Lie Group-Based Multi-Target Tracking Method for Intelligent Vehicle Stereo Vision Using an Improved JPDA Filter (Postprint)

Authors: Zhang Qi, Hu Guangdi, Zhu Xiaoyuan, Chen Yadong

Date: 2018-09-12T00:00:00+00:00

Abstract

Reliable environment analysis constitutes a core technology for intelligent vehicle autonomous driving, encompassing vehicle recognition, pedestrian detection, collision avoidance, among others, with the ultimate objective of achieving full driving automation. This paper proposes a stereo vision multi-target tracking methodology for intelligent vehicles based on an improved joint probabilistic data association filter. Stereo vision cameras are employed to acquire images and videos of vehicles and pedestrians; sensor uncertainty is modeled within the Lie group framework, and the Euclidean group algorithm is utilized for state filtering of preprocessed images; binocular vision is leveraged within potential vehicle regions to eliminate false detections and obtain vehicle positional information; measurement uncertainty and predicted target motion trajectories are validated through a Kalman filter; and an improved joint probabilistic data association filter is applied to optimize and rectify the tracking results for both vehicles and pedestrians. Experimental results demonstrate that the proposed method effectively addresses the multi-target tracking challenges in intelligent vehicles, substantially enhancing the automation and intelligence capabilities of the driving system. Compared with other relatively recent target tracking approaches, the proposed method exhibits distinct advantages in both tracking accuracy and speed, without introducing significant deviation during vehicle tracking or omitting pedestrian tracking.

Full Text

Stereo Vision Multi-Object Tracking for Intelligent Vehicle Using Joint Probabilistic Data Association in Lie Group

Zhang Qi[1], Hu Guangdi[2]†, Zhu Xiaoyuan[3], Chen Yadong[1]
[1]College of Mechanical Engineering, [2]Auto Research Institute, [3]School of Infor-

mation Science & Technology
Southwest Jiaotong University, Chengdu 610031, China

Abstract

Reliable scene analysis, which encompasses vehicle identification, pedestrian detection, and collision avoidance, constitutes a core technology for realizing fully automated intelligent vehicle driving. This paper proposes a stereo vision multi-object tracking method for intelligent vehicles based on an improved joint probabilistic data association (JPDA) filter. The method captures images and videos of vehicles and pedestrians using stereo vision cameras, models sensor uncertainty within the Lie group framework, and employs Euclidean group algorithms for state filtering of preprocessed images. False detections are eliminated and vehicle position information obtained using binocular vision within potential vehicle regions. Measurement uncertainty and predicted target trajectories are validated through a Kalman filter, while tracking results for both vehicles and pedestrians are optimized using an improved JPDA filter.

Experimental results demonstrate that the proposed method effectively resolves multi-object tracking challenges for intelligent vehicles, substantially enhancing the automation and intelligence level of driving systems. Compared with other state-of-the-art target tracking methods, our approach exhibits distinct advantages in both tracking accuracy and speed, producing no significant offset when tracking vehicles and avoiding missed pedestrian tracking.

Keywords: intelligent driving system; stereo vision; moving object tracking; joint probabilistic data association; Lie group

0 Introduction

Achieving fully automated vehicle driving while improving safety and reducing human errors that lead to road accidents represents a paramount objective. Among various enabling technologies, moving object tracking is a critical task for driver assistance systems. When a vehicle can detect dynamic objects in its environment and predict their future behavior, its intelligence level can be substantially enhanced.

Reliable perception of the surrounding environment under various uncertain conditions is a fundamental requirement for nearly any assisted or autonomous system application, particularly with the rising research interest in autonomous driving. Both academia and major technology companies are actively developing advanced driver assistance systems. The core technologies of driver assistance systems include adaptive cruise control, collision avoidance, lane change assistance, traffic sign recognition, and parking assistance, all aimed at achieving complete driving automation.

