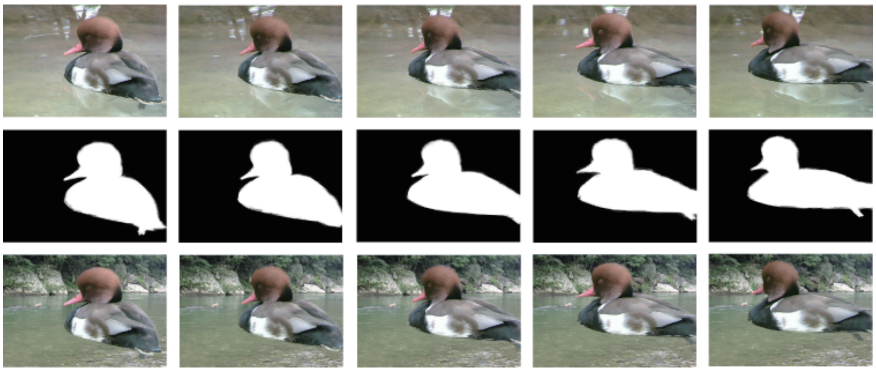

Figure 1: *Video matting results (a total of 80 frames, of which frames 20, 40, 60 and 80 are shown): The upper row of each sequence is the original video; the middle row is the alpha matte sequence, and the lower row is a sequence composited with a new background image.*

1 Introduction

Existing video matting approaches determine the alpha matte sequence frame-by-frame, which lead to flickering near the boundary of the foreground region. We reduce this effect by considering video data as a spatio-temporal cube, and extending a robust matting algorithm to a 3D solver. Our results demonstrate consistent and visually pleasing alpha mattes, and tend to preserve temporal coherence better than previous techniques.

2 Description

We extend an existing robust matting algorithm [Wang and Cohen] from 2D to a 3D lattice by considering the time axis as a third spatial dimension. We will assume that the trimap necessary for the video to be matted has already been generated. Our system can utilize any method of trimap generation or refinement. An accurate and temporally coherent series of trimaps will naturally improve the overall performance of the algorithm. In matting the current frame, we consider three consecutive frames, the previous, current, and next. We select several samples from each frame (20 samples per frame in all our examples). The confidence and alpha values for each pair of foreground and background samples are estimated using the technique described for image matting, and the three pairs with the highest confidence values are selected. We also need to extend the random-walk algorithm to three dimensions. We therefore construct a 3D graph in (x, y, t) space, and use a $3 \times 3 \times 3$ kernel to compute the edge weights W_{ij}.

Instead of the cubic $3 \times 3 \times 3$ kernel, we can use an anisotropic kernel which takes account of the movement of the foreground object. Using an optical flow vector field, we can obtain the *flow distance* $\mathbf{d}$, which is the vector difference between the optical flow vector and the edge vector of the node. The vector $\mathbf{d}$ measures how for the video object moves between the current frame and the next or previous frame. We can then distort the kernel in inverse proportion to the magnitude of $\mathbf{d}$. An anisotropic kernel produces more consistent results in the temporal axis. So Equation may be rewritten as:

$$W_{ij}=\frac{m}{|\mathbf{d}|}\sum_{k|(i,j)\in w_k}\frac{1}{|w_k|}\left(1+(C_i-\mu_k)\left(\Sigma_k+\frac{\epsilon}{|w_k|}I\right)^{-1}(C_j-\mu_k)\right), \quad (1)$$

where w_k is the kernel that contains pixels i and j, and k iterates over those kernels. The terms μ_k and Σ_k are respectively the mean and variance of the colors in each kernel, and $|w_k|$ is the number of pixels in the kernel w_k. We typically set ϵ to be 10^{-5}. Parameter m controls the influence of the *flow distance*. We use values of m in the range $[0.1, 2.5]$. If a video has a fast-moving foreground object, m is larges and the alpha values are more accurate, since $|\mathbf{d}|$ also becomes large.

3 Result and Future Work

A matting result is shown in Figure 1 as sequences of images sampled every few frames. (Please see also the results in the supplementary videos.) Our results demonstrate consistent and visually pleasing alpha mattes, and tend to preserve temporal coherence better than frame-by-frame techniques.

Our system could be further improved by employing an advanced color model and sampling method which were specialized to the time axis. In addition, when the user edits selected frames, it should be possible to generate new trimaps and mattes interactively, by taking advantage of the locality of the user editing procedure. We would like to develop an integrated tool with a user interface, and to construct it with the help of a users.

References

WANG, J., AND COHEN, M. F. Optimized color sampling for robust matting. In *IEEE Computer Vision and Pattern Recognition 2007*.

Acknowledgements

This work was supported by the IT R&D program of MKE/MCST/IITA. [2008-F-031-01, Development of Computational Photography Technologies for Image and Video Contents]

[*]e-mail: shepherd@cs.yonsei.ac.kr
[†]e-mail: media19@cs.yonsei.ac.kr
[‡]e-mail: iklee@yonsei.ac.kr

SIGGRAPH 2010, Los Angeles, California, July 25 – 29, 2010.
ISBN 978-1-4503-0210-4/10/0007