INTRODUCTION

In many robotics applications robots are required to autonomously explore unknown spaces without an accurate prior map or a global position reference.  
A fundamental challenge faced by the robot is to effectively localize itself using only the information extracted from the environment.  
For example, the capability of recognizing instances of objects and associating them with unique identifiers will enable the robot to build maps of the environment and localize itself within.  
The problem of constructing a global map and localizing the robot within is referred as simultaneously localization and mapping (SLAM).  
In occupancy based approaches, the world is represented by 2D/3D grids composed of free spaces and occupied spaces.  
In recent years, SLAM with 3D dense mapping and RGB-D cameras has become more and more popular. Incoming depth and color images are converted into volumes or deformation surfaces , then matched with previously constructed volumes or surfaces to incrementally build the map.  
Data association and SLAM are inherently coupled: good data association guarantees the convergence of SLAM, and good SLAM solution gives good initialization of data association.  
A factor graph is a different representation of the SLAM problem. Instead of using small units, such as grids, volumes, or surfaces, to represent the space, a factor graph encodes the poses of the robot and the observed landmarks along the trajectory.  
In a factor graph, each factor represents a constraint on the relative poses either between two consecutive robot poses or between a robot pose and a landmark. Factor graph SLAM scales much better than SLAM with occupancy grid maps or 3D dense maps. However, the convergence of factor graph SLAM algorithms relies heavily on correct data association of the landmarks.

The focus on of this work is on SLAM in unknown environment by recognizing objects and utilizing their positions. A factor graph is the natural representation, as objects can be easily represented as landmarks.  
Object SLAM requires the robot to be able to detect objects, generate measurements, and associate these measurements to unique identifiers. Some recent work on Region-based Convolutional Neural Networks gained significant success on training deep learning models to detect multiple objects instances within a single image. However, object detections only suggest the existence of objects of certain predefined object classes in an image, but provide no data association between images.  
This is problematic for SLAM especially when there are multiple objects of the same object class in an environment.

This paper proposes a novel world representation, the nonparametric pose graph, to jointly perform data association and SLAM. The inference of the data associations and the optimization of the the robot and object poses are performed alternatively in this algorithm. This coupled framework achieves better performance for both data association and SLAM.