2011 14th International IEEE Conference on
Intelligent Transportation Systems
Washington, DC, USA. October 5-7, 2011

Combining Monocular and Stereo-Vision for Real-Time Vehicle Ranging and Tracking on Multilane Highways

Sayanan Sivaraman and Mohan M. Trivedi

Abstract—**In this paper, we introduce a novel stereo-monocular fusion approach to on-road localization and tracking of vehicles. Utilizing a calibrated stereo-vision rig, the proposed approach combines monocular detection with stereo-vision for on-road vehicle localization and tracking for driver assistance. The system initially acquires synchronized monocular frames and calculates depth maps from the stereo rig. The system then detects vehicles in the image plane using an active learning-based monocular vision approach. Using the image coordinates of detected vehicles, the system then localizes the vehicles in real-world coordinates using the calculated depth map. The vehicles are tracked both in the image plane, and in real-world coordinates, fusing information from both the monocular and stereo modalities. Vehicles' states are estimated and tracked using Kalman filtering. Quantitative analysis of tracks is provided. The full system takes 46ms to process a single frame.**

Index Terms - **Active Safety, Driver Assistance, Real-time Vision, Multi-sensor Fusion, Machine Learning.**

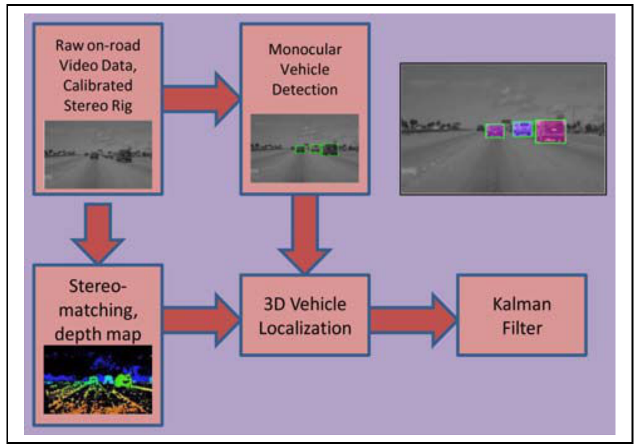


Fig. 1. Overview of the approach detailed in this paper. On-road video frames are grabbed using a calibrated stereo rig. A vehicle detector is applied to the monocular frame. Detected vehicles are then located in the depth map, and 3D coordinates are determined for each vehicle. Vehicles are tracked in both the image plane and real-world coordinates using Kalman filtering. The vehicles' colors are determined by their longitudinal distance, with respect to the ego vehicle. Tracked vehicles are displayed using a color gradient to indicated longitudinal distance, ranging between red for close vehicles, and blue for far away vehicles.

I. INTRODUCTION

The World Health Organization estimates that 1.2 million people are killed worldwide in road crashes each year, and a further 50 million are injured. The total economic cost associated with traffic accidents is estimated at a staggering $518 billion US per year. Stemming the tide requires significant actions by policy makers, auto manufacturers, and researchers [32].

In recent years, there has been an active research community devoted to developing the next generation of driver assistance systems. These include research associated with lane tracking [17], [23], [24], pedestrian detection [9], bicyclist detection [6], vehicle detection [28], and driver monitoring [8], [18].

Robust detection and tracking of other vehicles on the road using vision is a challenging problem. Roads are dynamic environments, with ever-changing backgrounds and illuminations. The ego vehicle and the other vehicles on the road are generally in motion, so the sizes and locations of vehicles in the image plane are diverse. Vehicles encountered vary widely in terms of their shapes, relative sizes, colors, and appearances [26].

Recent studies in on-road vehicle detection and tracking mainly focused on stereo based techniques [19], [4], and monocular techniques [13], [27], [22]. Stereo-based techniques have often used tracking of point clouds in 3D space to segment, detect, and track vehicles [4], [19]. Monocular approaches, by contrast, have generally utilized trained classifiers for vehicle detection and tracking in the image plane [13], [27], [22].

