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

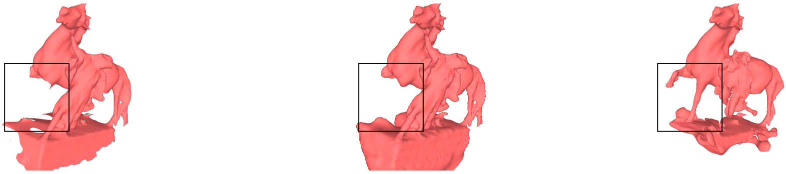
[1] Computer Vision Lab
ETH Zürich
Switzerland
[2] Department of Computer Science
University of British Columbia
Canada

Abstract

Manual 3D modelling can create clean complete models but takes time and expert knowledge. Image-based reconstructions of objects are easy to create, but are far from complete and clean. While small holes can be completed with a smoothness prior, large holes require a higher-level understanding of the object. We present the first method to complete large holes in articulating objects by reconstructing and aligning sets of objects of the same class, using the well-reconstructed parts in each model to complete holes in the others, resulting in a 'Frankenhorse' completion. Our proposed method is fully automatic, and still is able to handle articulation, intra-class variation, holes and clutter present in the reconstructions. This is achieved through our novel segmentation and clutter removal processes as well as by the use of a robust method for piecewise-rigid registration of the models. We show that our method can fill large holes even when only a small set of models with high variability and low reconstruction quality is available.

1 Introduction

Reconstruction of scene geometry and semantics are important problems in vision, and increasingly brought together [26, 27, 30]. The state of the art in Structure from Motion and Multi View Stereo (SfM+MVS) can already create accurate, dense reconstructions of scenes that satisfy a number of assumptions such as staticity and Lambertian reflectance. Systems such as Arc3D [32], 123D Catch [4], VisualSFM [33, 34], and CMPMVS [19] are freely available and produce impressive results automatically. However, problems remain when assumptions break down or there is insufficient data, resulting in noise, extraneous geometry and holes in the reconstruction. Where the main problem is missing data, additional data



(a) Incomplete reconstruction (b) Poisson reconstruction [20] (c) Our completion

Figure 1: Our method automatically completes models of articulating objects (a). Smooth completion is insufficient for large holes (b). Our method draws on a set of reconstructions of the same class of objects to create a more plausible completion (c). We refer to our completions as Frankenhorses, as they consist of parts from different objects.

can be collected, such as from the large datasets of images of the world available online (*e.g.* Flickr, Google Streetview). However, these datasets tend to have sparse coverage with most images taken from iconic viewpoints [14]. In any case, problems may still remain due to violation of assumptions.

In this work, we propose to solve these problems by introducing prior knowledge. For many objects such as buildings, low-level priors favouring smoothness and planarity can already improve the reconstructions [13, 15, 17]. We, however, focus on the more difficult class of articulating objects, such as people and animals as shown in Figure 1. Prior modelling of these classes is difficult due to the articulation and large intra-class variation.

We propose an automatic method for completing these objects which does not rely on learning a prior model of the deformation or training data captured under controlled conditions. Instead, given far from perfect reconstructions of a set of objects, we are the first to simultaneously complete each object using the well-reconstructed parts of the other objects.

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 and 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 in the completion that may be smoothed out in related learning-based approaches.

1.1 Related work

Methods for repairing 3D reconstructions have been extensively studied in computer vision and graphics. A comprehensive survey can be found in the recent work of Attene *et al.* [3]. Early work in the completion of meshes focused on closing small gaps and holes, for which an assumption of local smoothness was mostly sufficient to guide the completion. Although these methods can be efficient, they cannot satisfactorily complete large holes in structured shapes where the smoothness assumption no longer holds.

For larger holes, proposed methods exploit local self-similarity. Such methods include [28], which densifies a point cloud by copying points from similar but denser areas in the object, and [25], which uses detected repeating structures to complete the model.

