What Is in Front? Multiple-object Detection and Tracking with Dynamic Occlusion Handling

Junli Tao[1(✉)], Markus Enzweiler[2], Uwe Franke[2], David Pfeiffer[2], and Reinhard Klette[3]

[1] Department of Computer Science, The University of Auckland, Auckland, New Zealand
jtao076@aucklanduni.ac.nz
[2] Image Understanding, Daimler AG, Boeblingen, Germany
[3] Auckland University of Technology, Auckland, New Zealand

Abstract. This paper proposes a multiple-object detection and tracking method that explicitly handles dynamic occlusions. A context-based multiple-cue detector is proposed to detect occluded vehicles (occludees). First, we detect and track fully-visible vehicles (occluders). Occludee detection adopts those occluders as priors. Two classifiers for partially-visible vehicles are trained to use appearance cues. Disparity is adopted to further constrain the occludee locations. A detected occludee is then tracked by a Kalman-based tracking-by-detection method. As dynamic occlusions lead to role changes for occluder or occludee, an integrative module is introduced for possibly switching occludee and occluder trackers. The proposed system was tested on overtaking scenarios. It improved an occluder-only tracking system by over 10% regarding the frame-based detection rate, and by over 20% regarding the trajectory detection rate. The occludees are detected and tracked in the proposed method up to 7 seconds before they are picked up by occluder-only method.

1 Introduction

Multiple-object detection and tracking is a main subject in computer vision. Examples of considered objects are pedestrians [9] or vehicles [15]. Tracking-by-detection methods are developed for, e.g., surveillance, robotics, or autonomous driving [1,9,15]. These methods are mainly focusing on data association[1,7,21]. Occlusions pose difficulties for data association due to appearance changes. Because of occlusions, detection results (i.e. bounding boxes) for the occluded objects are noisy, containing partially their *occluders*. An occluder is a fully-visible vehicle; some fully-visible vehicles may not have any occludee but we still call them occluders in this paper.

By *occludee* we denote any partially occluded vehicle. Heavily occluded objects are often not detected at all, as the object model is designed or learned to detect non-occluded objects. Instead of taking occludees as exceptions, this paper proposes to detect occludees explicitly. The visible part of an occludee is obtained for further tracking, which is considered as being separated from its

G. Azzopardi and N. Petkov (Eds.): CAIP 2015, Part I, LNCS 9256, pp. 14–26, 2015.
DOI: 10.1007/978-3-319-23192-1_2

Fig. 1. *Left:* A sample frame with a tracked occluder (the filled pink rectangle). *Middle:* A partially-visible car. *Right:* Zoom-in view on an occludee (the yellow rectangle).

occluder. In this way, noise introduced by detection is suppressed in subsequent tracking. There are already methods [5,10] aiming at a detection of occludees separated from their occluders.

For a heavily occluded object, only weakly visible evidence complicates the detection task. Consider the scenario shown in Fig. 1. Without context knowledge it is even challenging for a human observer to recognize the partially-visible vehicle. But by identifying a vehicle as a (possible) occluder, it is more likely that we can also recognize occludees. Occludees act as valuable context information in a traffic scenario; they support the analysis of the behavior of their occluders. An occluder behaves possibly differently with or without an occludee. On the other hand, a visible vehicle, being potentially an occluder, also defines context for scanning for occludees. Simple as is, each occludee has at least one occluder. In this paper, we propose to detect occludees, using the occluders as priors.

We propose an integrated occluder-occludee object detection and tracking method Input sequences are recorded from a mobile binocular system. Occluders are detected and tracked independently. Occludees are explicitly detected and tracked, adopting occluders as priors. Finally, we integrate the occluder and occludee tracking systems. Figure 2, top row, shows three consecutive frames from

Fig. 2. *Top:* Input frames (intensity channel). *Middle:* A tracked vehicle acts as occluder (pink rectangles). *Bottom:* Tracked occludees (green rectangles) and occluders (pink rectangles).

a test sequence. The figure illustrates the following scenarios: An occluder switches to be an occludee (shown in the left and middle frames); an occludee is about to switch into an occluder (shown in the middle and right frames). The middle row illustrates results for occluder detection and tracking, shown by red filled rectangles; the bottom row shows detection and tracking results from the proposed system, with occluders and occludees shown by filled pink and green rectangles, respectively. In this paper we propose a context-based occludee detector, detecting occludees with occlusion portions up to 80%; we also apply the proposed occludee detector in an integrated occluder-occludee detection and tracking system to handle dynamic occlusions. Finally, we demonstrate the potential assistance for avoiding collisions in critical highway driving scenarios.