Stereo techniques feature the advantage of explicit computation of depth and location in real-world coordinates, but require additional specialized hardware, precise calibration, and additional computational cost. Monocular techniques have featured advantages of lower cost at runtime and established machine learning paradigms, but they do not immediately provide a 3D location for detected vehicles. While many prior studies have pursued a purely monocular approach, or a purely stereo-vision approach, few studies have tried to fuse information from both domains.

In this study, we introduce a novel approach to on-road vehicle localization and tracking using a combination of stereo and monocular vision. Using a calibrated stereo rig, on-road frames are grabbed and the depth map is computed. A monocular vehicle detector, based on prior work in [22] is applied to the monocular frame. Detected vehicles are then localized in 3D space by accessing the pre-computed depth map. The approach introduced in this paper fuses information from the monocular and stereo domains, treating them as complimentary modalities. The vehicles are tracked in both the image plane and in real-world coordinates. The tracked vehicles' states are estimated by means of Kalman filtering. The full system processes a given frame in 46ms.

The remainder of this paper is structured as follows. The following section details recent works related to on-road vehicle detection and tracking. Section 3 details the stereo-monocular fusion approach for on-road vehicle detection and tracking. Section 4 provides experimental results. Section 5 provides concluding remarks.

II. Related Research

In recent years, vehicle detection and tracking have been widely explored in the literature. The research community is well-represented in both stereo-vision and monocular approaches to vehicle detection.

A. Vehicle Detection and Tracking Using Stereo-Vision

Stereo-vision provides depth and relevant 3D information. As such, many stereo-vision works detect and track vehicles using geometric models, segmentation methods, and temporal filtering.

In [3], after calculating depth and optical flow, clustering is used to separate static from moving objects. Then, a modified Iterative Closest Point algorithm, using polar distance metric is used to fit a cuboid to the estimated vehicle. The system uses this methodology for estimating vehicle pose, and does not require temporal filtering.

In [4], the problem of tracking oncoming vehicles is addressed. In particular, as dangerous situations can arise when the oncoming vehicle turns, the paper addresses distinguishing between oncoming and turning vehicles. The proposed approach uses Interacting Multiple Models to address both the oncoming and turning modes. The respective motion models are estimated using temporal filtering. Comparison between the proposed IMM approach and standard single-mode tracking is provided.

In the recent work of [19], 3D position, velocity, and orientation information are provided by the fusion of depth maps with optical flow. The proposed methodology is shown to apply to detection and tracking of vehicles, pedestrians. The methodology provides a general approach to separating moving objects from static objects using stereo-vision.

In [14], vehicle detection and tracking are achieved by means of depth calculation and a two-stage mean-shift algorithm. The trajectories of objects are predicted using the object's history, relying on learned object trajectories using particle motion pattern.

B. Vehicle Detection and Tracking Using Monocular Vision

Many monocular approaches to vehicle detection are based in machine learning and pattern recognition. While most monocular approaches to vehicle detection have used discriminative classifiers, generative models have been used. In [5], a statistical model based on vertical and horizontal edge features was used for vehicle detection. The detected vehicles are then tracked using particle filtering.

Discriminative classifiers have been more common for vehicle detection in the literature. In [20], vehicle are detected using Haar-like rectangular features, and a cascade classifier using Adaboost, as was introduced in [31] for face detection. A similar formulation in used in [13], with enhanced feature sets. In [7], a vehicle detector using Haar-like features and Adaboost classification is also used. The detected vehicles are then tracked using a modified version of Lucas-Kanade tracking.

In [27], a deformable parts-based model is utilized for vehicle detection. Histogram of oriented gradient features are extracted at multiple scales, for the root vehicle, and parts such as tires. The Latent Support Vector Machine classifier is used for training the detector, based on the object detection work of [12]. The vehicles and parts are tracked using particle filtering.

C. Fusing Monocular and Stereo-vision

In [28], stereo and monocular vision were combined to detect and track the preceding vehicle in the ego lane. Stereo-vision based scene segmentation was used to identify vertical edges in 3D space. Then, monocular vision is used for detecting the preceding vehicle, using symmetry operators and imposing bounding box aspect ratio constraints. Tracking is achieved using the cross-correlation of vehicles across frames.