At large scales, however, structures are less likely to repeat, and a global perspective is required. The work of Pauly *et al.* [24] uses a large database of clean, complete models as templates. The templates are registered to the incomplete target globally, limiting the

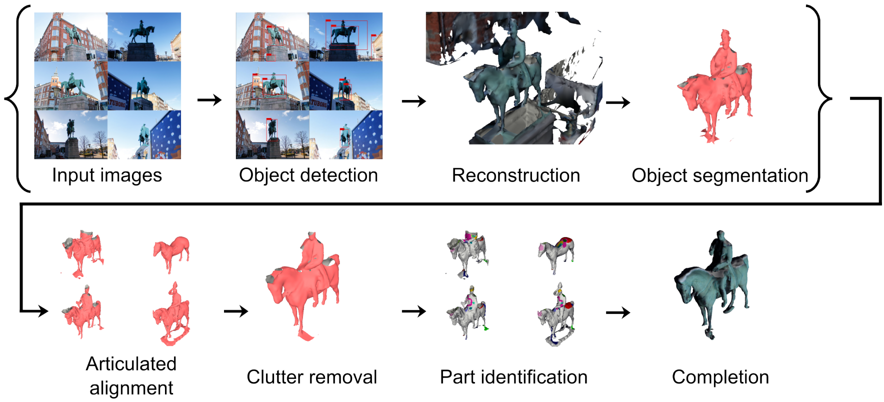


Figure 2: Our fully automatic pipeline takes at the input datasets of images, and processes each to obtain a segmented model of the object (upper row). Completion of a noisy target from SfM+MVS reconstruction draws on the whole set of segmented models (lower row).

class of objects for which this method is suitable to those of similar global shape. A similar restriction exists in recent work on semantic priors for reconstruction [5, 10].

For objects which undergo large non-rigid deformation, previous methods rely on user assistance, temporal consistency or large sets of training data. The work of [22] uses a single template model and user-specified correspondences to register models. The reconstruction of moving objects has been tackled [9, 23], but these methods rely on the temporal consistency to accurately combine geometry from each frame. For the specific class of the human body, effective methods have been developed [1, 2, 18] which rely on large carefully-captured training sets, limiting their scalability to other classes.

The work of Chang and Zwicker [8] offers an alternative approach. Given two models which differ largely in articulation, they estimate a piecewise-rigid registration, recovering the alignment of the models and their segmentation into rigid parts. This method is automatic and data-driven, but is only tested on clean models of the same object as it deforms.

Our work builds this approach into a full completion pipeline, and shows that we can align models well even in the presence of significant intra-class variation, holes and clutter as well as articulation. This allows us to perform completion fully automatically and without depending on datasets of clean models, in contrast to previous methods such as [24]. Furthermore, we additionally tackle the problems of segmenting the object from the reconstruction of the whole scene, and finding a single consistent completion given the aligned models.

2 Articulating shape completion

We propose a fully automatic method for the completion of articulating objects, which 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 2. Note that our method aims to exploit data containing noise and clutter. We are the first to propose explicit stages for segmentation and clutter removal which are not necessary in previous work working with clean data.

(a) (b) (c) (d)

Figure 3: Object segmentation takes the object detections (a) and 3D model, visualised here by the projected points, (b), and jointly estimates the object segmentation and best detection in each image (c). We refine the segmentation, re-attaching the head in this case (d).

In the first stage, each scene is reconstructed using a SfM+MVS pipeline (§2.1). In the second stage, we segment the objects from the scene by combining object detections in the images (§2.2). In the third stage, we align each of the segmented source models to the target model taking into account articulation (§2.3). In the fourth stage, we exploit these aligned source models to remove clutter from the target model (§2.4.1). In the fifth stage, we identify the holes in the target model (§2.4.2). Finally, we choose a completion for each hole from those proposed by the aligned source models (§2.4.3), and reconstruct the final result.

2.1 Scene reconstruction

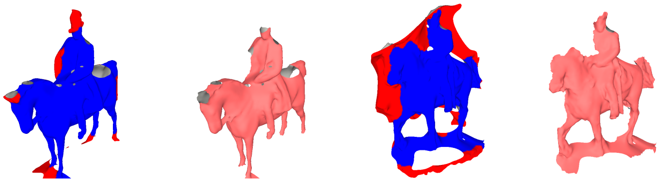
Our method takes as its input a series of image datasets, with each dataset containing images of a scene with an object of interest. We denote 'reconstruction' as the process of SfM [33, 34] and dense MVS [19] to obtain a polygonal mesh of the full scene and camera calibration.

2.2 Object-of-interest segmentation

Our method assumes the availability of an image-based object detector which proposes bounding boxes which contain the object with high recall, though not necessarily high precision. We run this detector on all the images in our datasets.

