2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Ambient Motion Estimation in Dynamic Scenes Using Wearable Visual-Inertial Sensors

Hongsheng He and Jindong Tan

***Abstract*— This paper proposes a method to estimate the motion of ambient objects including translational and rotational velocities by moving observers with hybrid visual-inertial sensors. Ambient motion is recovered from visual optical flows that represent ego and ambient dynamics. In this paper, each moving object is considered as a rigid body that has been segmented from the background using computer vision algorithms. In motion recovery, the fundamental challenge is to resolve the coupling between scene depths and translational velocities. Ambient rotational velocities are obtained following a depth-independent bilinear constrain. The scales of ambient translational velocities is computed using the proposed dynamics constraint with an assumption that ambient accelerations are negligible. A fix-point optimization scheme is further introduced to iteratively refine the recoveries of translational and rotational ambient motion until an expected precision is achieved or the maximal iteration is reached. During the optimization, translational ambient motion is precisely recovered and translational ambient motion is rescaled to the canonical amplitude. The results of the simulation study show the effectiveness of the proposed method in motion analysis and prediction.**

I. Introduction

Estimation of ambient motion using limited onboard or wearable sensors is a fundamental challenge in application of intelligent robots and assistive devices [1], [2], [3]. Increasing robotic applications require a robot to work in complex and dynamic environments with limited prior knowledge [4]. A robot should be aware of its surrounding environment both in static and dynamic aspects. In these circumstances, it is of vital importance for a robot to be able to estimate relative motion of ambient objects with respect to the robot and other objects using onboard sensors. For blind assistive devices, relative motion is the key parameter to guide the visually impaired to avoid a motion hazard during navigation in dynamic environments. Substantial research results have been achieved for path-finders [5] or object recognizers [6], but motion estimation is still an open problem in this filed with urgent needs. Ambient motion estimation also plays important roles in assistive driving where relative motion between cars is the most useful index to predict potential environmental collision. The available assistive driving systems generally depend on range finders such as 3D lasers [7], [8], which are cumbersome and require additional hardware supports. The problem of motion estimation using wearable lightweight sensors has not been fully addressed either in theory or in practical products.

Hongsheng He and Jindong Tan are with the Department of Mechanical, Aerospace and Biomedical Engineering, The University of Tennessee, Knoxville, TN, 37996, USA {hhe,tan}@utk.edu

*This work is partially supported by NSF grant 1309921. Hongsheng He is also partially supported by NSFC grant 61305114.

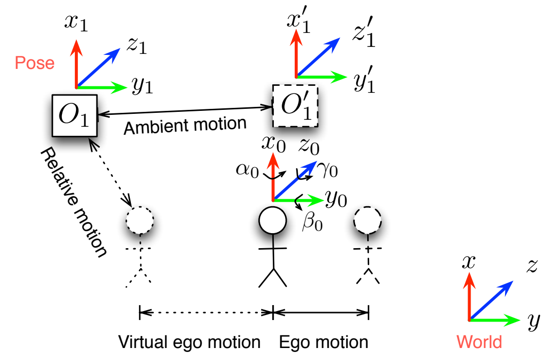


Fig. 1: Ambient motion estimation.

In this paper, we study the measurement of translational and rotational velocities of ambient objects based on a moving sensing platform with a camera and an inertial measurement unit (IMU). The concept of ambient motion estimation is presented in Fig. 1. The visual optical flows reflect the relative motion between the observer and the ambient object. We recover the ambient motion referred in world space from these visual observations. The composite motion is formulated in an optical-flow model. We first precisely recover rotational ambient motion based on the bilinear constraint. The major difficulty in recovering translational ambient motion is to deal with the scale ambiguity of scene depths and translational velocities. Without additional depth sensors, it is still impossible to recover scene depth in real time from visual observation not to mention scene reconstruction of a dynamic scene. Therefore, we introduce a dynamic constraint on the relationship between optical flows and motion dynamics. With this constraint, we are able to approximately estimate the scale of scene depths and hence the scale of translational velocities. Because of noises and outliers in optical-flow observations, we apply an iterative optimization mechanism to filter out outliers and to improve the estimation precision.