This paper is structured as follows. Section 2 provides a brief review of related work on occludee detection. Section 3 introduces our occluder detection and tracking method, followed by a proposed occludee detection and tracking method in Section 4. Section 5 describes the integration of both the occluder and occludee tracking systems. Experimental results and evaluations are given in Section 6. Section 7 concludes.

2 Related Work

Occlusions cause appearance changes and pose difficulties for data association in tracking. We review papers regarding occludee detection methods. Approaches used for fully-visible vehicles cannot be simply adapted for partially occluded vehicles. For example, Haar-like features, horizontal edges, visual symmetry, and corner density are properties used in [16] for detecting fully-visible vehicles. The visual appearance of partially occluded vehicles varies, edges might be too short to be identifiable. We cannot assume visual symmetry. This section reviews detection methods for occluded or general objects based on context information.

Single Object Model Occlusion Handling. [3] introduces a rich object representation for a deformable part model, extensively studied for object detection and pose estimation. For handling occlusion, [4] proposes to introduce a binary variable for each bounding box fragment, denoting whether it is from object or background; structured SVM and inference methods are used for learning and testing. In [5], a hierarchical deformable part model is proposed to explicitly handle occluded objects. Each part is further divided into subparts, and a modified structure SVM is adopted for learning. [20] discusses the training of two detectors (global bounding box-based or part-based); an occlusion map is generated by the global detector, and used for the part-based detector.

Occluder-occludee Pair Model. Occluders are often modelled together with occludee for detection. [17] proposes to train a pairwise object detector to detect occluder-occludee pairs explicitly. In [14], occluder-occludee occlusion patterns are explored for detection. A clustering method is adopted to obtain the occlusion patterns from a training set of pairs. Two joint deformable part models are proposed

for learning those occlusion patterns. [11,12] adopt an *and-or* graph model to couple the occluders and occludees based on structure SVM. The occluder-occludee models are manually designed for specific occlusions, e.g. on a parking lot.

Context-Based Methods. Contextual information, adopted for object class recognition tasks, leads to performance improvement [2,8,13,23]. [13] proposes to adopt a visual-cue surround to improve individual pedestrian detection. [2] adopts co-occurrence context evolution in a deformable part model. In [23], touch-codes are explored to model the interaction between two people in a photo.

Single model methods focus on occludees separately from their occluders. In order to handle dynamic occlusion patterns, designed occluder-occludee-pair models are not suitable. With promising results achieved by adopting contextual information, this paper proposes a context-based multiple-cue occludee detector. It uses occluders for extracting the context cue, the visible part for exploring appearance information, and stereo pairs for obtaining depth information. The combined verification of threes cues, context, appearance, and depth, is sufficient for robust occludee detection.

We handle dynamic occlusions by applying an occludee detector in a vehicle-detection and tracking system. Occluders are detected and tracked independently, and subsequently used as priors for occludee detection. Due to dynamic occlusions, occludees and occluders may change their roles. Thus, we propose an integrative module to switch occluder-occludee detection and tracking systems while processing an image sequence. The proposed integrated occluder-occludee tracking system handles dynamic occlusions efficiently; see Section 6 for experiments.

3 Occluder Detection and Tracking

Vehicles may appear fully visible (occluders) or partially occluded (occludees). Occluder detection and tracking is done independently from occludees. We use a stereo pair as input at each time step. A sliding window generates initial hypotheses $\mathcal{H}^\circ = \{h_i^\circ : 1 \leq i \leq N\}$, with $h_i^\circ = (x_i, y_i, W_i, H_i)$, where (x_i, y_i) are the top-left coordinates, and W_i and H_i the width and height of the bounding box of hypothesis h_i°.