In [9], a combination of stereo and monocular cues are used for pedestrian detection. Histogram of Oriented Gradients extraction is performed in the monocular image plane, the optical flow motion image, and the depth image, and SVM classification is used. In particular, this approach was shown to be robust to occlusions, which are common in urban scenes.

The approach presented in this paper differs from that in [9] in the nature of the stereo-monocular fusion. In this study, vehicles are detected using the monocular modality, and then tracked in both the image plane and 3D coordinates.

III. On-road Vehicle Detection and 3D Tracking Using Stereo-Monocular Fusion

A. Experimental Testbed and Data Acquisition

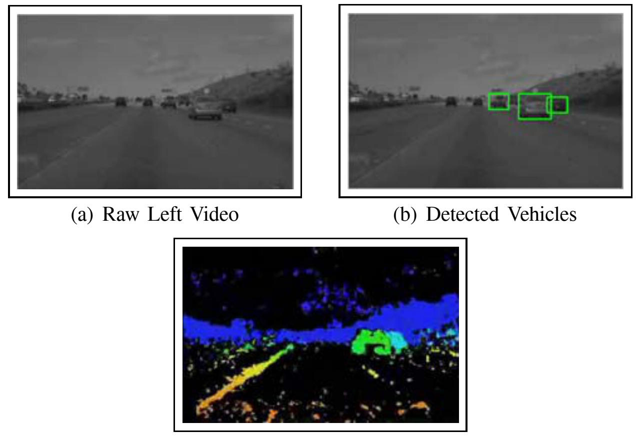
For this paper, data has been captured using a calibrated stereo rig, mounted on a vehicle platform, looking forward. Stereo matching is implemented using standard techniques, using commercially-available hardware made by Tyzx [30], [29]. A similar stereo setup has been used in [19]. Video is captured at an image resolution of 500×312. The implemented system operates on the left monocular image from the stereo rig, as well as the depth image that results from stereo matching.

B. Active Learning-Based Monocular Vehicle Detection

For detecting vehicles, we utilize monocular image data from the left camera of the stereo rig. We apply a monocular vehicle detector, trained using an active learning framework. This vehicle detector was part of the overall vehicle tracking system reported in [22].

Active learning for object detection has gained in popularity [10],[22], mainly for its ability to improve classifier performance by reducing false alarms [22], to increase a classifier's recall [21], to semi-automatically generate more training data [10]. A comparative survey of active learning for vehicle detection can be found in [25].

In general, batch active learning for object detection consists of two main stages: an initialization stage, and a stage of query and retraining [16]. Firstly, an initial classifier is trained



(a) Raw Left Video (b) Detected Vehicles

(c) Depth Image

Fig. 2. Noisy depth measurements in stereo-vision. a) Raw left monocular frame. b) Detected vehicles using monocular vehicle detection. c) Depth Image

using conventional supervised learning of a positive target class, and a negative class. Then, the initial classifier is used in conjunction with unlabeled data to update and improve classification performance. This process will generally add cost in terms of data and human labeling effort.

The vehicle detector in [22] was initialized using a labeled corpus of training examples. The resulting classifier was evaluated on independent, unlabeled on-road video data. A researcher used an interface for quick and efficient query and archival of informative independent training examples from the unlabeled data, with the learning methodology seeking inclusion of missed vehicle detections and false positives. After the second batch of learning, the resulting detector showed significant improvements in recall and precision. [22]

For the task of identifying vehicles, a boosted cascade of simple Haar-like rectangular features has been used, as was introduced by Viola and Jones [31] in the context of face detection. Various studies have incorporated this feature and classifier pairing for on-road vehicle detection systems [13],[20]. The set of Haar-like features is sensitive to edges, bars, vertical and horizontal details, and symmetric structures [31]. The resulting extracted values are then classified by Adaboost [11]. Modern implementations of the algorithm run at real-time speeds.

$$v_k = \begin{bmatrix} i_k & j_k & w_k & h_k \end{bmatrix}^T \tag{1}$$

Evaluating the vehicle detector on a given frame returns a list of bounding boxes. We denote one bounding box corresponding to a given detected vehicle as v_k, where k is a time-index. Bounding box v_k is parametrized by its $i-j$ pixel coordinate, and the width and height of the box, as given in equation 1.

