DEEP LEARNING PROJECT ON

EGO-MOTION COMPUTATION

INTERIM PROGRESS REPORT

to be submitted by

Group-11
Kushagra Singhal, Mamta Bhagia, Mehul Raj Kumawat

for the course
of

DEEP LEARNING AND ITS APPLICATIONS
CS-671

Indian Institute of Technology Mandi

SCHOOL OF COMPUTING AND ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MANDI, MANDI

March 2018

1 Introduction

Visual egomotion analysis or briefly Visual Odometry is one of the key building blocks of modern SLAM systems. Ego Motion estimation is a fundamental block for any robotic system, needed for localisation and route planning and for the more complex task of mapping an unknown environment. The task of estimating the egomotion using cameras is known as Visual Odometry. Egomotion computation means to detect the motion of the person wearing the camera from the videos captured. We are expected to compute or estimate the motion of the person wearing the camera. The detection proceeds by calculating the rotational and translations from the two consecutive frames from the video. All the estimations and calculations will be done in the two coordinate system framework.

2 Motivation

Traditionally, various sensors like rotary encoders are used to determine the motion of wheels to calculate the translational and rotational motion. But this method of estimating the motion fails when we want to analyse the motion of some non standard locomotion device such as legged machine or humans. Also some errors are bound to come due to slipping of wheels on smooth surface. So these odometry readings become unreliable when analysing motion over long time period as the errors tend to accumulate. Visual Odometry can be used here to minimize the error and enhance navigational accuracy for any type of locomotion on any surface.

In the last decade, huge research work has been put to the visual localization but no state of the art solution has been put forward. Ego Motion Estimation can be one of the solution to this problem. NASA uses visual odometry technique to know the trajectory of Mars Exploration Rovers.

3 Background

The methods to estimate the ego motion can be divided into two subcategories: Geometric Method and Learning Method. [1] Geometric Method involves analysing consecutive frames and detecting camera motion between these consecutive frames. This can be done by extracting salient features from the image and reconstructing the 3D points using the method of triangulation. Another geometric method is dense optical extraction method which calculates the displacement vector of each pixel by considering few assumptions. In Learning Method, network learns the features that it finds useful, on its own, from the frames available.The frames are divided into cells and then average optical flow is computed for each block and training is done. [2] The modern learning algorithms takes the plain input and starts learning the correct representation for the task with impressive results. Filters are stacked one over another and their coefficients are learned by supervised or unsupervised learning.

4 Problem Statement

In this project, we intend to compute the ego motion of a person. Ego motion, in all future mentions is meant to be a human head-mounted camera's motion. We are to estimate the translational and rotational motion of the head-mounted camera.

When a camera captures a video, a particular moving element in the picture has motion **X**. If the camera itself is moving, then its motion is **Y**. Given the frames from the cameras video footage, we only have **X** + **Y**. Using this, we have to find **Y** and **X**, separately.

5 Tentative Plan

5.1 Generating Training Data

To generate the training data set, we have two options at the moment:

1. Installing a survelliance camera on a fixed, unmovable point, which monitors the persons motion. Comparing the two videos and their frames at the same instant, will give us information on the change of camera video on persons motion.

2. Using motion sensors, monitor the persons translation and rotation. Knowing the persons motion, we can learn the cameras F2F estimates accordingly.

5.2 Segmentation of Background and Foreground Features.

In two consecutive frames, if we are able to differentiate which feature is not moving in reality and which feature is, then we can use the stable background features to find the cameras ego-motion. From most of the segmentation techniques, the Background Subtraction Techniques provides the most complete object data but is extremely sensitive to dynamic scene changes due to lighting and extraneous events. In advanced adaptive background modelling method, each pixel is modeled using a mixture of Gaussians and is updated by an on-line approximation. This segmentation method would alone suffice for applications where a rough estimate of the moving foreground, in the the form of irregular space blobs. Here the exact shape of the moving object need not be determined and only some post processing of the segmentation output using appropriate filters would give the desired blobs of interest.

5.3 Optical Flow

Optical Flow is a method to track the apparent motion of object in a video. So we can apply optical flow algorithm to the fixed background, to easily determine the motion of camera. We take into account few assumptions while applying Optical flow algorithm :

1. Pixel intensities do not rapidly change between consecutive frames.

2. Groups of pixels move together.

Based on these assumptions we can easily conclude that a pixel with Intensity I(x,y,t) after time dt remains same with intensity I(x+dx,y+dy,dt) between two image frames.

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \tag{1}$$

By applying calculus, these equations simplify to the following equation :

$$I_x u + I_y v + I_t = 0 \tag{2}$$

Where Ix is the frame shift in horizontal direction, Iy is the frame shift in vertical direction, u is the displacement of that pixel in horizontal direction, v is the displacement of that pixel in vertical direction and t is the time difference between two consecutive frames. Out of these we need to find u and v from this one equation which is unsolvable. This is called aperature problem of Optical flow algorithm. There are few methods for solving this problem some of them being LucasKanade method and HornSchunck method.

References

[1] G. Costante, M. Mancini, P. Valigi, and T. A. Ciarfuglia, "Exploring representation learning with cnns for frame-to-frame ego-motion estimation," *IEEE Robotics and Automation Letters*, vol. 1, pp. 18–25, Jan 2016.

[2] R. Roberts, H. Nguyen, N. Krishnamurthi, and T. Balch, "Memory-based learning for visual odometry," in *2008 IEEE International Conference on Robotics and Automation*, pp. 47–52, May 2008.