Since real-time, accurate tracking of various objects under different environmental conditions is required, no single sensing system can currently provide all necessary information for target tracking. Consequently, driver assistance systems typically employ composite sensing systems that integrate millimeter-wave radar, laser rangefinders, vision systems, and other devices to achieve accurate detection of moving targets. Radar devices can precisely measure object relative velocity and distance. Laser rangefinders offer higher lateral resolution than radar and can not only accurately detect object distances but also determine object occupancy areas and provide detailed scene representations. Vision-based sensing systems can provide precise lateral measurements and rich image information, serving as an effective complement to range-based sensor road scene analysis. Among these, stereo vision sensors enable target detection with high lateral resolution and relatively small defined ranges while typically providing sufficient information for object recognition and classification.

Regardless of the sensor employed, multi-target tracking must be addressed in traffic scenarios. This requires tracking the state of each target while processing measurements in cluttered environments and solving the data association problem for tracked targets.

Previous research has explored various approaches to target tracking. Literature [4] exploited the characteristic that grayscale features of deforming targets change at the same frequency, achieving accurate detection of target regions and predecessor information based on soft feature extraction, thereby enabling long-term stable tracking of deforming targets. Literature [5] pre-selected a reference region in each video scene where target appearance could be clearly distinguished to form training samples for model construction, while employing two symmetric, weight-sharing deep convolutional neural networks to achieve accurate target tracking. Literature [6] described step-invariant features as visual quanta with independence and constraint based on the quantum frequency step invariance characteristics at the foreground-background junction of moving targets, improving tracking accuracy and speed. Literature [7] designed a two-stage aggregation-based target localization algorithm on the basis of a defined grid network model, achieving shortest-path target localization for moving targets and improving localization accuracy. Literature [8] addressed the unsupervised multi-moving target detection and tracking problem, achieving automatic detection and tracking of multiple moving targets using a method based on particle filtering and background subtraction. Literature [9] fused color and texture features of moving targets based on DST and PCR5, establishing an adaptive visual tracking model for multi-feature information fusion under complex scenes within a particle filtering framework. Literature [10] proposed a distributed tracking method based on gregariousness, achieving accurate tracking of heterogeneous multi-target systems while considering target inertia. Literature [11] proposed a metric for evaluating multi-target tracking algorithm performance—deficiency-aware subpattern assignment—providing important support for scientific evaluation and optimization of tracking algorithms. Literature [12] proposed a new region extraction method based on autonomous targets, achieving efficient segmentation of high-resolution long-range perception images and consequently precise tracking of moving targets. Literature [13] achieved reliable tracking of moving targets under conditions lacking precise data based on random finite sets. Literature [14] proposed a Bayesian tracking framework and achieved rapid target detection and tracking based on a belief propagation scheme and Markov chain. Literature [15] organically combined adaptive genetic multi-model with cardinal probability hypothesis density to achieve adaptive multi-target tracking.

Through careful examination of these research achievements, this paper proposes a multi-target detection and tracking method that: first, estimates the relative displacement of intermediate vehicles using a visual stereo ranging algorithm, treating objects that do not conform to the estimated motion as moving targets and sending their measurements to the tracking algorithm; second, constructs state uncertainty representation and motion models based on an extended Kalman filter in Lie groups; and finally, achieves accurate multi-target tracking through an improved joint probabilistic data association (JPDA) approach. To validate the effectiveness of the proposed method, moving target detection and tracking experiments were conducted on actual roads, with results demonstrating high tracking accuracy and favorable dynamic performance.

1 Moving Target Stereo Detection

Effective detection of moving objects is achieved through stereo imaging, though the continuous motion of the observer introduces considerable difficulty to the detection task. To accomplish moving object detection, fusion of sensor modality and object estimation is required. First, stereo-based moving object detection must be implemented. A corner detector is used to detect semi-dense feature points, with the position and velocity of each detected feature estimated in three-dimensional Euclidean space. Consequently, correspondence between features in the current and previous frames must be established for both left and right images. To this end, this paper employs optical flow methods and stereo block matching algorithms to compute feature correspondences.