C. Estimating 3D Coordinates Using Stereo-Vision

The calibrated stereo rig is aligned such that the depth image and the left image share the same coordinate system. This allows us to use detections from the monocular vehicle detector to estimate the distance to a given vehicle. Inferring depth from the stereo image is not entirely straight-forward, as the main error in stereo measurements is the depth component [2].

Figure 2(a)-2(c) demonstrate this difficulty. Figure 2(a) shows the raw video frame. Figure 2(b) shows monocular detection results on the frame. Figure 2(c) shows the depth image for this frame. We note that there are outliers and noise throughout the depth image. While one would expect that the rear face of a vehicle would have roughly constant longitudinal distance from the ego-vehicle, it's evident from the closest vehicle, shown in bright green in 2(c), that the raw stereo measurements do not deliver this.

There is a rich literature on improving stereo depth measurements. Major contributions have explored improved stereo matching algorithms [15], and the use of temporal filtering and smoothing of the depth map [2]. In this study, we calculate the depth measurement by spatial smoothing. For a given detected vehicle v_k, we take the sample mean depth of a given bounding box from the depth image. Equation 2 shows this calculation, where Z_k is the depth measurement, and D is the depth image.

$$Z_k = median\ (D(i,j)), \forall\ i,j\ \in v_k \tag{2}$$

Taking the median over the bounding box v_k improves the depth measurement for detected vehicles [9], but does not require us to smooth depth calculations over the entire image.

We then proceed to calculate X_k, and Y_k, the lateral and vertical positions of the vehicle respectively. We use the following equations, using the $i-j$ pixel coordinates of a given detected vehicle's bounding box, v_k. We first find the centroid of the rectangle, v_k, and then apply equation 3 to solve for X_k, and Y_k. The variables c_i and c_j are the center pixel coordinates in the i and j directions, respectively. Λ_x and Λ_y are scaling constants, intrinsic to the camera.

$$\begin{aligned} X_k &= (i_k + \frac{1}{2}w_k - c_i)\Lambda_x Z_k \\ Y_k &= (j_k + \frac{1}{2}h_k - c_j)\Lambda_y Z_k \end{aligned} \tag{3}$$

The complete measurement for a given detected vehicle, for a given time instant k is then obtained by concatenating the monocular bounding box v_k and the vehicle's full 3D position, as solved in equations 2 and 3 [30].

D. Tracking Vehicle State Using Kalman Filtering

While prior works have tracked vehicles in 3D world coordinates [4] or image coordinates [5], in this study, we track each vehicle across both the image plane and the 3D world. Each vehicle is tracked using a single Kalman filter. The measurement is a combination of their detected bounding boxes, as parametrized by v_k and the full 3D coordinates X_k, Y_k, Zk.

$$V_k = \begin{bmatrix} i_k & j_k & w_k & h_k & Xk & Y_k & Z_k & \Delta X_k & \Delta Y_k & \Delta Z_k \end{bmatrix}^T \tag{4}$$

The full state vector also includes the differences in 3D. The full state-space system is given in equation 5, where η_k and

ξ_k are the plant and observation noise, respectively. V_k is the full state of the tracked vehicle, and M_k is the measurement taken each frame, as detailed in equations 1- 4.

$$\begin{aligned} V_{k+1} &= AV_k + \eta_k \\ M_k &= CV_k + \xi_k \end{aligned} \tag{5}$$

The variables η_k and ξ_k are the plant and observation noise, respectively. The state transition matrix A and the observation matrix C are given below.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$
$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

The tracking formulation presented estimates the vehicle's state in both image and 3D coordinates. This tracking formulation incorporates the measurements from both sources of information for the system. Tracking in the image plane alone would ignore 3D measurements coming from stereo-vision. Tracking purely in 3D coordinates would belie the fact that our initial measurements come from the image plane. Given that our depth measurements are dependent on the estimated bounding box v_k and equation 2, it works to the system's advantage to track the state of the vehicle in both modalities.

