

Ego-motion estimation concepts, algorithms and challenges: an overview

Naila Habib Khan¹ · Awais Adnan¹

Received: 17 December 2015 / Revised: 11 July 2016 / Accepted: 5 September 2016 /

Published online: 14 September 2016

© Springer Science+Business Media New York 2016

Abstract Ego-motion technology holds great significance for computer vision applications, robotics, augmented reality and visual simultaneous localization and mapping. This paper is a study of ego-motion estimation basic concepts, equipment, algorithms, challenges and its real world applications. First, we provide an overview for motion estimation in general with special focus on ego-motion estimation and its basic concepts. For ego-motion estimation it's necessary to understand the notion of independent moving objects, focus of expansion, motion field, and optical flow. Vital algorithms that are used for ego-motion estimation are critically discussed in the following section of the paper. Various camera setups and their potential weakness and strength are also studied in context of ego-motion estimation. We also briefly specify some ego-motion applications used in the real world. We conclude the paper by discussing some open problems, provide some future directions and finally summarize the entire paper in the conclusions.

Keywords Camera motion · Ego-motion · Motion estimation · Visual odometry

1 Introduction

Ego-motion technology is widely used nowadays for autonomous navigation, video conferencing, remote surveillance, visual odometry, human computer interaction and short term control applications. The highest accomplishment and effectiveness of ego-motion technology can be known from the fact that it was used by NASA for the famous mars rover project [18, 60]. Ego-motion technology facilitated in robot navigation on another planet, i.e. Mars.

✉ Naila Habib Khan
naila.khans@yahoo.com

Awais Adnan
awais.adnan@imsciences.edu.pk

¹ Department of Computer Science, Institute of Management Sciences, Peshawar, Pakistan

The term “ego-motion” can be traced back to the work done by [30] in the field of psychology, he explains his views on theories of visual space. Six years later Warren [106] gives his idea of ego-motion perception on the basis of visual signals. These two studies marked the entry of ego-motion in the field of psychology, many other researchers, for instance, Neisser [70], Brandt, et al. [9], Prazdny [75], Prazdny [76] followed and produced their own concepts for ego-motion.

In computer vision domain, the first computational model for ego-motion was developed by Prazdny [76]. Prazdny believed that optical flow fields are generated at an observer’s retina as he moves in a 3D world. The ego-motion parameters of the observer and the relative depth map of the stationary environment can be computed by using the Instantaneous Positional Velocity Fields (IPVF). Only local properties of optical flow are enough for computation of ego-motion. Prazdny developed a computer model for investigation, analysis and performance evaluation of the proposed method. The computer model provided a simulation of 3D world where an observer is moving through the environment. Prazdny’s final result confirmed that the method was reasonable and feasible for ego-motion estimation from optical flow.

Many computer vision researchers have focused on providing solutions to ego-motion technology. These solutions were based on diverse algorithms [12, 36, 49, 59, 76, 79, 97, 98, 116], ego-motion computation using stereo sequences [4, 20, 22, 23, 26, 42, 55, 62, 65, 72–74, 82, 84, 90] and monocular [48, 66, 80, 90, 110, 111, 115] sequences. On the other hand, several researchers have also focused on methods that learn visual representations from videos, where knowledge of ego-motion can be used for feature learning [1, 44, 89]. To learn visual representations, the videos are normally captured using vehicle mounted cameras. Most notably, the method suggested in [44] predicts feature responses to a distinct collection of observer motions and discrete ego-motion in one image is estimated using linear transformations. Olson et al. [72] used probabilistic approach to develop an algorithm for improving the robustness of ego-motion estimation methods for long distance navigation. They claim that their approach can achieve an error below 1 % for long distance travels. Ess et al. proposed a method in [25] for tracking pedestrians in the presence of occlusions and outliers. They use a novel feedback connection from the object detector to visual odometry which utilizes the semantic knowledge of detection to stabilize localization.

1.1 Outline of the research paper

Our goal with this review paper is threefold: (1) For computer vision readers, we expect this paper to serve as an introduction to ego-motion and its related concepts, and believe that our work will be a substantial contribution to the computer vision domain. (2) For ego-motion researchers, we believe this review paper will provide a handy opportunity to study the ego-motion estimation methods/algorithms in diverse scenarios. (3) For visual odometry readers, we hope this review to contribute and extend the significance of ego-motion technology in the computer vision domain.

The paper is structured as follows: Section 2 begins by introducing the motion estimation in general, ego-motion technology and provides an understanding towards some basic motion estimation concepts such as independent moving objects, focus of expansion, motion field, and optical flow. Section 3 discusses three major applications of ego-motion technology, i.e. autonomous vehicles, visual odometry and visual SLAM (Simultaneous Localization and Mapping). Section 4 critically reviews the previous literature and specifies various algorithms that have been successfully used for ego-motion estimation. Next, Section 4 presents several types of camera setups along with its strengths and weaknesses in computer vision. Section 5

enlightens the major ego-motion estimation complexities of occlusion, noise and camera calibration. Section 6 lists some open problems found during this literature review. Finally, section 7 summarizes and concludes the research paper and provides some future directions. Figure 1 summarizes the contents of this research article.

2 Motion estimation concepts

An image sequence is composed of a chain of moving images which are learned at discrete instances of time [99]. When sub-sampling image sequences, it should be made sure that the time interval is small enough, it guarantees that discrete sequence represent an actual continuous image sequence evolving over time [99]. There are two motions associated with any image sequence i.e. the camera motion and the object motion in the scene [40]. Further, there are three possibilities for relative motion within an image sequence. The observing camera can move in front of a static scene, or objects can move in front of a fixed camera, or the camera and the object can be moving simultaneously at any given time [99]. 3D motion estimation is challenging task since a total of seven parameters are required to compute motion at each and every pixel [40].

Understanding motion estimation concepts in computer vision is of great importance. In the sub-sections, we have discussed all basic concepts when estimating motion in a scene. Firstly, in section 2.1, we have discussed ego-motion which is the true motion of moving camera also known as observer motion. Ego-motion has three parameters associated with it translation, rotation and depth. In the perspective of camera motion if there are moving objects in the scene they are known as independent moving objects. Section 2.2 discusses independent moving objects and the complexities it poses in estimating ego-motion. Focus of Expansion is also an important concept when dealing with motion estimation. Using FOE, it can be found if the camera or observer is moving towards the objects or away from it, see section 2.3 for more details. Section 2.4 and 2.5 discusses motion field, Motion field is the true motion of objects in

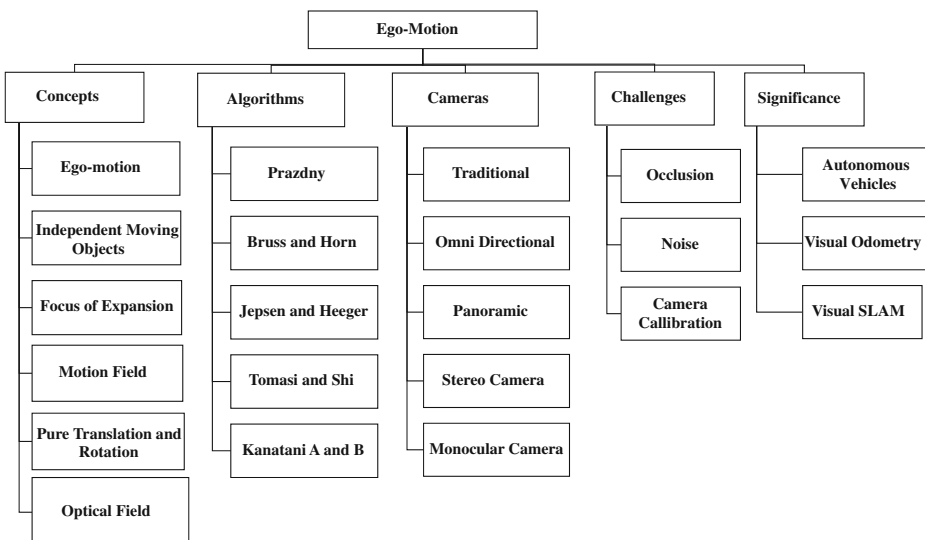


Fig. 1 Ego-motion framework

a scene, and motion vector is assigned to each and every point in the image. These motion vectors can be pure translational or rotational. Finally, section 2.6, we discuss the optical flow which is the motion of brightness or luminance patterns in an image.

2.1 Ego-motion

Raudies and Neumann [78] defined ego-motion as a sequence of images recorded using a moving camera, having depth of the pictured environment along with all the data about the 3D movement of the camera. Similarly, Baik, et al. [6] defined visual ego-motion as a continuous process where 2D image sequences captured by a camera are used to estimate 3D camera pose. Generally speaking, ego-motion is the camera's motion within an environment, relative to a rigid scene, where the motion can be 3D.