Initially, after image rectification, all feature points from the previous frame are projected into the 3D world frame through a standard pinhole camera model. This position is then composited with the obtained motion matrix and back-projected into the current camera frame, connecting to corresponding 3D points from the previous frame to form a vector field, where each vector represents the motion of the corresponding 3D point relative to the world frame. Second, because measurement uncertainty in 3D space is highly anisotropic, accurately determining motion intensity along the optical axis direction becomes difficult. By projecting vectors onto the image plane where uncertainty is uniformly distributed and assigning thresholds to each point' s motion magnitude, remaining vectors are clustered based on translation and rotation parameters. Finally, if at least three vectors appear within a cluster, each cluster corresponds to a moving object, described by the centroid point of all corresponding points.

2 Mobile Target Motion Model

To establish a motion model for moving targets, it is necessary to first briefly introduce relevant knowledge regarding Lie groups and Euclidean groups used in the modeling process.

Definition 1. A Lie group GL is a differentiable group with a smooth manifold structure.

Definition 2. A Lie algebra g is an open neighborhood in the tangent space of GL at the identity matrix, where the matrix exponential $\exp(GL)$ and matrix logarithm $\log(GL)$ can form a local diffeomorphism, as shown in equation (1):

$$\begin{cases} \exp : g \to GL \\ \log : GL \to g \end{cases}$$

To effectively utilize Lie groups, it is necessary to introduce a concentrated Gaussian distribution, which involves connected unimodular Lie groups.

Definition 3. The probability density function of sample X is:

$$P(X) = \lambda \cdot \exp\left(-\frac{1}{2}\log(X)^T \cdot \log(X)\right)$$

where λ is a normalization constant.

At this point, X can be transformed into GL using Lie group transformations:

$$X = \exp(\log(X) \cdot \delta)$$

where δ is an allocation coefficient.

Definition 4. The Euclidean group S describes rigid body motion in two-dimensional space, expressed as:

$$S = \begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix}$$

where $a \in \mathbb{R}^{1\times 1}$ and $A \in \mathbb{R}^{2\times 2}$.

2.1 State Space

Vehicles are typical rigid bodies, necessitating the use of rigid body motion equations to describe their state. Furthermore, when considering vehicle velocity, higher-order state changes can also be represented through the same motion equations. Based on the equivalent principle of constant velocity rigid body motion models, the state space GL is constructed as the Cartesian product of two matrix Lie groups S:

$$GL = S \times S$$

where the first S represents the position component and the second S represents the velocity component.

Since the model must be applied to vehicle tracking applications, compared with other motion models (constant velocity, constant angular velocity and speed, constant curvature and speed), the motion model in equation (5) offers superior flexibility. This flexibility is manifested through velocities in three components—longitudinal, lateral, and rotational—providing the capability to describe objects such as vehicles, motorcycles, and pedestrians appearing in various scenarios.

Matrix composition and inversion in Lie groups are simply matrix multiplication and inversion. Therefore, for all operations on GL, notation can be constructed using two S blocks placed diagonally.

2.2 Motion Model

The vehicle motion model is:

$$X_{k+1} = \alpha_k \cdot X_k \cdot \beta_k$$

where $X_k \in GL$ is the system motion state at time k; α_k is a nonlinear function; and β_k is Gaussian white noise.

If the posterior distribution at step $k-1$ satisfies a Gaussian distribution on the Lie group, the motion state shown in equation (6) can be predicted according to:

$$X_{k+1|k} = \exp\left(\int_k^{k+1} \log(\alpha_\tau) d\tau\right) \cdot X_{k|k}$$

The motion equation shown in equation (6) can be remodeled as:

$$X_{k+1} = \exp\left(\begin{bmatrix} v_{1k} & v_{2k} & \omega_k \end{bmatrix}\right) \cdot X_k \cdot \beta_k$$

where v_{1k}, v_{2k}, and ω_k are the longitudinal, lateral, and rotational velocities, respectively; and β_{1k}, β_{2k}, and $\beta_{\omega k}$ are the Gaussian noise components in the longitudinal, lateral, and rotational directions.

3 Mobile Target Tracking

In this paper, we assume that multiple tracked targets are $\{T_1, \dots, T_k\}$, where the number of tracked targets k varies over time, meaning targets may appear or disappear from the sensor' s field of view at any time.

