Deformable Modeling for Human Body Acquired from Depth Sensors

Vamshhi Pavan Kumar Varma Vegeshna

This paper presents a novel approach to reconstruct complete 3D deformable models over time by a single depth camera. These are the steps employed for deforming objects from single depth camera. The partial surfaces reconstructed from various times of capture are assembled together to form a complete 3D surface. A mesh warping algorithm is used to align different partial surfaces based on linear mesh deformation. A volumetric method is then applied to combine partial surfaces, fix missing holes and smooth alignment errors.

1 Introduction

In general deformation model completion is an ill-posed problem as the occluded part can be in any shape at any instant and most dynamic cases have short temporal interval. This paper focus on how to fuse the partial deformable surfaces over time to form a complete model.

In this paper, entire modelling pipeline is separated into three steps. An image sequence is captured using a depth camera. Each captured depth map defines a partial surface of a deforming object at each instant of time. Temporal point correspondences are located using image sequence. In the second step the temporal correspondences referred as anchor points to wrap partial surfaces using global deformation algorithm \[1\]. Later they become part of the same object surface at the same instant. In the third step, partial surfaces are assembled together into a complete watertight
surface using volumetric method.

The work done by the authors is closely related to the modeling and motion tracking techniques by Pekelny and Gotsman. They both assumed the deformation as piecewise and rigid. The estimated rigid transformation components are used to merge partial surfaces over time using Iterative Closest Point (ICP) method. Our method can deal with both rigid and non-rigid smooth deformations and does not require manual segmentation of different components.

2 Related Work and preliminary results

The proposed framework was based on Structures from Motion (SFM) techniques. Which was originally limited to static scenes. It has recently extended to reconstruct dynamic non-rigid scenes by making extra assumptions about shape deformation. The motion of a non-rigid time varying object can be decomposed into a rigid transformation and non-rigid deformation. Represented by a set of sparse feature points and their motions, shape deformation has been successfully reconstructed using different models like Gaussian distribution or based on probabilistic principal component analysis.

In order to obtain a complete dynamic model from dynamic scenes, a camera array system is usually deployed to capture objects from different views. Surface reconstruction is done using either multi view stereo algorithms or shape from silhouette techniques. Unfortunately, missing regions caused by occlusions are still hardly avoidable no matter how many cameras are used in the most real cases. Hence dynamic reconstruction remained as an open problem in filling these
missing regions\textsuperscript{1, 5}.

Considering the problem of Hole filling high quality static models that are acquired using laser range scanner with relatively small missing parts. Dynamic reconstruction from sparse cameras has become an active research topic recently in both graphics and vision, due to its usability in many future applications\textsuperscript{4, 9}. In addition to Pekelny and Gotsmans work deforming objects are modelled as 4D hype-surface with spatial temporal smoothness. Missing regions can be filled by sampling the 4D surface.

3 The major steps of the algorithm

This work focuses on how to assemble surface patches captured at different time instants for the same dynamic object into a complete 4D space-time model. These are the steps employed in this algorithm. The inputs to the system are multiple color depth image pairs captured at different time instants. In the initialization step correspondences from different frames are established by any tracking algorithms, typically SIFT is employed in this paper.

The depth maps are triangulated into 3D meshes. The second step estimates a rigid transformation for each frame to map surface patches from all frames into a reference global coordinate, where they roughly align with each other. A linear mesh deformation method is then applied in the next step to wrap one mesh to another.
1. Initial alignment  The motion of deformable object is decomposed into rigid part and non-rigid part. The goal is to separate a potentially large rigid translation and rotation from a relatively small surface deformation. A rough rigid transformation is achieved in this step. The feature correspondences between frames are mapped to 3D point correspondences. The transformation between two non-overlapping frames will be estimated by accumulating the Pairwise ones between consecutive frames.

2. Warping between two consecutive frames  Due to the constraints by traditional epipolar geometry, the object is assumed as an arbitrary, nonlinear deformation in long term, but linearly continues in short term. Therefore, surface patches can be wrapped to shapes in neighboring frames using linear mesh deformation, given sufficient feature point correspondences. In the mesh deformation, surface shapes are usually described locally by Laplacian coordinates for vertices. When deformation is large, the shape undergoes large rotation or scaling, the Laplacian coordinate is not a good descriptor since it is well known as affine-variant.

If the mesh performs rotation, the Laplacian coordinates of its vertices should also rotate with the same angle along the same axis. Unfortunately, this cannot be properly handled by Laplacian coordinates. In order to overcome this issue, The authors proposed an explicit affine transformation together with the warping procedure to account for any large affine transformation.