Two layers of classifiers are adopted for classification. The first layer uses a cascaded AdaBoost classifier as commonly used for face, pedestrian, or vehicle classification [18,19]. We adopt it for rejecting 'easily' identifiable false hypotheses. Remaining hypotheses are fed into a small convolutional network. For details see [22]. Verified hypotheses define the subset $\mathcal{B}^\circ \subseteq \mathcal{H}^\circ$; they are passed on for tracking. We note that our overall approach is independent from the actual type of classifiers used.

We assign a tracker T_j^{er} (using tracking-by-detection; superscript "er" for "occluder", j denotes the tracker ID) to each verified hypothesis $b_j^\circ \in \mathcal{B}^\circ$, $1 \leq j \leq M \leq N$, which uses a Kalman filter for tracking the vehicle 3D position (X_j, Z_j). X denotes the lateral position, Z the longitudinal position, and (X_j, Z_j) is the mid-point at the bottom of the vehicle, assuming vertical position $Y = 0$ (vehicles are not flying). The 3D location is provided by a disparity map generated by

a semi-global matching technique [6]. We assume that disparity values inside b_j° are all identical (i.e. we use the mean disparity). To calculate the initial state $\mathbf{x}_j^{er} = (X_j, Z_j)^T$, we use the mean disparity and the mid-bottom point coordinates (x_j, y_j) of the vehicle bounding box.

The sketched detection-by-tracking method uses the trackers $\mathcal{T}^{er} = \{T_j^{er} : 1 \leq j \leq M\}$ for generating hypotheses, denoted by $\mathcal{H}^{er}$. $\mathcal{H}^{er}$ are verified by occluder classifiers. In this way we include the trackers' capabilities into the detection process which improves the detection rate. The process is robust with respect to jittering and small scale changes.

By verification we obtain a subset $\mathcal{B}^{er} \subseteq \mathcal{H}^{er}$ of hypotheses, where each $b_j^{er} \in \mathcal{B}^{er}$ is flagged with its tracker ID j. Thus, data association can be done by matching tracker IDs. Currently active trackers are updated by detections obtained in initial (detected) hypotheses $\mathcal{B}^\circ$ or in tracked hypotheses $\mathcal{B}^{er}$ when applying the occluder trackers. Some detected bounding boxes b_j° may overlap with some tracked boxes b_g^{er}. Mean-shift based non-maximum suppression is used to merge multiple detection responses. New trackers are initialized if there are unmatched boxes b_j°. Overall, our approach effectively combines the tracking-by-detection and detection-by-tracking paradigms.

4 Occludee Detection and Tracking

This section describes a new occludee detection and tracking method. We employ multiple cues, occlusion context, appearance, and disparity. Each cue poses individually a weak constraint which is not yet sufficient for detecting a vehicle in general based on a small fragment of its rear side.

Detected occludees, denoted by b_{kj}^{ee}, are tagged with their occluder track ID. Let j denote the occluder tracker ID, and k the index of the occludee (superscript "ee" for "occludee"). We assume that each occluder has a maximum of two occludees, with $k = l$ if on the left, or $k = r$ if on the right. We adopt again the tracking-by-detection method to track detected occludees. Different to occluders, where we apply detection-by-tracking, occludees are detected by our proposed context-based multiple-cue detector. We do not use tracking of predictions in this case. The prediction from occludee trackers may not be as reliable, due to small visible regions. The appearance may change considerably caused by dynamic occlusions. For occluder j, the occludee trackers T_{kj}^{ee} define a set $\mathcal{T}_j^{ee}$ (of up to two elements).