Define Y_k as the set of all detections at time k:

$$Y_k = \{y_{k,1}, \dots, y_{k,n_k}\}$$

Define $Y_{1:k}$ as the history of all measurements:

$$Y_{1:k} = \{Y_1, \dots, Y_k\}$$

In addition to initial target measurements, vector Y_k also contains random variables (clutter) satisfying a Poisson distribution. The problem to be solved is how to appropriately assign the received measurement sets to targets in motion trajectories and how to manage the appearance and disappearance of targets to be detected. The JPDA method addresses this by estimating the posterior density of each T_i.

Equation (11) indicates that the density of target state X_k^i and its existence χ_k^i are measurements of all Y_k and are related to k.

For the probability of target existence, a Markov chain model is used:

$$P(\chi_k^i = 1 | Y_{1:k-1}) = \rho \cdot P(\chi_{k-1}^i = 1 | Y_{1:k-1})$$

where ρ represents the probability that target i continues to exist at time k given that it existed at time $k-1$.

Based on the obtained measurements Y_k, the posterior density of target T_i scanned at time k can be inferred through the total probability formula:

$$P(X_k^i, \chi_k^i | Y_{1:k}) = \sum_{j=0}^{n_k} P(X_k^i, \chi_k^i, \theta_{k,j}^i | Y_{1:k})$$

where $\theta_{k,j}^i$ represents the posterior data association probability of object existence; $\eta_{k,j}^i$ represents the probability hypothesis.

The probability of detected target existence is:

$$P(\chi_k^i = 1 | Y_{1:k}) = \sum_{j=0}^{n_k} P(\chi_k^i = 1, \theta_{k,j}^i | Y_{1:k})$$

To calculate $\sigma_{k,j}^i$ and $\eta_{k,j}^i$, measurement-to-object association within the object set must be considered. At this point, assume θ_k^i consists of all feasible joint events F, where each trajectory has zero or one measurement, and each measurement is assigned to zero or one trajectory. Then:

$$P(\theta_{k,j}^i|Y_{1:k}) = \sum_{F\in\mathcal{F}} P(F|Y_{1:k}) \cdot \delta(\theta_{k,j}^i \in F)$$

The probability that T_i exists but is not detected by measurements within the cluster is:

$$P(\chi_k^i = 1, \theta_{k,0}^i|Y_{1:k}) = P(\chi_k^i = 1|Y_{1:k-1}) \cdot (1 - P_d) \cdot P_{FA}$$

To calculate $P(F|Y_{1:k})$ corresponding to each joint event F, it is necessary to assign a measurement set C to target set T and a track set D to measurement set C. Then:

$$P(F|Y_{1:k}) = \frac{1}{c} \cdot \prod_{i\in D} \left[P_d \cdot \frac{P(y_{k,j}|x_k^i)}{P_{FA}}\right] \cdot \prod_{i\notin D}(1 - P_d)$$

where P_d is the probability of detection; P_{FA} is the probability of false alarm; and P_{FA} is the prior clutter measurement density.

Based on equation (18), the rotation probability can be calculated:

$$P(\theta_{k,j}^i|Y_{1:k}) = \frac{P_d \cdot P_g \cdot P(\chi_k^i = 1|Y_{1:k-1})}{\sum_{l=0}^{n_k} P_d \cdot P_g \cdot P(\chi_k^i = 1|Y_{1:k-1})}$$

To satisfy the gating threshold, the rotation probability shown in equation (19) is normalized through P_d and P_g.

The above formulas provide all elements needed to determine target existence probability. The joint data probability can be further calculated as:

$$P(F|Y_{1:k}) = \prod_{i\in T} \left[P_d \cdot P_g\right]^{\tau_i(F)} \cdot \left[1 - P_d \cdot P_g\right]^{1-\tau_i(F)}$$

where $\tau_i(F)$ indicates whether target i is assigned a measurement in joint event F.

