Frankenhorse: Automatic Completion of Articulating Objects from Image-based Reconstruction

Alex Mansfield[1]
mansfield@vision.ee.ethz.ch

Nikolay Kobyshev[1]
nk@vision.ee.ethz.ch

Hayko Riemenschneider[1]
hayko@vision.ee.ethz.ch

Will Chang[2]
wychang1@cs.ubc.ca

Luc Van Gool[1]
vangool@vision.ee.ethz.ch

[1] Computer Vision Lab
ETH Zürich
Switzerland
[2] Department of Computer Science
University of British Columbia
Canada

Reconstruction of scene geometry and semantics are important problems in vision, and increasingly brought together. The state of the art in Structure from Motion and Multi View Stereo (SfM+MVS) can already create accurate, dense reconstructions of scenes. Systems such as CMPMVS [2] are freely available and produce impressive results automatically. However, when assumptions break down or there is insufficient data, noise, extraneous geometry and holes appear in the reconstruction.

We propose to solve these problems by introducing prior knowledge. We focus on the difficult class of articulating objects, such as people and animals. Prior modelling of these classes is difficult due to the articulation and large intra-class variation. We propose an automatic method for completion which does not rely on a prior model of the deformation or training data captured under controlled conditions. Instead, given far from perfect reconstructions, we simultaneously complete each using the well-reconstructed parts of the others.

This is enabled by the data-driven piecewise-rigid 3D model alignment method of Chang and Zwicker [1]. This method estimates local coordinate frames on the meshes and proposes correspondences by matching local descriptors. Each correspondence determines a rigid alignment, which is used as a label in a graph labelling problem to determine a piecewise-rigid alignment which brings the meshes into correspondence while penalising stretching edges.

Our main contributions are as follows. We present a novel, fully automatic method for the completion of noisy real SfM+MVS reconstructions which (1) exploits a set of noisy reconstructions of objects of the class, rather than relying on a large clean training set which is expensive to collect, (2) handles the articulation structure in the class of objects, allowing larger holes to be filled and with greater accuracy than a generic smoothness prior and (3) is exemplar-based, allowing details to be maintained that may be smoothed out in related learning-based approaches.

Our method takes as its input sets of images of scenes each containing an object of a specific class. For each input image set, initially yielding an incomplete and cluttered reconstruction of the whole scene, the output is a completed model of the object, created using the other reconstructions. Our method consists of a pipeline of several stages, visualised in Figure 1.

In the first stage, each scene is reconstructed using a SfM+MVS pipeline [2]. We then segment the objects from the scene by combining object detections in the images. In the third stage, we align each of

Figure 1: Our fully automatic pipeline takes at the input datasets of images, and processes each to obtain a segmented model of the object (upper row). Completion of a noisy target model from SfM+MVS reconstruction draws on the whole set of segmented models (lower row).
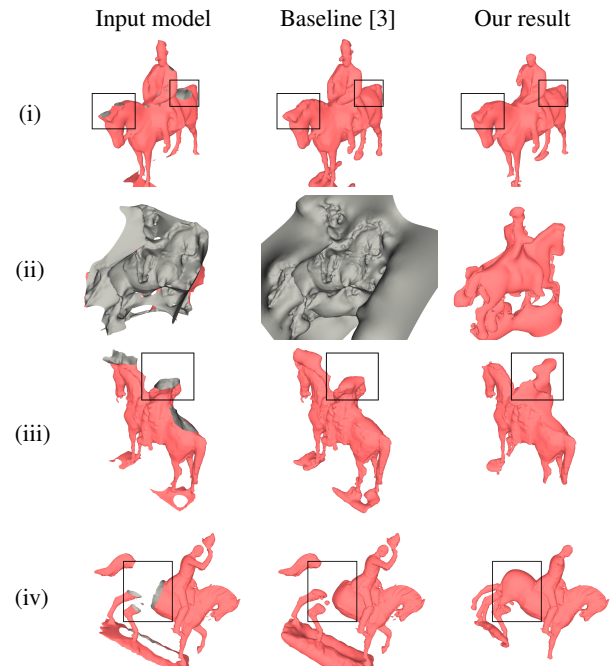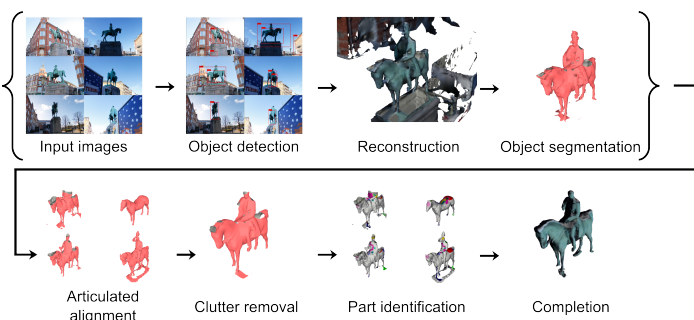

Figure 2: Results (i) – (iii) show the completion of real holes, in (i) the back, (ii) half of the horse, and (iii) the rider. Result (iv) shows the completion of synthetically created holes in the back. For small holes, the baseline also produces good results (i), but for larger holes, the smooth completion rounds off the hole, while our method can complete the part.

the segmented source models to the target model taking into account articulation using the method of Chang and Zwicker [1]. We exploit these aligned source models to remove clutter from the target model, and hence correctly identify the holes. Finally, we choose a completion for each hole from those proposed by the aligned source models, and reconstruct the final result filling small holes using screened Poisson reconstruction [3].

As our method performs completion as a post-process, we expect it to produce a plausible reconstruction of the real object. Given the large holes, there is a large variety of appropriate solutions. This is hard to model quantitatively and so visual inspection provides the best evaluation method. Using visual inspection, we perform a qualitative evaluation of our full set of our results, and show typical results in Figure 2. We also perform a quantitative analysis that proves effectiveness of our method.

We demonstrate that while small holes can be completed with local smoothness priors, completing large holes requires a global perspective. We successfully add missing parts like heads, legs and horse riders which are otherwise just smoothed out stumps. Our failure modes occur due to the registration of the models and confusing locally similar parts.

[1] Will Chang and Matthias Zwicker. Automatic registration for articulated shapes. *Computer Graphics Forum (Proc. SGP)*, 27(5):1459–1468, 2008.

[2] M. Jancosek and T Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *Proc. CVPR*, 2011.

[3] M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *Proc. SIGGRAPH*, 32(3), 2013.