3. Warping all frames simultaneously  Sequential warping algorithm is employed to obtain a 3D shape by assembling two surface patches each time. For example in order to create the shape surface at frame I, the surface patch at frame is combined with frame 2 into a new surface 1-2
which is combined with frame and so on. By this all surfaces are combined to a desired shape at frame I. Since local temporal correspondences between two adjacent frames determining the warping procedure in each step, errors can be easily accumulated from frame to frame, causing misalignment between surface patches.

As the sequential warping cannot deal with occlusion authors employed a global warping algorithm in order to warp surface patches in all frames altogether to the destination frame in a single step. A global warping system is the extension from local warping algorithm which lists all Laplacian constraints as diagonal sub matrices. This gives a single linear system with unknown as the final positions of vertices in the destination frame.

4. Occlusion Handling To complete the 3D model of frame the occluded part of the frame needs to be recovered with the information from its neighboring frames. One possible way is that occluded part is topologically connected to the visible surfaces, and will be pulled to certain position under the Laplacian constraint. This approach resulted in incorrect warping when the occluded part itself is moving during these frames.

The more sophisticated approach used to comprise the tracked features to predict the occluded positions. The features are first extracted and stored for each frame. The first step consists of a global feature pool by searching the feature set of each frame for those that are visible in multiple frames. The global features are recorded along with their frame numbers and corresponding 3D positions. With those globally tracked features, the occluded part can be interpolated or extrapolated under the continuous motion assumption.
5. Smoothing and Refinement  After warping procedure surface patches are aligned together to over the shape of an object at the same instant. In this section, they are merged together to form a single surface. Instead of manipulating meshes directly, the authors used an Eulerian approach by volumetric representation. As the volumetric representation easily handles topological changes among different meshes as well as it can fix missing holes and misalignments. Eulerian approach is straightforward to implement and it does not require the complicated re-meshing process.

4 Experimental Results

These algorithms are tested on both synthetic data and real data. The synthetic data is generated with the 4D model. As shown in figure 1 for each frame the renderer outputs the depth map and
tracked 3D points specified in advance. 400 out of 20000 vertices are specified as tracking points specified as tracking points, which are uniformly distributed on the object surface. Figure 1 shows the results with correct correspondences. The complete 3D model is recovered II.

The next set of experimental results have been taken from another set of research conducted using 3D cameras, a CamCube ToF camera and Kinect sensor.

The above figure shows the results and the depth measurements were based on time-of-flight principle which allows parallel measurement of its phase, offset and amplitude. The difference between ToF cameras and Kinect camera is their resolution. Kinect displays VGA resolution whereas the former on has still have a limited resolution of (200 x 200), however both can work in illuminated conditions as they are auto-illuminated. The basic experimental results is to compare both the cameras in different configurations.
They have briefly described their perception and manipulation of deformable objects primarily based on textiles and plants. A good measure of graspability for clothes lying on a flat surface has been presented. Another interesting experimental setup which shows the segmentation of the leaf using the 3D tracking information, so when the leaf is manually selected the robot arm tries to keep the leaf into the image area.

5 Conclusion

The authors have developed a novel approach to reconstruct complete 3D surface deformation over time by a single camera. The deformable surface patches are stitched together by mesh deformation in a global manner, and merged into a complete model by a volumetric method. Tests had been conducted on both synthetic and real data demonstrated that this approach works well with even large deformation. This approach will help to simplify the challenging task of creating time varying models for dynamic objects.

1. M. Botsch and O. Sorkine. On linear variational surface deformation methods. In Transcation on visualization and Computer Graphics. IEEE, 2008.

2. Y. Pekelny and C. Gotsman. Articulated object reconstruction and markerless motion capture from depth video. In Eurographics, 2008.

3. L. Torresani, A. Hertzmann, and C. Bregler. Learning nonrigid 3d shape from 2d motion. In In proceedings of NIPS, 2003. 1
4. B. Curless and M. Levoy. A volumetric method for building complex models from range images. In SIGGRAPH, 1996.

5. J. Shen and S. C. S. Cheung, “Layer Depth Denoising and Completion for Structured-Light RGB-D Cameras,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 1187-1194, 2013.

6. B. K. Horn. Closed-form solution of absolute orientation using unit quaternions. In Journal of the Optical Society of America, 1987.

7. O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Ross, and H.-P. Seidel. Laplacian surface editing. In Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing, 2004.

8. C. Tomasi and T. Kanade. Shape and motion from image streams: a factorization method. In Technical Report CMUCS-91-105 Carnegie Mellon University, 1991.

9. J. Shen, P. C. Su, S. C. S. Cheung, J. Zhao, “Virtual mirror rendering with stationary rgb-d cameras and stored 3-d background,” IEEE Transactions on Image Processing, vol. 22, no. 9, pp. 3433-3448, 2013.

10. S. Osher and R. Fedkiw. Level Set Methods and Dynamic Implicit Surfaces. Springer-Verlag, 2002.

11. E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun. Performance capture from sparse multi-view video. In Siggraph. ACM, 2008.