There are many difficult problems to address in measurement of ambient motion using visual-inertial sensors. Inertial sensors can provide precise local self-motion information but cannot measure ambient motion, whereas visual sensors can estimate single movement in a scene but cannot work well in dynamic environments without self-positioning. With their complementary properties, visual-inertial sensors compose the minimal sensing system to measure ego and ambient motion. In a complex environment, dynamics of individual objects may overlap and generate composite optical flows on visual captures. Though computer vision algorithms serve to perform motion segmentation, it is infeasible to precisely isolate a moving rigid in a dynamic environment, especially when reference objects or observers are moving as well. In addition, visual features are vulnerable to ambient conditions and image noises. Inertial sensors generally measure translational accelerations and suffer from accumulated errors in computation of translational velocities. Ambient motion estimation using visual-inertial sensors deserves more research efforts owing to its signification in real applications and research challenges in robust recovery.

Compute vision techniques have been investigated to estimate ego motion from visual optical flows with developed additional constraints [9], [10], [11]. However, these methods could not be directly applied to ambient motion estimation in that they cannot recover the scale of translational velocities without scene depths. Measurements of scene depths require addition hardware [12]. In addition, pure vision technologies may not applicable to dynamic environments where moving objects contribute dominant optical flows observations. In these circumstances, ego motion and dynamics should be measurable or known [13].

The main contribution of the paper is to recover ambient motion including translational and rotational velocities in a dynamic environment:

(i) the problem of ego and ambient motion estimation is formulated in an optimization framework of optical flows;
(ii) a dynamics constraint is proposed to approximate the dynamics of an observer and an ambient object; and
(iii) an iteration framework is introduced to optimize the estimation.

II. Ambience-Motion Estimation

In this section, we will model the estimation problem of ambient motion from optical flows in an optimization framework. To solve this optimization problem, a depth independent constraint and a dynamics constraint are proposed in the recovery of translational and rotational ambient velocities. The estimations of ambient motion with these constraints are iteratively refined in a fix-point manner by muting noisy observations.

The projection of 3D points on an image is described with a pinhole model,

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} f\frac{X_i}{Z_i} \\ f\frac{Y_i}{Z_i} \end{bmatrix}$$

where $[x_i, y_i]^T$is projection of a spacial point $[X_i, Y_i, Z_i]^T$ by a camera with a focal length f. Suppose that the ego motion is with a translational speed $\mathbf{v}^e = \begin{bmatrix} v_x^e & v_y^e & v_z^e \end{bmatrix}^T$ and a rotational speed $\omega^e = \begin{bmatrix} \omega_x^e & \omega_y^e & \omega_z^e \end{bmatrix}^T$. The optical flows in the image plane resulted by this motion are [14]

$$o_i^e = \begin{bmatrix} \dot{x}_i^e \\ \dot{y}_i^e \end{bmatrix} = \frac{1}{Z_i} A_i \mathbf{v}^e + B_i \omega^e \tag{1}$$

where

$$\begin{aligned} A_i &= \begin{bmatrix} -f & 0 & x_i \\ 0 & -f & y_i \end{bmatrix} \\ B_i &= \frac{1}{f} \begin{bmatrix} x_i y_i & -(f^2 + x_i^2) & f y_i \\ (f^2 + y_i^2) & -x_i y_i & -f x_i \end{bmatrix}. \end{aligned}$$

In a conjugate manner, given the ambient motion of a rigid body with a translational speed $\mathbf{v}^a = \begin{bmatrix} v_x^a & v_y^a & v_z^a \end{bmatrix}^T$ and a rotational speed $\omega^a = \begin{bmatrix} \omega_x^a & \omega_y^a & \omega_z^a \end{bmatrix}^T$ in world space, the optical flows caused by ambient motion are

$$o_i^a = -\frac{1}{Z_i} A_i \mathbf{v}^a - B_i \omega^a. \tag{2}$$

The relative motion of a moving object relative to a camera frame is $\mathbf{v}^a - \mathbf{v}^e$, and therefore the resulted optical flows are

$$o_i = \frac{1}{Z_i} A_i \left(\mathbf{v}^e - \mathbf{v}^a\right) + B_i \left(\omega^e - \omega^a\right). \tag{3}$$

With observed optical flows o_i of image points and measured ego motion $\mathbf{v}^e$ and ω^e, the true rotation ω^a and translation $\mathbf{v}^a$ follow (3). Thus an ambient motion can be recovered with observations of optical flows by minimizing the cost function

$$\left((\mathbf{v}^a)^*, (\omega^a)^*\right) = \arg\min_{\mathbf{v}^a, \omega^a} \sum_{i=1}^{n} \left\| f_i\left(\mathbf{v}^a, \omega^a\right) \right\|^2 \tag{4}$$

where the error function of observations is defined as $f_i\left(\mathbf{v}^a, \omega^a\right) = o_i - \left(\frac{1}{Z_i} A_i \left(\mathbf{v}^e - \mathbf{v}^a\right) + B_i \left(\omega^e - \omega^a\right)\right)$.

