Automatic Inference of Dynamic Shape Models from Depth Data

1 Summary

Recent developments in depth sensing technology has opened up new possibilities for scene analysis and understanding. The availability of accurate depth data in a wide range of indoor lighting conditions makes it much easier to segment and identify objects. Already there are promising commercial applications: game controllers are replaced with simple gestures and movements of our hands, clothes shopping involves less guesswork by virtually "trying them on" using scanned body shapes, and museums display virtual tours and exhibits of artwork on computer screens. Depth sensing technology also has the potential to be very useful for automatic building exploration by a robot, building parametric models of everyday objects (e.g. furniture), and monitoring movement in scientific experiments and in construction sites. The medical field provides more serious applications. Images and scans of internal organs are used to detect abnormalities and assist surgery by visualization. Contributing to this area offers the potential to directly impact the welfare of our lives.

In the near future, we expect such depth sensing cameras to gradually become as ubiquitous as digital cameras. New applications of this scanning technology will better enable exploration and understanding of the world around us, and it will inspire novel techniques for human-computer interaction. Therefore, there is an increasing need of new algorithms for the processing, storage, retrieval, and display of scan data.

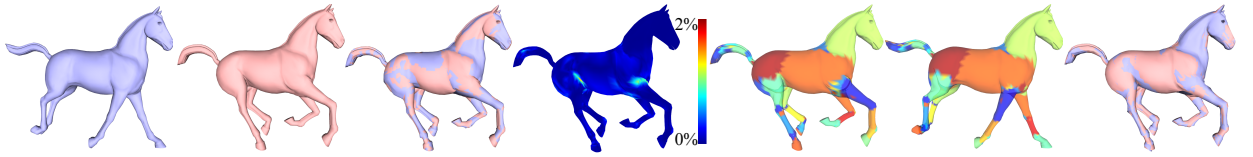
I am interested in the acquisition of geometric models of shape and motion from range scans. While techniques in this subject has been largely limited to static subjects in the past, my research goal is to extend them for dynamic objects. **Through this research, we will ultimately be able to automatically infer models of shape variation, kinematics, and dynamics on a wide variety of scales and motions.** I present a concrete plan to advance the border of knowledge and techniques in this area. Specifically, my work will contribute by

1. Reducing the assumptions on the input data by removing the need for template, markers, user-specified segmentation, etc.,
2. Simplifying or eliminating parameters that need to be manually adjusted,
3. Making existing algorithms work on datasets of 10 to 100 times larger,
4. Extending the motion model to express and fit a wider variety of deformations, and
5. Seeking new applications in natural human-computer interaction.

2 State of Research

A key processing step in capturing any subject is the registration or alignment of multiple range scans. This is because, in most cases, the entire surface of an object cannot be observed from a single viewpoint. Multiple range scans taken from different viewpoints must be merged together to build a model that covers the entire surface.

While much previous work has focused on aligning scans of a static subject, aligning scans of a deformable subject is a relatively new topic. The recent development and widespread availability of real-time scanning systems has sparked even more interest in the topic as well. Compared to the static case, the problem is more difficult since the surface changes its shape in every scan. A successful alignment algorithm must estimate and compensate for the motion of the surface, while being robust to missing surface data caused by occlusions in the scanning process.



(a) Source (b) Target (c) Aligned result (d) Registration Error (e) Assigned labels (source, target) (f) With simplification

Figure 1: Automatic articulated registration performed on a deforming model of a horse.

2.1 Shape capture and model reconstruction

Perhaps the most popular approach to track range scans is to fit a template to the scan data. The template provides a very strong prior that gives several advantages in tracking and fitting the data. Earlier work tend to rely on tracked marker locations to automatically fit a template model to the scanned point cloud data [Allen et al., 2002, 2003; Anguelov et al., 2005; Pauly et al., 2005]. The work by Anguelov et al. [2004b] is one of the first techniques to perform unsupervised registration which does not require markers to register to a template. While this method is desiged to cope with large pose differences between scans via a globalized optimization strategy, the temporal coherence provided by scans of video frame rates enables easier optimization based on local fitting. This has been demonstrated by multiple systems that efficiently capture human faces [Zhang et al., 2004; Weise et al., 2009].

