3D Object Surface Tracking Using Partial Shape Templates Trained from a Depth Camera for Spatial Augmented Reality Environments

Kazuna Tsuboi and Yuji Oyamada and Maki Sugimoto and Hideo Saito
Graduate School of Science and Technology
Keio University
3-14-1, Hiyoshi, Kohoku-ku, Yokohama, Kanagawa,Japan 223-8522
kazunat@hvrl.ics.keio.ac.jp

Abstract

We present a 3D object tracking method using a single depth camera for Spatial Augmented Reality (SAR). The drastic change of illumination in a SAR environment makes object tracking difficult. Our method uses a depth camera to train and track the 3D physical object. The training allows maker-less tracking of the moving object under illumination changes. The tracking is a combination of feature based matching and frame sequential matching of point clouds. Our method allows users to adapt 3D objects of their choice into a dynamic SAR environment.

Keywords: **S**patial Augmented Reality, Depth Camera, 3D object, tracking, real-time, point cloud

1 Introduction

Augmented Reality (AR) visually extends the real environment by overlaying virtual objects on to the real world. Spatial Augmented Reality (SAR) aims to enhance the visual perception of the real environment by projecting virtual objects directly on to the real environment with a projector. The projection allows multiple users see the fusion with the naked eye which is effective for cooperative scenes like product designs.

An AR system becomes convincing when the geometric consistency between the real environment and the virtual object is kept in real time. This is achieved by real- time determination of the relation between the camera and the real environment. Solving this problem is the most fundamental and challenging task for AR research.

This paper presents a marker-less tracking method of a moving object under SAR environment using single depth camera.

2 Related Works

The illumination of a SAR environment changes drastically by the projected light. The common approach to avoid interference of light during the tracking is using sensors unaffected by illumination as in Dynamic Shader Lamps (Bandyopadhyay 2001). However, special sensors must be prepared and attached to target objects. Instead, colour camera and computer vision techniques were used to avoid physical sensors. Audet proposed a marker-less projector camera tracking system which analyse the position of the projecting image and the actual projection

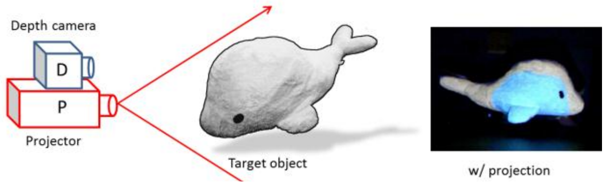

Figure 1 System overview

(Audet 2010), but is limited to planner surfaces. A depth camera is robust to illumination change and computer vision technique is applicable to the depth image.

A depth image can be converted to point clouds. The ICP algorithm (Besl & McKay 1992) can align point clouds by assuming that the corresponding points are close together and works in real time as in Kinect Fusion (Izadi. 2011). A feature based algorithm is also used which automatically finds the corresponding points with similar feature. It enables global matching and also gives a good initial registration for ICP algorithm (Rusu 2009).

3D objects have different views when perceived from a variety of viewpoints. Prior 3D object tracking methods commonly track objects by matching 3D model of the target with the input data. Azad matched the input colour image and a set of 2D images of the object from different views generated from the 3D model (Azad 2011). However, this conversion of the 3D model takes time and resources. By using depth camera, direct matching of the model and the input is possible.

The proposed method trains the shape of the object as point clouds and tracks the object using a depth camera. The method is illumination invariant and handles the view change by direct matching of the model and the input point cloud. The tracking is in real-time by matching point clouds combining the feature based algorithm and ICP algorithm.

3 Proposed method

The proposed method tracks a moving rigid object using single depth camera under SAR environments. The setup of the system is as shown in Figure 1 .The system tracks the moving object in front of the camera. The virtual object is rotated and translated based on the result of tracking and projected onto the object. The proposed method can be divided into two step, off-line training which trains several point clouds of the object from different viewpoints and on-line tracking which is a combination of feature based and frame sequential matching.

