Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 89 (2016) 209 – 212

Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Point Cloud Mapping Measurements using Kinect RGB-D Sensor and Kinect Fusion for Visual Odometry

N. Namitha[a], S. M. Vaitheeswaran[b,*], V. K. Jayasree[a] and M. K. Bharat[b]

[a] *College of Engineering, Alappuzha Dt., Kerala State 688 541, India*
[b,] *National Aerospace Laboratories, Bengaluru, India*

Abstract

RGB-D camera like Kinect make available RGB Images along with per-pixel depth information in real time. This paper uses the Kinect Fusion developed by Microsoft Research for the 3D reconstruction of the scene in real time using the MicroKinect Camera and applies it as an aid for Visual Odometry of a Robotic Vehicle where no external reference like GPS is available.

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016

Keywords: Kinect; Kinect Fusion; Odometry; Robotics; Vision.

1. Introduction

Visual Motion Research is at the heart of Computer Vision techniques for advancing a variety of critical technologies. These include three dimensional reconstruction, tracking, surveillance, recognition, navigation and control to name a few. In the domain of reconstruction methods, using structure from motion techniques, optical flow[1] provides information about the motion of a three dimensional scene using two dimensional projections. For a single view camera using optical flow limitation however exists, preventing the reliable estimation of motion of a three dimensional scene. The solution to this is provided by using stereo or multiple cameras which concurrently estimate both structure and motion increasing the robustness of these techniques. The downside of the methods is the complexity, requiring mapping over a number of frames, iterative refinement steps such as bundle adjustment, reduction of uncertainty, lack of smoothness due to 2D parameterization from surfaces and so on.

With the availability of time of flight RGB-D (colour-depth) cameras using structured light sensing, in recent years, access to three dimensional information in real time and frame rates has become possible. This warrants a relook in the reconstruction methodologies and formulations. Since a depth camera is available, the sensor provides structure information and surface estimation are not needed. The earlier methods were based on the displacements of the colour pattern. There are several other advantages, one of these is the introduction of the Kinetic Fusion Algorithm[2] and its extensions that enable 3D dense scanning using a moving volume and the representation to an octree which does not use the standard Iterative Closest Point (ICP) making it possible to work in real time.

*Corresponding author. Tel.: 918025086704; Fax: 918025268546.
E-mail address: smvaithu@nal.res.in

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016
doi:10.1016/j.procs.2016.06.044

In this paper, we propose to use the Kinect Fusion developed by Microsoft Research for the 3D reconstruction of the scene in real time using the Micro Kinect Camera and apply it as an aid for Visual Navigation of a Robotic Vehicle where no external reference like GPS is available.

2. Related Work

Vision sensors have shown potential to provide pose information (through dead reckoning) in structured areas and in cluttered environments where the traditional GPS can get degraded and/or get denied. Among the different approaches three common methods have received increased attention: Vision Odometry[3], Vision Based Simultaneous Location and Mapping[4] and Structure from Motion[6] applications. Among these techniques used, Vision Odometry is a low latency and low cost approach and outperforms the other two approaches in terms of computational complexity and hardware. In contrast, the latter two methods are computationally intensive and require mapping over a number of frames, iterative refinement steps such as bundle adjustment, reduction of uncertainty and so on.

Visual Odometry can be split into two categories: a) Sparse Odometry and b) Dense Odometry. The sparse methods extract a set of sparse points using feature detectors like Harris, FAST or feature descriptors such as Speeded up Robust features (SURF), Scale Invariant Feature Transform (SIFT). Correspondences between the features are established between successive frames in time. A match is assumed if the error between the patch of points is minimal. From a set of good correspondences the Transformation matrix describing the set of translations and rotations are computed, frame to frame.

In contrast, the Dense Odometry approaches use a dense set of data or the whole image data. The Transformation data is obtained from the photometrical error between the frames. Alternately, in a different approach, the geometrical error between the surfaces is obtained to describe the rigid body transformations. A drawback of this approach is that they need structured surfaces and a further problem is that they require a computationally expensive nearest neighbour search to create point correspondences. To overcome this issue, the 3D surfaces are represented as 2D depth maps. The correspondence for one point in a second depth map is found by applying the rigid body motion and projecting it to 2D coordinates. These methods suffer from long term drift and consequently accumulate errors in the motion estimation process.

Fortunately the availability of Microsoft Kinect and other RGB-D sensors using structured light sensing has resulted access to three dimensional information leading to more reliable information about the structure reconstruction process and at in real time. This has opened up new possibilities and opportunities in the field of tracking and navigation. One of the significant advances is the development of the Kinetic Fusion Algorithm[5] which has demonstrated its potential for real time dense scanning of indoor static scenes. A further gain has been the representation of the algorithm in terms of hierarchical octrees (pyramidal structure) to achieve faster computations and parallelization. Taking advantage of this, the present paper proposes to use a RGB-D camera, readily available for visual odometry application. Since the structure information is readily available, we compute the camera motion from two consecutive frames.

3. Methodology

The present work is based on the Kinect Fusion to generate a 3D robust reconstruction of the environment of a robot in real-time. This is obtained by moving the Microsoft Kinect sensor around the real scene. The input to the Kinect Fusion algorithm is a temporal sequence depth maps from the Kinect sensor. This algorithm uses only the depth maps and no color information, and do not interfere with lighting conditions, allowing Kinect Fusion to function even in complete darkness. The following Fig. 1 shows the obtained Kinect Fusion result.

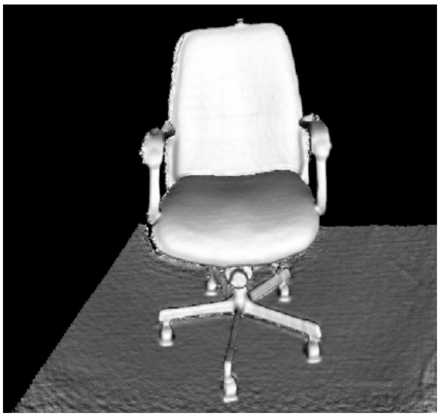

Fig. 1. A chair reconstructed with Kinect Fusion.

The algorithm runs at 30 fps in real time and provides a surface representation for each current depth frame refining a global model