Li et al. [2009, 2011] describe a general non-rigid shape capture pipeline of non-rigid shapes. An important idea is to track the template using a coarse representation expressed as a graph. It is also possible to capture deforming garments in detail, by automatically tracking a few key locations to fit a template [Bradley et al., 2008]. These recent methods also incorporate a detail synthesis step to give fine-scale geometric details to an otherwise coarse template model.

Relaxing the requirement of having a template and a high frame rate results in a more ill-posed and challenging reconstruction problem. To reconstruct without without a template, many have modeled the scans as a four-dimensional space-time surface. Mitra et al. [2007] use kinematic properties of this 4D space time surface to track points and register multiple frames of a rigid object. Süßmuth et al. [2008] and Sharf et al. [2008] explicitly model and reconstruct the 4D space-time surface using an implicit surface representation. The alternative is to use numerical optimization. The work by Wand et al. [2009] aligns range scans by solving the surface motion in terms of an adaptive displacement field, and the scans are processed in a hierarchical fashion. These methods still require the surface to be sampled densely in both space and time.

Researchers have also discovered that the articulated prior is perhaps enough for registration, although this imposes constraints on the movement of the surface. The work by Pekelny and Gotsman [2008] is able to register scans without a template nor video frame rates, but it requires additional user input in the form of a manual part segmentation. Zheng et al. [2010] fit a skeleton to scans automatically. This is subsequently used to assist a registration using the articulated movement of the skeleton.

2.2 Deformation Modeling from Examples

Having a motion model of the shape means that we understand how it moves. We can use the information of this model to create new poses and animations of the shape. A classic example of using a motion model is inverse kinematics (IK). While IK has traditionally been used for robot manipulation and skeletal animation, recent systems have extended this for meshes [Zhang et al., 2004; Sumner et al., 2005]. These techniques extrapolate a set of examples to match user constraints, and the model is the set of examples given to the system. It is also possible to explicitly impose and model the parameters of surface motion. A popular representation is linear blend skinning (LBS), for which a variety of techniques are available to extract

parameters from a set of complete or incomplete examples [Anguelov et al., 2004a; Schaefer and Yuksel, 2007; de Aguiar et al., 2008; Zheng et al., 2010].

2.3 Author's work

My work on articulated registration [Chang and Zwicker, 2008] effectively aligns pairs of deforming range scans under the assumption that the movement of the surface is articulated (see example in Figure 1). Previous approaches assumed that the shape of the entire object is known in advance (i.e. a template shape), but this algorithm demonstrates that one can align deforming objects without the need for a template shape. The key insight in this work is to first sample the movement of the surface (rigid transformations) in many locations, and cast the problem as an optimal assignment of the transformations such that the surfaces align. The key benefit, compared to previous approaches, is that the approach is completely automatic: no template or manual correspondences are required to compute a registration.

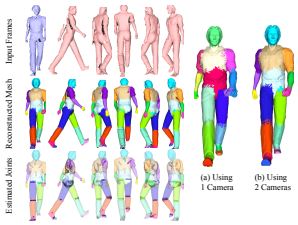


Figure 2: Multiple frame registration on scans of a walking man.

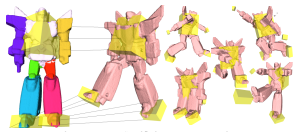


Figure 3: Articulated reconstruction enables easy manipulation.

An interesting consequence of the articulated registration is that it produces a segmentation of the surface into its constituent rigid parts. This enables the possibility of automatically producing a fully rigged, skinned model just by aligning range scans with movement. I address this aspect in a follow-up work, where I explicitly fit the parameters of a linear blend skinning (LBS) model in the process of scan alignment [Chang and Zwicker, 2009]. This work demonstrates that it is indeed possible to construct fully rigged, poseable models from range scans. Although the initial version of this work is only able to process a pair of scans, I further extend this work to successfully align multiple scans [Chang and Zwicker, 2011]. This algorithm produces a single, unified model of the geometry along with skinning weights (Figure 2). It also automatically derives the joint relationship between neighboring rigid parts. The result can be directly plugged into an inverse kinematics (IK) system to create new poses and animations of the reconstructed model (Figure 3).

