Multibody VSLAM with Relative Scale Solution for Curvilinear Motion Reconstruction

Rahul Kumar Namdev, K Madhava Krishna and C. V. Jawahar

***Abstract*— A solution to the relative scale problem where reconstructed moving objects and the stationary world are represented in a unified common scale has proven equivalent to a conjecture. Motion reconstruction from a moving monocular camera is considered ill posed due to known problems of observability. We show for the first time several significant motion reconstruction of outdoor vehicles moving along non-holonomic curves and straight lines. The reconstructed motion is represented in the unified frame which also depicts the estimated camera trajectory and the reconstructed stationary world. This is possible due to our Multibody VSLAM framework with a novel solution for relative scale proposed in the current paper. Two solutions that compute the relative scale are proposed. The solutions provide for a unified representation within four views of reconstruction of the moving object and are thus immediate. In one, the solution for the scale is that which satisfies the planarity constraint of the object motion. The assumption of planar object motion while being generic enough is subject to stringent degenerate situations that are more widespread. To circumvent such degeneracies we assume that the object motion to be locally circular or linear and find the relative scale solution for such object motions. Precise reconstruction is achieved in synthetic data. The fidelity of reconstruction is further vindicated with reconstructions of moving cars and vehicles in uncontrolled outdoor scenes.**

I. INTRODUCTION

With the advent of outdoor robotics [1] in a prominent way the need for solutions that are able to provide for a geometric understanding of the scene in terms of three dimensional reconstructions of the stationary world and moving objects cannot be overemphasized. The multibody Structure from Motion (SFM) framework where both stationary world and moving objects are reconstructed comes across as an appropriate framework for providing such an understanding. However one of the pertinent problems in multibody SFM is the problem of relative scale while representing both the moving object and the stationary world in an unified frame of reference. The problem of relative scale is difficult to solve because of the lack of correspondences between the moving object and the stationary world. In other words there is no easy way to associate a point on the reconstructed moving object with a point in the stationary world. The need for an accurate relative scale estimate is indeed critical. A unified representation of the stationary and dynamic objects at wrong relative scales results in meaningless portrayals such as a vehicle sinking beneath the ground plane or floating in space.

Rahul Kumar Namdev, K Madhava Krishna and C. V. Jawahar all are with IIIT Hyderabad, India.
rahul.namdev@research.iiit.ac.in,
{mkrishna,jawahar}@iiit.ac.in

The previous prime solution to the relative scale problem [2] is non-incremental and uses many camera views thereby not applicable in a robotic setting. It imposes a planarity constraint to solve relative scale problem while assuming a non-accidentalness criteria. The non-accidental criteria minimally involves a search through various scales. The verification of such a criteria could be quite involved and is affected by degeneracies. [2] also proposes a solution by assuming independence between camera and object motion, but as mentioned by the authors themselves, this independence criteria does not hold in typical outdoor road scenarios.

In this paper we present two approaches that determines the relative scale within four views of reconstruction of the moving object. Called the four view solution this provides for immediate availability of the unified representation for further robot action such as collision avoidance. The first method assumes planar object motion, henceforth called as planar method. It does not approximate continuous curvature trajectories as circles or straight lines during reconstruction. However, the degeneracies that arise by assuming planar motion are stringent. Degenerate situations are those for which the solution becomes independent of scale or infinite values of scale satisfy the planar trajectory assumption. In other words degeneracy arises if for every possible scale the reconstructed trajectory is planar. In this case it becomes impossible to find a unique scale solution that satisfies the planarity assumption. Degeneracy occurs in the planar method if the object and camera motion are coplanar or if object and camera moves in parallel planes. These situations typically arise both outdoors and indoors such as when the camera and object move parallel to the floor or the road. However in the presence of an active camera that can be controlled not to move in a plane parallel to the object degeneracies can be avoided. One common example is a hand held camera that is controlled by the human to prevent a degenerate situation. Unlike [2] the proposed solution is incremental and involves only four camera views.

Typically, most of the outdoor non-holonomic trajectories can be modelled through a combination of circular arcs and straight line segments [3]. If we solve relative scale problem for such a combination of locally circular arcs or locally linear segments, the aforementioned difficulties and degeneracies of planar case can be surmounted. In this regard, we employ a four view solution for circular case and a three view solution for linear case. We term this four view method which provide solution of relative scale problem for circular case as circular method of relative scale estimation. The three view solution which assumes a linear motion is similar to

Preprint submitted to 2013 IEEE International Conference on Robotics and Automation. Received September 17, 2012.

[4]. As shown in several synthetic and outdoor experiments reliable and accurate reconstruction is achieved on various object trajectories with this assumption of modelling non-holonomic curves as a combination of circular arcs and straight lines. For locally circular motion the only degenerate situation as discerned by us, is when the object and camera move in parallel concentric arcs. Such situations are a lot more rare than the degenerate situations for planar motion. By exploiting specific properties of circular trajectories, the broad spectrum of planar degeneracies is averted. In this case we make use of the fact that the perpendicular bisectors of the chords of the circle meet at the center as the unique feature that is used to solve for the scale.

Our solution also provides for a seamless model switching based on scale drift values. Object motion consists of straight lines and arcs. The relative scale computation should be able to switch between linear and circular models of computing scale accordingly. The entailment of a model switch is detected through disproportionately high drift in scale values. We show results in synthetic and real world scenes where trajectories consisting of both linear and curve segments get robustly reconstructed.

The main contribution of the current work include the following.

First, we present a novel four view solution of relative scale problem for locally circular object motion. This circular four view solution is assisted by three view solution of [4] for handling most of the outdoor anholonomic trajectories. The scale solution makes use of the Multibody VSLAM framework introduced in [5].