Context-Based Multiple-cue Occludee Detector. Occludees do not appear everywhere in the image. The occluder gives hints for scanning for its occludees. Assuming a (nearly) planar road surface, the occludees are located further away (in longitudinal direction), and the occludee is occluded by this occluder in the image plane. The bounding box of an occludee b_{kj}^{ee} and that of its occluder b_j^{er} are expected to be overlapping, or adjacent to each other. Considering real-world applications (e.g. autonomous driving), a range of possible positions of an occludee can be estimated according to the position of its occluder. Given a candidate occludee position (X_i, Z_i) in a defined 3D region, the corresponding

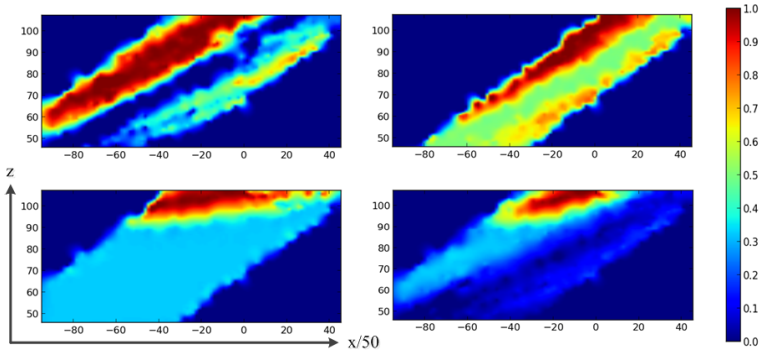

Fig. 3. Multiple-cue responses shown in heat-color maps. The corresponding 3D top-view is shown in Fig. 4. A more reddish color denotes large response values, meaning more likely a 3D position (X, Z) of an occludee. The combined response map (shown on the bottom, right) denotes that there is possibly an occludee at a distance of about 100 m ahead. *Top, left*: Quarter-width classifier response. *Top, right*: Half-width classifier response. *Bottom, left*: Disparity response. *Bottom, right*: Combined multiple-cue response.

occludee hypothesis h_i^{ee} in the image plane is obtained with a defined vehicle size. The context, i.e. occlusion with its occluder in the image plane, is adopted as context cue. More context cues, e.g. lane detection results, could be included to further improve the robustness.

Intuitively, since the occludees are partially-visible vehicles, we propose to train partial-vehicle classifiers. Those classifiers are applied for recognising that the occluded object is a vehicle, instead of, for example, a traffic sign or any other object in a traffic scene. We train a quarter- and a half-width classifier. Both classifiers' training data are cropped from a fully-visible vehicle training set used for training of the occluder classifier.

Using an occlusion check, the classifier can be applied to various occlusion patterns. Adopting occluders as priors, with a given candidate occludee at 3D position (X_i, Z_i), the visible part of the occludee is known by occlusion check in the image plane as mentioned above. When the occludee's bounding box h_i^{ee} is visible more than half of the usual width, both the quarter- and the half-width classifier are adopted to classify a quarter or half of h_i^{ee} in the intensity image.

We apply a local convolutional neural network; it could be replaced by any bounding-box-based classifier. The quarter- or the half-width classifiers' responses are taken as appearance cues.

Given a candidate occludee's 3D position (X_i, Z_i), assuming the disparity value for a vehicle (considered to be a vertically planar object), the measured disparity value within the corresponding h_i^{ee} region should be aligned with the expected disparity value. We model disparities by a Gaussian distribution with respect to differences between expected disparity and measured disparity values.

So far we have multiple weak cues, context priors, classifiers, and disparities. Multiple cues are combined in a particle filter framework. Each particle is the 3D

location of a candidate occludee. The confidence for each particle presents those multiple cues. A higher confidence value denotes a larger likelihood of an existing occludee. The most confident particle is selected as an occludee detection. Left and right occludees are detected independently.