The twin problem to recover ambient motion is estimation of scene depths. In general, it is impractical to reconstruct precise depths Z_i of ambient objects without additional rangefinder sensors. Vision-based depth reconstruction methods are not competent to measure scenes in distance. In addition, visual disparity based reconstruction methods fail in a scenario that ambient motion caused by moving objects is prominent to the ego motion of an observer. Therefore, we derive a depth-independent constraint in recovering ambient motion.

A. Depth Independent Constraint

The bilinear constraint is obtained by optimizing (4) with respect to depths Z_i as

$$f_i\left(\mathbf{v}^a, \omega^a\right) = \left[B_i\left(\omega^a - \omega^e\right) + o_i\right]_\times^T A_i \left(\mathbf{v}^a - \mathbf{v}^e\right) = 0 \tag{5}$$

where $[\cdot]_\times$ represents a projection matrix of a crossing vector satisfying $\mathbf{x} \times \mathbf{y} = [\mathbf{x}]_\times \mathbf{y}$. The cross product of two-dimensional scalars is $\mathbf{x} \times \mathbf{y} = x_1 y_2 - x_2 y_1$. The projective relation between $\mathbf{v}^a$ and ω^a is intuitively demonstrated in Fig. 2. The ambient motion $(\mathbf{v}^a)^*$ and $(\omega^a)^*$ will be optimally recovered from multiple observations of these relationships.

It holds after expansion of (5) that

$$\begin{aligned} &\left((\omega^a)^T [B_i]_\times^T A_i - (\omega^e)^T [B_i]_\times^T A_i + [o_i]_\times^T A_i\right) \\ &\times \left(\mathbf{v}^a - \mathbf{v}^e\right) = 0 \end{aligned} \tag{6}$$

It should be noted that the coefficient is symmetric

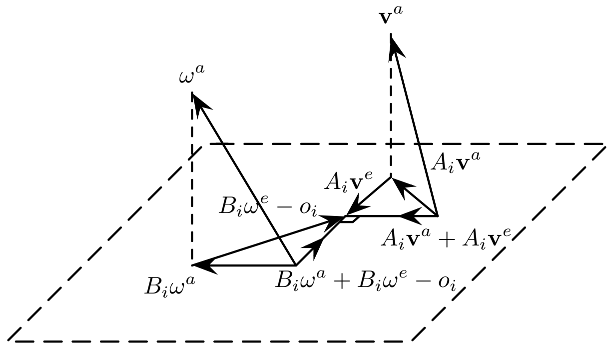


Fig. 2: Geometric relation between $\mathbf{v}^a$ and ω^a in (5).

$[B_i]_\times^T A_i = A_i^T [B_i]_\times$, which is straightforward by plugging in (1)

$$\begin{aligned} A_i^T [B_i]_\times &= \begin{bmatrix} -f & 0 \\ 0 & -f \\ x_i & y_i \end{bmatrix} \begin{bmatrix} -f^2 - y_i^2 & x_i y_i & f x_i \\ x_i y_i & -f^2 - x_i^2 & f y_i \end{bmatrix} \\ &= f \begin{bmatrix} f^2 + y_i^2 & -x_i y_i & -f x_i \\ -x_i y_i & f^2 + x_i^2 & -f y_i \\ -f x_i & -f y_i & -x_i^2 - y_i^2 \end{bmatrix}. \end{aligned} \tag{7}$$