Visual motion fields have been used as a primary model for ego-motion estimation over the past 30 years [78]. At low power consumption, rich information can be extracted from the cameras, but there are certain issues with handling all the information from a camera before the ego-motion can be inferred. Ego-motion detection is a challenging task; a reliable and accurate technique is yet to be discovered. Image motion, ego-motion and depth of scene are tightly coupled with each other [91]. The rotation, translation and depth parameters for ego-motion estimation is given below [69, 108].

$$T = (T_x, T_y, T_z)^T \quad (1)$$

$$R = (R_x, R_y, R_z)^T \quad (2)$$

$$Z(x, y) \quad (3)$$

Here T represents the observer translation, R represents the observer rotation and $Z(x, y)$ is depth at any location. See Fig. 2 for translational parameter and Fig. 3 for rotational parameter respectively.

Visual ego-motion estimation is a traditional problem in the computer vision literature. Raudies and Neumann [78] lay emphasis on the following three estimation problems: First, how to compute the optical flow. Second, how to estimate ego-motion using optical flow when combined with the visual image motion model. Third, how to notice the translation speed of the observer for relative depth estimation. According to Gluckman and Nayar [32], the main problem with ego-motion is recovery of translation direction and observer rotation as it moves through the environment. Numerous algorithms have been developed by the vision researchers to provide a solution to this ego-motion problem [32]. Baik, et al. [6] specified ego-motion estimation as a state estimation problem, that can be solved effectively using Bayesian filtering and particle filtering methods. Camera projection always has problems on nonlinearity, Bayesian filtering methods are capable to deal with this nonlinearity. Apart from these problems, visual ego-motion holds great significance for computer vision applications, robotics, augmented reality and vSLAM (Visual Simultaneous Localization and Mapping) [60, 71, 74].

2.2 Independent moving objects

As an observer moves through the environment the scene changes. If the scene is static and no objects are moving, essential information about the motion of the observer can be inferred

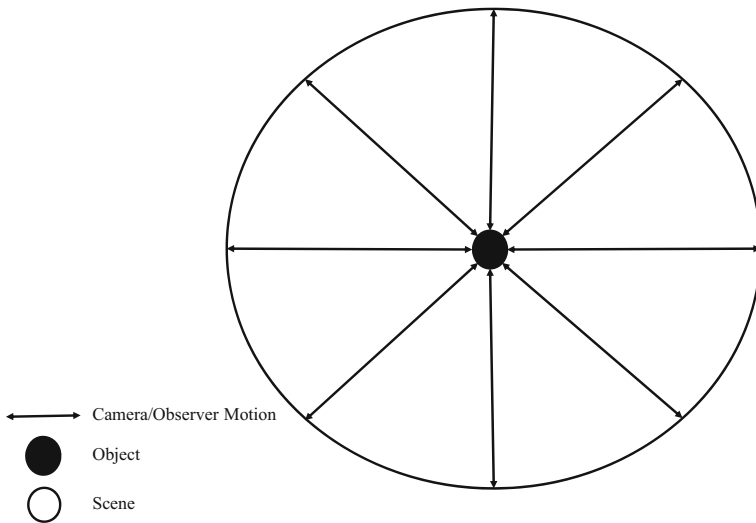


Fig. 2 Translational parameter for ego-motion estimation

from such image sequences [78]. Objects in the scene that are non-static are referred to as “moving independently” [78]. Moving object detection in videos is a key step for information extraction in several vision based applications such as semantic annotation, video surveillance, traffic monitoring and people tracking or more generally objects tracking. It is challenging to extract observer information from scenes with moving objects. It becomes difficult to interpret whether the objects within the scene are moving or the observer is moving.

When attempting to construct any vision system, it’s important to address the visual motion too [78]. The task of identifying Independent Moving Objects (IMO’s) by an observer that is moving

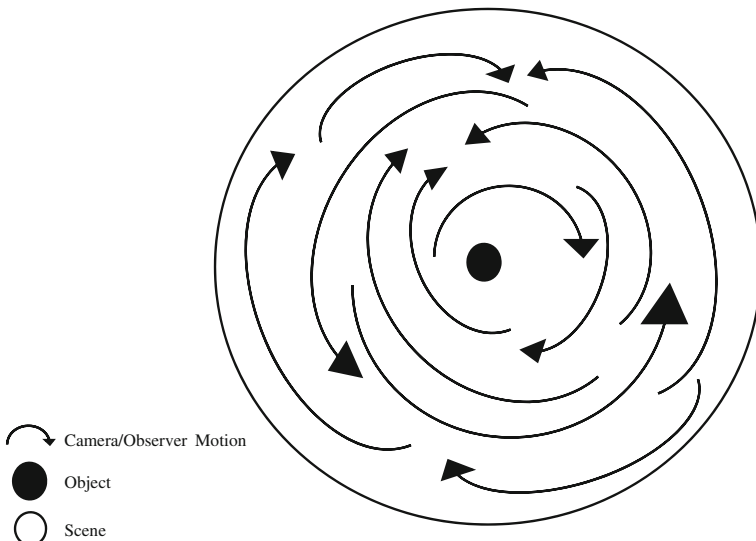


Fig. 3 Rotational parameter for ego-motion estimation

is of remarkable significance and is effortlessly achieved by humans. Developing algorithms that can achieve such tasks is challenging for computer vision researchers. Even for navigational purposes, it's necessary that the observer is capable of identifying moving objects in the scene, in order to avoid collisions [59].

The fastest algorithm for moving object detection in a static camera environment is frame differencing [47]. In frame differencing two consecutive image frame difference is computed and moving objects are found. However, if in a scene both camera motion (ego-motion) and object motion is present, then differencing is not applicable since two independent motions are involved [47].

Independent movement of one or more objects gives rise to several issues in ego-motion estimation. Famous work was done by MacLean, et al. [59] for dual motion estimation i.e. object and ego-motion. MacLean, et al. [59] used Heeger and Jepson subspace method and expectation maximization (EM) algorithm for ego-motion estimation. Several constraints were modified so that a sphere camera can be used instead of a pinhole camera. Bilinear constraint of Gaussian distribution function was suggested as a constraint [59].

2.3 Focus of expansion

In pure camera translation conditions, when the camera is looking forward, all image features seem to diverge from a specific image location known as the focus of expansion (FOE) [13, 14]. In backward camera motion image features converge from a specific image location called the focus of contraction. In other words FOE can be described as the intersection point of translation vector with the image plane [60, 64]. If the object present in the scene is stationary, then all the velocity vectors emerge from the FOE. In scenarios of independent moving objects velocity vectors will have different directions of image flow.

Finding the focus of expansion from the frames is significant for the computation of 3D camera translation [40, 60]. Focus of expansion cannot be determined when using traditional cameras, however it is covered when using wide-angle imaging systems [32]. Optical flow can also be used to find the FOE. Once FOE is determined, it can be used to estimate distances to points in the scene. These distances can be used for finding the time to impact or collision between objects and camera in the scene. The rate of expansion gives the estimated time for collision. This rate of expansion feature is used in applications for the control of moving vehicles and robotics, collision warning system and obstacle avoidance [64].

2.4 Motion field

Visual motion fields are used for projection of 3D scene points onto camera surface [78]. Visual motion fields are also used to observe movements in any scene, movement can be rotational or translational. Motion field can also be used for determining the relative depth in a scene, which, when combined with ego-motion produces information beneficial for visual navigation [78]. The motion field solves problems like recovering 3D velocity field, image segmentation for moving object identification or reconstruction [104].

Illustration of 3D motion projection onto a camera image can be achieved using motion field [99]. In the motion field each pixel in the image is assigned with a velocity vector. The relative motion between the 3D scene and the camera allows these velocities to be induced. Each point of a 3D scene is projected as a point in the image. However the location of projection of a fixed point may vary with time.

Variations in image brightness patterns, spatial and temporal, is known as ‘optical flow’ and can be used for motion field estimation [104]. The motion field can also be represented using a mapping function that maps image coordinates to a 2D vector. The image of a scene point, P , is the point p given using the equation below [99].

$$p = f \left(\frac{P}{Z} \right) \tag{4}$$

Where $P = [X, Y, Z]^T$ is a 3D point in the typical camera reference frame. The Z axis is the optical axis and the focal length is denoted by f . The projection of the scene point is given by p , such that $p = [x, y, f]$. The third coordinate z , is always equal to f , hence the projection p can be written as $p = [x, y]^T$. The relative motion between the camera and the scene point P can be expressed using the Eq. (5) [99].

$$V = -T - \omega \times P \tag{5}$$

Where T is the translational component such that $T = (T_x, T_y, T_z)^T$ and ω is the angular velocity such that $\omega = (\omega_x, \omega_y, \omega_z)^T$. In expanded form the equation (5) can be written as,

$$V_x = -T_x - \omega_y Z + \omega_z Y \tag{6}$$

$$V_y = -T_y - \omega_z X + \omega_x Z \tag{7}$$

$$V_z = -T_z - \omega_x Y + \omega_y X \tag{8}$$

The time derivative of both sides of the projection Eq. (4) is taken in order to obtain the relation between the velocity of P and the corresponding velocity of p .

$$v = f \frac{ZV - V_z P}{Z^2} \tag{9}$$

In expanded form the Eq. (9) can be written as

$$v_x = \frac{T_z x - T_x f}{Z} - \omega_y f + \omega_z y + \frac{\omega_x x y}{f} - \frac{\omega_y x^2}{f} \tag{10}$$

$$v_y = \frac{T_z y - T_y f}{Z} + \omega_x f + \omega_z x + \frac{\omega_y x y}{f} - \frac{\omega_x y^2}{f} \tag{11}$$

Finally, it can be concluded that motion field is the sum of two main components i.e. translation and rotation. The part of a motion field that depends on angular velocity does not carry any depth information [99].