Second, we present a new four view solution of relative scale for planar object motion. Unlike [2] this solution is incremental. This solution is specifically applicable when object and camera do not move in parallel planes.

Third, conditions of degeneracy for a linear case are derived. These degeneracies are different from degenerate conditions that arise when the object is constrained to a plane, wherein degeneracy occurs if either the camera moves in the same plane as the object or the object and camera motions are planar but in different parallel planes as reported in [2]. Planar degeneracy thus becomes a common phenomena in many outdoor and indoor robotic settings where camera and object motions are either coplanar or are in parallel planes. These conditions are the subject of discussion in section IV-D.

Finally, we show results on various publicly available datasets wherein often the camera and object motion can be co-planar. Reconstruction is shown for such potentially degenerate situations confirming the fidelity of the proposed method. Each of these datasets are challenging in their own way and consists of outdoor vehicles, indoor robots and drones.

The relative scale estimation is the final module in the pipeline that includes motion detection and segmentation along with the VSLAM framework [5], [6]. The framework provides for both sparse and dense segmentation using a combination of optical flow and multi view geometry cues. We explain the overall pipeline briefly in later parts of the paper as we begin first by presenting a brief review of related literature.

II. Related Literature

Motion reconstruction from a single moving camera is considered ill posed for it is quite impossible to triangulate a moving object without some assumptions regarding trajectory or camera motion or both. There are broadly two paradigms that have appeared in literature. In the first paradigm, often called trajectory triangulation, the motion of the moving camera is considered well known. In other words it does not attempt the SLAM problem in dynamic environments but focuses on triangulating a moving point from a sequence of known camera matrices. The seminal work in this first appeared in [7] for linear and conic trajectories. However this method cannot triangulate a moving point if the camera motion is linear or is coplanar with the moving object. Very recently [8] showed how to reconstruct trajectories that can be represented as a linear combination of basis functions. They analysed and showed in detail that when the object trajectory can be represented as a linear combination of camera trajectory and a constant vector, reconstruction is not possible. The reconstruction was over several views. Unlike a SLAM framework real results were from multiple cameras observing motion from known locations. In another paper they present a method to reconstruct articulated trajectories [9] given a set of image projection and the parent trajectory in 3D.

In the second paradigm motion is reconstructed by explicitly providing for camera motion estimation. This has typically taken the form of multibody extension to multi-view geometry that tackles multiple moving rigid objects using classical SfM formulations. This appeared in [10], [11], [12]. These methods either used factorization techniques [10], [12] or statistical method [11] to segment multiple moving objects in two views. They assumed known correspondences. While initial papers showed results over few views, often with known correspondences and manually segmented objects, the practical aspects relating to implementation of such a multibody SfM over longer sequences is discussed in [13]. [2] devoted itself to the relative scale problem. In [14] moving objects are reconstructed through multibody multiview stereo. However their work does not address the relative scale problem since the scales were not so crucial from the point of view of segmentation, one of the main focus in that paper.

Within the robotic community the number of approaches that perform MonoSLAM within a dynamic environment and as well as provide some information about the target has been rather sparse. The pioneering work has been due to [15] that used a Bearing Only Tracker (BoT) within a Visual SLAM and Object Tracking (VSLAMMOT) framework with inverse depth parametrization. It presented comparisons with stereo SLAMMOT and showed superior performance with stereo SLAMMOT vis-a-vis VSLAMMOT due to the problems of observability. A similar approach that combines

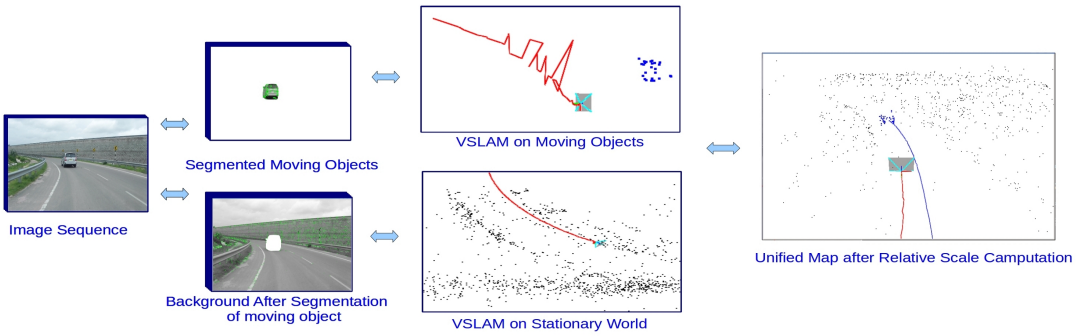


Fig. 1. This figure depicts overview of our complete multibody VSLAM system. Any one of two kind of motion segmentation, sparse or dense motion segmentation can be used in multibody VSLAM. The final result is unified map of the scene which includes, 3D-structure of moving object, 3D-structure of stationary world, camera trajectories and trajectories of moving objects at correct relative scale.

moving object tracking through BoT and MonoSLAM done on stationary parts of the environment, which does not reconstruct the moving object was proposed in [17] . A stereo or bicameral method of SLAM cum motion tracking that overcomes the observability problem was also presented in [18]. An approach that keeps or filters the dynamic features out of the SLAM framework without resorting to provide additional information in form of BoT of such features was presented in [19]. [16] demonstrated a technique for simultaneous co-operative localization and moving object tracking.

In contrast to most approaches in the Multibody setting the current approach invokes the incremental Multibody VSLAM framework introduced in [5] and shows explicit trajectory reconstruction in challenging outdoor scenarios over long sequences. The reconstruction is an outcome of several individual components such as motion detection and segmentation [5], [20], [6], the Multibody framework integrated with relative scale computation. The method of solving scale for circular trajectories in four views is novel and the solution based on planarity constraints is also different from [13]. Such reconstruction of stationary and moving elements along with camera trajectory estimation in uncontrolled scenes previously appears apart from our earlier effort [5] in [13].