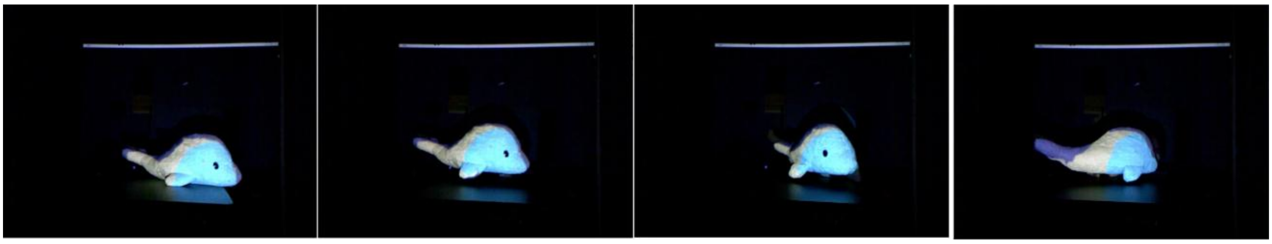

Figure 2 Result of Tracking

3.1 Training of Partial Shape Templates

The point clouds of the object from different views are first trained. The relations between each template are labelled and also trained with the templates. These point clouds are called partial shape templates. This relationship is used to guess the new template when the object view changes and re-initialization is necessary. The trained point clouds are meshed to polygon model for surface projection. The tracking method only uses the point cloud of the templates to track the surface of the object.

3.2 Feature Based matching

The feature used in the feature based matching is FPFH (Rusu 2009) which describes the relationship between the point and the surrounding points.

By using the feature based matching, all the templates are matched to the first input point cloud of tracking. The best matching template is chosen as the template of the current view. The chosen template cloud is rotated and translated to the position of the object. This result becomes the initial position of the following Frame Sequential Matching.

3.3 Frame Sequel Matching

Frame to frame matching starts from the initialized point cloud. The shape and the position of the point clouds are similar in the consecutive frame which can be easily matched by ICP algorithm.

In order to track in real-time, Generalized ICP (Segal 2009)is used. The result of ICP algorithm, which are the rotation and translation of the consecutive frame, is calculated and applied to the template. The position of the template indicates the current position of the 3D object.

3.4 Re-initialization and Selection of New Template

The appearance of the object changes as the object rotates which eventually will not match the template currently used. A new template is chosen from the trained templates by analogizing the movement and the relation of templates. The template is initialized again to the current input point cloud and frame sequential matching starts again.

4 Experiment Result

We implemented a texture projection SAR application to test our method under illumination changing environment. The texture is augmented by projecting a surface polygon made from the trained templates to the position of the tracking result. The camera and the projector are pre-calibrated. The used devices are the following:

- Depth camera: Kinect
- CPU : Intel(R)Core(TM)i7-2600 3.40GHz
- Memory : 3.49GB

The result is shown in Figure 2. We projected yellow and blue texture on the object. The texture followed the object while it moved and turned. There was a small delay on the projection. The average tracking speed was about 10fps for this object.

5 Conclusion

We presented a 3D object tracking method for SAR application using a depth camera. The proposed method has off-line training and on-line tracking. Experiment result show that 3D objects were tracked in video rate speed and under illumination changing environment. For future extension, we will implement a application which user could change the texture of the object while tracking which would be useful for designing such as products.

6 References

Audet, S., Okutomi, M., Tanaka, M.: "Direct image alignment of projector-camera systems with planar surfaces," *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* , 303-310

Azad, P., Munch, D., Asfour, T. & Dillmann, R.(2011), 6-dof model-based tracking of arbitrarily shaped 3d objects, in ‘Robotics and Automation (ICRA), 2011 IEEE International Conference on’,pp. 5204 –5209.

Bandyopadhyay, D., Raskar, R. & Fuchs, H. (2001),Dynamic shader lamps : painting on movable objects, in ‘Augmented Reality, Proceedings.IEEE and ACM International Symposium on’,pp. 207 –216.

Besl, P. J. & McKay, N. D. (1992), ‘A method for registration of 3-d shapes’, IEEE Trans. Pattern Anal. Mach. Intell. 14(2), 239–256.

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman D., Davison, A. & Fitzgibbon, A. (2011), Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera, in ‘Proceedings of the 24th annual ACM symposium on User interface software and technology’, UIST ’11, ACM, New York, NY, USA, pp. 559–568.

Rusu, R. B., Blodow, N. & Beetz, M. (2009), Fast point feature histograms (fpfh) for 3d registration, in ‘Robotics and Automation, 2009. ICRA ’09. IEEE International Conference on’, pp. 3212 –3217.

Segal, A., Haehnel, D. & Thrun, S. (2009), Generalized-ICP, in ‘Proceedings of Robotics: Science and Systems’, Seattle, USA.