IV. Experimental Evaluation

The presented system has been evaluated on real-world on-road data, captured using the calibrated vehicle-mounted stereo-rig. For a given stereo pair frame, stereo matching is performed and the depth map is generated. Then, the vehicle detector is evaluated over the left camera's frame. We use equations 1-3, to solve for a given vehicles' bounding box in image coordinates, and the vehicle's full 3D position. These seven parameters comprise the measurement vector for a given vehicle in a given frame. We then use equations 4-5 to estimate the full state of the vehicle using Kalman filtering.

We quantify the performance of the system on a highway sequence consisting of 1500 frames, captured at 25 frames per second. The sequence contains 5457 vehicles to be detected. Figure IV plots the true positive rate vs. false positives per frame, comparing the system to that presented in [22]. We note that tracking in both the stereo and monocular domain yields improvement in performance. This is because the tracker searches for samples in both the stereo and monocular domain; if a track is dropped in the image plane due to missed detection, the search in the stereo domain will maintain this track.

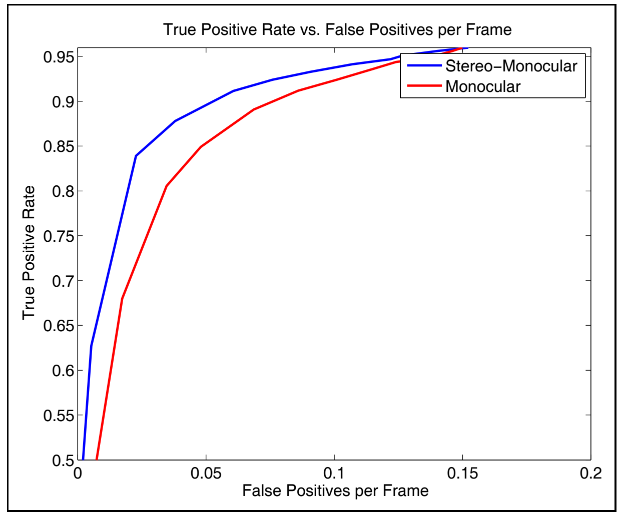


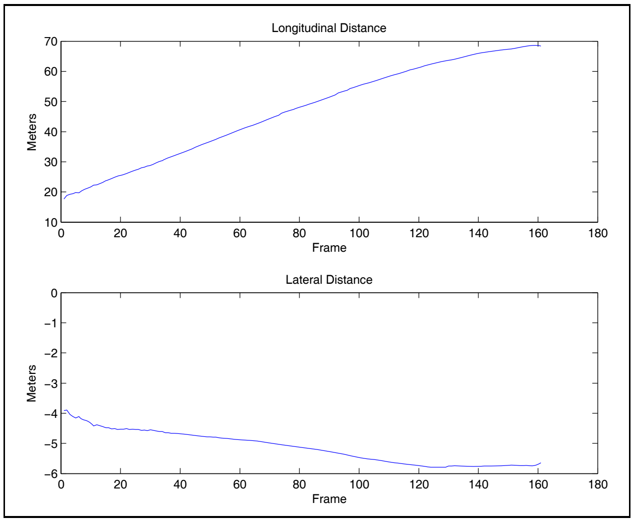
Fig. 3. True positive rate vs. false positives per frame, stereo-monocular tracker compared with the system from [22]

In addition to sample per-frame outputs, we provide the trajectories of three vehicles that were tracked.

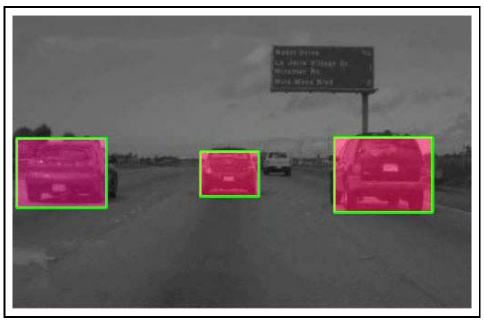
Figure 4(a) plots the Xk and Zk of a given vehicle, respectively, over the course of 120 frames. In this case, the indexing variable k corresponds to the frame number over which we track the vehicle. This is a vehicle that quickly overtook the ego vehicle. Figures 4(b)-4(d) show the vehicle's progression while it was in the system's tracking range. In figure 4(b), the tracked vehicle is the closest vehicle in the left lane, colored red. In figure 4(c) the tracked vehicle is still in the left lane, and is colored purple. At this point, it's midway through its trajectory, and Z_k is roughly 25 meters. In figure 4(d), the vehicle as at the end of its trajectory, about 40 meters away, and colored blue.