2.5 Pure translation and rotation motion field

The motion field is the sum of two components, one of which depends on the translation and the other depends on the rotation [99]. In pure translation, rotational component ω of the motion field is considered to be zero.

Since $\omega=0$, both Eqs. (10) and (11) becomes,

$$v_x = \frac{T_z x - T_x f}{Z} \quad (12)$$

$$v_y = \frac{T_z y - T_y f}{Z} \quad (13)$$

There are two cases associated with the pure translational component. Let p be any point in the scene, if $T_z < 0$ then p is the focus of expansion, the motion field vectors point away from it. If $T_z > 0$ then motion field vectors point towards p and is called the focus of contraction. In the first two cases when $T_z \neq 0$, motion field is said to be radial. If $T_z = 0$, the motion field vectors become parallel.

In pure rotation translational component T of the motion field is considered to be zero. In such a case, pure rotation can be computed by modifying the Eqs. (10) and (11),

$$v_x = -\omega_y f + \omega_z y + \frac{\omega_x x y}{f} - \frac{\omega_y x^2}{f} \quad (14)$$

$$v_y = \omega_x f + \omega_z x + \frac{\omega_y x y}{f} - \frac{\omega_x y^2}{f} \quad (15)$$

2.6 Optical flow

The recovery of motion field, which is a perspective projection onto the image plane of true 3-D velocity field, is thought to be a crucial step. Only two forms of data are available, the spatial and the temporal. The temporal variations in the image are referred to as brightness pattern [104]. From these variations it is possible to derive an estimate of the motion field called the optical flow [104].

Optical flow is the distribution of apparent velocities of image brightness patterns in an image [37]. This optical flow can be generated due to the relative motion of the viewer or the objects. Optical flow can provide data about the spatial arrangement of objects and its rate of change. The velocity measurements must be accurate and dense for the inference of ego-motion [7].

Many methods have been proposed for the computation of optical flow [7]. Over thirty years have passed and most of the latest methods resembles the original technique of Horn and Schunck [96]. Therefore, Horn and Schunck method can be correctly termed as classical [96]. Horn and Schunck optical flow depends on both spatial smoothness and brightness, but it's not robust to the outliers [96]. Black and Anandan [8] formulated a robust framework to deal with outliers in both the spatial and temporal terms. In order to generate piecewise smooth results the quadratic regulariser in Horn and Schunck model was replaced by a smoothness constraint [2, 11, 21, 67, 83, 107]. To preserve flow discontinuities L1 penalty was used by Shulman and Herve [88]. Some of the best methods that can be used to find optical flow include coarse-to-fine estimation, texture decomposition, incremental warping, warping with bi-cubic interpolation, graduated non-convexity, temporal averaging of image derivatives and median filtering [96].

3 Significance of ego-motion estimation

Ego-motion technology is beneficial for many applications such as autonomous navigation, video conferencing and remote surveillance [20, 29, 32, 71, 93, 100]. According to Tsao, et al. [100], ego-motion is extremely useful for human computer interaction and short term control applications.

For several years the primary goal of computer vision researchers has been directed towards effective use of video sensors for navigation and obstacle detection [71]. Using image sensor for motion estimation is an interesting idea, its compact, low-cost and a self-contained device [54]. Also, these sensors can be very important if used as a component in navigation systems. Various types of sensors are used for the purpose of navigation and motion estimation, the primary ones are GPS and inertial measurement units (IMUs) [54].

There are several drawbacks associated with these sensors and therefore ego-motion estimation is a better option [54]. For high-precision applications, these sensors can be quite costly and prone to error. If low-cost IMUs are used, they degrade quickly unless corrected. On the other hand, GPS sensors are incapable of working indoors or under tree shelters. Therefore the conclusion drawn is to use visual ego-motion estimation in combination with the traditional methods [54].

3.1 Autonomous vehicles

Vehicle's ego-motion estimation provides the capability for advanced driving assistance systems and mobile robot localization [51]. Autonomous driving and computer vision based driving assistance requires accurate ego-motion estimation [93].

Vehicle analysis systems can be divided into two groups i.e. offline and online applications [29]. Offline applications are useful when processing video sequences to study the behavioral patterns of the driver. These behavioral patterns are used for development of tools and applications for driver assistance. On the other hand an online system holds great significance for intelligent driver support.

Omnidirectional cameras have always been preferred for ego-motion estimation due to its panoramic view support and capability to deal with ambiguities [29]. When an omnidirectional camera is mounted on an automobile it can provide a complete panoramic view which is very appropriate for both offline and online vehicle analysis applications [29].

Several researchers [16, 23, 29, 31, 41, 86, 87, 92–94, 102, 114] opted to use ego-motion technology for autonomous vehicles. Investigations have discussed several challenges that are associated with autonomous vehicles, which include and are not limited to uneven terrains, translation motions, relative rotation motion, low number of feature points on a typical road, obstacle and lane detection [93]. Various sensors have been proposed to tackle these issues such as radar, laser or GPS-based systems. Radar and laser are specifically of great importance in extreme conditions when there is no presence of light. On the other hand GPS-based systems have reliability issues in environments where they have no direct line of sight with a satellite. Using computer vision based systems, ego-motion, rather than sensors reduces maintenance and cost associated with the sensors. It also eliminates the need to calibrate between the sensors and don't have other drawbacks associated with sensors [33, 93].

3.2 Visual odometry

Visual odometry deals with motion estimation of a robot [65]. The motion is estimated from the visual input only [71]. Visual odometry has been proved to be an extremely effective tool

for vehicle safety applications when driving near obstacles, on inclines, performing slip checks, making tough drive tactics and ensuring precise imaging [60]. Visual odometry provides position knowledge, this position information allows additional autonomous vehicle capability and better analysis during planetary maneuvers [60].

As mentioned earlier, one of the most famous and outstanding experiment in stereo visual odometry is the Mars Exploration Rovers (MER) by Cheng, et al. [18] and Maimone, et al. [60]. NASA sent its two MERs and each rover had accurate knowledge of its position, which allowed autonomous detection and compensation for an unforeseen slip during the drive [60]. Several computer vision researchers [18, 54, 57, 60, 71, 105] have worked on visual odometry.

Decades of research in the field of visual odometry have proved that accurate localization and motion estimation is extremely important for navigational purposes [18, 54, 57, 60, 71, 105]. Active beacons and GPS-based navigation provides absolute positioning, but with high costs. On the other hand visual odometry reduces the maintenance and cost associated with sensors by replacing it with an inexpensive camera. It also allows more reliable and safer navigation system, which can be used in close proximity to human beings [15]. This inexpensive camera used in visual odometry provides a broader field of view and simultaneous sensing capabilities of the range and appearance [65].

3.3 Visual SLAM

In localization a robot estimates its own location relative to its surroundings [113]. In pure localization environments the robot needs to be delivered with a precise map and is incapable of adapting to the environmental changes. Manual labor, specific expertise and cost are associated with the map building for the robot. To achieve an advanced degree of independence, the robot needs to construct a map using its sensors and then recognize where it is, this is called Simultaneous Localization and Mapping (SLAM) [113]. When a robot is capable enough to travel across an environment without user-involvement, form a map, and also localize itself in the map, it is believed to achieve full autonomy [50].

SLAM has typically been used with landmarks that are too sparse. Hence, these sparse map representations are insufficient to be effectively and efficiently used for tasks like path planning, obstacle avoidance and autonomous navigation. SLAM for ground vehicles has mostly been done with sensors rather than using visual cameras [71]. If a global positioning sensor (GPS) is used and something makes the external and data beacons unavailable, then the robot must be capable of finding its position based on reference points and build a map [50]. One of the main approaches to move away from costly sensor is, camera usage and some optical encoders to locate suitable reference points, this is referred to as Visual SLAM (vSLAM), [50, 113]. Recent improvements in sensors and hardware have made vision based processing more practical and mature [71].

Visual SLAM is cost-efficient for consumer robotics in particular [50]. It improves the overall performance of localization and mapping. Its algorithm is also very robust to dynamic changes in the environment which may be caused by moving objects and/or people and/or lighting changes [113]. The primary purpose of visual SLAM is to combine the image and odometry data such that it enables robust map-building and localization. This robustness allows the sensor data acquired from the mobile robot to remove the difficult-to-model noise [50].