III. System Architecture

We here delineate the Multibody VSLAM architecture and its pipeline(figure 1). The pipeline consists of a motion detection and segmentation module that segments independent motion. Each segmented moving object is given to a separate VSLAM module and another VSLAM module processes the static content (stationary world) in the image sequence. The output of each such VSLAM module is either the camera trajectory with respect to the stationary world or the moving object and the reconstruction of the stationary world or the moving objects. Each reconstructed moving object is then given to the module that finds the relative scale of that object with respect to the stationary world, which has been elaborated in detail in the section IV.

The motion detection framework which is sparse model of tracking and segmentation uses fast corners as means of tracking. Each such track is given to a module which estimates either epipolar constraint or flow vector bound constraint [20]. The output of these constraints is then fed to a probabilistic Bayes filter. The output of this Bayes filter is classfication of features into moving and non-moving. The technical details of this work can be found in [20].

The dense motion segmentation is that of [6]. It is an incremental framework in which dense optical flow features are tracked and motion potentials based on geometry are computed for each of these dense tracks. A graph based clustering algorithm then clusters and segments various moving objects.

The VSLAM module is that of bundle adjustment based optimization framework [21], [22], [23], [24] than filter based approaches [25], [26]. Our VSLAM system closely sembles with [21], [22], [24]. In brief, a five point algorithm [23] is used to estimate initial structure and camera parameters. A resection algorithm [27] subsequently estimates structure and motion parameters. Global and local optimizations are performed on key frames in two different threads to robustify structure and camera estimates.

IV. Solving for Relative Scale

Consider a moving object B and the frame fixed on it as D. The multi-body VSLAM/SfM outputs reconstructed points and cameras which see these points. Let one such point is P(see figure 2) which has a position vector ${}^{D}P$ with respect to moving frame D. Let the camera C which sees this point P has transformation given by ${}^{D}_{C}T = [{}^{D}_{C}R, {}^{D}t_C]$. Where ${}^{D}_{C}R$ represents orientation and ${}^{D}t_C$ represents position vector of origin of camera frame C, with respect to dynamic frame D. Then P is represented in the camera's frame C as

$$ {}^{C}P = {}^{D}_{C}R^{-1} \cdot ({}^{D}P - {}^{D}t_C) \tag{1} $$

Since the pose of the camera C with respect to the stationary/ground frame G is also known through the multibody framework, let this be represented as ${}^{G}_{C}T = [{}^{G}_{C}R, {}^{G}t_C]$, where ${}^{G}_{C}R$ is the rotation of the camera frame C with respect to G and ${}^{G}t_C$ represents position vector of origin of camera frame C with respect to ground frame G, then the point P can be represented in the frame G as,

$$ {}^{G}P = s\,{}^{G}_{C}R \cdot {}^{C}P + {}^{G}t_C \tag{2} $$

Preprint submitted to 2013 IEEE International Conference on Robotics and Automation. Received September 17, 2012.

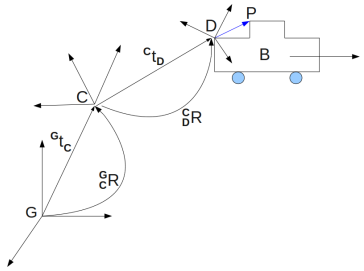


Fig. 2. This figure shows how a points P on the moving object B, measured in the coordinate frame D fixed on the object gets transformed to the frame G of the stationary world. The transform involves a scale apart from rotation and translation.

where s is the relative scale between the object/dynamic and the stationary world frame.

A. For Planar Motion

Herein we present solution of relative scale when the moving object undergoes a planar motion. Let the point P on the reconstructed moving object be represented as ${}^{G}P_{n}$, ${}^{G}P_{n+1}$.... ${}^{G}P_{n+m-1}$, in ground frame G, at m consecutive time instances then,

$$ {}^{G}P_{n+r} = s\ {}^{G}_{C_{n+r}}R \cdot {}^{C_{n+r}}P + {}^{G}t_{C_{n+r}} \forall r \in (0, m) \quad (3) $$

Suppose if the points on the moving object undergoes a planar motion then we propose the following methods of solving for scale:

Search Based Solution: This solution is similar to cross product scale search solution mentioned in [4]. We need four camera views to solve for relative scale using this solution. Let us assume that,

$$ {}^{G}P_{n+r} = x_r\hat{i} + y_r\hat{j} + z_r\hat{k} \forall r \in (0, m) $$

where each of

$$ x_r = x_{1r} + s\, x_{2r}\ \&\ y_r = y_{1r} + s\, y_{2r}\ \&\ z_r = z_{1r} + s\, z_{2r} $$

Then equation of plane from first three points ${}^{G}P_{n+0}$,${}^{G}P_{n+1}$ and ${}^{G}P_{n+2}$ can be given by

$$ \begin{vmatrix} x & y & z & 1 \\ x_0 & y_0 & z_0 & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \end{vmatrix} = \begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0 $$

Now if the fourth points also lies on this plane then we have,

$$ \begin{vmatrix} x_3 - x_0 & y_3 - y_0 & z_3 - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0 \quad (4) $$

We can search for all possible value of scale and find the solution of s such that it satisfy equation 4. In practice we need a solution such that L.H.S of equation 4 has its non-zero minimum value. It should be noted that instead of considering first four camera views we can consider any four camera views.