Given these bounding boxes, the 3D model and the camera calibrations, we jointly estimate the 3D segmentation and best bounding box in each image (Figure 3) with a RANSAC-like approach [12]. For P iterations, we randomly choose two distinct images and a bounding box in each of them, and segment from the 3D model all of the points which reproject into these two bounding boxes. Then, for all other images, we find their best bounding box as that which is most similar to the bounding box of the reprojected segmented points, where we measure similarity using the intersection over union. If the similarity is greater than a threshold T_D, we add this bounding box to the inlier set. From all of the iterations, we return the set with the largest number of inliers, and its corresponding segmentation.

Then, we re-estimate the segmentation from the full set of selected bounding boxes by voting for each point in the point cloud by the number of selected bounding boxes it projects into. We segment all points with at least T_X votes. After, we add back to the segmentation any parts of the mesh connected to the segmented object, and which form new isolated connected components when it is removed (see Figure 3(d)). In order to prevent adding back the whole scene but only small missed details, we only add back these parts if their number of points as a fraction of the number of points in the segmented object is less than a fraction T_S.



(a) Original (b) Clutter-free (c) Original (d) Clutter-free

Figure 4: Clutter is removed to allow correct hole identification. We define a graph labelling problem to classify points on the mesh as clutter when they have little support in the aligned source models (coloured red) (a,c), which we remove to obtain a clutter-free model (b,d).

2.3 Articulated alignment

Given the set of segmented models, we begin completing them by aligning all pairs of models. This alignment is very challenging as it must be able to handle the articulation and other intra-class variation between the models, while being robust to holes and clutter.

We use the piecewise-rigid alignment method of Chang and Zwicker [8] to achieve this. In outline, this method estimates local co-ordinate frames for every point 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.

2.4 Shape completion

Given each of the source models aligned to the target, we proceed to complete the target model. Firstly, we identify areas of the target model which are likely to be clutter or extraneous geometry, and remove these. We then identify holes in the target and the completions proposed by each of the aligned sources. Finally, we choose the best completion for each hole. We now describe each of these stages in detail.

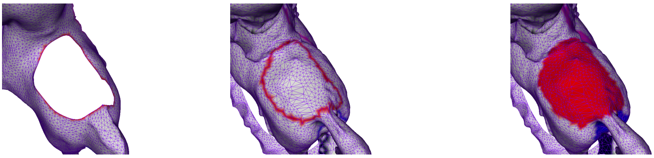
2.4.1 Clutter removal

Our target models often contain clutter and extraneous geometry: parts of the surface that did not belong to the real shape, but were created by errors in the reconstruction process or left in by the segmentation. To clean the models, we examine the support from the aligned source models. Statistically, regions of the target to which no part of any of the source models could be aligned are likely to be clutter, assuming that the source models cover well the intra-class variation and the articulated alignment is accurate. We formulate the segmentation of the clutter as the binary graph labelling problem

$$\underset{f_p \in \{0,1\}, \forall p \in \mathcal{P}}{\operatorname{argmin}} \sum_{p \in \mathcal{P}} D(p, f_p) + \sum_{(p,q) \in \mathcal{E}} V(f_p, f_q) \ , \tag{1}$$

for target mesh points $\mathcal{P}$ and edges $\mathcal{E}$, where inlier points take label 0 and clutter label 1. We find the globally optimal labelling with a single binary graph cut [6, 7, 21]. An example result is shown in Figure 4.

We define the unary to label a point as an inlier if the distance to the closest point in any of the sources is 'low'. For robustness, we use the first quartile of the distances. We compare



(a) Target (b) Source boundary (c) Source part segmented

Figure 5: To identify parts for completion, we find open boundaries in the target model (a), map these to the source model (b) and extract the contained part (c).

them to a constant proportional to their median over the whole model:

$$D(p, f_p) = \begin{cases} \text{Quartile}\left(\bigcup_{s \in \mathcal{S}} \min_{q \in \mathcal{P}^s} \left\|\mathbf{x}_p - \mathbf{x}_q^s\right\|\right) & f_p = 0 \\ c & f_p = 1 \end{cases} , \tag{2}$$

where Quartile$(\mathcal{A})$ returns the first quartile element of a set $\mathcal{A}$, $\mathcal{S}$ is the set of aligned source models, $\mathcal{P}^s$ is the set of points in the s^{th} source model, and α as a constant weight parameter:

$$c = \alpha \text{Median}\left(\bigcup_{p \in \mathcal{P}} D(p, 0)\right) \tag{3}$$

For the pairwise energy, we use a Potts model with fixed penalty c (using in Equation 3):

$$V(f_p, f_q) = c\,[f_p \neq f_q] \quad , \tag{4}$$

where $[A]$ is the Iverson bracket, taking value 1 when statement A is true.