Let N denote the number of particles (3D locations around an occluder); see Fig. 4. Colored dots identify particles. For a given occluder, the occludees are located in a range further away with valid occlusion to their occluder. The particles are denoted by $\{(X_{ij}, Z_{ij}) : 1 \leq i \leq N\}$. Each position (X_{ij}, Z_{ij}), identifying the middle-bottom point of a candidate occludee, can be projected into a hypothesis h_{ij}^{ee} in the image plane with a defined 3D size (width and height). The occlusion between the candidate hypothesis and its occluder is valid if their bounding boxes h_{ij}^{ee} and b_j^{er} are non-disjoint. In other words, the candidate occludee is actually occluded by its occluder. Non-valid hypotheses are excluded from further processing, formally represented by

$$C_i^{cont} = \begin{cases} 1 \text{ if } \tau_1 > f(h_{ij}^{ee}, b_j^{er}) > \tau_2 \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

where $f(h, b)$ is a function which returns an overlap-ratio for input boxes h and b, and τ_1 and τ_2 are upper and lower thresholds for the overlapping ratio.

The appearance of the visible part is verified by a quarter- or half-vehicle classifier. According to occlusion patterns, we derive a visible part of an candidate occludee from the occluder. Two partial vehicle classifiers are applied to obtain classifier responses C_{quar} and C_{half}. The appearance cue is defined by

$$C_i^{class} = \begin{cases} \omega_1 C_{quar} + \omega_2 C_{half} & \text{if } f(h_{ij}^{ee}, b_j^{er}) < \tau_3 \\ C_{quar} & \text{otherwise} \end{cases} \tag{2}$$

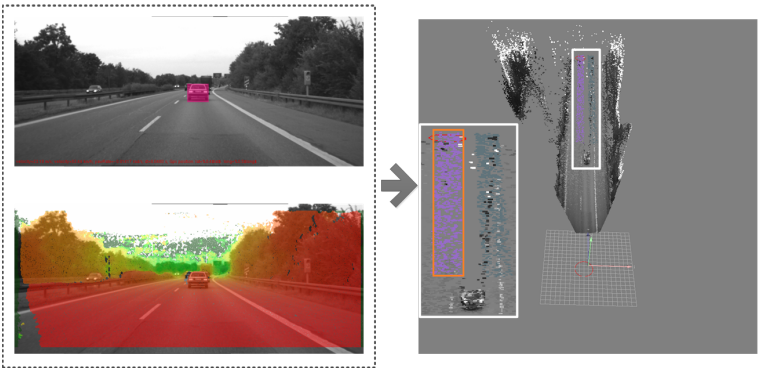

Fig. 4. *Left, top*: Intensity image with tracked occluder (the pink filled rectangle). *Left, bottom*: Corresponding disparity map, with close to far away encoded by red to green. *Right*: Top view of the shown 3D scene with disparity map shown in lower-left. The zoom-in region (the light-grey rectangle) shows two sample regions overlaid with colored dots. Each colored dot denotes a sample. The region highlighted in orange denotes the same 3D position as show in Fig. 3.

If more than half of the width of a candidate occludee appears, we adopt the half-width classifier along with the quarter-width classifier. The ratio threshold τ is constant. Weights ω_1 and ω_2 define the applied contributions of the two classifiers.

If there is an occludee then the measured disparity value from hypothesis h_{ij}^{ee} is aligned to the expected disparity value. Even if being verified by context prior and classifiers, hypotheses with high scores are still shattered across different distances, which corresponds to different scaled bounding boxes in the image plane; see Fig. 3.

Given a candidate occludee (X_{ij}, Z_{ij}), the expected disparity d_i^{exp} for the occludee is obtained by assuming a vertical position $Y = 0$. The measured disparity d_i^{mea} is obtained by averaging disparity values in the central subregion of h_{ij}^{ee}. A Gaussian distribution is adopted to model the disparity cue with respect to the difference between d_i^{mea} and d_i^{exp}. The value of σ is obtained by measuring the uncertainty of disparity in a statistical manner. The disparity-cue response is defined by

$$C_i^{disp} = \frac{1}{\sqrt{2\sigma\pi}} \exp^{-\frac{(d_i^{mea} - d_i^{exp})^2}{2\sigma^2}} \tag{3}$$