Figure 5(a) plots the trajectory of a tracked vehicle in the right lane over a period of 550 frames. The vehicle remained in the system's field of view for quite a while. Initially, the ego vehicle approached it from behind, at which point the two vehicles maintained relatively close relative distance. Figure 5(b) shows the vehicle early in its trajectory, far from the ego vehicle. It is located in the right lane, and is colored blue. Figure 5(c) shows the vehicle roughly a third through its trajectory, in the right lane, colored red. Figure 5(d) shows the vehicle towards the end of the plotted trajectory.

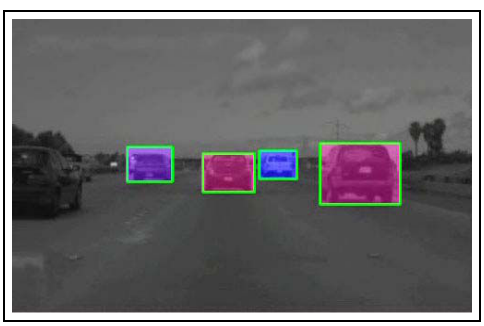
The fully implemented system processes a single frame in 46ms. This time includes stereo matching, vehicle detection, 3D localization, and Kalman tracking of multiple vehicles. The system has been implemented using an Intel Core Duo 2.4GHz architecture. As noted earlier, the longitudinal range is roughly 40m.



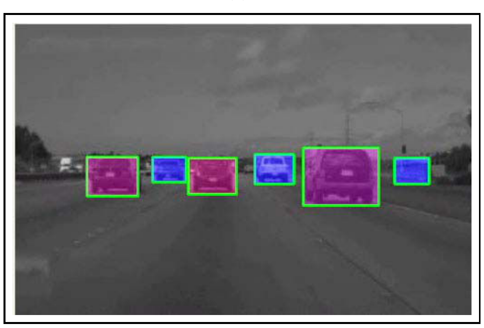
(a)



(b)

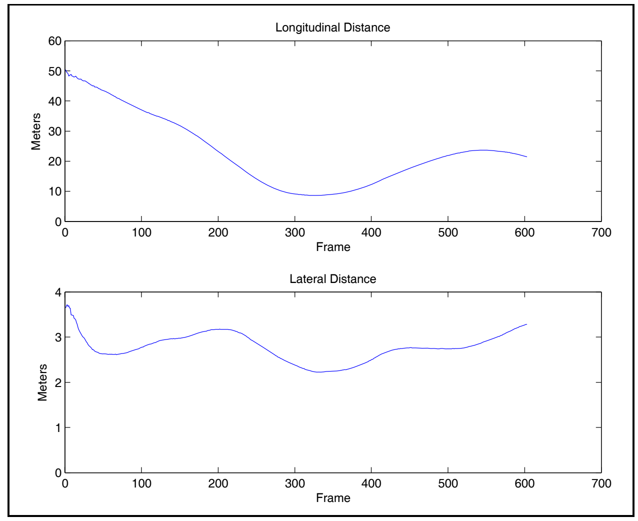


(c)



(d)

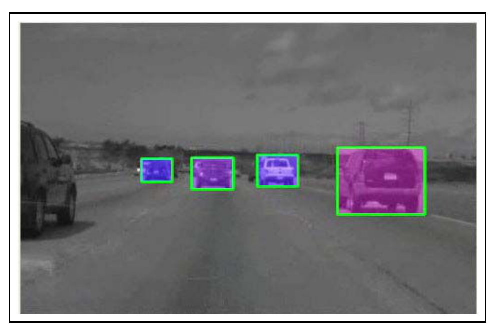
Fig. 4. a) Sample tracked vehicle trajectory over time. The vehicle entered left of the ego lane and overtook the ego vehicle with high relative velocity. b) Snapshots of vehicle trajectory plotted in Figure 4(a). 4(b) The tracked vehicle is the closest vehicle in the left lane, colored red. 4(c) The tracked vehicle is still in the left lane, and is colored purple. 4(d) The vehicle as at the end of its trajectory, and colored blue.



