

High-Quality Depth Recovery via Interactive Multi-View Stereo

Supplementary Document

Weifeng Chen¹ Guofeng Zhang^{1*} Xiaojun Xiang¹ Jiaya Jia² Hujun Bao¹
¹State Key Lab of CAD&CG, Zhejiang University ²The Chinese University of Hong Kong

1. Geometry Coherence Constraint

The geometric coherence constraint $p_v(\cdot)$ is defined as

$$p_v(\mathbf{x}, D_t(\mathbf{x})) = \frac{\sigma_d^2}{\sigma_d^2 + \|P_{t' \rightarrow t}(D_{t'}(\mathbf{x}')) - D_t(\mathbf{x})\|^2}, \quad (1)$$

where σ_d is a parameter, and \mathbf{x}' is the projected point in frame t' . $P_{t' \rightarrow t}(D_{t'}(\mathbf{x}'))$ denotes the transformed disparity value according to the camera parameters between frames t' and t . The transformed disparity is denoted as $P_{t' \rightarrow t}(D_{t'}(\mathbf{x}'))$. We denote the projective matrix from frame t to frame t' as $[R|t]$, and the intrinsic matrix of frame t' as K . The 2D position of \mathbf{x} is denoted as (u, v) . Then $P_{t' \rightarrow t}(D_{t'}(\mathbf{x}'))$ can be computed by

$$P_{t' \rightarrow t}(D_{t'}(\mathbf{x}')) = \frac{1}{(\frac{1}{D_{t'}(\mathbf{x})}RK^{-1}(u, v, 1)^\top + T)[3]}, \quad (2)$$

where $[3]$ denote the third element of the vector. The geometric coherence constraint requires that $P_{t' \rightarrow t}(D_{t'}(\mathbf{x}'))$ equal to $D_t(\mathbf{x})$. Different from the method of [1], our geometric coherence term is defined in the disparity space instead of image space, thus the computation is faster and more robust for forward motion or small translational motion sequences.

2. Depth Repairing with Imperfect Segmentation Results

Here, we would like to demonstrate that our method does not require very fine object segmentation with precise boundaries. Our spatio-temporal depth optimization can automatically correct it and achieve spatio-temporally consistent depth values around discontinuous boundaries. One example is shown in Figure 1. Figure 1(b) shows that the segmentation result contains obvious artifacts and not very consistent in some frames. Even with this problematic segmentation result, the erroneous depth information can be effectively corrected by our method, as shown in Figure 1(d).

3. Intermediate results of Area Masking

Here, we show some intermediate results to demonstrate the effectiveness of the proposed area masking method. The yellow strokes in the figures denote the drawn foreground strokes, and the blue ones denote the background strokes. The red strokes are used to indicate the occlusion boundary. Generally, the user select one frame to start masking a repaired region. The segmented area mask will then be automatically propagated to the neighboring frames. If the propagation result contain significant error, the user can draw a few additional strokes to correct it and then propagate the result to the following frames again.

Figures 2 and 3 show all the edited key frames for two key repaired regions in “Stela” example. As shown in Figure 2, the repaired region between two stelas is occluded seriously. We select the 99th frame as the starting edited frame (highlighted with red rectangle) and draw a red stroke to indicate the occlusion boundary. With occlusion handling, the user only needs to edit 22 frames in total for this sequence with 143 frames. Figure 3 shows the edited 12 key frames for another repaired region.