Because of the symmetry, the upper triangular matrix together with the diagonal of $[B_i]_\times^T A_i$ is chose to construct linearly independent equations [10]. An augmented vector is defined as

$$\mathbf{e} = \left[\mathbf{v}_x\omega_x^a, \mathbf{v}_x\omega_y^a, \mathbf{v}_x\omega_z^a, \mathbf{v}_y\omega_y^a, \mathbf{v}_y\omega_z^a, \mathbf{v}_z\omega_z^a\right]^T \tag{8}$$

with $\mathbf{v} = \mathbf{v}^a - \mathbf{v}^e$, and the reformulated coefficients are

$$\mathbf{q} = \left[f^2 + y_i^2, -x_i y_i, -f x_i, f^2 + x_i^2, -f y_i, -x_i^2 - y_i^2\right]^T \tag{9}$$

and

$$\mathbf{p} = (\omega^e)^T [B_i]_\times^T A_i + [o_i]_\times^T A_i. \tag{10}$$

Then the optimization problem (4) becomes

$$\min \sum_{i=1}^{n} \left(\mathbf{v}^T \mathbf{p}_i + \mathbf{e}^T \mathbf{q}_i\right)^2. \tag{11}$$

Computation of partial derivatives of (11) with respect to $\mathbf{v}$ and $\mathbf{e}$ and evaluating them to zero generates nine equations and nine variables. A homogenous translational system obtained by canceling variables $\mathbf{e}$ is [10], [14]

$$\mathbf{v}^T G = 0 \tag{12}$$

where $G_{r,s} = \sum_{i=1}^{n} l_{r,i} l_{s,i}$ with

$$\begin{aligned} l_{r,i} &= \mathbf{p}_{r,i} - \sum_{r=1}^{6}\sum_{s=1}^{6} D_{r,s}\mathbf{q}_{s,i} \sum_{j=1}^{m} \mathbf{q}_{r,j}\mathbf{p}_{j,l} \\ & \quad l = \{1,2,3\}, i = \{1,2,\cdots,m\} \\ D &= \left(\sum_{i=1}^{m} \mathbf{q}_i \mathbf{q}_i^T\right)^{-1}. \end{aligned} \tag{13}$$

A solution for the direction of composite motion $\mathbf{v}$ is the eigenvectors corresponding to the smallest eigenvalue of G using singular value decomposition (SVD). The amplitude of the translational velocity is still unknown though the rotational velocity ω^a can been completely recovered from (6).

B. Velocity Scale Estimation

With the depth independent constraint, we can only recover ambient translational velocities up to a scale due to coupling of translational velocities and depths in (5). In this section, we will discuss on the method to estimate the absolute value of ambient translational velocities.

The derivatives of the optical flows and motion states are

$$\frac{d}{dt}o_i = \frac{1}{Z_i}A_i\left(\frac{d}{dt}\mathbf{v}^e - \frac{d}{dt}\mathbf{v}^a\right) + B_i\left(\frac{d}{dt}\omega^e - \frac{d}{dt}\omega^a\right). \tag{14}$$

Assuming ambient objects are with uniform rectilinear motion i.e. $\frac{d}{dt}\mathbf{v}^a = 0$ or the ambient accelerations are negligible compared with camera ego-motion, the below approximation holds during a very short sampling span

$$\frac{d}{dt}o_i = \frac{1}{Z_i}A_i\frac{d}{dt}\mathbf{v}^e + B_i\left(\frac{d}{dt}\omega^e - \frac{d}{dt}\hat{\omega}^a\right) \tag{15}$$

where $\hat{\omega}_a$ is the prediction of rotational ambient motion from the depth independent constraint. Thus the depth is recovered as

$$\frac{1}{Z_i}I = \left(\dot{o}_i - B_i\dot{\omega}^e + B_i\dot{\hat{\omega}}^a\right) \oslash A_i\dot{\mathbf{v}}^e \tag{16}$$

where $\oslash$ represents element-wise vector product and $I = [1,1]^T$.

With the estimated scene depth, the optimization solution of the cost function (4) can be obtained from a series of translational system equations in vector form with n observations of optical flows

$$K\begin{bmatrix}\mathbf{v}^a \\ \omega^a\end{bmatrix} = \mathbf{b} \tag{17}$$

where

$$\begin{aligned} K &= \begin{bmatrix} \frac{1}{Z_i}A_i & B_i \end{bmatrix}_{i=1}^{n} \\ \mathbf{b} &= \begin{bmatrix} \frac{1}{Z_i}A_i\mathbf{v}^e + B_i\omega^e - o_i^a \end{bmatrix}_{i=1}^{n}. \end{aligned} \tag{18}$$

The optimal solution of the translational system is

$$\begin{bmatrix}(\mathbf{v}^a)^* \\ (\omega^a)^*\end{bmatrix} = VS^{-1}U\mathbf{b} \tag{19}$$

with K as the singular value decomposition (SVD) of $K = USV^T$. The scale $\left\|(\mathbf{v}^a)^*\right\|$ of translational velocity is treated as a canonical value in further estimation, where only the direction of the velocity is optimized.

