Video Patch Replacement

Igor Guskov*
EECS Department, University of Michigan, Ann Arbor

Abundance of video data calls for more easy-to-use video editing tools. A basic editing operation is to paste a user-specified portion of another image or video onto the existing video footage. This works well for images, however for video sequences the problem is harder because usually one has to estimate the motion of the underlying object shapes, changes in lighting conditions, and possible occlusions. This sketch introduces a simple user-controlled video editing system that tracks and replaces patches of video sequences with a specified replacement texture.

Our main contribution is to introduce a multi-level template tracking procedure extending the approach of [Guskov 2006], and show its applications within a video patch pasting system with simple user interface. Our system can handle large frame-to-frame motions of the underlying shapes and imperfect matching of the template regions. The final rendered result uses Poisson-based gradient fitting method to blend the replaced portion seamlessly with the rest of the sequence.

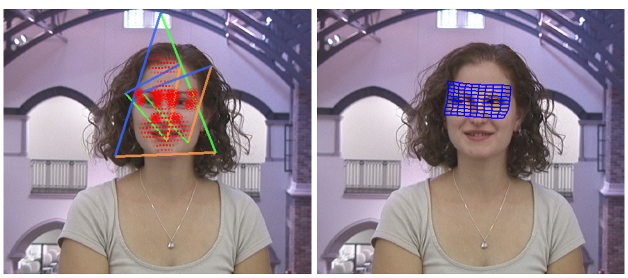

Figure 1: Left: tracking targets, right: warped replacement grid.

Recently, there appeared several approaches for superimposing of images onto a video sequence. Most of them track the underlying geometry and warp a mesh based on the estimated motion. The main difference lies in the tracking capabilities of the proposed methods: Pilet et al. [2005] use a fast keypoint detection procedure for robust tracking of non-rigid surfaces, with impressive real-time performance. However this approach requires a lengthy preprocessing stage for every template, which reduces its attractiveness for interactive editing. It is not clear how the method performs for imperfectly matched texture regions. Bartoli et al. [2004] propose to use a non-rigid factorization for the set of correspondences established by point and curve feature tracking. For near-regular textures, the work of Lin [2005] uses an MRF based approach for superimposing images onto tracked geometry. The main difference between these approaches and our work is the ability of our system to handle noisy and imperfectly matching features during the tracking process.

In our approach, a user session starts by specifying *template target masks* for the image regions to be tracked and *a replacement cutout mask* for the image region to be replaced both done using a brush tool on a frame from the input video sequence. Then, a tracking routine is invoked that performs template alignment for the specified template collection for all the frames of the input video sequence. The tracking results can be visualized by the motion of individual templates, or by the corresponding warping of the replacement patch. The warp is computed by blending the transformations of all the relevant templates, with weights determined locally by the distances from the particular cutout patch point to each template area. A grid covering replacement patch is warped by storing the barycentric coordinates of each grid vertex with respect to each target triangle. In each consecutive frame, new image plane positions of grid vertices are computed with respect to warped targets, and blended using the weights determined in the first frame, similar to the morphing approach of [Beier and Neely 1992]. Some of the tracking targets do not affect the cutout warping directly; instead they are used for a more robust tracking of the collection of targets. For instance, in Figure 1 the two smaller templates are used for warping the replacement patch, whereas the bigger template with coarsely sampled blob texture is used to improve the robustness of tracking under large motions.

Once the user is happy with the tracked collection, she can replace the cutout image by an arbitrary part of another image. The user specifies a warp between the replacement image and the cutout patch with the help of a warping grid adjusted on the replacement image plane. Finally, the system renders the video with the cutout region replaced by an adjusted version of the replacement image patch. The adjusted color mask matches the gradients of the warped replacement image inside the warped cutout region and satisfies the boundary conditions coming from the input video frame [Pérez et al. 2003]. We also find it necessary to perform a user-specified rescaling of the gradients to adjust the contrast of the replacement image to the contrast of the input video sequence. Since our motion estimation procedure is able to handle blurred video sequences, we also added the capability to add motion blur to the rendered replacement patch.

We evaluated performance of our system on several replacement tasks shown in the accompanying video. The current bottleneck of the sequence is the tracking stage: currently a single template propagation in one frame takes 1.5 seconds on average (measured on a 3GHz Pentium 4 Windows XP workstation). On a sixty-frame video clip, a typical user session would include a minute of interactive mask specialization, followed by a minite or two of fully automatic tracking. After the tracking, the user can try several replacement images or video sequences in a fully interactive mode, varying the parameters of the Poisson fill and motion blur.

Acknowledgements This work was supported in part by NSF (CCR-0133554), and by the NSF ERC-RMS (award EEC-9529125). Example video sequences are courtesy of Iain Matthews and Aseem Agarwala.

References

BARTOLI, A., VON TUNZELMANN, E., AND ZISSERMAN, A. 2004. Augmenting images of non-rigid scenes using point and curve correspondences. In *Proceedings of CVPR*, 699–706.

BEIER, T., AND NEELY, S. 1992. Feature-based image metamorphosis. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, 35–42.

GUSKOV, I. 2006. Kernel-based template alignment. To appear in Proceedings of CVPR 2006.

LIN, W.-C. 2005. *A Lattice-Based MRF Model for Dynamic Near-Regular Texture Tracking and Manipulation*. PhD thesis, Carnegie Mellon University.

PÉREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Trans. Graph. 22*, 3, 313–318.

PILET, J., LEPETIT, V., AND FUA, P. 2005. Real-time non-rigid surface detection. In *Conference on Computer Vision and Pattern Recognition, San Diego, CA*.

*e-mail: guskov@eecs.umich.edu