*email: zhangguofeng@cad.zju.edu.cn

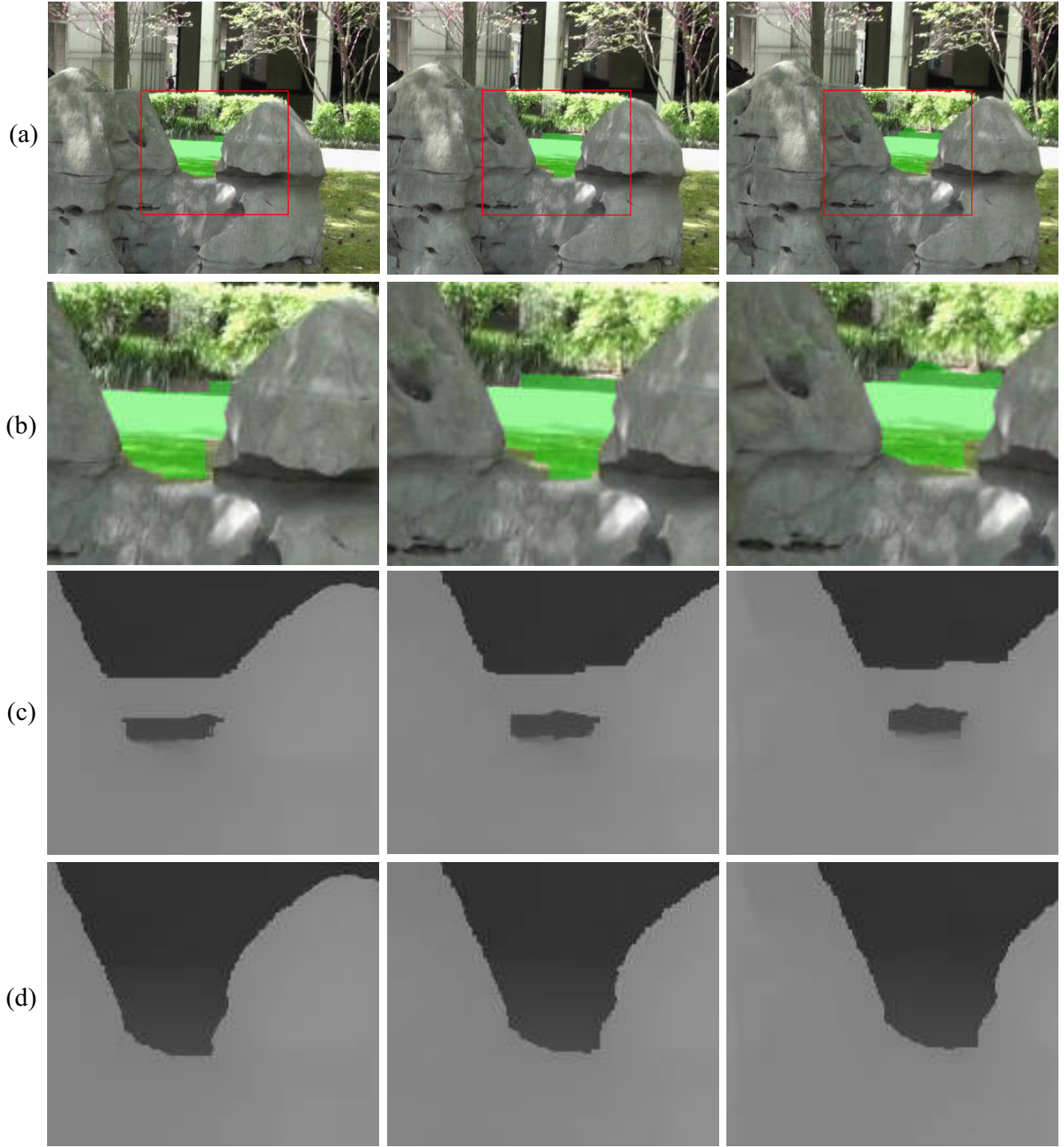


Figure 1. Depth repairing with imprecise segmentation results. (a) Some selected frames with marked regions. (b) The magnified regions of (a). (c) The depth maps recovered by ACTS. (d) The refined depth maps by our system.

Figure 4 shows the edited key frames for “Car” example. The black car has strong specular reflection and similar colors with the background, which is very challenging for segmentation propagation. Even in this challenging case, the user only needs to draw a few additional strokes in 24 key frames to mask the black car for this sequence with 121 frames.

Figures 5 and 6 show the edited key frames for two repaired regions in the KITTI sequence 000070.



Figure 2. Edited keyframes for one repaired region in “Stela” sequence. Top: the edited 22 key frames. The frame with red rectangle is the starting edited frame. Bottom: the segmentation results.

References

- [1] G. Zhang, J. Jia, T.-T. Wong, and H. Bao. Consistent depth maps recovery from a video sequence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(6):974–988, 2009. **1**