2.4.2 Part identification

Part identification process is visualised in Figure 5. We first identify holes in the target model. We find all edges on the boundary of the mesh (they are contained in only one polygon of the mesh). We identify connected components within these boundary edges and each connected component is closing a hole.

We then identify the corresponding boundaries in the source meshes. For each boundary and source mesh in turn, we find the closest points in the source to each point in the boundary. We connect these corresponding points by finding the shortest path through the source mesh between each of them to obtain a corresponding closed boundary in the source.

Finally, we identify the part that each source model proposes to complete the target hole. We identify the smaller of the new connected components created within the mesh when all edges connected to boundary vertices are broken. Note that this assumes that the part is smaller than the rest of the mesh, and also that the regions of the mesh on either side of the boundary are disconnected when connections to the boundary vertices are removed. Where these assumptions do not hold, such as where there are loops in the mesh, a good proposal for the completion is not found. To ensure a good overlap, we add to the part all polygons that contain at least one point in the part.

2.4.3 Hole completion

For each hole, given the set of proposed completions, we choose the best completion in a way that is robust to holes and clutter in the source models, as well as to errors in the alignment and other preceding stages of the pipeline. We choose the part from source s^* such that

$$s^* = \underset{s}{\operatorname{argmin}} C_{\text{fit}}(s) + \beta C_{\text{incompleteness}}(s) \quad . \tag{5}$$

The first term penalises parts which do not have a tight fit to the target boundary $\mathcal{B}$ where $\mathcal{P}^s$ is the set of points in the s^{th} part as:

$$C_{\text{fit}}(s) = \text{Median}\left(\bigcup_{p\in\mathcal{B}}\left\{\min_{q\in\mathcal{P}^s}\left\|\mathbf{x}_p - \mathbf{x}_q^s\right\|\right\}\right) \quad . \tag{6}$$

The second term penalises holes within the source part where $\mathcal{B}^s$ is the set of internal open boundary points in the s^{th} part. We define this term as

$$C_{\text{incompleteness}}(s) = |\mathcal{B}^s| \quad . \tag{7}$$

To reconstruct a single watertight output model we perform screened Poisson reconstruction [20] on the set of points in the completed model.

3 Evaluation

We evaluate our method on a dataset that consists of sets of images for 16 equestrian statues from cities across Europe, with 25 to 216 images for each statue. The variation in articulation, style and contained elements is quite large. We also created 28 test target models by introducing large synthetic holes into some of the reconstructed models, making a total of 44 incomplete target models. We also make use of 3 synthetic models of horses as sources: two from [29], and one from [31], resulting in 19 source models.

3.1 Implementation details

Our pipeline was implemented in a mixture of MATLAB and C++.

SfM pipeline. We obtain the camera calibrations using VisualSFM [33, 34] and use CMP-MVS [19] for a dense mesh reconstruction. The resulting full scene meshes contain on the order of 10^5 triangles.

Object detection. We perform the object detection using the Deformable Part Model detector trained on the VOC 2007 dataset [11, 16]. To achieve high recall, we use a low threshold of -0.9. For the horse statues, we find that the existing equestrian detector works sufficiently well, even though it is trained on real horses and not statues.

Articulated alignment. For scale invariance, we normalise the scale of the models after segmentation with a Procrustes analysis. To get good results in feature matching step, we normalise the resolution of the meshes by multiple iterations of mesh subdivision followed by a mesh simplification to approximately 50,000 triangles. We use $5,000$ sampled features per model during the feature matching, and the non-symmetric cost function only, to improve robustness to clutter.