3 Proposed Research Objective

A common theme of my research involves automatically fitting a parametric model of the scanned subject's motion. The key idea is to use movement between scans as source of new information to build the model. Pushing these algorithms to the limit, we will be able to ultimately build surface and motion models simultaneously, all in real-time. I envision the final result to be similar to the KinectFusion system [Izadi et al., 2011], except that it will be able to model deforming surfaces as well.

I propose to extend this research agenda in several concrete directions: (1) improving models or methods to compensate for the noise inherent in the scanning process, (2) applying the algorithms for datasets that are one or two orders of magnitude larger, and (3) extending the parameteric model itself to include non-rigid deformations. I outline detailed steps of this plan in the following several sections and an estimated timeline for completion in Figure 4.

3.1 Robustness to noise and outliers

Surface matching is an integral part of any scene reconstruction algorithm. Inevitably there is noise in the scanning process, so the surface matching metric must take the sensor noise into account and handle outlier points appropriately. While current methods are largely based on heuristics such as distance or normal angle

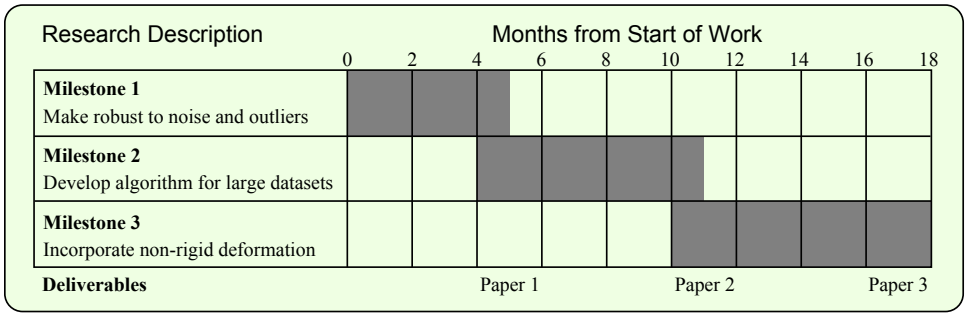


Figure 4: Schedule of completion for each research milestone.

thresholds, the focus of this first research phase is to incorporate new models that explicitly model the noise within the metric.

Using probabilistic models as a different point distance metric is able to deal with range scanning noise better. One example is the work by Myronenko and Song [2010], which models the registered point cloud as centroids in a Gaussian Mixture Model (GMM). We will review the literature and give a systematic treatment and evaluation of each surface matching metric based on actual matching experiments on deformable surfaces. These experiments will be the basis of an extensive evaluation on how to build an accurate registration method that is robust to sensor noise.

Milestone 1: Systematically analyze and report the robustness of deformable registration algorithms under different surface metrics and input data noise.

3.2 Handling larger datsets

A key factor in determining the broad applicability of registration algorithms is the speed and the ability to handle moderate to large sized datasets. Current algorithms report registration results for 50 to 200 frames, with a total of 0.1 to 5 million points [Chang and Zwicker, 2011; Li et al., 2011]. The goal is to extend to much larger datasets on the order of 10 million to 100 million points.

We will attack this problem with two main approaches. The first is to perform the registration hierarcically. We will first consider adapting the octree data structure for bookkeeping of changes within the scene. This will yield an adaptive surface registration algorithm that focuses the optimization in places where the scene has not changed. The second approach is to identify parallelizable portions of the algorithm and map them to GPU computation. One of the most time consuming steps is the closest point matching step, which is a prime candidate for acceleration. Others include linear system solving and discrete optimization (e.g. graph cuts), for which we can apply and extend existing GPU implementations.