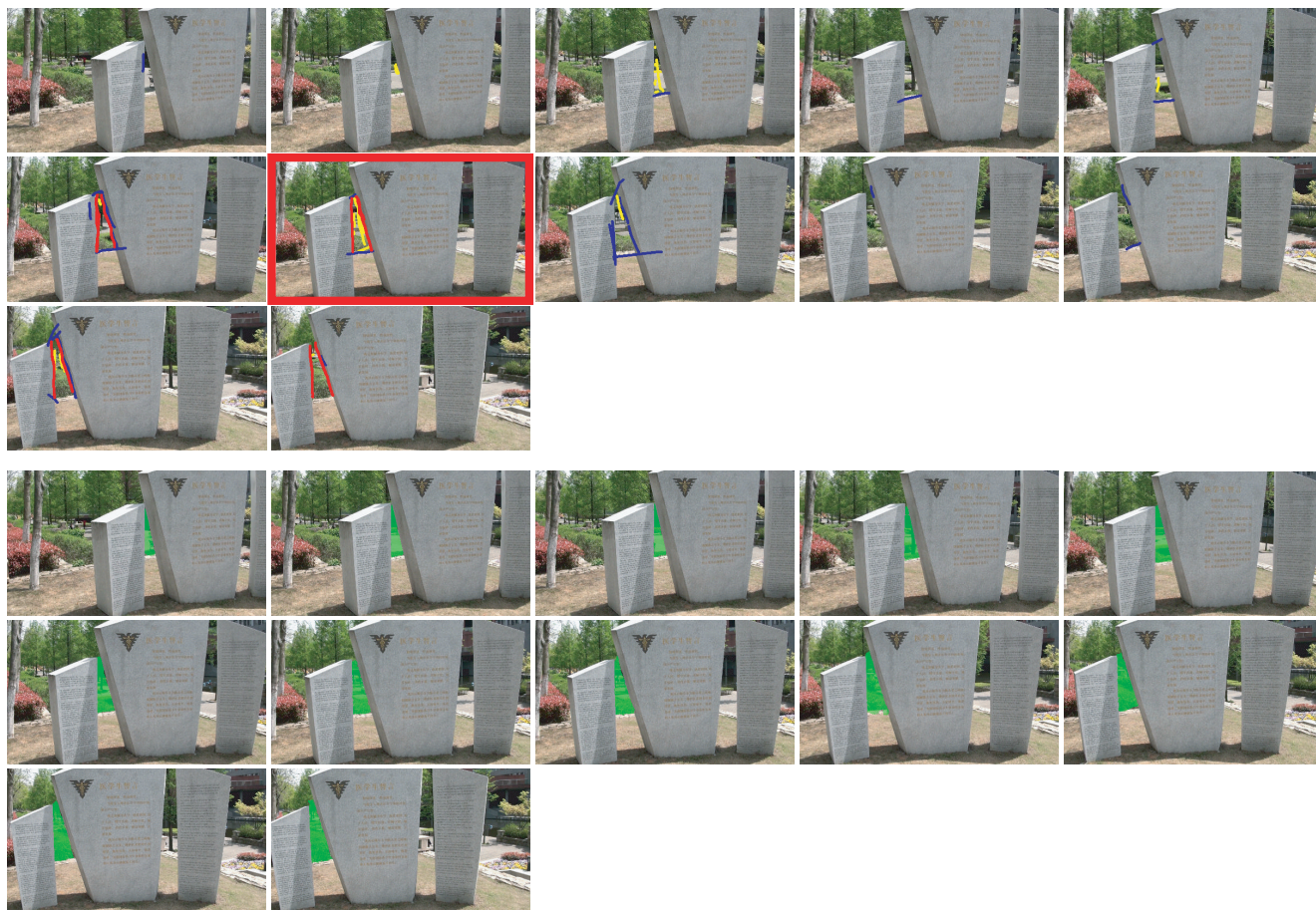


Figure 3. Edited keyframes for another repaired region in “Stela” sequence. Top: the edited 12 key frames. The frame with red rectangle is the starting edited frame. Bottom: the segmentation results.