Table 1: Comparison of performance with different source datasets

Dataset	Average reconstruction quality			
	Legs (1–4)	Body (1–4)	Head (1–2)	Total (1–10)
Raw reconstruction models	3.23	2.70	1.28	7.15
Repaired by baseline method [20]	3.02	3.36	1.50	7.89
Repaired by synthetic data	2.99	3.55	1.60	8.14
Repaired by reconstruction data	2.90	3.50	1.74	8.14
Repaired by synthetic + reconstruction data	2.99	3.77	1.75	8.51

Parameters. In the object-of-interest segmentation, we set the number of iterations $P = 50$, the threshold on the Jaccard index for a bounding box to be an inlier $T_D = 0.6$, the threshold on the number of votes required for the foreground $T_X = 3$ and the threshold for adding back connected components $T_S = 75\%$. In the clutter removal, we use $\alpha = 1.5$, and in the completion, weighting factor $\beta = 0.001$.

Timings. The bottleneck in the computation is the articulated alignment. For each pair of models, this takes on the order of 1 hour, so for our dataset around 19 hours per target model. The other stages are significantly faster. Per model, the object-of-interest segmentation takes on the order of 30s, the clutter removal around 30s and the completion around 10 minutes.

3.2 Results

As our method performs completion as a post-process, we expect our method to produce a completed plausible reconstruction of the real object. Given the large holes, there is a large variety of plausible solutions, and hence no unique 'ground truth' which could be used to easily evaluate performance fully objectively. Hence visual inspection is the best evaluation method available to us. Using visual inspection, we perform a quantitative evaluation of our full set of our results, and show typical results in Figure 6 and 7.

As our only baseline, since we are the first to work on noisy SfM+MVS data, we complete the segmented input mesh using screened Poisson reconstruction [20]. This provides a watertight reconstruction of the input points, smoothly filling in any holes in the original mesh efficiently. As Poisson-based methods require good segmentation of the object, we use our object-of-interest segmentation.

We quantitatively evaluate the reconstruction quality by an expert grading the results from 0 to 10. The instructions are to rate each salient part for presence and alignment. Each leg was given 1 point, the head 2 points, the overall body torso with two sides, tail and torso 4 points. If a part is fully present and properly aligned, it gets the full score; if it has some minor holes, alignment problems or clutter, it gets half the points; otherwise, it gets none. In order to evaluate the contribution of the two types of source models, we tested our method on the SfM+MVS and clean models separately as well as together. The results are shown in Table 1, averaged over the 44 models, with in total 132 ratings per method (row).

Our method improves the completeness (for the input data 71.5%) and outperforms the baseline [20] (78.9%). Using the synthetic or real data alone results in the same total score of 81.4%, but with different partial results: synthetic data best completes the legs and body, while for the heads, the amount of variability captured in the source set is critical. Here better results are obtained using the SfM+MVS data. Overall, the best results of 85.1% are