4 Algorithms and hardware comparisons

Section 4.1 we briefly explain some popular algorithms that are used for ego-motion estimation. In section 4.2, we overview different camera setups and specify its drawbacks in context of ego-motion estimation.

4.1 Critical review of major algorithms

This section primarily focuses on instantaneous-time algorithms for ego-motion estimation. The extensions and modifications to these algorithms by other renowned computer vision researchers are also studied.

Over the past 30 years extensive research has been devoted towards ego-motion estimation. The most common and notable method for its estimation is using the optical flow. With the passage of time several techniques evolved and point correspondence method were introduced. The main difference between optical flow and point correspondence method is of motion representation. The point correspondence method is capable of presenting large motions while optical flow presents small motions.

Overall the algorithms for ego-motion computation can be classified as discrete time algorithms or instantaneous-time algorithms [97]. In instantaneous-time algorithms, input is the image velocity while discrete time methods concentrate on image displacement [97]. In a rigid scene under perspective projection, image velocity is produced due to camera motion. As mentioned in [97], image velocity can be expressed using Eq. (16).

$$v(X) = \begin{bmatrix} 1 & 0 & -x_1 \\ 0 & 1 & -x_2 \end{bmatrix} \left(\frac{T}{Z(X)} + \Omega \times X \right) \quad (16)$$

The image velocity is $v(X)$, image position is $X = (x_1, x_2, Z)$, T is the translational velocity, Ω is the rotational velocity and Z is the depth. The focal length is taken as 1.

Prominent work on instantaneous-time algorithms was done by Bruss & Horn, Jepson & Heeger, Tomasi & Shi, Kanatani A, Kanatani B and Prazdny (see Fig. 4). All these proposed

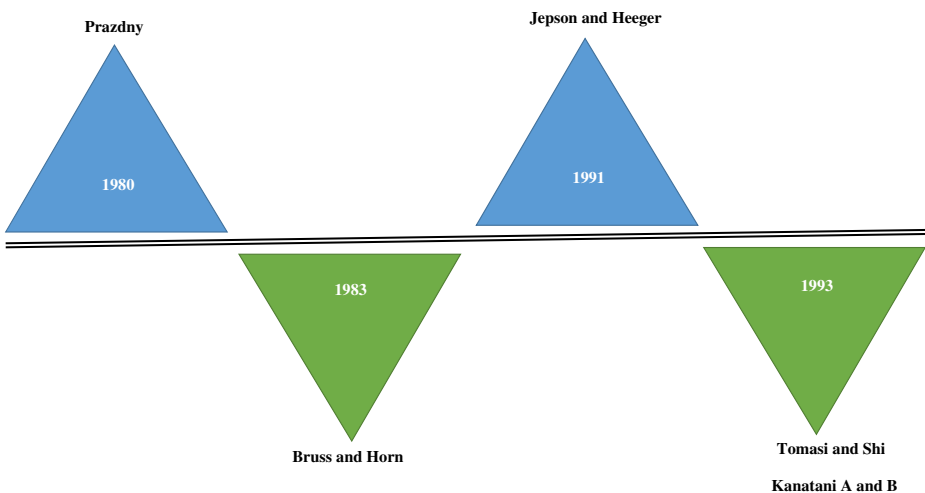


Fig. 4 Timeline for major ego-motion algorithms

algorithms differ from each other on four factors [97]. First, algorithms that calculate rotation-first. Second, algorithms that calculate translation-first. Third, algorithms that require numerical optimization requirement or not. Finally, motion parallax based algorithms versus epipolar constraint.

4.1.1 Prazdny

In computer vision domain, the first computational model for ego-motion was developed by Prazdny [76]. Prazdny believed that optical flow fields are generated at an observer's retina as he moves in a 3D world, and these ego-motion parameters of an observer can be computed. He proposed an algorithm that estimated rotation first, which was independent of translation and depth. Prazdny's implementation consisted triple image points and rotation constraint. Later Tian, et al. [97], combined all constraints across the image. It was also found that each triple of points came from Delaunay triangulation and that different triangulations lead to inconsistent estimates. In order to avoid inconsistent estimates and fix triangulation, Tian, et al. [97] used a fixed uniform sampling grid. The algorithms discussed in section 0, 4.1.3, 4.1.4 and 4.1.5; all begins by estimating the translational factor first (see Table 1).

Prazdny method assumes that surfaces in scene are smooth. Prazdny model is very sensitive to noise and works best in situations where there is low or no noise (falling leaves, trees, blowing snow etc.) in the optical flow field. There is no mechanism used to overcome external and internal noise.

4.1.2 Bruss and horn

Bruss and Horn removed depth and acquired a bilinear constraint to use on each sole image pixel [12]. The same bilinear constraint was later derived by MacLean and Jepson, but by using a different algebraic manipulation [59]. Bruss and Horn algorithm was checked in many different simulations and performed quite well, the main drawback was of numerical optimization [97]. Bruss and Horn also worked on direct methods for motion estimation [38]. In direct methods neither point correspondence, nor optical flow estimation is required. In cases of pure translation and rotation, the first derivative of brightness in an image region is used for motion estimation. Bruss and Horn algorithm works best in a situation where we have a planar environment and there is no depth parameter involved.

4.1.3 Jepson and Heeger

Rieger and Lawton [79] suggested a technique for ego-motion estimation based on motion parallax. Different depths might be associated with two 3D points on a same image location.

Table 1 Ego-motion estimation algorithms facts

Algorithm	Rotation First	Translation First	Constraint	Drawback
Prazdny	✓		Rotation	Worst Noise Sensitivity
Bruss and Horn		✓	Bilinear	Numerical optimization
Jepson and Heeger		✓	Linear-subspace	Sensitive to Noise
Tomasi and Shi		✓	Bilinear	Failure to converge
Kanatani A and B		✓	Epipolar	Low Performance

The difference of the flow vectors is oriented toward the FOE. Hence Rieger and Lawton [79] algorithm is used to locate the focus of expansion from the dissimilarities of flow vector. Later Hildreth [36] revised algorithm of Rieger and Lawton, to increase its performance. In both these algorithms, it was challenging to measure flow vectors near occlusion boundaries [97]. Jepson and Heeger studied these preceding efforts and suggested their own set of solutions. Various subspace approaches were proposed by Jepson and Heeger for ego-motion estimation [35, 45, 46]. The main benefit of using linear subspace method was that there was no requirement for iterative numerical optimization. Jepson and Heeger algorithm is best suited for scenes where there is rich depth structure involved and have less to no noise.

4.1.4 Tomasi and Shi

Tomasi and Shi [98] use the motion parallax information differently. The translation is estimated from image distortions, which is the variation in the angular distance amid a pair of image points. Image deformations are independent of camera rotations. Tomasi and Shi algorithm works best for the scenes which have straight ahead motion and some sideways motions.

4.1.5 Kanatani A and B

The epipolar constraint is the foundation for several linear discrete time algorithms. There is also an instantaneous time form of the epipolar constraint for ego-motion estimation which was proposed by Zhuang, et al. [116]. In future this instantaneous-time version of epipolar constraint was reformulated by Kanatani [49] in terms of parameters and twisted flow. Kanatani's and Zhuang's algorithm are similar and equivalent to one another [97]. This algorithm is referred to as Kanatani A.

The Least-squares estimates of translation vector are scientifically biased [97]. Kanatani introduced a renormalization method that automatically removed the bias and unknown noise. This bias was analyzed by Kanatani using a simple Gaussian noise model and hence its named Kanatani B [49]. Kanatani A and B are both used for scenes where there are non-rigid and small motions involved.

4.1.6 Other methods

The algorithms discussed in sections 4.1.1 to 4.1.5 are the most famous and novel methods used for ego-motion estimation. Several extensions and modifications has also been proposed to these methods. Kanatani's method was further extended to study its effect on brightness of moving objects [61]. As an extension to Kanatani, orthogonal subspace decomposition was proposed by Wu, et al. [109].

Horn and Schunck model is one of the most classical models for optical flow estimation. Several researchers have adapted and modified this rich model [2, 8, 11, 21, 67, 83, 88, 107]. Black and Anandan [8] formulated a robust framework to deal with outliers in both the spatial and temporal terms as an extension to the Horn and Schunck model. In order to generate piecewise smooth results the quadratic regulariser in Horn and Schunck model was replaced by a smoothness constraint [2, 11, 21, 67, 83, 107]. To preserve flow discontinuities L1 penalty was used by Shulman and Herve [88].

Optical Flow methods have also been considered for ego-motion estimation by Raudies and Neumann [77] and Briod, et al. [27]. Combination of Optical Flow method and Random

Sample Consensus (RANSAC) was used for ego-motion estimation and efficient separation of translation and rotation parameters [77]. Recently, Briod, et al. [10] proposed a novel ego-motion estimation algorithm that uses only optical flow directions and not its scale, making the method immune to inertial sensor drifts.