A Linear Solution for Scale: It should be noted that equation 4 which is in determinant form can be represented as third order polynomial in relative scale s as,

$$ \alpha_r\ s^3 + \beta_r\ s^2 + \gamma_r\ s + \delta_r = 0 \quad (5) $$

Instead of four if we consider five views, it is possible to obtain a linear and exact solution for relative scale which will not have any scale search criteria. Let us consider first five views as, ${}^{G}P_{n+0}$,${}^{G}P_{n+1}$...${}^{G}P_{n+4}$ then, we can consider any 3 or 4, four view combinations from $\binom{5}{4}$=5 combinations of these views. Similar to equation 5 we formulate a 3rd order polynomial from each of these four view combinations. Let the polynomials be,

$$ \alpha_r\ s^3 + \beta_r\ s^2 + \gamma_r\ s + \delta_r = 0 \forall r \in (1, 5) \quad (6) $$

Now considering the above equation 6 for r=1,2,3,4 and eliminating s^3 and s^2 term from above equations we will have a linear solution of relative scale.

B. For Circular Motion

We now present a method of estimating relative scale for the scenario when moving points undergoes a circular motion. This solution requires four camera views. Let us consider the first four views to be ${}^{G}P_{n+0}$,${}^{G}P_{n+1}$...${}^{G}P_{n+3}$, then,

Let ${}^{G}P_{n+0}$ and ${}^{G}P_{n+1}$ be the end points of first chord of circle and let ${}^{G}P_{n+1}$ and ${}^{G}P_{n+2}$ be the end point of the second chord of circle. Let,

$$ \overrightarrow{A} = {}^{G}P_{n+1} - {}^{G}P_{n+0} $$
$$ \overrightarrow{B} = {}^{G}P_{n+2} - {}^{G}P_{n+1} $$

Let the mid point of first and second chord be $\overrightarrow{I}$, $\overrightarrow{J}$ respectively, where

$$ \overrightarrow{I} = ({}^{G}P_{n+1} + {}^{G}P_{n+0})/2 $$
$$ \overrightarrow{J} = ({}^{G}P_{n+2} + {}^{G}P_{n+1})/2 $$

The perpendicular bisector of first chord has the direction of $\overrightarrow{A} \times (\overrightarrow{A} \times \overrightarrow{B})$. Now, the equation of perpendicular bisector of first chord can be given by,

$$ \overrightarrow{r} = \overrightarrow{I} + t_1\ \overrightarrow{A} \times (\overrightarrow{A} \times \overrightarrow{B}) $$

Similarly the equation of perpendicular bisector of second chord can be given by,

$$ \overrightarrow{r} = \overrightarrow{J} + t_2\ \overrightarrow{B} \times (\overrightarrow{A} \times \overrightarrow{B}) $$

The point of intersection of above two perpendicular bisector is the center of the circle. Let it be represented by $\overrightarrow{Cen}$. Now let,

$$ f = |radius - dis(Cen, {}^{G}P_{n+3})| $$

then, for fourth point (${}^{G}P_{n+3}$) to dwell on the circle, we have,

$$ f = 0 $$

In principal we need a value of scale at which this function f attains a non-zero minima.

Preprint submitted to 2013 IEEE International Conference on Robotics and Automation. Received September 17, 2012.

C. RANSAC and Temporal Smoothing

It is quite possible that, in real datasets the estimates from VSLAM/SfM system could be noisy. In such situations we compute the scale for multiple points on the objects over three or four views and resort to RANSAC to estimate the most likely scale.

Though, it requires only four views for estimating the relative scale, we adapted a temporal smoothing and initialization schema for improving the accuracy of initial unified reconstruction. Instead of first four views first n (6 in our case) views are considered. Now there will a total of $\binom{n}{3}$ (for linear method) or $\binom{n}{4}$ (for circular and planar method) combinations of camera views. For each such combination of camera views, a relative scale was estimated and RANSAC computed value of all these estimated scales was used for precise reconstruction. For example, if we consider first 6 views for initialization then we will have $\binom{6}{4}$ (assuming only circular method is required) values of scale and RANSAC computed value over all these scales was considered for initialization.

D. Degeneracies

Since our reconstruction framework handles motion along lines and curves we describe degenerate conditions for reconstruction of a line. The degenerate situations that arise for linear motion has not been elaborated elsewhere and forms an important contribution of this work.

Let the moving point P be represented as ${}^{G}P_{n+0}$,${}^{G}P_{n+1}$ and ${}^{G}P_{n+2}$ in ground frame G at three time instances. Then, considering equation 3 for $r = 0, 1$ and doing simple algebraic manipulations gives,

$$\begin{aligned} {}^{G}P_{n+1} - {}^{G}P_{n+0} = s\,[{}^{G}_{C_{n+1}}R \cdot {}^{C_{n+1}}P - {}^{G}_{C_{n+0}}R \cdot {}^{C_{n+0}}P] \\ +[{}^{G}t_{C_{n+1}} - {}^{G}t_{C_{n+0}}] \end{aligned}$$

With simple notations this equation can also be written as,

$${}^{G}P_{n+1} - {}^{G}P_{n} = \gamma_1 \hat{p_1} + s\,\delta_1 \hat{r_1} \tag{7}$$

Similar we can have,

$${}^{G}P_{n+2} - {}^{G}P_{n+1} = \gamma_2 \hat{p_2} + s\,\delta_2 \hat{r_2} \tag{8}$$

The left hand side of above two equations can be interpreted as the displacement of the object between two time instances, as represented in the stationary/global frame G.

The right hand side consists of a combination of unit vectors $\hat{p_1}$, $\hat{p_2}$, $\hat{r_1}$ and $\hat{r_2}$. This is a combination of the camera displacement as represented in G and the displacement of the object with respect to camera rotationally aligned with G. More conveniently $\hat{p_1}$ and $\hat{p_2}$ represent the unit vector in the direction of the camera velocity in time instances [t0,t1] and [t1,t2], while $\hat{r_1}$, $\hat{r_2}$ represents the direction of the relative velocity vector of the object with respect to camera aligned with frame G.