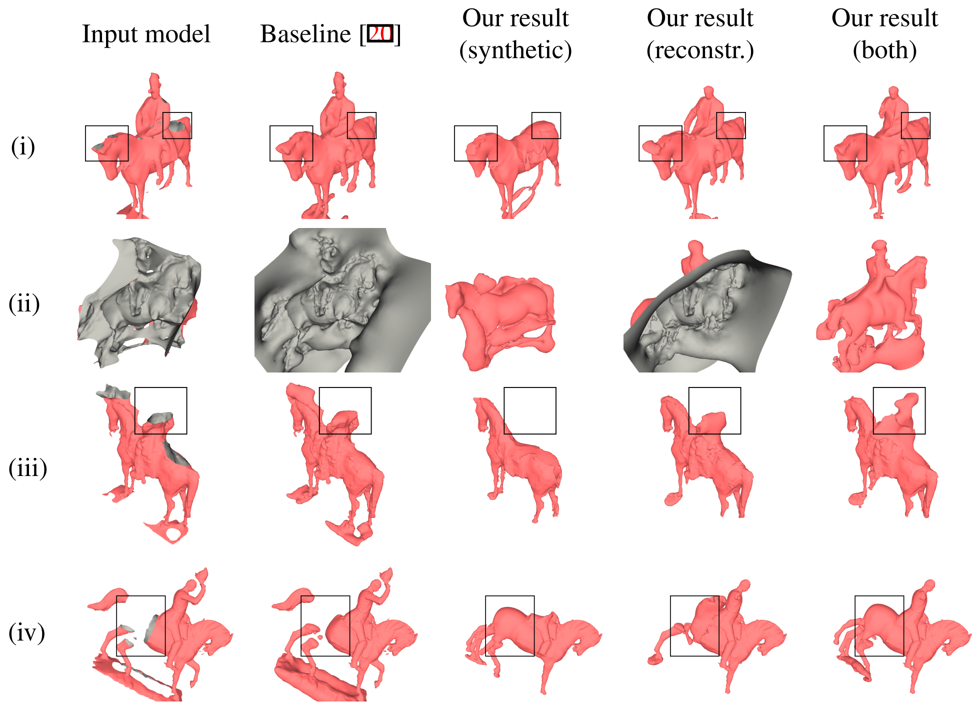


Figure 6: Results of our method for different source datasets. Results (i) through (iii) show the completion of real holes, in (i) on the back, (ii) one half of the horse, and (iii) the rider. Result (iv) shows the completion of synthetically created holes in the back of the horse. For small holes, the baseline also produces good results (i), but for larger holes, the smooth completion simply rounds off the hole, while our method can complete the part. Note that when the completion is based only on synthetic data, the rider (not present in the synthetic models) is cut off during clutter removal. Combining both sources of data clearly yields the best results in all cases.

obtained using the full set, showing our method can successfully exploit the advantages of both data types.

We found that 76% of the added points came from the SfM+MVS models, and the remaining from the synthetic models, in proportion to their contribution to the source set.

While the quality of the head is improved and the body much improved by our algorithm, the quality of the legs is a little reduced. This shows that parts of the legs are removed in the segmentation processes. This could be prevented by identifying the legs explicitly, which would require semantic understanding of the object, a problem we leave to future work.

We show a number of our results in Figure 6. The figure demonstrates that our method is able to handle significant variation in the input models and complete very large holes. Failure cases are discussed in Figure 7 where high-level pose estimation or better part segmentation would improve the results.

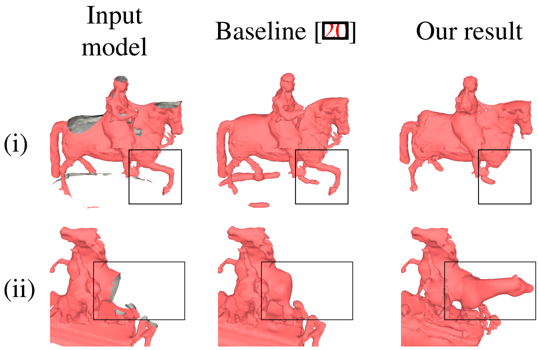


Figure 7: Failure cases of our method. In some cases, correct parts of the model are classified as clutter and removed as in (i), and our method does not always choose the best part to complete the resulting holes, leaving, for example, the front right leg removed in (i). Symmetries can cause the alignment to fail, resulting in semantically inconsistent completions (ii).

4 Conclusion

In this paper, we propose a novel automatic method for the completion of 3D reconstructions of articulating objects. Our method draws on reconstructions of objects of the same class and uses the best parts of each to overcome incompleteness and clutter without relying on a clean exemplar set, creating 'Frankenobjects' consisting of parts from multiple other objects.

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.

For future work, we will investigate detecting and separating multiple instances of the object class from large image datasets. We will also investigate learning and incorporating higher level object knowledge such as pose estimation, to disambiguate in difficult cases. Another line of future work is to analyse reconstructions of dynamic scenes, where articulation is present for a larger range of object classes. Finally, we intend to close the loop and combine our exemplar-based understanding as a prior with the original data, in order to achieve greater accuracy.

Acknowledgements. We would like to thank Sandeep Kakani and Angelika Garz for their assistance in data collection and Matthias Zwicker for helpful discussions.

References

[1] B. Allen, B. Curless, and Z. Popović. The space of all body shapes: reconstruction and parameterization from range scans. *Proc. SIGGRAPH*, 22(3):587–594, 2003.

[2] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and Jamespatch Davis. SCAPE: shape completion and animation of people. *Proc. SIGGRAPH*, 24(3):408–416, 2005.

[3] Marco Attene, Marcel Campen, and Leif Kobbelt. Polygon mesh repairing: An application perspective. 45(2):15:1–15:33, 2013.

[4] Autodesk. 123D Catch, 2012. URL http://www.123dapp.com/catch.

