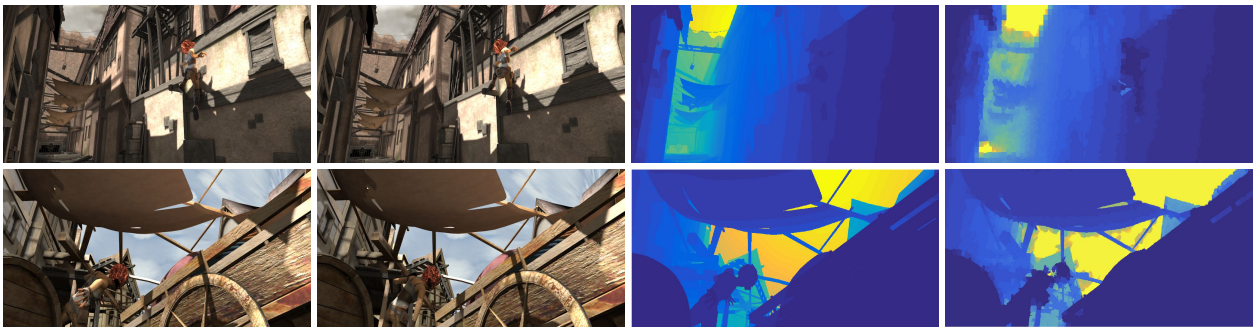
# Supplementary Material: Monocular Dense 3D Reconstruction of a Complex Dynamic Scene from Two Perspective Frames

Suryansh Kumar[1] Yuchao Dai[1,2] Hongdong Li[1,3]
[1]Australian National University, Australia.
[2]Northwestern Polytechnical University, China.
[3]Australian Centre for Robotic Vision.



a) b) c) d)

**Figure 1:** (a)-(b) are the reference frame and the next frame. It is a very challenging case for proper relative scale recovery with monocular images with dynamic motion. In both cases, the motion of the girl between two consecutive frames is very large and therefore the neighboring relations of planes (say superpixels in image domain) in the consecutive frames get violated. In such cases, our method may not be able to provide correct relative scales for each moving planes in 3D. As a result, the complicated motion of the feet of the girl in this example cannot be explained correctly. In the second example, the cart along with girl is moving w.r.t the camera. The hand of the girl has a substantial motion in consecutive frames. (c)-(d) are the ground-truth and obtained depth map respectively. (Best Viewed on Screen)

## Abstract

*In this supplementary material, we provide a brief discussion of the possible failure cases. We also provide a concise overview of the challenging future work or extension of our proposed algorithm.*

## 1. Discussion on failure cases

The success of our method depends on the effectiveness of the piece-wise planar and rigid motion assumption. Our method may fail if the piece-wise smooth model is no longer a valid approximation for the dynamic scene.

Furthermore, our approach may also disappoint if the motions of the dynamic objects in the scene between consecutive frames are significantly large such that the neighboring relations defined in the reference frame get violated in the next frame. A couple of examples for such situations are illustrated in Fig.1. Other possible situations of failure may arise in the case of textureless surfaces. Interested readers, researchers, and critics may refer to some new source of information, such as examining surface shading for surface description [1]. However, we would like to argue that our algorithm assumes that reasonable dense feature correspondences are provided as input.

## 2. Future work and possible extension

One possible extension of our present work is to exploit the current formulation for multiple frames. Other viable and challenging problem is to simultaneously solve for dense optical flow estimation and dynamic 3D reconstruction. A detailed discussion on the aforementioned problems is beyond the scope of this work, however, we want to posit that the solution to these challenging problems is important for the development of sophisticated dense reconstruction algorithm.

## References

[1] A. Blake, A. Zisserman, and G. Knowles. Surface descriptions from stereo and shading. *Image and Vision Computing*, 3(4):183–191, 1985. 1