Provided that the coefficient matrix is with rank six, the motion of the rigid body can be recovered from multiple observations of optical-flows. In theory, the minimal number of observed points is three to satisfy the rank condition. In practice, more points are beneficial to improve the matrix condition of K_i and to filter out noisy observations.

Remark 1 (Recoverability): Uniform rectilinear ambient motion is recoverable from visual optical flows when ego dynamics are involved and measured.

The optical flows resulted by ambient motion on a stable observer is modeled in (3). It is infeasible to recover absolute scales of translational speeds due to the coupling between depths and translational motion. By introducing the dynamics constraint on ego and ambient motion, it is possible to virtually evaluate moving speed of ambient objects by (15). However, this method is unable to precisely estimate arbitrary ambient motion without additional sensors.

C. Fix-Point Optimization

The depth estimation from (15) is an approximated result. In addition, the algorithm is vulnerable to selection of candidate optical flows points and noise in observations. We will hence introduce an optimization mechanism to further improve the accuracy of the motion estimations and to suppress influence of optical-flow outliers that may be introduced by different moving objects.

The optimal estimation of ambient motion $(\mathbf{v}^a)^*$ and $(\omega^a)^*$ can be computed by minimizing the cost function of the depth independent constraint (5) with n observations of optical flows

$$\min f = \min \sum_{i=1}^{n} \left\| f_i \left(\mathbf{v}^a, \omega^a\right)\right\|^2 . \tag{20}$$

The optimization is achieved through a fix-point schema that iterates estimations for velocity $\mathbf{v}^a$ and ω^a. The estimation of $(\mathbf{v}^a)^*$ in the $k+1$ iteration is determined by

$$\mathbf{v}^a_{k+1} = \arg\min_{\mathbf{v}^a} f\left(\mathbf{v}^a_k, \omega^a_k\right). \tag{21}$$

For simplicity of representation, the relative translational velocity is defined as $\mathbf{v} = \mathbf{v}^e - \mathbf{v}^a$. The optimization is a least-square problem and the solution is given by translational and homogeneous system of equations

$$G_v \mathbf{v} = \mathbf{d}_v \tag{22}$$

where

$$\begin{aligned} G_v &= \left[\frac{1}{Z_i} A^a_i\right]^n_{i=1} \\ \mathbf{d}_v &= \left[\frac{1}{Z_i} A_i \mathbf{v}^e + B_i\left(\omega^e - \omega^a\right) - o_i\right]^n_{i=1} . \end{aligned} \tag{23}$$

The resulting system is solvable by computing pseudo-inverse of G_v. The resulted relative velocity $\mathbf{v}$ is normalized to a canonical form with $\bar{\mathbf{v}} = \frac{\mathbf{V}}{\|\mathbf{V}\|}$, and the ambient motion is recovered from

$$\begin{aligned} \|\bar{\mathbf{v}}\| &= 1 \\ \alpha_{k+1}\bar{\mathbf{v}} &= \mathbf{v}^e - \mathbf{v}^a_{k+1} . \end{aligned} \tag{24}$$

Similarly, the rotational part $(\omega^a)^*$ is estimated by the second optimization problem that is with the same formulations in (20) to search with respect to the variable ω^a as

$$\left(\omega^a_{k+1}, Z\right) = \arg\min_{\omega^a} f^{bp}\left(\mathbf{v}^a_{k+1}, \omega^a_k\right). \tag{25}$$

This optimization problem for extrema brings a homogeneous system of equations

$$G_\omega \begin{bmatrix} \left[\frac{1}{Z_i}\right]^n_{i=1} \\ \omega^a \end{bmatrix} = \mathbf{d}_\omega \tag{26}$$

where

$$\begin{aligned} &G_\omega = \\ &\begin{bmatrix} A_1\left(\mathbf{v}^e - \mathbf{v}^a\right) & \mathbf{0} & \mathbf{0} & \mathbf{0} & -B_1 \\ \mathbf{0} & A_2\left(\mathbf{v}^e - \mathbf{v}^a\right) & \mathbf{0} & \mathbf{0} & -B_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & A_n\left(\mathbf{v}^e - \mathbf{v}^a\right) & -B_n \end{bmatrix} \\ &\mathbf{d}_\omega = \left[o_i - B_i \omega^e\right]^n_{i=1} . \end{aligned} \tag{27}$$

In computation of the pseudo-inverse of G_v and G_ω, a good condition number of the matrix is favorable depending on the observations of optical flows in the image plane. In practical numerical computing, the number of observations is chose large enough to guarantee good conditions of the coefficient matrices.