Thus the above equations represent the true object velocity as a combination of the camera velocity and object's relative velocity. For locally linear motion,

$${}^{G}P_{n+1} - {}^{G}P_{n} = k({}^{G}P_{n+2} - {}^{G}P_{n+1})$$

or

$$\gamma_1 \hat{p_1} + s\,\delta_1 \hat{r_1} = k(\gamma_2 \hat{p_2} + s\,\delta_2 \hat{r_2})$$

For degeneracy we seek situations wherein above equation holds for all s. Two cases arise,

Case1: The case of *velocity degeneracy*. If $\hat{p_1} = \hat{p_2}$ and $\hat{r_1} = \hat{r_2}$ then,

$$\gamma_1 \hat{p_1} + s\,\delta_1 \hat{r_1} = k(\gamma_2 \hat{p_1} + s\,\delta_2 \hat{r_1})$$

In such a situation we are able to equate components as $\gamma_1 = k\gamma_2$ and $\delta_1 = k\delta_2$ independent of scale resulting in degeneracy. We denote this condition as velocity degeneracy. Velocity degeneracy occurs when the camera velocity and object's relative velocity do not change direction and the ratio of their magnitude remains constant.

Case2: The case of *parallel degeneracy* If $\hat{p_1} = \hat{p_2} = \hat{r_1} = \hat{r_2} = \hat{p}$ then, we have a situation where the locally linear condition holds for all s. In this situation camera and object's relative velocity are parallel to each other, which in effect imply the camera and the object velocity are parallel to each other.

However, practical occurrences of such precise conditions of linear degeneracy is extremely rare in a real world scenario. For example a camera mounted on a vehicle would change its velocity ever so slightly, nor is it possible for the moving objects to maintain their ratio of velocities constant over time. Even very minute alterations from these degeneracy conditions will results in successful solution of relative scale.

At this point it is worthwhile to note the difference in degenerate conditions obtained above vis-a-vis planar degeneracies of [2]. The degeneracy of [2] prevents coplanar object and camera motion as well as object and camera to move in parallel planes.

For circular motion degeneracy occurs only when the camera and object both moves in parallel concentric arcs. This kind of degeneracies are far more rare than planar degeneracies.

V. Results

We show results of our relative scale estimation algorithm and the contingent unified representation of object and stationary world in various indoor, outdoor and synthetic datasets. In various ways we argue how the relative scale computed is precise or close to true scale.

A. Synthetic Data

We generated a set of 200 3D points. Some of these points are stationary while rest of them moved on various trajectories to simulate moving points. Pinhole camera model with a fixed focal length was assumed. Random extrinsic matrix were used to generate camera translations and rotations. These cameras were used to project 3D points to generate synthetic 2D image points. Our synthetic SfM or

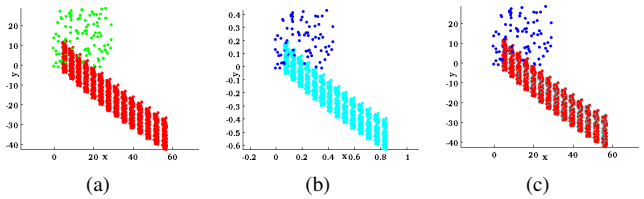


Fig. 3. This figure portrays synthetic simulation results for Circular method. (a) shows assumed ground truth points. Stationary points are shown in green and moving in red. (b) shows multibody SfM output. Blue points show stationary points and cyan points show reconstructed moving points in ground frame of Multibody SfM. As SfM gives upto scale results, this result is scaled version of original simulated structure. (c) shows result of (a) and result of (b) scaled to ground truth values in one single image. The purpose of (c) is to verify the accuracy of circular method of relative scale solution. The cyan points are invisible as they have completely coincided with the ground truth red points of (a).

VSLAM closely resembles with TorrSAM [28] and VLG [29]. Figure 3 portrays synthetic results for circular method

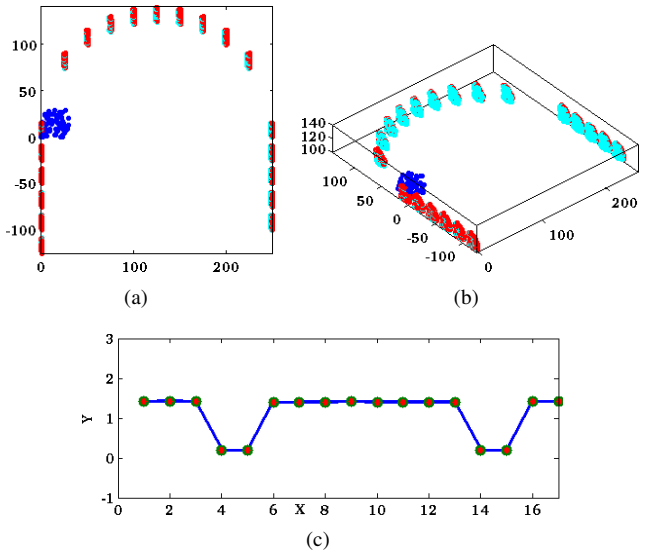


Fig. 4. This figure depicts results of our relative scale solution when object moves on a path which has straight line segment followed by a circular arc which is followed another line segment. (a) and (b) shows result of Multibody SFM. (c) shows results for variation in scale with time for this dataset. Scale breaks only at the instances when model switching takes place and it is close to accurate except for these instances. At these instances median of all the previous scale computations is used to represent the moving object. From these results the accuracy of our system to switch from circle to straight line (and vice versa) stands vindicated.