Inspired by recent developments in deep learning networks, Convolutional Neural Networks has been used to learn feature representations from images for frame-to-frame motion estimation [19]. The proposed method is extremely robust with respect to image anomalies and imperfections. There have been other methods proposed such as branch-and-bound methods [27, 28, 52]. These methods are robust in handling outliers, but only deals with pure translational camera motion environments.

4.2 Types of camera

The main equipment required for ego-motion estimation is the camera. Over the course of years, numerous cameras have been developed, some cameras provide small field views while others provide panoramic and wide views. In this section we first overview the traditional camera setup and specify its drawbacks in context of ego-motion estimation. Later we view omnidirectional, stereo and monocular camera setups which have been widely used in ego-motion estimation environments.

4.2.1 Traditional camera

Traditional cameras typically have a cone of 45 degrees and limited field of view [32, 68]. It comprises of a video camera attached to a lens. The projection model is perspective with a single center of projection. These conventional imaging systems are of finite sizes and therefore the incoming rays are occluded by the camera lens. Instead of a hemisphere the lens has a small field of view that corresponds to a cone [68]. Traditional cameras make computation of camera motion sensitive to noise, because the direction of translation may lie outside the field of view [32].

Using multiple traditional cameras to overcome the limited field of view is infeasible because the centers of projections will reside inside their respective lenses [68]. Another method to overcome this limited field of view is by using rotating imaging systems. The problem with rotating imaging systems is that it requires precise positioning, use of moving parts, more time for large field view computation and thus can only be used with static scenes [68].

4.2.2 Omnidirectional and panoramic camera

Omnidirectional cameras provide a 360° panoramic view of the scene [29, 63, 81]. When compared to traditional cameras it has an enhanced field of view [17, 63]. It is ideal for 3D vision tasks such as obstacle detection and motion estimation [95]. Wide field of view can be attained by using a combination of a camera and a mirror, a camera with wide-angle lens or several synchronized panoramic cameras [63]. Omnidirectional camera's panoramic view helps in dealing with ambiguities that are associated with ego-motion estimation [29].

Omnidirectional cameras show a great potential in intelligent vehicle applications, autonomous navigation applications, remote surveillance, video conferencing and scene recovery [29, 68]. Another advantage of 360° horizontal panoramic view is that one can track feature

point for longer distances with less constraint [17]. Omnidirectional cameras are very advantageous for robot vision system applications for e.g. when surrounding scene is stored as a single image frame [17, 29, 63]. 360° view ensures that no object will escape the camera's view [63].

Various algorithms have been proposed by the vision researchers to solve the ego-motion problem [32]. Typically two methods are used for solving ego-motion. The first method includes the optical flow computation. The second method is based on motion field analysis and extraction of camera translation and rotation from the optical flow. The main problem with the computation of ego-motion is the sensitivity of optic flow to noise, omnidirectional camera systems seek to overcome this problem [32]. There are also several drawbacks associated with omnidirectional cameras. One drawback is lower image resolution for a larger field of view [17]. Many researchers have preferred and used omnidirectional camera for ego-motion estimation [29, 32, 34, 55, 56, 63, 81, 103].

4.2.3 Stereo and monocular camera

Many methods have been proposed for ego-motion computation using stereo [4, 20, 22, 23, 26, 42, 55, 62, 65, 72–74, 82, 84, 90] and monocular [48, 66, 80, 90, 110, 111, 115] sequences. The main difference among these methods is feature tracking and the transformation applied for ego-motion estimation [65].

Monocular camera setup has low runtime cost and well-developed machine learning paradigms [90]. Due to translation scale ambiguity, monocular ego-motion estimation is only reliable for measuring translational velocity [20]. Monocular cameras do not provide a 3D location for detecting vehicles directly [90].

Whenever possible the stereo camera setup is preferred over a monocular camera setup [20]. The main advantage of using a stereo camera setup is that it provides explicit computation of location in real world coordinates and depth [90]. One of the most outstanding work on stereo visual odometry, Mars Exploration Rovers, was done by Cheng, et al. [18] and Maimone, et al. [60]. Mars Exploration Rovers (MER) project by NASA successfully demonstrated the usefulness of visual odometry on another planet. The MER system used its 45 degree mounted navigation cameras to obtain stereo pairs, which were compared using an on-board software [60]. The MER system was also able to determine 6 Degrees-of-Freedom rover pose (roll, pitch, yaw, x, y, z). Stereo camera setup's main drawback is that it requires additional specialized hardware, precise calibration, and additional computational cost [90].

5 Ego-motion estimation challenges

Difficulties and ambiguities in ego-motion estimation arise due to unwanted camera motion, occlusion, noise, lack of image texture, illumination changes and the aperture problem [20, 85]. In this section we discuss these challenges and specify the avoidance and recovery methods for it.

5.1 Occlusion

Occlusions arise in video streams when a portion of scene is visible in one image but not another [3]. Depth discontinuities may also introduce occlusion in video streams [3]. Image formation is often limited due to occlusion [20].

Determining various occlusion regions is challenging [39]. Occlusion can be easily detected if the motion field is known [3]. The computation of motion field might also be disrupted when a scene becomes occluded and dis-occluded over a course of a video [39].

Self-occlusions are produced when the shape of the scene changes and results in significant deformations [101]. In forward motion scenarios shape of scene is affected, scale changes in image domain and produces self-occlusions [101]. Tsotsos, et al. [101] described an ego-motion estimation system specifically for humanoid robots with specific emphasis on the challenges of scale changes due to forward motion.

Determination of occluded region is important for one-to-one correspondence of image scene points [39]. These image scene points are required for video segmentation, tracking, segmentation and reconstruction algorithms. In order to manage occlusion regions some applications treat them as outliers [39]. However, outliers may also pose as problems, triggered by two reasons, namely, incorrect matches and the correct matches [5]. Incorrect matches are produced by temporal or erroneous spatial formation of point-to-point correspondence. On the other hand, correct matches are caused by misrepresentation model and assumption that a point is static [5].

Occlusion boundaries are very beneficial for motion direction, ordered depth and scene context [39]. Recently Yamaguchi, et al. [112] proposed a method for recovering occlusion boundaries from a static scene given two frames of a stereo pair captured from a moving vehicle.

5.2 Noise

Vision applications are inherently noisy, but are also rich sources of information [58]. One of the primary challenges associated with the ego-motion computation is noisy estimates with optic flow [32]. Whenever signals of translation lie outside the field of view, signals are seriously corrupted, making the camera motion computation highly sensitive to noise [32, 100]. It's also challenging to compute the translation parameters in the presence of noise [100]. Noise likewise affects stereo and feature tracking parameters [5].

Visual odometry is typically computed incrementally from frame to frame and with each step small errors are introduced due to noise [82]. Many of the visual odometry applications such as robot navigation, autonomous vehicle driving and dealing with high-speed traffic scenes are extremely challenging [58]. In order for these real-time applications to work accurately, robustness to noise proportions needs to be attained in constant time [58].

Ego-motion estimation requires some previous analysis of the noise involved in the imaging process [5]. One of the most widely used models in simulation studies is the Gaussian noise model [78]. It's useful to know how the noise propagates as the data is processed [5].

5.3 Camera calibration

Accurate camera calibration is necessary for any computer vision task, such as ego-motion estimation that requires extraction of metric information from 2D images [81]. The camera calibration problem comprises both the interior and exterior orientation complications [43]. In order to join the image plane coordinates to absolute coordinates, orientation, camera position and camera constant must be determined [43].

In ego-motion estimation each pixel is imaged using perspective projection [43]. The camera calibration problem is to relate the position of pixels in the image array to the points in the scene [43]. Lens distortions, aspect ratio and the location of the principal point needs to be found, to relate pixels in image array to a point in scene [43].

Intrinsic camera calibration is the calibration of an individual camera and involves parameters such as principal point, focal length and distortion model [24]. Whereas extrinsic camera calibration is the relative offset between two sensors [24].

Numerous methods have been developed for planar camera calibration as compared to omnidirectional cameras [81]. The overall methods for omnidirectional camera calibration can be divided into two groups. The first group exploits previous knowledge about the scene, for example presence of plumb lines or calibration patterns. The second group includes techniques that do not exploit this prior knowledge, such as using point correspondence or epipolar constraint for pure rotation, self-calibration procedures, and planar motion of camera [81].

6 Open problems

In the literature that we have reviewed, we have come across the following open problems for ego-motion estimation.

- (1) Ego-motion, Object and Background Motion: How is the user supposed to determine both the camera and object(s) motion captured in extreme constraints and conditions?
- (2) Scalability: How does each algorithm behave for extreme camera movements? How is the accuracy and algorithm time affected?
- (3) Evaluation: How to decide which algorithm and camera setup works best when used together for ego-motion estimation?
- (4) Application: How to know that a vision system is good enough to be used for a non-trivial robotic application?
- (5) Proportion: How many motion parameters need to be considered to detect a moving object reliably?

