Detection of moving objects from moving platform

N.G Kiewiet*, W.A Clarke†,
HyperVision Research Laboratory
University of Johannesburg
South Africa
*Email: kiewietn@gmail.com
†Email: willemc@uj.ac.za

Abstract—A method of detecting moving objects from an image sequence acquired by a single moving camera is presented. No assumptions about the movement of the camera platform is made. Motion detection from a moving platform is a non-trivial problem as the moving camera induces apparent motion in the entire image. The epipolar geometric constraint is used between matched points in consecutive image frames to estimate the camera’s ego-motion, described by the fundamental matrix. Criterion for rigidity detection are used to measure the confidence of the estimated fundamental matrix. Point correspondences which do not conform with the estimated ego-motion indicate dynamic objects. A GPU implementation is proposed to achieve significant performance increases compared to the standard CPU implementations.

I. Introduction

The detection of moving objects in a video sequence is a fundamental issue in computer vision. The aim of motion detection is to segment the video frames into static background and dynamic foreground regions. Detecting motion from a moving platform is computationally demanding with respect to motion detection from a stationary camera. The GPU implementation is proposed to accelerate the implementation for real-time applications.

Detecting motion from stationary cameras have been extensively researched and have been found to be computationally efficient. The background can be estimated at pixel level using statistical approaches [1]. This estimated background is then subtracted from the image to produce any foreground or moving objects.

Estimating the background from a moving camera requires compensating for the motion of the camera prior to updating the background model [2]. One approach is to approximate the camera motion by a 2D parametric transformation such as a $3 \times 3$ homography [3]. This assumption is exact for PTZ (Pan-Tilt-Zoom) cameras and a good approximation where the scene is far from the optic center and the depth is small compared to the distance. If this is not the case the homography model cannot account for pixel displacement due to 3D depth in the scene, known as parallax [4].

Geometric constraints within multiple frames provide a framework for detecting independent motion. The epipolar constraint can define independent motion when a point does not lie on the epipolar plane. Objects moving on the epipolar plane cannot be detected using only the epipolar constraint since the epipolar plane does not define the exact 3D position of a point. An additional constraint is needed to remove this ambiguity. This is given by a structure consistency constraint implemented within a “Plane+Parallax” framework [5] which recovers 3D information relative to a planar surface in the scene. The magnitude of each parallax displacement is directly related to the point’s relative depth.

We propose a GPU-based motion detector which implements geometric constraints from multiple views to identify independently moving objects.

II. Theory

A. Epipolar Geometry

Epipolar geometry between two views can be defined as the geometry of the intersection of the image planes with the plane joining the camera centers [6]. This is shown in Fig. 1 where:

- The camera centers are coplanar and define the epipolar plane, denoted as $\pi$;
- The two cameras are indicated by their centers $C$ and $C'$. 

![Fig. 1. Epipolar plane and point correspondences.](image-url)
The 3D point $X$ is represented by $x$ and $x'$ on each image plane respectively.

The epipolar constraint given by two views of the same 3D point is represented by the fundamental matrix $F$ such that:

$$x_2^T F_{21} x_1 = 0$$

Where,

- $x_2$ and $x_1$ represents the position of the 3D point $X$ in image 1 and 2 respectively,
- and $F_{21}$ is the fundamental matrix representing the epipolar geometry between views 1 and 2.

Estimation of the fundamental matrix will be done by applying probabilistic criterion defined by [7] to rate the meaningfulness of a rigid set of points as a function of the number of point correspondences and the accuracy of the matches. This method guarantees that the expected number of rigid sets found in a random distribution of points is as small as desired.

### B. ASIFT

In order for us to estimate the epipolar geometry of two views it is first necessary to be able to obtain point correspondences between these two views. It is desirable to obtain a large number of correct matches even with great variation in the camera axis orientation. A framework for a fully affine invariant extension of David Lowe’s SIFT feature descriptor satisfies these requirements [8]. The ASIFT algorithm is proven to be rotational, scale and affine invariant. These parameters are modeled using the affine camera model, Fig. 2. The small parallelogram on the top-right represents a camera looking at the image plane $u$. The zoom parameter is given by $\lambda$. The angles $\theta$ and $\phi$ represent the latitude and longitude, while $\psi$ represents the camera spin.

Each image is transformed by simulating all possible affine distortions, depending on the longitude $\phi$ and latitude $\theta$, caused by the change of camera optical axis from a frontal position, Fig. 3. The latitude sampling follows a geometric series $1,a,a^2,...,a^n$ with $a > 1$ while the longitudes are for each tilt an arithmetic series $0,b/t,...,kb/t$. Choosing the value of $b = 72^\circ$ and $k$ as the last integer such that $kb/t < 180^\circ$, is a good compromise between accuracy and sparsity. The value for $a$ is selected as $\sqrt{2}$ while $n$ can go up to 5 or more. All simulated images are then compared to each other using the SIFT feature descriptor [8].

### III. Future Work

We are currently in the process of creating a GPU-based implementation of ASIFT. This is followed by using a parallel architecture to implement the geometric constraints on matched feature points in consecutive images. The use of a particle filter for matched feature points and a sliding window to avoid accumulated errors will be investigated.

### References


Nico Kiewiet received his Electrical Electronic Engineering degree with IT endorsement from the UJ in 2010. He is currently doing his Masters in Electrical Engineering at the UJ in machine vision at the high performance computer vision research lab. Research interests include high performance computing and machine vision, DSP, texture analysis and GPGPU programming.