of relative scale estimation. From this figure the fidelity of circular method stands substantiated as the multibody VSLAM reconstructed points almost blends with original ground truth structure. 3(a) shows assumed ground truth points. Stationary points are shown in green and moving in red. 3(b) shows multibody SfM output. Blue points shows stationary points, in ground frame of Multibody SfM and cyan points depict reconstruction of moving points at the accurate scale. 3(c) shows result of 3(b) scaled to ground truth scale and merged with ground truth structure shown in 3(a). In 3(c) cyan points are invisible as they completely coincide with the ground truth red points. Figure 3(c) substantiate the accuracy and soundness circular method. Figure 4 delineates result for a scenario when object moves in a path which has a straight line segment followed by a circular arc which is further followed another line segment. This figure is of importance as it depicts a scenario where straight line relative scale solution and circular method of relative scale solution are used as and when needed. These results are for a scenario of planar object and camera motion hence inheriting planar degeneracy. But, we are able to solve for relative scale using our circular and linear method of relative scale solution. Variation in computed scale with time for this kind of motion is shown in figure 4(c). In this graph, the first three scales corresponds to linear object motion and scale was estimated using linear method. After this scale breaks at two instances. This is the time when model switching takes place. Neither of the two methods namely linear and circular method works while model switching takes places. At these instances median of all the previous scale computations is used to represent the moving object. After these two instances scale computations are rectified and remains close to accurate until the time when next model switching takes place. Meanwhile, figure 5 portrays result for simulated serpentine motion using circular and linear method wherein once again multiple model switching between circles occur.

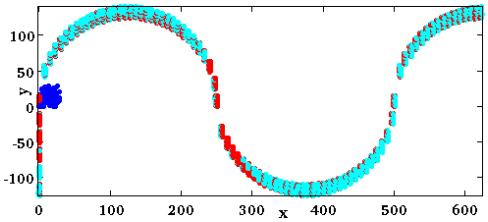


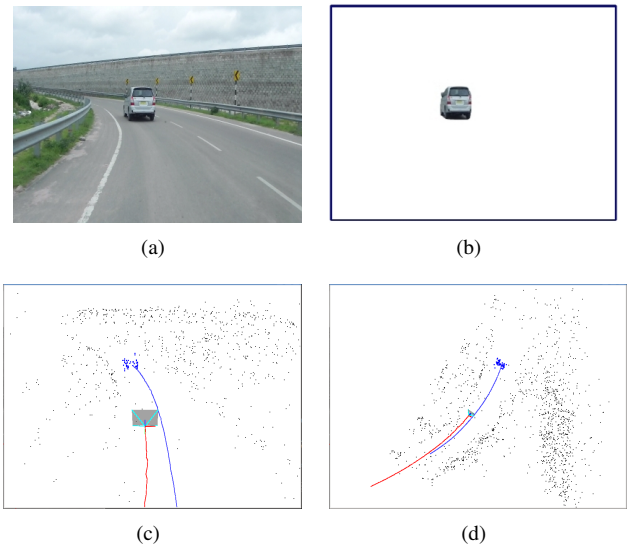
Fig. 5. This figure depicts results on simulated serpentine motion of our relative scale estimation. This result vindicate the efficacy our system to switch from one circle to other.

B. Real Results for various datasets

We now present real results on various publicly available and dataset collected outdoor. All the datasets are very important as they present very common real life scenario. We present in detail the reconstruction and the verification of all of the results.

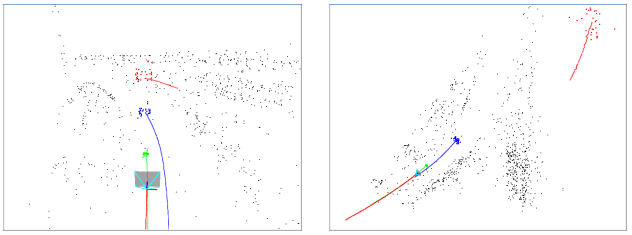
1) Moving Car Dataset: This dataset was collected by a high resolution camera. In this dataset moving car was moving on a circular arc while ascending on a slight acclivity. This dataset depicts a highly challenging scenario where planar degeneracy exists. We are able to successfully reconstruct the moving car at correct relative scale. We used circular method of relative scale estimation for reconstructing the moving car at correct relative scale. The results for this dataset are shown in figure 6 and figure 7.

Figure 7 shows one of the cars at the correct relative scale of 0.15 (blue) and also at scales of 0.03 and 0.7. Qualitatively as well it is possible to discern from these plots that the scale returned by the algorithm (0.15) ought to be closest to the



(a) (b) (c) (d)

Fig. 6. Results on MovingCar Dataset. In this dataset the moving car was moving almost on a circular arc while ascending on a slight acclivity. (a) An instance of the image sequence. (b) depicts segmented moving car. (c) delineates structure and trajectory of moving car in blue. In this unified map black dots show stationary world. (d) shows results of unified reconstruction from top view. The camera trajectory is shown in red.

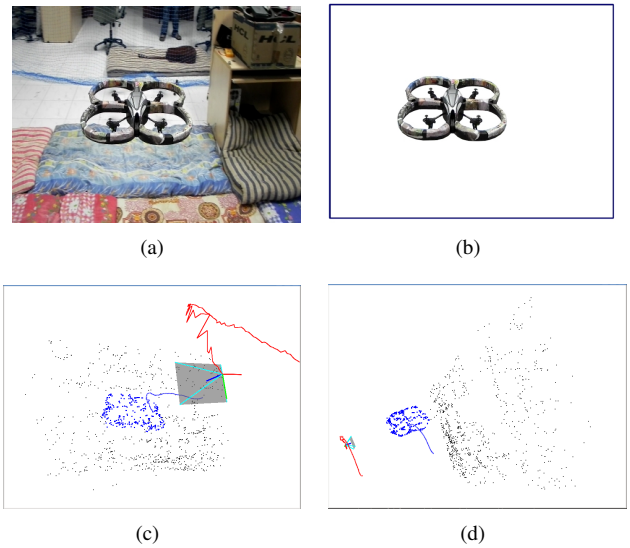


(a) (b)

Fig. 7. Moving object trajectory and its structure for three different scales of 0.03, 0.15 and 0.7 where 0.15(blue) being the correct scale. Qualitatively as well it is possible to discern from these plots that the scale returned by the algorithm (0.15) ought to be closest to the correct scale. At scale 0.05 the car shown in green lies very close to the moving camera. At the scale of 0.7 the moving car assumes a bigger structure than it should be and it lies beyond the road through which it travelled. (a) depicts result from front and (b) from top view.