Independent movement of one or more objects in extreme conditions gives rise to several issues in ego-motion estimation. Even some of the state of the art algorithms fails to deliver quality results. To tackle this problem efficiently, we may need to use deep complex structured learning algorithms for feature analysis in-combination with optical flow. The scalability of algorithms with extreme camera movements can be tested using computer simulations or models. The problem of evaluation can be eased by creating a benchmark set that allow comparison between different algorithms and camera setups. Applications using ego-motion will need to perform extensive tests on real world images, most important consideration is installing and calibrating the entire experimental equipment that is capable of moving the camera in a controlled fashion. Finally the last open problem discusses the number of motion parameters. Estimation results are more reliable in cases where the number of motion parameters is reduced, but this raises another problem of detecting moving objects effectively from these reduced number of motion parameters.

7 Conclusions and future directions

We have reviewed the literature for ego-motion estimation in order to provide solutions to different computer vision issues. The ego-motion literature reviewed can be grouped based on the algorithms, stereo camera setup, monocular camera setup, omnidirectional camera, autonomous vehicles, visual odometry and visual SLAM as shown in Table 2.

Table 2 Motion estimation classified according to different domains

Domain	Related Work
Algorithms	Prazdny [76], Bruss and Horn [12], Zhuang, et al. [116], Rieger and Lawton [79], MacLean, et al. [59], Hildreth [36], Kanatani [49], Tomasi and Shi [98], Tian, et al. [97]
Omnidirectional Camera	Koyasu, et al. [55], Gluckman and Nayar [32], Gandhi and Trivedi [29], Scaramuzza, et al. [81], Markovic, et al. [63], Vassallo, et al. [103], Lauer [56], Hariyono, et al. [34]
Monocular Camera Setup	Sivaraman and Trivedi [90], Yang, Sun [115], Scaramuzza and Siegwart [80], Munguia and Grau [66], Jung, et al. [48], Yamaguchi, et al. [110], Yang, et al. [115], Yamaguchi, et al. [111]
Stereo Camera Setup	da Silva [20], Schmid and Hirschmuller [82], Sivaraman and Trivedi [90], Domaika and Sappa [23], Badino [4], Seki and Okutomi [84], Milella and Siegwart [65], Franke, et al. [26], Olson, et al. [74], Domaika and Chung [22], Koyasu, et al. [55], Olson, et al. [73], Olson, et al. [72], Mandelbaum, et al. [62], Jain, et al. [42]
Autonomous Vehicles	Sivaraman and Trivedi [90], Domaika and Sappa [23], Yang, et al. [115], Scaramuzza and Siegwart [80], Yamaguchi, et al. [110], Yamaguchi, et al. [111], Gandhi and Trivedi [29], Stein, et al. [93], Stephens, et al. [94], Gillner [31], Shakernia, et al. [86], Shakernia, et al. [87], van der Mark, et al. [102], Yang, et al. [114], Stein and Mano [92], Iyer, et al. [41], Cao, et al. [16], Klappstein, et al. [53]
Visual Odometry	Cheng, et al. [18], Maimone, et al. [60], Scaramuzza and Siegwart [80], Munguia and Grau [66], Konolige, et al. [54], Levin and Szeliski [57], Wang, et al. [105]
Visual SLAM	Munguia and Grau [66], Karlsson, et al. [50]

We have also pointed out the main challenges and issues that need to be taken into account when estimating camera motion. Noise, occlusion and calibration problems need to be addressed for proper and effective ego-motion estimation.

In all the literature that we have reviewed we have come to know that this research area is widely open to new research and development. Here we offer a few possible suggestions for future research.

- In future research, we suggest that one should try to implement 6-DOF ego-motion parameters for autonomous vehicles. This approach can be fruitful for ego-motion estimation of large-scale rough terrain.
- For vehicle collision detection and safety systems a fusion of classifiers and ego-motion based approaches should be used in fusion to promote the detection rate. To avoid collision, object learning can also be carried out with different viewing angles.
- Optical flow holds great significance in motion estimation. Researchers should investigate the problem of dynamic backgrounds in the future. Dynamic backgrounds include those objects that tend to have fixed locations, but adds a great deal of noise, for example windblown trees.
- Overall the ego-motion process can also be hastened by using hardware resources such as field-programmable gate array (FPGA). By using the hardware we mean that some part of the ego-motion algorithm should be transferred on to the hardware.
- The use of a car speedometer can be investigated in future for estimating magnitude of forward motion in case of long stretches of new highway. In such a highway case there is enough vertical texture for computing rotation estimates, but no horizontal texture for computing the forward magnitude.

Acknowledgments We would like to thank, Dr. Saleem Gul, Institute of Management Sciences, Peshawar, Pakistan, for providing help in structuring the research article. We also appreciate the efforts of, Dr. Muhammad Haseeb, Department of Computer Science, University of Peshawar, Pakistan.

References

1. Agrawal P, Carreira J, Malik J (2015) Learning to see by moving, In Proceedings of the IEEE International Conference on Computer Vision, pp 37–45
2. Álvarez León LM, Esclarín Monreal J, Lefébure M, Sánchez Pérez J (1999) A PDE model for computing the optical flow, Proceedings of CEDYA XVI. - Las Palmas: University of Las Palmas, pp 1349–1356
3. Ayvaci A, Raptis M, Soatto S (2012) Sparse occlusion detection with optical flow. *Int J Comput Vis* 97:322–338
4. Badino, H (2007) A robust approach for ego-motion estimation using a mobile stereo platform, In *Complex Motion*, ed: Springer, pp 198–208
5. Badino H (2009) Binocular ego-motion estimation for automotive applications, Goethe University Frankfurt am Main
6. Baik YK, Kwon J, Lee HS, Lee KM (2013) Geometric particle swarm optimization for robust visual ego-motion estimation via particle filtering. *Image Vis Comput* 31:565–579
7. Barron JL, Fleet DJ, Beauchemin SS (1994) Performance of optical flow techniques. *Int J Comput Vis* 12: 43–77
8. Black MJ, Anandan P (1996) The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Comput Vis Image Underst* 63:75–104
9. Brandt T, Büchele W, Arnold F (1977) Arthrokinetic nystagmus and ego-motion sensation. *Exp Brain Res* 30:331–338
10. Briod A, Zufferey J-C, Floreano D (2016) A method for ego-motion estimation in micro-hovering platforms flying in very cluttered environments. *Auton Robot* 40:789–803
11. Brox T, Bruhn A, Papenbergh N, Weickert, J (2004) High accuracy optical flow estimation based on a theory for warping, In *Computer Vision-ECCV 2004*, ed: Springer, pp 25–36
12. Bruss AR, Horn BK (1983) Passive navigation. *Computer Vision, Graphics, and Image Processing* 21:3–20
13. Burger W, Bhanu B (1989) On computing a fuzzy focus of expansion for autonomous navigation, In *Computer Vision and Pattern Recognition, 1989 Proceedings CVPR '89, IEEE Computer Society Conference on*, pp 563–568
14. Burger W, Bhanu B (1990) Estimating 3D egomotion from perspective image sequence. *IEEE Trans Pattern Anal Mach Intell* 12:1040–1058
15. Campbell J, Sukthankar R, Nourbakhsh I, Pahwa A (2005) A robust visual odometry and precipece detection system using consumer-grade monocular vision, In *Robotics and Automation, 2005 ICRA 2005 Proceedings of the 2005 I.E. International Conference on*, pp 3421–3427
16. Cao Y, Cook P, Renfrew, A (2007) Vehicle ego-motion estimation by using pulse-coupled neural network, In *Machine Vision and Image Processing Conference, 2007. IMVIP 2007. International*, pp 185–191
17. Chang P, Hebert M (2000) Omni-directional structure from motion, In *Omnidirectional Vision, 2000 Proceedings IEEE Workshop on*, pp 127–133
18. Cheng Y, Maimone M, Matthies L (2005) Visual odometry on the Mars exploration rovers, In *Systems, Man and Cybernetics, 2005 I.E. International Conference on*, pp 903–910
19. Costante G, Mancini M, Valigi P, Ciarfuglia TA (2016) Exploring representation learning with cnns for frame-to-frame ego-motion estimation. *IEEE Robotics and Automation Letters*, 1:18–25
20. Da Silva HMG (2014) A probabilistic approach for stereo visual egomotion, Ph.D. dissertation, Instituto Superior Technico, Universidade De Lisboa
21. Deriche R, Kornprobst P, Aubert G (1996) Optical-flow estimation while preserving its discontinuities: A variational approach, In *Recent Developments in Computer Vision*, ed: Springer, pp 69–80
22. Dornaika F, Chung C-KR (2003) Stereo geometry from 3D ego-motion streams. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 33:308–323
23. Dornaika F, Sappa AD (2007) Real-time vehicle ego-motion using stereo pairs and particle filters, In *Image Analysis and Recognition*, ed: Springer, pp 469–480
24. Endres F, Sprunk C, Kummerle R, Burgard W (2014) A catadioptric extension for RGB-D cameras, In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pp 466–471
25. Ess A, Leibe B, Schindler K, Gool LV (2008) A mobile vision system for robust multi-person tracking, In *Computer Vision and Pattern Recognition, 2008 CVPR 2008 I.E. Conference on*, pp 1–8
26. Franke U, Rabe C, Badino H, Gehrig S (2005) 6d-vision: Fusion of stereo and motion for robust environment perception, In *Pattern Recognition*, ed: Springer, pp 216–223