The segmentation of ambient moving objects depends on the performance of vision algorithms. However, it is impractical to precisely segment different motion especially when they are overlapped. Therefore, candidate optical-flow observations are selected by evaluating conditions that

$$\begin{cases} Z_i\left(\left(\mathbf{v}^a\right)^*, \left(\omega^a\right)^*\right) > 0 \\ f_i^2\left(\left(\mathbf{v}^a\right)^*, \left(\omega^a\right)^*\right) < f_t \end{cases} \tag{28}$$

where f_t is the threshold to classify a candidate observation as a outlier, and depths Z_i are calculated from (16). The RANSAC scheme is employed to select data points from optical-flow observations. The iteration procedure stops when maximal numbers of iterations or minimal fitting errors are met for a set of selected optical-flow observations.

The whole procedure for ambient motion estimation is summarized as below:

- ◇ Initialization: Initial estimation $\left(\mathbf{v}^a_0, \omega^a_0\right)$ of ambient motion is estimated using (12) and (16) based on a small number of observations of optical flows.
- ◇ Estimation of $\mathbf{v}^a$ direction: Compute $\mathbf{v}^a_{k+1} = \arg\min_{\mathbf{v}^a} f\left(\mathbf{v}^a_k, \omega^a_k\right)$. This problem is formatted to be a linear system (23) and the solution is achieved by pseudo-inverse.
- ◇ Rescale $\mathbf{v}^a$: Compute the absolute value of $\mathbf{v}^a$ according to (16).
- ◇ Estimation of ω^a: Compute $\omega^a_{k+1} = \arg\min_{\omega^a} f\left(\mathbf{v}^a_{k+1}, \omega^a_k\right)$. This problem is again formatted to be a linear system (27).
- ◇ Refinement: Outliers or noisy data are filtered out according to the criterion (28).
- ◇ Iteration: The optimization continues until a minimal fitting error is obtained.

Remark 2 (Global optimal): The globally optimal estimation of ambient motion is not guaranteed using the fix-point optimization (21) and (25). Nonetheless, with the suboptimal starting point (19), there is rationale to suppose that the suboptimal initializations are sharply concentrated around the globally optimal.

III. Experimental Analysis

The experiments are conducted with a virtual camera of the focal length 0.0187 m and CCD size 0.01×0.01 m^2. In the simulation, optical flows are generated using the virtual camera for 50×50 observations. The scene depth is predefined by random generation or from scene reconstruction.

We evaluate the performance of the method by measuring the amplitude and directional errors for ambient motion estimation. The relative amplitude error is defined as

$$E^a = \left\| \frac{\hat{x} - x}{x} \right\|$$

with $\hat{x}$ as the estimation, x as the ground truth. The directional errors is measured by

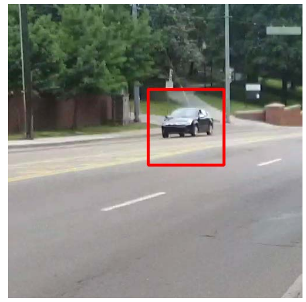
$$E^d = \arccos\left(\frac{\hat{x}^T x}{\|\hat{x}\|\,\|x\|}\right).$$