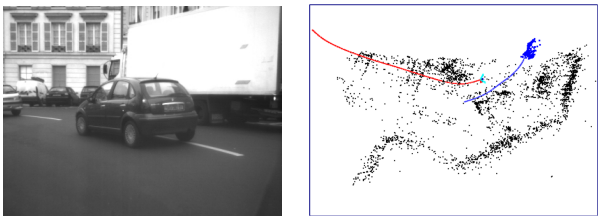
correct scale. At 0.03 the object is too close and almost lies on the camera while at 0.7 it goes beyond the road in which the car moved.

2) Drone Dataset: In this dataset a flying quad-copter (drone) was captured from a hand-held camera. The drone hovered almost in a plane. As this dataset was captured from a hand-held camera the stringent degeneracies of object and camera moving in parallel planes do not apply here. For the hand held camera can be made to move in a non planar trajectory or along a plane not parallel to the plane of drone's motion. Therefore we used planar method to estimate the relative scale of this drone. The results for this dataset are shown in figure 8. From this result the accuracy and efficacy of our VSLAM system to handle the reconstruction of flying vehicles undergoing highly dynamic motion stands corroborated. This dataset is of prime importance as this kind



(a) (b) (c) (d)

Fig. 8. Results on Drone Dataset. In this dataset a flying quad-copter (drone) was captured by a hand-held camera. This hand-held camera introduced human error thereby avoiding planar degeneracies and making it possible to use planar method for estimating relative scale. (a) shows an image from the dataset. (b) shows segmented Drone. (c) and (d) portray results of Multibody VSLAM. Structure of reconstructed drone along with its trajectory is shown in blue. Black dots represents stationary world. Trajectory of moving camera is shown in red.



(a) (b)

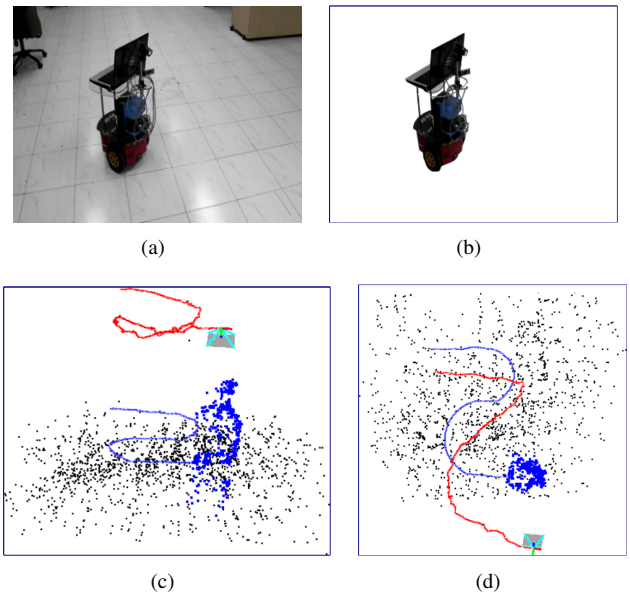
Fig. 9. (a) An image from the Versailles Rond sequence. (b) shows the an instance of unified map from Multibody VSLAM. Black dots represents stationary world. Structure and trajectories of moving car is shown in blue.

of results of reconstruction of a flying drone with a moving monocular camera is not seen anywhere in earlier literature.

3) Versailles Rond Dataset: We have shown our results on publicly available Versailles Rond dataset [30]. Only right images from the stereo pair have been used. We show reconstruction of one of the moving car along with the stationary world at the correct relative scale for the car vis-a-vis stationary world. Figure 9(a) shows an image from the sequence while 9(b) depicts final unified map from multibody VSLAM.

4) Line-Circle-Circle-Line (LCCL) Dataset: The purpose of this dataset is to capture serpentine object motion. This dataset starts with a straight line object motion which is followed by a serpentine segment made of two semicircles of almost same radius. This serpentine segment is further followed by a straight line. The result for this dataset are shown in figure 10. The accurate results on this dataset vindicate the ability and efficacy of the system to deal with difficult motion involving multiple model switches. The

Preprint submitted to 2013 IEEE International Conference on Robotics and Automation. Received September 17, 2012.



(a) (b) (c) (d)

Fig. 10. Results on LCCL dataset. In this dataset a pioneer (P3DX) robot moved in path which is a combination of 2 circles and 2 straight line simulating serpentine kind of trajectory of the object points. (a) shows an image from the dataset. (b) shows segmented robot. (c) and (d) portray results of Multibody VSLAM. Structure of reconstructed robot along with its trajectory is shown in blue.

results are that of a P3DX robot moving along a serpent like non holonomic curve and the camera moved in a plane parallel to the motion of P3DX.

Video sequences of results are attached in the video provided as supplementary material. A high resolution version of the video could be found at `http://web.iiit.ac.in/~rahul_namdev/videosequence`.