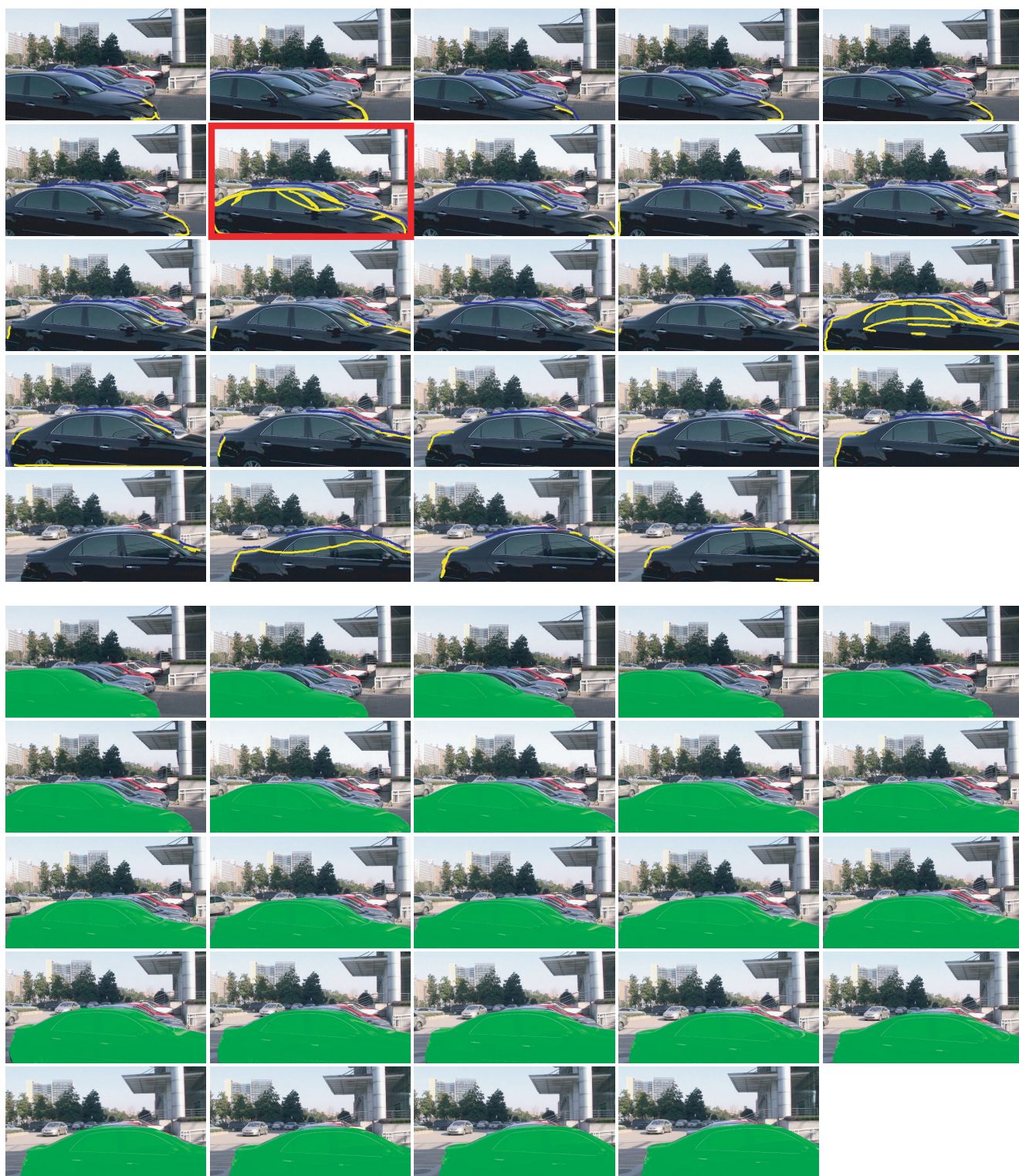


Figure 4. Edited key frames for masking the black car in “Car” sequence. Top: the edited 24 key frames. The frame with red rectangle is the starting edited frame. Bottom: the segmentation results.