4 Experiments

4.1 Experimental Platform and Conditions

To verify the effectiveness of the proposed method, a sensor experimental platform was built in an urban environment for relevant experiments. The experimental platform primarily consists of two stereo camera systems. Stereo images were recorded using a monochrome camera system with 1.3-megapixel global shutter sensors, an image resolution of 1280$\times$960 pixels, and a 50° horizontal field of view. Since the expected target measurement range is 40 meters, experiments were conducted using a maximum 20 cm baseline at a maximum frame rate of 20 Hz. Stereo image synchronization is performed internally within the system, and experiments were recorded under the camera’ s automatic exposure mode.

Because the proposed method depends on asynchronous state estimation performed by two sensors, it is necessary to ensure clock synchronization between them. Although both sensors operate at similar frequencies, transformations may occur in practice. Consequently, the prediction step directly depends on the time period T, i.e., the time between two consecutive moments k and $k+1$. Therefore, clock drift or significant delays in data acquisition may substantially affect algorithm performance.

4.2 Comparison and Analysis

To fully verify algorithm adaptability, experimental tests were conducted containing multiple highly dynamic scenes, simultaneously tracking vehicles, trams, and pedestrians. In the experiments, once an object’ s existence probability exceeded 0.93, its trajectory was confirmed as a truly existing object; if the existence probability fell below 0.15, the trajectory was deleted.

The experiment duration was 90 seconds, involving not only moving vehicles proceeding straight and turning right but also including stationary vehicles, moving pedestrians, and stationary pedestrians. Experimental results are shown in Figures 1 [Figure 1: see original paper] through 4. Figures 1 and 3 display the tracking results from the left camera for the proposed method and the method from literature [13], respectively, while Figures 2 [Figure 2: see original paper] and 4 [Figure 4: see original paper] show the tracking results from the right camera for the two methods. Tables 1 and 2 present data comparisons of tracking accuracy and speed between the two methods.

During the experiments, after preprocessing the raw sensor data, an average of 2.12 stereo camera detections per frame were obtained, generating 305 trajectories. Comparative analysis of experimental results reveals that compared with the method proposed in literature [13], our method demonstrates obvious advantages in both tracking accuracy and speed, does not miss pedestrian tracking, and produces no significant offset when tracking vehicles. Moreover, the experimental results from the left and right cameras are consistent, enabling accurate detection of different targets in the same scene; when targets move continuously, the proposed method can also achieve accurate tracking of both.

4.3 Uncertainty Testing

Figures 5 [Figure 5: see original paper] and 6 [Figure 6: see original paper] display the results of LG-EKF filter state uncertainty updates for two different sensor types. In the experiments, filter predictions follow the S motion model, producing banana-shaped state uncertainty. Figure 5 shows the results of updating the filter with an elliptical measurement uncertainty sensor, similar to "classical" Gaussian uncertainty. Figure 6 [Figure 6: see original paper] shows the results based on stereo vision sensor range, where orientation has greater uncertainty. Note that the predicted uncertainty is biased to the right, indicating a higher probability of the vehicle turning right than left. These examples demonstrate how the filter processes different measurement uncertainties and effectively fuses them with information from the prediction step.

Experimental results show that the proposed method enables assisted driving systems to achieve precise tracking of moving objects in complex environments with stereo vision assistance. However, since available datasets for these sensors do not contain ground truth data, quantitative practical evaluation of the proposed method is difficult to ensure. Nevertheless, through the proposed experiments, the method incorporating two fundamental multi-target tracking building blocks—state estimation and probabilistic data association schemes, both based on the geometric structure of Lie groups—has been validated.

From an estimation perspective, the proposed method' s advantage lies in its flexible modeling of sensor and tracking object uncertainties and motion. This is beneficial for predicting future object motion and its uncertainty, as well as for implementing collision avoidance and autonomous vehicle motion planning functions. Additionally, the stereo camera detection process does not depend on object-specific appearance and can detect arbitrary motion, including movements of cars and pedestrians as shown in Figures 1-4.