VI. Conclusion

This paper presented solution to the relative scale problem in a multibody setting for non-holonomic motions within four views of reconstruction of the stationary world and moving object. Two solutions are proposed based on planarity and circular constraints of object motions. The specific advantages of either of them have also been well argued. The solution differs from the recent probabilistic approaches, which involve many views as well as earlier approaches that involved an exhaustive search in scale space. The proposed method for handling circular motions in four views is novel and does not appear in literature earlier. That the method is also able to seamlessly switch between multiple motion models is vividly depicted in synthetic and real world scenarios. The analysis of degeneracies especially in the context of linear object motion seems to be the first of its kind to have appeared in literature. The method has been verified on publicly available datasets and the unified representation of the stationary and dynamic worlds are shown accurate through qualitative visual appeal by contrasting the scene when objects are represented at wrong scales. Quantitative verification with ground truth on synthetic data confirms the fidelity of the formulation. The method works in presence of high degrees of correlation between camera and object trajectories as well as when the object and camera trajectories are coplanar or move in parallel planes. Such extensive results portrayed on outdoor vehicles, indoor ground and aerial robots is also an unique aspect of this effort.

References

[1] C. Urmson and et al., "Autonomous driving in urban environments: Boss and the urban challenge," *Journal of Field Robotics Special Issue on the 2007 DARPA Urban Challenge, Part I*, vol. 25, no. 1, pp. 425–466, June 2008.

[2] K. Egemen Ozden, K. Cornelis, L. Van Eycken, and L. Van Gool, "Reconstructing 3D trajectories of independently moving objects using generic constraints," *CVIU*, vol. 96, no. 3, pp. 453–471, 2004.

[3] P. Soures and J.-P. Laumond, "Shortest path synthesis for a car-like robot," *IEEE Transactions on Automatic Control*, vol. 41(5): 672688, 1996.

[4] Linear method of relative scale solution. http://web.iiit.ac.in/~rahul_namdev/technicalreport.pdf.

[5] A. Kundu, K. M. Krishna, and C. V. Jawahar, "Realtime multibody visual slam with a smoothly moving monocular camera," in *ICCV*, 2011.

[6] R. K. Namdev, A. Kundu, K. M. Krishna, and C. V. Jawahar, "Motion segmentation of multiple objects from a freely moving monocular camera," in *ICRA*, 2012.

[7] S. Avidan and A. Shashua, "Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence," *PAMI*, vol. 22, no. 4, pp. 348–357, 2002.

[8] H. S. Park, I. Matthews, and Y. Sheikh, "3d reconstruction of a moving point from a series of 2d projections," in *ECCV*, 2010.

[9] H. S. Park and Y. Sheikh, "3d reconstruction of a smooth articulated trajectory from a monocular image sequence," in *ICCV*, 2011.

[10] S. Rao, A. Yang, S. Sastry, and Y. Ma, "Robust Algebraic Segmentation of Mixed Rigid-Body and Planar Motions from Two Views," *IJCV*, 2010.

[11] K. Schindler and D. Suter, "Two-view multibody structure-and-motion with outliers through model selection," *PAMI*, vol. 28, no. 6, pp. 983–995, 2006.

[12] R. Vidal, Y. Ma, S. Soatto, and S. Sastry, "Two-view multibody structure from motion," *IJCV*, vol. 68, no. 1, pp. 7–25, 2006.

[13] K. E. Ozden, K. Schindler, and L. V. Gool, "Multibody structure-from-motion in practice," *PAMI*, vol. 32, pp. 1134–1141, 2010.

[14] J. J. Guofeng Zhang and H. Bao, "Simultaneous multi-body stereo and segmentation," in *ICCV*, 2011.

[15] K. Lin and C. Wang, "Stereo-based Simultaneous Localization, Mapping and Moving Object Tracking," in *IROS*, 2010.

[16] C.-H. Chang, S.-C. Wang, and C.-C. Wang, "Vision-based cooperative simultaneous localization and tracking," in *ICRA*, 2011.

[17] D. Migliore, R. Rigamonti, D. Marzorati, M. Matteucci, and D. G. Sorrenti, "Avoiding moving outliers in visual SLAM by tracking moving objects," in *ICRA'09 Workshop on Safe navigation in open and dynamic environments*, 2009.

[18] J. Sola, "Towards visual localization, mapping and moving objects tracking by a mobile robot: a geometric and probabilistic approach," Ph.D. dissertation, LAAS, 2007.

[19] S. Wangsirpitak and D. Murray, "Avoiding moving outliers in visual slam by tracking moving objects," in *ICRA*, 2009.

[20] A. Kundu, K. M. Krishna, and C. V. Jawahar, "Realtime motion segmentation based multibody visual slam," in *ICVGIP*, 2010.

[21] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Real time localization and 3d reconstruction," in *CVPR*, 2006.

[22] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *ISMAR*, 2007.

[23] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," in *CVPR*, 2004.

[24] H. Strasdat, J. Montiel, and A. Davison, "Scale Drift-Aware Large Scale Monocular SLAM," in *RSS*, 2010.

[25] A. Davison, I. Reid, N. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *PAMI*, vol. 29, no. 6, pp. 1052–1067, 2007.

[26] J. Civera, A. Davison, and J. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 932–945, 2008.

[27] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *IJCV*, vol. 81, no. 2, pp. 155–166, 2009.

[28] A structure and motion toolkit in matlab. http://cms.brookes.ac.uk/staff/PhilipTorr/Beta/torrsam.zip.

[29] Vision lab geometry library. http://vision.ucla.edu/vlg/.

[30] A. Comport, E. Malis, and P. Rives, "Accurate quadri-focal tracking for robust 3d visual odometry," in *ICRA*, 2007.

Preprint submitted to 2013 IEEE International Conference on Robotics and Automation. Received September 17, 2012.