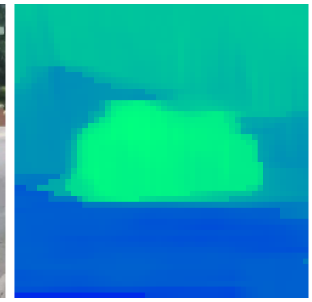
A. General Performance

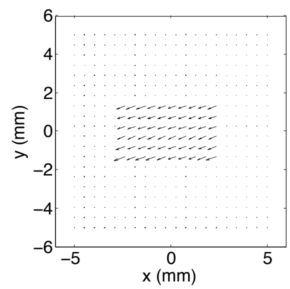
We perform ambient motion estimation on a real world scene as shown in Fig. 3. The car is running with translational speed of $[-13, -6, 0]$ meter/frame and rotational speed of $[0, 0, 1]$ degree/frame, and the observing camera is moving at $[1, 0, 0]$ meter/frame and $[0, 0, 3]$ degree/frame. The scene depth given in Fig. 3b is the ground truth from scene reconstruction of the region-of-interest around the car in Fig. 3a. With these configurations, the optical flows are generated as given in Fig. 3c. From the generated clean optical-flow observations, the recovered ambient motion is $[-13.000, -6.000, -0.000]$ meter/frame and $[0.008, -0.008, 1.000]$ degree/frame when the car is running in a constant speed. The reconstructed scene depths from (16) are presented in Fig. 3d for comparison with the ground truth. The simulation results in the ideal experiment conditions prove the capability of the proposed method to precisely recover ambient motion from segmented optical flows.

B. Performance from Noisy Observations

We conduct an experiment to investigate the robustness of the proposed method. Visual optical flows suffer from noises and errors in numerical computation in practical applications. We examine the method on visual optical flows with different scales of added Gaussian noises with zero mean. We adjust the scale of noise variance with respect to the camera focal length in the experiment. The results shown in Fig. 4 reveal that the estimation of the translational velocity is more robust to noises in visual optical flows than the estimation of the rotational velocity. The proposed method is able to precisely recover ambient motion when the variance of the optical-flow noises is up to 10% of the camera focal length. Recovery precision degrades as the level of noises increases. The proposed method behaves similarly to optical-flow outliers in recovery precision as shown in Fig. 5.


(a) A car running on a road.


(b) Ground truth scene depth from stereo reconstruction.



(c) Simulated observations of optical flows.

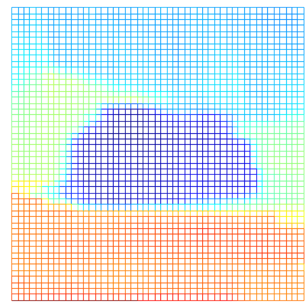
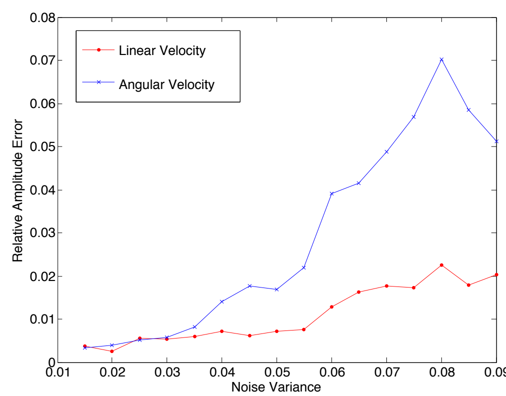

(d) Reconstructed scene depth.

Fig. 3: Ambient motion estimation and resulted scene depths.



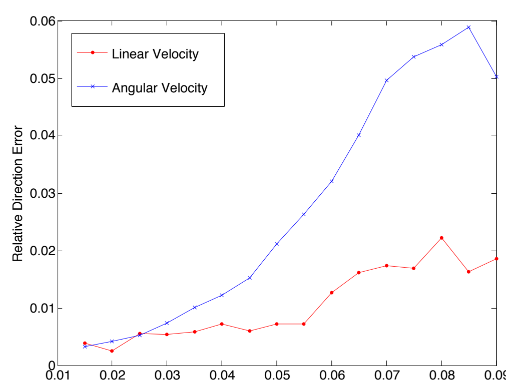


Fig. 4: Recovery precision with noisy optical flows.

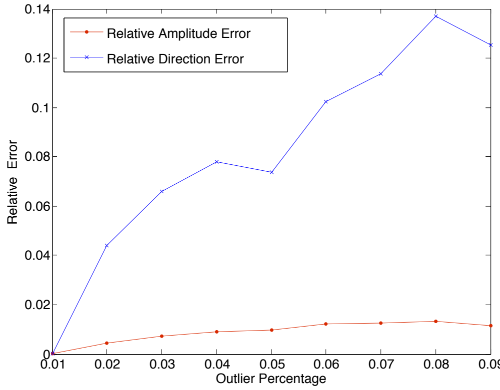


Fig. 5: Recovery precision from optical flows with outliers.