5 Conclusion

Addressing the tracking problem in the context of driver assistance systems, this paper proposes a method for estimating the relative displacement of vehicles based on stereo camera ranging, generating measurement results from clustering centers of optical flow vectors that do not conform to estimated motion. Due to frequency synchronization issues between the two cameras, an asynchronous Kalman filter on Lie groups is used to fuse their detection information. The proposed moving target representation enables accurate description of sensor measurements and vehicle state uncertainty. To solve the multi-target tracking problem, the JPDA filter is improved to work collaboratively with the Kalman filter on Lie groups. Real data recorded in urban traffic scenes were used to test the proposed method, and experimental results prove that the method can achieve precise and rapid detection of moving targets, thereby enhancing the intelligence level of driver assistance systems.

In this paper, detection and tracking problems have been addressed in the context of advanced driver assistance systems. However, statistical analysis of vehicle and pedestrian motion trajectories has not been considered. Future work could investigate this aspect to enable prediction of vehicle and pedestrian motion trends.

References

[1] Han Feilong, Ying Jie, Zhu Dandan. A new vehicle aided driving dynamic obstacle detection and classification method [J]. Application Research of Computers, 2017, 34(6): 1909-1912.

[2] Zhai Gen, Wu Cui, Wang Yu. Target tracking based on millimeter wave radar in complex scenes [J]. International Journal of Performability Engineering, 2018, 14(2): 232-244.

[3] Dekan M, Duchoň F, Babinec A, et al. Moving obstacles detection based on laser range finder measurements [J]. International Journal of Advanced Robotic Systems, 2018, 15(1): 1-18.

[4] Jiang Wentao, Liu Wanjun, Yuan Heng. Target tracking based on soft feature theory [J]. Journal of Computer Science, 2016, 39(7): 1334-1355.

[5] Gao Junyu, Yang Xiaoshan, Zhang Tianzhu, et al. Robust visual tracking method based on depth learning [J]. Journal of Computer Science, 2016, 39(7): 1419-1434.

[6] Jiang Wentao, Liu Wanjun, Yuan Heng, et al. Visual quantum target tracking method [J]. Software Journal, 2016, 27(11): 2961-2984.

[7] Ren Qianqian, Liu Hongyang, Liu Yong, et al. Wireless sensor network based on 2 stage aggregation of target tracking algorithm [J]. Computer Research And Development, 2017, 54(9): 2001-2010.

[8] Li Mingjie, Liu Xiaofei, Zhang Fuquan, et al. Multi-target detection and tracking algorithm based on particle filter and background subtraction [J]. Application Research of Computers, 2018, 32(6): 134-142.

[9] Wang Yi, Fang Yingwu. Multi-target adaptive visual tracking method based on DST-PCR [J]. Application Research of Computers, 2018, 35(2): 632-636.

[10] Shen Chen, He Pei, Qin Lai, et al. Multitarget tracking control for coupled heterogeneous inertial agents systems based on flocking behavior [J]. IEEE Trans on Systems Man & Cybernetics Systems, 2018, 14(9): 1-7.

[11] Oksuz K, Cemgil A T. Multitarget tracking performance metric: deficiency aware subpattern assignment [J]. IET Radar, Sonar & Navigation, 2018, 12(3): 373-381.

[12] Tan Wu, Luo Jin, Zhou Xue, et al. Automatic newly increased built-up area extraction from high-resolution remote sensing images using line-density-based visual saliency and PanTex [J]. Journal of Applied Remote Sensing, 2018, 12(1): 1-12.

[13] Cano P, Ruiz-Del-Solar J. Robust tracking of multiple soccer robots using random finite sets [J]. IEEE Intelligent Systems, 2018, 32(6): 22-29.

[14] Meyer F, Braca P, Hlawatsch F, et al. Scalable adaptive multitarget tracking using multiple sensors [C]//Proc of IEEE Globecom Workshops. 2017: 1-6.

[15] Li Bing, Zhao Jin, Pang Fang. Adaptive genetic MM-CPHD filter for multitarget tracking [J]. Soft Computing, 2017, 21(16): 4755-4767.

Note: Figure translations are in progress. See original paper for figures.