(a)

(b)

(c)



(d)

Fig. 5. a) Sample tracked vehicle trajectory over time. The ego vehicle caught up to this vehicle, which was in the right lane. The vehicle and the ego-vehicle maintained small relative speed for a while. b) Snapshots of the tracked vehicle, whose trajectory is plotted in 5(a). 5(b) The tracked vehicle far from the ego vehicle, in the right lane, colored blue. 5(c) The tracked vehicle roughly a third through its trajectory, in the right lane, colored red. 5(d) Tracked vehicle in the right lane, colored red.

V. Conclusion and Future Works

In this paper, we have introduced a novel approach to vehicle localization and tracking, combining stereo-vision cues with monocular vehicle detection. The contribution of this work entails fusing information from the monocular modality for vehicle detection with information from stereo-vision for 3D localization, and tracking the tandem measurements as a single state. Vehicles are initially detected in the monocular frame using a robust vehicle detector that was trained using active learning[22]. Using the detection results, the vehicle's 3D location is solved using stereo-vision. The full vehicle state is tracked both in the image plane, and in 3D coordinates using Kalman filtering. The fully deployed system processes a single frame in 46ms, including stereo matching, vehicle detection, vehicle localization, and tracking. Future areas of research include learning and modeling of vehicle trajectories, and fusion with other modalities.

VI. ACKNOWLEDGMENTS

The authors would like to thank colleagues Dr. B. Morris, Dr. A. Doshi, Mr. Cuong Tran, Mr. Ashish Tawari, Mr. Derick Johnson, and Mr. Jacoby Larson. This work was supported in part by Audi and Volkswagen Electronics Research Lab, Palo Alto, CA.

References

[1] B, Aytekin and E. Altug, Increasing driving safety with a multiple vehicle detection and tracking system using ongoing vehicle shadow information, *IEEE Systems Man and Cybernetics Conference* , 2010.

[2] H. Badino, R. Mester, J. Wolfgang, U. Franke, Free Space Computation using Stochastic Occupancy Grids and Dynamic Programming, *ICCV Workshop on Dynamical Vision*, 2007.

[3] B. Barrois, S. Hristova, C. Wohler, F. Kummert, C. Hermes, 3D pose estimation of vehicles using a stereo camera, *IEEE Intell. Veh. Symp.*, 2009.

[4] A. Barth and U. Franke, Tracking Oncoming and Turning Vehicles at Intersections, *Intell. Transp. Syst. Conf.*, 2010.

[5] Y. Chan, S. Huang, L. Fu, P. Hsiao, Vehicle Detection under Various Lighting Conditions by Incorporating Particle Filter, *IEEE Intell. Transp. Syst. Conf.*, 2007.

[6] H. Cho, P. E. Rybski, W. Zhang; , Vision-based bicyclist detection and tracking for intelligent vehicles, *IEEE Intell. Veh. Symp.*, 2010.

[7] J. Cui, F Liu, Z. Li, Z. Jia, Vehicle Localisation Using a Single Camera, *IEEE Intell. Veh. Symp.*, 2010.

[8] A. Doshi and M. M. Trivedi, On the Roles of Eye Gaze and Head Pose in Predicting Driver's Intent to Change Lanes, *IEEE Trans. on Intell. Transp. Syst.*, 2009.

[9] M. Enzweiler, A. Eigenstetter, B. Schiele, D. M. Gavrila, Multi-Cue Pedestrian Classification with Partial Occlusion Handling, *IEEE Comp. Vis. Patt. Recog.*, 2010.

[10] M. Enzweiler and D.M. Gavrila, A Mixed Generative-Discriminative Framework for Pedestrian Classification, *IEEE Conf. Comp. Vis. Pattern Recog.*, 2008.

[11] Y. Freund and R. E. Schapire, A Short Introduction to Boosting, *J. Japanese Soc. Artif. Intell.*, 14(5):771-780, September, 1999.

