Articulated Motion Models for Scene Understanding

Suren Kumar  
SUNY Buffalo  
Buffalo, NY  
surenkum@buffalo.edu

Vikas Dhiman, Jason J. Corso  
University of Michigan  
Ann Arbor, MI  
{dhiman@, jjcorso@eecs}.umich.edu

Venkat N. Krovi  
SUNY Buffalo  
Buffalo, NY  
vkrovi@buffalo.edu

Abstract

We hypothesize a articulated model world view of scene understanding by representing scene as collection of objects/bodies (humans, chairs, doors, walls, floor etc.) which are connected to each other by various types of articulated joints (prismatic, revolute, static, motion along a plane etc.). Spatial structure of scene is due to the nature of these articulated joints, i.e. objects do not exhibit motions with full degree of freedom (6 for 3-dimensional world) but instead have a structure afforded by the articulated joint (door moves about a hinge). Along with this spatial structure, we hypothesize motion in world is temporally structured by finite order of motion. We use this spatio-temporal structure to represent the scene and demonstrate the proposed framework for the problem of Simultaneous localization and Area Mapping in dynamic environments.

1. Introduction

Imagine a robot moving in a typical living room environment which encounters indoor objects such as doors, drawers and chairs etc. We posit that in order for the robot to understand, map or interact with such objects, the robot needs to be able to understand the articulation.

2. Motion Models in Literature

There is significant interest in modeling motion for object tracking in computer vision literature [1]. However the modeling of motion is often done in image space with simplistic models such as move left/move right which does not explain the motion of objects in physical world. Yan and Pollefeys [6] proposed a model to detect the articulated structure of a object by using Structure from motion and then clustering trajectories to find articulation. But their approach only models revolute joint and doesn’t model the temporal structure.

There has been tremendous interest in robotics community to learn the kinematic structure such as the type of articulation from depth data [2]. However the primary limitation of such methods is the restricted work space as they do not model the appearance/disappearance of objects from the scene because they do not make any map of the world.

SLAM deals with dynamic environment in two different ways, 1) State Augmentation [4, 3] which adds the state of the moving objects to the overall state being estimated, 2) Outlier Rejection which aims to detect and remove the objects/landmarks from the mapping and localization estimation. However, both these approaches fail to model the motion of moving features/landmarks in a robust manner which precludes the possibility of using moving landmarks for both localization and mapping.

The proposed approach uses robust and physically inspired models of spatio-temporal motion which can be used to accurately describe the motion and hence enabling moving landmarks to be used in localization and mapping.

3. SLAM for Dynamic World

Figure 1 shows the graphical model of the most general SLAM problem, where $x_k, u_k, z_k, m_k, v_k$ represents the robot state, input to the robot, observation by robot, state of the world and action of various agents in the environment.

3.1. Time update

The time update models the evolution of state according to the motion model. To write equation concisely, let $A = \{Z_{0:k-1}, U_{0:k}, V_{0:k}, x_0, m_0\}$

$$P(x_k, m_k|A) = \int \int P(x_k|x_{k-1}, u_k)P(m_k|m_{k-1}, v_{k-1})P(x_{k-1}, m_{k-1}|A)dx_{k-1}dm_{k-1}$$ (1)

The independence relationship in derivation of time update in Equation 1 are due to the Bayesian networks in Figure 1a in which each node is independent of its non-descendants given the parents of that node. Given the structure of time update, we need two motion models, one for robot: $P(x_k|x_{k-1}, u_k)$ and another one for the world $P(m_k|m_{k-1}, v_{k-1})$. It can be clearly observed that
\[ P(m_k|m_{k-1}, v_{k-1}) \] for a static map is dirac delta function and integrates out in Equation [1]

### 3.2. Measurement Update

Measurement update uses the bayes formula to update the state of the estimation problem given a new observation \( z_k \) at time step \( k \). To write the equations concisely, let \( B = \{ Z_{0:k}, U_{0:k}, V_{0:k}, x_0, m_0 \} \)

\[
P(x_k, m_k|B) = \frac{P(z_k|x_k, m_k)P(x_k, m_k|A)}{P(z_k|A)} \tag{2}
\]


### 4. Dynamic World Representation

Dropping the notation for scene part, in current formulation, we assume a uniform prior \( \mu_j(0) = P(C_j), \sum_{j=1}^{N} \mu_j(0) = 1 \) over different motion models for each scene part. However, this prior can be modified appropriately by object detection such as doors are more likely to have revolute joints etc.. Motion model probability is updated as more and more observations are received [5] as

\[
\mu_j(k) \equiv P(C_j|Z_{0:k}) = \frac{P(z_k|Z_{0:k-1}, C_j)\mu_j(k-1)}{\sum_{j=1}^{N} P(z_k|Z_{0:k-1}, C_j)\mu_j(k-1)}
\]

The probability of the current observation \( z_k \) at time step \( k \), conditioned over a specific motion model and all the previous observation can be represented by various method. In the current work, we filter the states using Extended Kalman Filter, for which this probability is the probability of observation residual w.r.t a normal distribution distributed with zero mean and innovation covariance [5].

### 5. Preliminary experiment and Results

We test our algorithm in a simulated environment of 3 landmarks generated with each of the first order motion models: static, prismatic and revolute. We assume that robot egomotion is known and we use bearing and distance observations from the robot to estimate different motion models over time. Our experimental results in Figure [1b] show that our EKF based motion model and parameter estimation works as expected.

### References