[5] Sid Ying-Ze Bao, Manmohan Chandraker, Yuanqing Lin, and Silvio Savarese. Dense object reconstruction with semantic priors. In *Proc. CVPR*, pages 1264–1271, 2013.

[6] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. PAMI*, 23(11):1222–1239, 2001.

[7] Y.Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. PAMI*, 26(9):1124–1137, 2004.

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

[9] Will Chang and Matthias Zwicker. Global registration of dynamic range scans for articulated model reconstruction. *ACM Trans. Graph.*, 30(3), 2011.

[10] Amaury Dame, Victor Adrian Prisacariu, Carl Yuheng Ren, and Ian Reid. Dense reconstruction using 3D object shape priors. In *Proc. CVPR*, pages 1288–1295, 2013.

[11] Pedro F. Felzenszwalb, Ross B. Girshick, David A. McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. PAMI*, 32:1627–1645, 2010.

[12] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communcations of the ACM*, 24(6):381–395, 1981.

[13] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. Manhattan-world stereo. In *Proc. CVPR*, 2009.

[14] Y. Furukawa, B. Curless, S.M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *Proc. CVPR*, pages 1434–1441, 2010.

[15] D. Gallup, J. Frahm, and M. Pollefeys. Piecewise planar and non-planar stereo for urban scene reconstruction. In *Proc. CVPR*, 2010.

[16] R. B. Girshick, P. F. Felzenszwalb, and D. McAllester. Discriminatively trained deformable part models, release 5, 2012. URL http://people.cs.uchicago.edu/~rbg/latent-release5/.

[17] C. Häne, C. Zach, B. Zeisl, and M. Pollefeys. A patch prior for dense 3D reconstruction in man-made environments. In *Proc. 3DIMPVT*, 2012.

[18] Nils Hasler, Carsten Stoll, Martin Sunkel, Bodo Rosenhahn, and Hans-Peter Seidel. A statistical model of human pose and body shape. *Computer Graphics Forum (Proc. EUROGRAPHICS)*, 28(2):337–346, 2009.

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

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

[21] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. PAMI*, 26(2):147–159, 2004.

[22] Vladislav Kraevoy and Alla Sheffer. Template-based mesh completion. In *Proc. SGP*, 2005.

[23] Hao Li, Linjie Luo, Daniel Vlasic, Pieter Peers, Jovan Popović, Mark Pauly, and Szymon Rusinkiewicz. Temporally coherent completion of dynamic shapes. *ACM Trans. Graph.*, 31(1), 2012.

[24] M. Pauly, N. J. Mitra, J. Giesen, M. Gross, and L. Guibas. Example-based 3D scan completion. In *Proc. SGP*, pages 23–32, 2005.

[25] M. Pauly, N. J. Mitra, J. Wallner, H. Pottmann, and L. Guibas. Discovering structural regularity in 3D geometry. *Proc. SIGGRAPH*, 27(3), 2008.

[26] Bojan Pepik, Peter Gehler, Michael Stark, and Bernt Schiele. $3D^2$PM - 3D deformable part models. In *Proc. ECCV*, 2012.

[27] S. Savarese and L. Fei-Fei. 3D generic object categorization, localization and pose estimation. In *Proc. ICCV*, 2007.

[28] Andrei Sharf, Marc Alexa, and Daniel Cohen-Or. Context-based surface completion. *Proc. SIGGRAPH*, 23(3):878–887, 2004.

[29] Oana Sidi, Oliver van Kaick, Yanir Kleiman, Hao Zhang, and Daniel Cohen-Or. Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering. *Proc. SIGGRAPH Asia*, 30(6), 2011.

[30] H. Su, M. Sun, L. Fei-Fei, and S. Savarese. Learning a dense multi-view representation for detection, viewpoint classification and synthesis of object categories. In *Proc. ICCV*, 2009.

[31] Robert W Sumner and Jovan Popović. Deformation transfer for triangle meshes. *Proc. SIGGRAPH*, 23(3):399–405, 2004.

[32] Maarten Vergauwen and Luc Van Gool. Web-based 3D reconstruction service. *Machine Vision and Applications*, 17(6):411–426, 2006.

[33] Changchang Wu. VisualSFM: A visual structure from motion system, 2011. URL http://ccwu.me/vsfm/.

[34] Changchang Wu. Towards linear-time incremental structure from motion. In *Proc. 3DV*, 2013.