The resulting algorithms will be evaluated using longer recording sequences of real-time depth scans. The amount of points will be increased to 10 to 100 times more by accumulating and superresolving depth data acquired over longer periods of time. We expect the research results of this phase to also be applicable for systems that automatically scan interiors of buildings.

Milestone 2: Improve the performance of the registration algorithms for larger scan datasets, by investigating the use of hierarchical surface registration and GPU parallelization.

3.3 Extending the parametric model for non-rigid deformations

Reconstructing a parameteric model of motion provides many useful applications for the captured shape. However, the model fitting process that I developed is currently limited to piecewise rigid movement. We can

relax this restriction to allow non-rigid deformation within each part, in order to allow complex deformations such as muscle bulging or the movement of clothing. This will bring more expressiveness and accuracy for the registration process.

While it is possible to fit non-rigid motion directly, fitting locally rigid parts results in a more robust algorithm in the case of large deformation [Huang et al., 2008]. Therefore, it makes sense to combine the articulated model with local non-rigid deformations. One possible strategy is to relax the graph-based articulated shape representation and assign affine transformations to each graph node. This will be able to capture the scale and shear within each articulated part.

To perform the optimization, I plan to use a global-local strategy. The global step will define the overall articulated movement, while the local step will define local non-rigid deformations. The challenge is to be careful not to overfit the local deformation. To deal with this problem, I plan to investigate methods to limit the amount of total local deformation. A successful algorithm will be robust to occasional temporal incoherence and avoid overfitting in the case of a registration failure.

Finally, when this step becomes robust, I am interested in finding patterns that correlate the articulated movement with the local non-rigid deformations. This has the potential to encode highly detailed deformations in a compact fashion.

Milestone 3: Extend the parameteric model to include local non-rigid surface deformation.

4 Other applications and conclusion

I believe that an important area for making real-world impact is motion tracking and natural user interface applications. Integrating these reconstruction algorithms with a real-time pose detection system will help to improve the robustness and accuracy of the tracked surface movement. In addition, registration will help to build more accurate shape and motion models that are used during the tracking process. To this end, I have developed a collaborative relationship with a tech startup for real-time hand tracking applications, working with Dr. Robert Wang (MIT / 3Gear Systems) and Dr. Hao Li (Columbia / Princeton). My personal hope is that the algorithms I develop will leave the research lab and help people in their practical, everyday situations.

This research builds on the results of my Ph.D. thesis work on "Reconstruction of Dynamic Articulated Models from Range Scans" conducted at the University of California, San Diego. All of the related publications were disseminated at the leading conferences and journals in computer graphics and geometry processing (ACM SIGGRAPH, ACM Transaction on Graphics, Eurographics, and SGP).

Other than research in shape capture, I have also worked on hair reconstruction and rendering [Paris et al., 2008]. This experience sharpened my skills with numerical algorithms and appearance modeling for applications in free-viewpoint capture and capturing complex geometries. During my military service in Korea, I have gained industry experience in embedded systems software, GPGPU, and graphics pipeline optimization. This work allowed me to develop in various aspects of software engineering, project & team management, and the software business. These experiences will help me persue this research plan and work collaboratively with other research groups.

References

B. Allen, B. Curless, and Z. Popović. Articulated body deformation from range scan data. *ACM SIGGRAPH*, 21(3):612–619, 2002.

B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *ACM SIGGRAPH*, pages 587–594, 2003.

D. Anguelov, D. Koller, H. Pang, P. Srinivasan, and S. Thrun. Recovering articulated object models from 3d range data. In *Uncertainty in Artificial Intelligence Conference (UAI)*, 2004a.

D. Anguelov, P. Srinivasan, H.-C. Pang, D. Koller, S. Thrun, and J. Davis. The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. In *NIPS*, 2004b.

D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. In *ACM SIGGRAPH*, pages 408–416, 2005.

D. Bradley, T. Popa, A. Sheffer, W. Heidrich, and T. Boubekeur. Markerless garment capture. *ACM SIGGRAPH*, 27, 2008.

W. Chang and M. Zwicker. Automatic registration for articulated shapes. *Comput. Graph. Forum (Proc. SGP)*, 27(5):1459–1468, 2008.