[12] P. Felzenszwalb, D. McAllester, D. Ramanan, A Discriminatively Trained, Multiscale, Deformable Part Model, *IEEE Comp. Vis. Patt. Recog.*, 2008.

[13] A. Haselhoff, A. Kummert, A Vehicle Detection System Based on Haar and Triangle Features, *IEEE Intell. Veh. Symp.*, 3-5 June 2009.

[14] C. Hermes, J. Einhaus, M. Hahn, C. Wohler, F. Kummert, Vehicle Tracking and Motion Prediction in Complex Urban Scenarios, *IEEE Intell. Veh. Symp.*, 2010.

[15] H. Hirschmuller, Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information, *IEEE Comp. Vis. Patt. Recog.* , 2005.

[16] M. Li and I. K. Sethi, Confidence-Based Active Learning, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1251-1261, August, 2006.

[17] J. McCall, M. M. Trivedi, Video Based Lane Estimation and Tracking for Driver Assistance: Survey, System, and Evaluation, *IEEE Trans. Intell. Transp. Syst.*, vol. 7, No. 1, Mar. 2006.

[18] J McCall, D Wipf, M. M. Trivedi, B. Rao, Lane Change Intent Analysis Using Robust Operators and Sparse Bayesian Learning, *IEEE Trans. on Intell. Transp. Syst*, Sept 2007.

[19] C. Pantilie and S. Nedevschi, Real-time Obstacle Detection in Complex Scenarios Using Dense Stereo Vision and Optical Flow, *IEEE Intell. Transp. Syst. Conf.*, 2010.

[20] D. Ponsa, A. Lopez, F. Lumbreras, J. Serrat, T. Graf, 3D Vehicle Sensor Based on Monocular Vision, *IEEE Intell. Transp. Syst. Conf.*, pp. 1096-1101, Sept. 2005.

[21] P. M. Roth and H. Bischof, Active Sampling Via Tracking, *IEEE Conf. Comp. Vis. Pattern Recog. Workshops* , 2008.

[22] S. Sivaraman and M. M. Trivedi, A General Active Learning Framework for On-road Vehicle Recognition and Tracking, *IEEE Trans. Intell Transp. Syst.*, 2010.

[23] S. Sivaraman and M. M. Trivedi, Improved Vision-Based Lane Tracker Performance Using Vehicle Localization, *IEEE Intell. Vehicle Symp.*, June 2010,

[24] S. Sivaraman and M. M. Trivedi, Integrated Lane and Vehicle Detection and Tracking for Driver Assistance: a Synergistic Approach, *IEEE Trans. Intell. Transp. Syst*, *to appear*.

[25] S. Sivaraman and M. M. Trivedi, A Comparative Study of Active Learning for On-Road Vehicle Detection, *Machine Vision and Applications*, *to appear*.

[26] Z. Sun, G. Bebis, R Miller, On-road Vehicle Detection: A Review, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006.

[27] A. Takeuchi, S. Mita, and D. McAllester, On-road Vehicle Tracking Using Deformable Object Model and Particle Filter with Integrated Likelihoods, *IEEE Intell. Vehicle Symp.*, 2010.

[28] G. Toulminet, M. Bertozzi, S. Mousset, A. Bensrhair, A. Broggi, Vehicle Detection by Means of Stereo Vision-based Obstacles Features Fxtraction and Monocular Pattern Analysis, *IEEE Trans. Image Proc.* , Aug. 2006.

[29] J. I. Woodlill,R. Buck, D. Jurasek, G. Gordon, T. Brown , 3D Vision: Developing an Embedded Stereo-Vision System, *IEEE Computer*, vol.40, no.5, pp.106-108, May 2007

[30] Tyzx Dense Stereo, http://www.tyzx.com, accessed Jan 27, 2011

[31] P. Viola and M. Jones, Rapid Object Detection Using a Boosted Cascade of Simple Features, *IEEE Computer Vision and Pattern Recognition Conference* , vol. 1, pp. 511-518, Dec 2001.

[32] World Health Organization, World report on road traffic injury prevention, http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/summary_en_rev.pdf, accessed Jan 27, 2011.