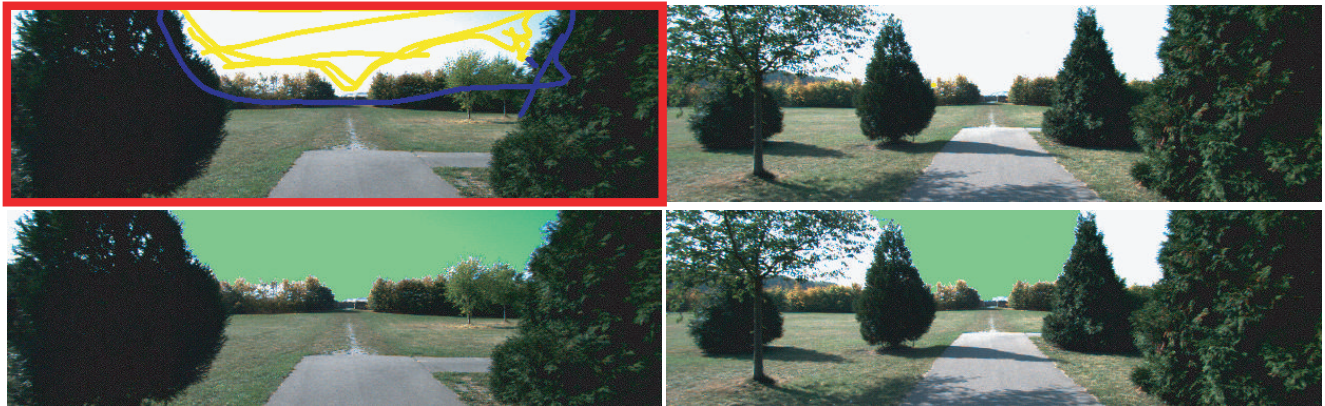


Figure 5. Edited key frames for masking the sky region in the KITTI sequence 000070. Top: the edited two key frames. The frame with red rectangle is the starting edited frame. Bottom: the segmentation results.

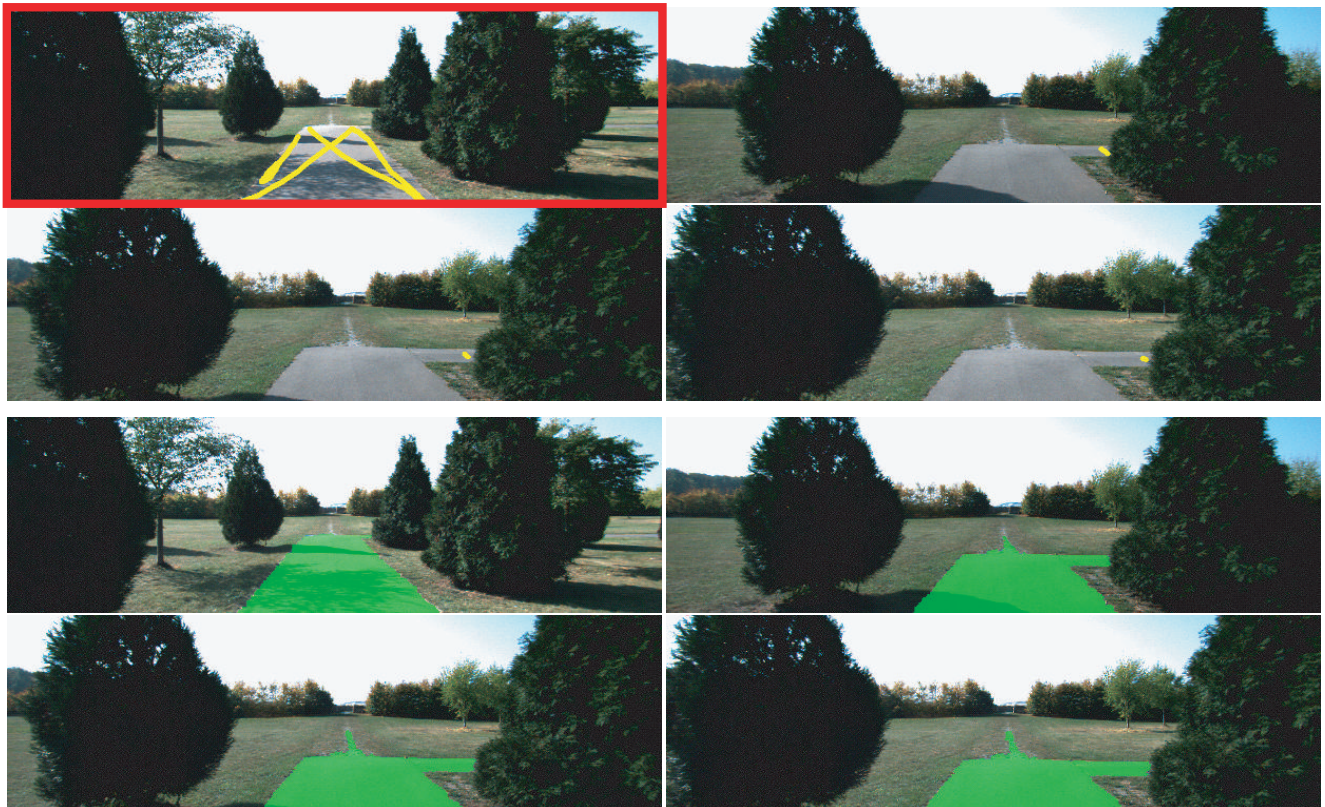


Figure 6. Edited key frames for masking the road region in the KITTI sequence 000070. Top: the edited four key frames. The frame with red rectangle is the starting frame. Bottom: the segmentation results.