W. Chang and M. Zwicker. Range scan registration using reduced deformable models. *Comput. Graph. Forum (Proc. Eurographics)*, 28(2):447–456, 2009.

W. Chang and M. Zwicker. Global registration of dynamic range scans for articulated model reconstruction. *ACM Transactions on Graphics*, 30, 2011.

E. de Aguiar, C. Theobalt, S. Thrun, and H.-P. Seidel. Automatic conversion of mesh animations into skeleton-based animations. *Computer Graphics Forum (Proceedings of Eurographics)*, 27(2):389–397, 2008.

Q.-X. Huang, B. Adams, M. Wicke, and L. J. Guibas. Non-rigid registration under isometric deformations. *Computer Graphics Forum (Proceedings of SGP)*, 27(5):1449–1457, 2008.

S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. *UIST*, 2011.

H. Li, B. Adams, L. J. Guibas, and M. Pauly. Robust single view geometry and motion reconstruction. In *ACM SIGGRAPH ASIA, to appear*, 2009.

H. Li, L. Luo, D. Vlasic, P. Peers, J. Popovic, M. Pauly, and S. Rusinkiewicz. Temporally coherent completion of dynamic shapes. *ACM Transactions on Graphics*, 31(1), 2011.

N. J. Mitra, S. Flory, M. Ovsjanikov, N. Gelfand, L. J. Guibas, and H. Pottmann. Dynamic geometry registration. In *SGP*, pages 173–182, 2007.

A. Myronenko and X. Song. Point set registration: Coherent point drift. *IEEE TPAMI*, 32(12), 2010.

S. Paris, W. Chang, O. I. Kozhushnyan, W. Jarosz, W. Matusik, M. Zwicker, and F. Durand. Hair photobooth: geometric and photometric acquisition of real hairstyles. *ACM Transactions on Graphics*, 27(3), 2008.

M. Pauly, N. J. Mitra, J. Giesen, M. Gross, and L. J. Guibas. Example-based 3d scan completion. In *SGP*, page 23, 2005.

Y. Pekelny and C. Gotsman. Articulated object reconstruction and markerless motion capture from depth video. *Computer Graphics Forum (Proceedings of Eurographics)*, 27(2), 2008.

S. Schaefer and C. Yuksel. Example-based skeleton extraction. In *SGP*, pages 153–162, 2007.

A. Sharf, D. A. Alcantara, T. Lewiner, C. Greif, A. Sheffer, N. Amenta, and D. Cohen-Or. Space-time surface reconstruction using incompressible flow. *ACM SIGGRAPH ASIA*, 2008.

R. W. Sumner, M. Zwicker, C. Gotsman, and J. Popovic. Mesh-based inverse kinematics. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 24(3):488–495, 2005.

J. Süßmuth, M. Winter, and G. Greiner. Reconstructing animated meshes from time-varying point clouds. *Computer Graphics Forum (Proceedings of SGP)*, 27(5):1469–1476, 2008.

M. Wand, B. Adams, M. Ovsjanikov, A. Berner, M. Bokeloh, P. Jenke, L. Guibas, H.-P. Seidel, and A. Schilling. Efficient reconstruction of non-rigid shape and motion from real-time 3d scanner data. *ACM Transactions on Graphics*, 28, 2009.

T. Weise, H. Li, L. V. Gool, and M. Pauly. Face/off: Live facial puppetry. In *Eighth ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 2009.

L. Zhang, N. Snavely, B. Curless, and S. M. Seitz. Spacetime faces: high resolution capture for modeling and animation. In *ACM SIGGRAPH*, pages 548–558, 2004.

Q. Zheng, A. Sharf, A. Tagliasacchi, B. Chen, H. Zhang, A. Sheffer, and D. Cohen-Or. Consensus skeleton for non-rigid space-time registration. *Computer Graphics Forum (Proceedings of Eurographics)*, 29(2), 2010.