27. Fredriksson J, Enqvist O, Kahl F (2014) Fast and reliable two-view translation estimation, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1606–1612
28. Fredriksson J, Larsson V, Olsson C (2015) Practical robust two-view translation estimation, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 2684–2690
29. Gandhi T, Trivedi M (2005) Parametric ego-motion estimation for vehicle surround analysis using an omnidirectional camera. *Mach Vis Appl* 16:85–95
30. Gibson JJ (1970) On theories for visual space perception. *Scand J Psychol* 11:75–79
31. Gillner WJ (1995) Motion based vehicle detection on motorways, In Intelligent Vehicles' 95 Symposium, Proceedings of the, pp 483–487
32. Gluckman J, Nayar, SK (1998) Ego-motion and omnidirectional cameras, In Computer Vision, 1998. Sixth International Conference on, pp 999–1005
33. Goecke R, Asthana A, Pettersson N, Petersson L (2007) Visual vehicle egomotion estimation using the fourier-mellin transform, in Intelligent Vehicles Symposium, 2007 IEEE, pp 450–455
34. Hariyono J, Hoang V-D, Jo, K-H (2014) Human detection from mobile omnidirectional camera using ego-motion compensated, In Intelligent Information and Database Systems, ed: Springer, pp 553–560
35. Heeger DJ, Jepson AD (1992) Subspace methods for recovering rigid motion I: algorithm and implementation. *Int J Comput Vis* 7:95–117
36. Hildreth EC (1992) Recovering heading for visually-guided navigation. *Vis Res* 32:1177–1192
37. Horn BK, Schunck BG (1981) Determining optical flow, In 1981 Technical Symposium East, pp 319–331
38. Horn BK, Weldon Jr E (1988) Direct methods for recovering motion. *Int J Comput Vis* 2:51–76
39. Humayun A, Mac Aodha O, Brostow GJ (2011) Learning to find occlusion regions, In Computer Vision and Pattern Recognition (CVPR), 2011 I.E. Conference on, pp 2161–2168
40. Irani M, Rousso B, Peleg S (1994) Recovery of ego-motion using image stabilization, In Computer Vision and Pattern Recognition, 1994 Proceedings CVPR'94, 1994 I.E. Computer Society Conference on, pp 454–460
41. Iyer RV, He Z, Chandler PR (2006) On the computation of the ego-motion and distance to obstacles for a micro air vehicle, In American Control Conference, 2006, p 6 pp
42. Jain R, Bartlett SL, O'Brien, N (1987) Motion stereo using ego-motion complex logarithmic mapping, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, pp 356–369
43. Jain R, Kasturi R, Schunck BG (1995) *Machine vision*. McGraw-Hill, New York
44. Jayaraman D, Grauman K (2015) Learning image representations tied to ego-motion, In Proceedings of the IEEE International Conference on Computer Vision, pp 1413–1421
45. Jepson AD, Heeger DJ (1991) A fast subspace algorithm for recovering rigid motion, In Visual Motion, 1991, Proceedings of the IEEE Workshop on, pp 124–131
46. Jepson AD, Heeger DJ (1992) Linear subspace methods for recovering translational direction, *Spatial Vision in Humans and Robots*, pp 39–62
47. Jung B, Sukhatme GS (2004) Detecting moving objects using a single camera on a mobile robot in an outdoor environment, In International Conference on Intelligent Autonomous Systems, pp 980–987
48. Jung S-H, Eledath J, Johansson S, Mathevon V (2007) Egomotion estimation in monocular infra-red image sequence for night vision applications, In Applications of Computer Vision, 2007 WACV'07 I.E. Workshop on, pp 8–8
49. Kanatani K (1993) 3-D interpretation of optical flow by renormalization. *Int J Comput Vis* 11:267–282
50. Karlsson N, Di Bernardo E, Ostrowski J, Goncalves L, Pirjanian P, Munich, ME (2005) The vSLAM algorithm for robust localization and mapping, In Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 I.E. International Conference on, pp 24–29
51. Kellner D, Barjenbruch M, J. Klappstein, J. Dickmann, and K. Dietmayer, (2014) Instantaneous ego-motion estimation using multiple Doppler radars, In Robotics and Automation (ICRA), 2014 I.E. International Conference on, pp 1592–1597
52. Kim J-H, Li H, Hartley R (2010) Motion estimation for nonoverlapping multicamera rigs: linear algebraic and ∞ geometric solutions. *IEEE Trans Pattern Anal Mach Intell* 32:1044–1059
53. Klappstein J, Stein F, Franke, U (2006) Monocular motion detection using spatial constraints in a unified manner, In Intelligent Vehicles Symposium, 2006 IEEE, pp 261–267
54. Konolige K, Agrawal M, Sola J (2011) Large-scale visual odometry for rough terrain, In Robotics Research, ed: Springer, pp 201–212
55. Koyasu H, Miura J, Shirai Y (2002) Recognizing moving obstacles for robot navigation using real-time omnidirectional stereo vision. *Image* 2:1
56. Lauer, M (2007) Ego-motion estimation and collision detection for omnidirectional robots, In RoboCup 2006: Robot soccer world cup X, ed: Springer, pp 466–473
57. Levin A, Szeliski R (2004) Visual odometry and map correlation. *Computer Vision and Pattern Recognition, 2004 CVPR 2004 Proceedings of the 2004 I.E. Computer Society Conference on* 1:1:611–I-618
58. Lim JJK (2010) Egomotion estimation with large field-of-view vision, PhD thesis

59. MacLean WJ, Jepson AD, Frecker RC (1994) Recovery of egomotion and segmentation of independent object motion using the EM algorithm, *BMVC*, pp 1–10
60. Maimone M, Cheng Y, Matthies L (2007) Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics* 24:169–186
61. Maki A, Wiles C (2000) Geotensity constraint for 3D surface reconstruction under multiple light sources, In *European Conference on Computer Vision*, pp 725–741.
62. Mandelbaum R, Salgian G, Sawhney H (1999) Correlation-based estimation of ego-motion and structure from motion and stereo, In *Computer Vision, 1999 The Proceedings of the Seventh IEEE International Conference on*, pp 544–550
63. Markovic I, Chaumette F, Petrovic I (2014) Moving object detection, tracking and following using an omnidirectional camera on a mobile robot, In *IEEE Int. Conf. on Robotics and Automation, ICRA'14*
64. McQuirk IS, Lee H-S, Horn B (1997) An analog VLSI chip for estimating the focus of expansion, In *Solid-State Circuits Conference, 1997 Digest of Technical Papers. 43rd ISSCC, 1997 I.E. International*, pp 40–41
65. Milella, A, Siegwart, R (2006) Stereo-based ego-motion estimation using pixel tracking and iterative closest point, In *Computer Vision Systems, 2006 ICVS'06. IEEE International Conference on*, pp 21–21
66. Munguia R, Grau A (2007) Monocular SLAM for visual odometry, In *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*, pp 1–6
67. Nagel H-H, Enkelmann W (1986) An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pp 565–593
68. Nayar SK (1997) Catadioptric omnidirectional camera, In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 I.E. Computer Society Conference on*, pp 482–488
69. Negahdaripour S, Horn BK (1989) A direct method for locating the focus of expansion. *Computer Vision, Graphics, and Image Processing* 46:303–326
70. Neisser U (1977) Gibson's ecological optics: consequences of a different stimulus description. *J Theory Soc Behav* 7:17–28
71. Nistér D, Naroditsky O, Bergen J (2004) Visual odometry, *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 I.E. Computer Society Conference on* 1:1-652–I-659
72. Olson CF, Matthies LH, Schoppers M, Maimone MW (2000) Robust stereo ego-motion for long distance navigation, In *Computer Vision and Pattern Recognition, 2000 Proceedings IEEE Conference on*, pp 453–458
73. Olson CF, Matthies LH, Schoppers M, Maimone MW (2001) Stereo ego-motion improvements for robust rover navigation, In *Robotics and Automation, 2001 Proceedings 2001 ICRA IEEE International Conference on*, pp 1099–1104
74. Olson CF, Matthies LH, Schoppers M, Maimone MW (2003) Rover navigation using stereo ego-motion. *Robot Auton Syst* 43:215–229
75. Prazdny K (1979) Motion and structure from optical flow. *Proceedings of the 6th International Joint Conference on Artificial Intelligence* 2:702–704
76. Prazdny K (1980) Egomotion and relative depth map from optical flow. *Biol Cybern* 36:87–102
77. Raudies F, Neumann H (2009) An efficient linear method for the estimation of ego-motion from optical flow, In *Joint Pattern Recognition Symposium*, pp 11–20
78. Raudies F, Neumann H (2012) A review and evaluation of methods estimating ego-motion. *Comput Vis Image Underst* 116:606–633
79. Rieger J, Lawton D (1985) Processing differential image motion. *JOSA A* 2:354–359
80. Scaramuzza D, Siegwart R (2008) Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE Trans Robot* 24:1015–1026
81. Scaramuzza D, Martinelli A, Siegwart R (2006) A flexible technique for accurate omnidirectional camera calibration and structure from motion, In *Computer Vision Systems, 2006 ICVS'06 I.E. International Conference on*, pp 45–45
82. Schmid K, Hirschmuller H (2013) Stereo vision and IMU based real-time ego-motion and depth image computation on a handheld device, In *Robotics and Automation (ICRA), 2013 I.E. International Conference on*, pp 4671–4678
83. Schnorr C (1994) Segmentation of visual motion by minimizing convex non-quadratic functionals, In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, pp 661–663
84. Seki A, Okutomi M (2006) Ego-motion estimation by matching dewarped road regions using stereo images. In *Robotics and Automation, 2006 ICRA 2006 Proceedings 2006 I.E. International Conference on*, pp 901–907
85. Shafait F, Grimm M, Grigat R-R (2004) Low-complexity camera ego-motion estimation algorithm for real time applications, In *Multitopic Conference, 2004 Proceedings of INMIC 2004. 8th International*, pp 131–136
86. Shakernia O, Ma Y, Koo TJ, Hespanha J, Sastry, SS (1999a) Vision guided landing of an unmanned air vehicle, In *Decision and Control, 1999 Proceedings of the 38th IEEE Conference on*, pp. 4143–4148.