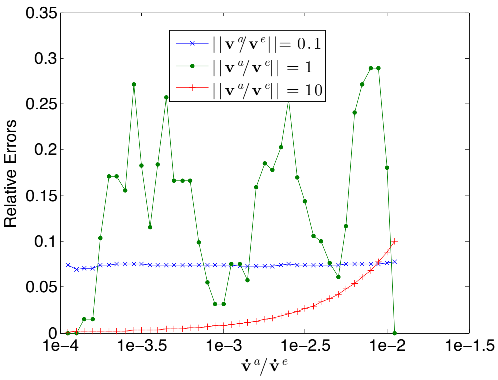


Fig. 6: Different configurations of ambient motion and ego motion.

C. Dependence on the Dynamics Constraint

The recovery precision of the scale of translational ambient motion depends on the extent to which the dynamic constraint (15) is satisfied in real applications. We compare the recovery precision of the proposed method with different configurations of ambient motion $\mathbf{v}^a$ and ego motion $\mathbf{v}^e$. The experimental result given in Fig. 6 shows that the recover precision generally drops as the ratio between $\mathbf{v}^a$ and $\mathbf{v}^e$ increases. The performance of the proposed method is unstable when the scales of $\mathbf{v}^a$ and $\mathbf{v}^e$ are comparable, while the performance is most stable when the scale of $\mathbf{v}^a$ is smaller than that of $\mathbf{v}^e$. The performance degrades sharply when the scale of $\mathbf{v}^a$ is larger than that of $\mathbf{v}^e$. In conclusion, the best recover precision is achieved when the acceleration of ego motion is greater than that of ambient motion while the velocity of the former is smaller than that of the latter.

IV. Conclusion

This paper has laid the theoretical and computational foundations to estimate ambient motion including translational and rotational speeds. We have proposed a method to estimate ambient motion using visual and inertial measurements. Though the depth and ambient translational motion are coupled in the model, we showed that it is possible to recover the ambient motion if it is uniform rectilinear motion.

References

[1] T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *Robotics, IEEE Transactions on*, vol. 28, no. 1, pp. 61–76, 2012.

[2] A. Martinelli, "Vision and imu data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination," *Robotics, IEEE Transactions on*, vol. 28, no. 1, pp. 44–60, 2012.

[3] S.-W. Yang and C.-C. Wang, "Simultaneous egomotion estimation, segmentation, and moving object detection," *Journal of Field Robotics*, vol. 28, no. 4, pp. 565–588, 2011.

[4] H. He, S. S. Ge, and Z. Zhang, "A saliency-driven robotic head with bio-inspired saccadic behaviors for social robotics," *Autonomous Robots*, vol. 36, no. 3, pp. 225–240, 2014.

[5] J. M. Coughlan and H. Shen, "The crosswatch traffic intersection analyzer: a roadmap for the future," in *Computers Helping People with Special Needs*. Springer, 2012, pp. 25–28.

[6] R. Manduchi, S. Kurniawan, and H. Bagherinia, "Blind guidance using mobile computer vision: a usability study," in *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2010, pp. 241–242.

[7] S. Pietzsch, T. D. Vu, J. Burlet, O. Aycard, T. Hackbarth, N. Appenrodt, J. Dickmann, and B. Radig, "Results of a precrash application based on laser scanner and short-range radars," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, no. 4, pp. 584–593, 2009.

[8] S. Cherng, C.-Y. Fang, C.-P. Chen, and S.-W. Chen, "Critical motion detection of nearby moving vehicles in a vision-based driver-assistance system," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, no. 1, pp. 70–82, 2009.

[9] A. Chiuso, R. Brockett, and S. Soatto, "Optimal structure from motion: Local ambiguities and global estimates," *International journal of computer vision*, vol. 39, no. 3, pp. 195–228, 2000.

[10] F. Raudies and H. Neumann, "An efficient linear method for the estimation of ego-motion from optical flow," in *Pattern Recognition*. Springer, 2009, pp. 11–20.

[11] K. Pauwels and M. M. Van Hulle, "Optimal instantaneous rigid motion estimation insensitive to local minima," *Computer Vision and Image Understanding*, vol. 104, no. 1, pp. 77–86, 2006.

[12] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments," in *the 12th International Symposium on Experimental Robotics (ISER)*, vol. 20, 2010, pp. 22–25.

[13] Y. Tian and J. Tan, "A fast adaptive-gain orientation filter of inertial/magnetic data for human motion tracking in free-living environments," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 6760–6763.

[14] F. Raudies and H. Neumann, "A review and evaluation of methods estimating ego-motion," *Computer Vision and Image Understanding*, vol. 116, no. 5, pp. 606–633, 2012.