87. Shakemia O, Ma Y, Koo TJ, Sastry S (1999b) Landing an unmanned air vehicle: vision based motion estimation and nonlinear control. *Asian Journal of Control* 1:128–145
88. Shulman D, Herve J-Y (1989) Regularization of discontinuous flow fields, in *Visual Motion*, 1989, Proceedings Workshop on, pp 81–86
89. Singh KK, Fatahalian K, Efros AA (2016) KrishnaCam: using a longitudinal, single-person, egocentric dataset for scene understanding tasks. *Chance* 43:48.5
90. Sivaraman S, Trivedi MM (2011) Combining monocular and stereo-vision for real-time vehicle ranging and tracking on multilane highways, In *Intelligent Transportation Systems (ITSC)*, 2011 14th International IEEE Conference on, pp 1249–1254
91. Srinivasan N, Roberts R, Dellaert, F (2013) High frame rate egomotion estimation, In *Computer Vision Systems*, ed: Springer, pp 183–192
92. Stein GP, Mano O (2004) System and method for estimating ego-motion of a moving vehicle using successive images recorded along the vehicle's path of motion, ed: Google Patents
93. Stein GP, Mano O, Shashua A (2000) A robust method for computing vehicle ego-motion, In *Intelligent Vehicles Symposium*, 2000 IV 2000 Proceedings of the IEEE, pp 362–368
94. Stephens M, Blissett R, Chamley D, Sparks E, Pike J (1989) Outdoor vehicle navigation using passive 3D vision, In *Computer Vision and Pattern Recognition*, 1989. Proceedings CVPR'89., IEEE Computer Society Conference on, pp 556–562
95. Strelow D, Mishler J, Koes D, Singh S (2001) Precise omnidirectional camera calibration. *Computer Vision and Pattern Recognition*, 2001 CVPR 2001 Proceedings of the 2001 I.E. Computer Society Conference on 1:1-689–I-694
96. Sun D, Roth S, Black MJ (2014) A quantitative analysis of current practices in optical flow estimation and the principles behind them. *Int J Comput Vis* 106:115–137
97. Tian TY, Tomasi C, Heeger, DJ (1996) Comparison of approaches to egomotion computation, In *Computer Vision and Pattern Recognition*, 1996 Proceedings CVPR'96, 1996 I.E. Computer Society Conference on, pp 315–320
98. Tomasi C, Shi J (1993) Direction of heading from image deformations, In *Computer Vision and Pattern Recognition*, 1993 Proceedings CVPR'93, 1993 I.E. Computer Society Conference on, pp 422–427
99. Trucco E, Verri A (1998) Introductory techniques for 3-D computer vision, vol 201. Prentice Hall, Englewood Cliffs
100. Tsao A-T, Hung Y-P, Fuh C-S, Chen Y-S (1997) Ego-motion estimation using optical flow fields observed from multiple cameras, In *Computer Vision and Pattern Recognition*, 1997. Proceedings., 1997 I.E. Computer Society Conference on, pp 457–462
101. Tsotsos K, Pretto A, Soatto, S (2012) Visual-inertial ego-motion estimation for humanoid platforms, In *Humanoid Robots (Humanoids)*, 2012 12th IEEE-RAS International Conference on, pp 704–711
102. van der Mark W, Fontijne D, Dorst L, Groen, FC (2002) Vehicle ego-motion estimation with geometric algebra, In *Intelligent Vehicle Symposium*, 2002 IEEE, pp 58–63
103. Vassallo RF, Santos-Victor J, Schneebeil HJ (2002) A general approach for egomotion estimation with omnidirectional images, In *Omnidirectional Vision*, 2002 Proceedings Third Workshop on, pp 97–103
104. Verri A, Poggio T (1989) Motion field and optical flow: qualitative properties. *IEEE Trans Pattern Anal Mach Intell* 11:490–498
105. Wang H, Yuan K, Zou W, Zhou Q (2005) Visual odometry based on locally planar ground assumption, In *Information Acquisition*, 2005 I.E. International Conference on, p 6 pp
106. Warren R (1976) The perception of egomotion. *J Exp Psychol Hum Percept Perform* 2:448
107. Weickert J, Schnörr C (2001) A theoretical framework for convex regularizers in PDE-based computation of image motion. *Int J Comput Vis* 45:245–264
108. Weishaupt A (2010) Tracking and Structure from Motion, Master's Thesis, Signal Processing Laboratory – Section of Electrical Engineering School of Engineering Swiss Federal Institute of Technology. Lausanne, Switzerland
109. Wu Y, Zhang Z, Huang TS, Lin JY (2001) Multibody grouping via orthogonal subspace decomposition. *Computer Vision and Pattern Recognition*, 2001 CVPR 2001 Proceedings of the 2001 I.E. Computer Society Conference on 2:II-252–II-257
110. Yamaguchi K, Kato T, Ninomiya Y (2006a) Moving obstacle detection using monocular vision, In *Intelligent Vehicles Symposium*, 2006 IEEE, pp 288–293
111. Yamaguchi K, Kato T, Ninomiya Y (2006b) Vehicle ego-motion estimation and moving object detection using a monocular camera, In *Pattern Recognition*, 2006 ICPR 2006 18th International Conference on, pp 610–613
112. Yamaguchi K, McAllester D, Urtasun R (2014) Efficient joint segmentation, occlusion labeling, stereo and flow estimation, In *Computer Vision–ECCV 2014*, ed: Springer, pp 756–771

113. Yamamoto Y, Pirjanian P, Munich M, DiBernardo E, Goncalves L, Ostrowski J et al. (2005) Optical sensing for robot perception and localization, In Advanced Robotics and its Social Impacts, 2005 I.E. Workshop on, pp 14–17
114. Yang M, Dong B, Wang H, Zhang, B (2002) Laser radar based real-time ego-motion estimation for intelligent vehicles, In Intelligent Vehicle Symposium, 2002. IEEE, pp 44–51
115. Yang D, Sun F, Wang S, Zhang J (2014) Simultaneous estimation of ego-motion and vehicle distance by using a monocular camera,” *Science China. Inf Sci* 57:1–10
116. Zhuang X, Huang TS, Ahuja N, Haralick RM (1988) A simplified linear optic flow-motion algorithm. *Computer Vision, Graphics, and Image Processing* 42:334–344



Naila Habib Khan received her BCS degree from Institute of Management Sciences, Peshawar, Pakistan in 2011 and MS (Information Technology) degree in 2014. She is a double Gold Medalist and has been awarded numerous merit scholarships during her academic career. Currently, she is a PhD student and works as a Research Assistant under the supervision of Dr. Awais Adnan at Institute of Management Sciences. Her areas of interest are Document Image Understanding, Pattern Recognition and Multimedia.



Awais Adnan is Coordinator of IT/CS for Morning Computer Science programs, Assistant Professor and Director ORIC at Institute of Management Sciences Peshawar. He teaches different programs at graduate and post-graduate level. Supervision of students at MS-IT, MS-CS and BSC level are also part of his duties. Major areas of his research is multimedia, digital image processing and Network on chip (NOC). He has also been working as a trainer in HRDC where he gives training on computer packages and data analysis tools to different professionals from different government other public sector organization.