Each sample (X_{ij}, Z_{ij}) is measured with context, classifier response, and disparity cues, obtaining responses C_i^{cont}, C_i^{class}, and C_i^{disp}. The higher the responses value, the more likely that there is an occludee located at the sample position. The *confidence* (i.e. combined response) of a sample (X_{ij}, Z_{ij}) that contains an occludee is defined by

$$C_i = C_i^{cont} C_i^{class} C_i^{disp} \tag{4}$$

All cues are required for a response with high confidence, as just individual cues are insufficient. The occludee is detected by a greedy selection of that sample which has the highest confidence, formally

$$b_{kj}^{ee} = arg \max_{h_{ij}^{ee}} C_i \tag{5}$$

A low-pass filter is employed to reduce false positives. Figure 3 shows multiple-cue responses of an occludee candidate; a corresponding 3D top-view is shown in Fig. 4. The intensity image with a tracked occluder (the pink rectangle) and a disparity map are shown in Fig. 4. In Fig. 3, the more reddish color denotes large response values, meaning a larger likelihood of an occludee at that position. The combined response map (bottom, right) indicates that there is possibly an occludee at distance 100 m ahead.

Occludee Tracking. We detected occludees in $\mathcal{B}^{ee} = \{B_j^{ee} : 1 \leq j \leq M\}$. A Kalman filter-based tracking-by-detection method is adopted for occludee tracking. Similar to occluder tracking, the middle-bottom 3D position of an occludee is defined as tracking state (X, Z). A constant-velocity assumption is adopted. The detections from a multiple-cue detector are used for updating the state.

Using the proposed context-based occludee detector, occludee detections are tagged with their occluder tracker IDs. Instead of doing data association for each occludee against the occludee trackers, an occludee detection, b^{ee}_{kj}, is associated with occludee tracker T^{ee}_{kj}. The occluder tracker ID j and occludee tracker ID k specify the corresponding occludee detection to its tracker.

5 Integration of Occluder-Occludee Tracking

In real world scenarios, a vehicle was fully visible (occluder) may be partially occluded (occludee) in a few seconds. On the other hand, a partially-visible vehicle may become fully visible. We introduce the proposed integration of occluder-occludee tracking.

Case 1: An occluder is gradually occluded by another occluder. The detection-by-tracking method will fail to verify the predicted bounding box, due to occlusion. This vehicle is lost even if it is still partially visible. *Case 2*: A tracked occludee is shifting away to another lane and becomes gradually more visible. The occludee tracker can generate a hypotheses for the occluder classifier for verification.

To interactively integrate occluder and occludee tracking systems, we propose to switch occludee and occluder trackers when conditions apply.

Occluders Switch to Occludees. To switch an occluder tracker T^{er}_{j} to an occludee tracker T^{ee}_{kg}, the occluder has a valid occluder T^{er}_{g} that causes the occlusion. The occluder j is located further away from its potential occluder. The overlap in image plane between b^{er}_{j} and b^{er}_{g} is over a given threshold. We conclude an occludee detection from having occluder g matched to occluder j, with b^{ee}_{kg} overlapped by b^{er}_{j}. The conditions are formulated as

$$T^{er}_{j} \Rightarrow T^{ee}_{kg}, \quad \text{s.t. } f(b^{er}_{j}, b^{er}_{g}) > \tau_4, \quad \exists b^{ee}_{kg}, f(b^{er}_{j}, b^{ee}_{kg}) > \tau_5, \quad z_j > z_g \tag{6}$$

Occludees Switch to Occluders. To switch an occludee tracker T^{ee}_{kj} to an occluder tracker T^{er}_{g}, the occludee is tracked for a while and overlap ratio with its occluder are low, defined by

$$T^{ee}_{kj} \Rightarrow T^{er}_{g}, \quad \text{s.t. } g(T^{ee}_{kj}) > \tau_6, \quad f(b^{ee}_{kj}, b^{er}_{j}) < \tau_4 - \varepsilon \tag{7}$$

where $g(T)$ denotes the tracked frames for the tracker T, and ε is a positive constant, adopted to prevent a switching loop between occluders and occludees.

6 Experiments

The proposed system is evaluated on two types of sequences, `Dynamic` and `Dense`; see Tab. 1. `Dynamic` contains four sequences of 1650 frames, approximately one minute each. 6259 vehicles (occluder and occludees), 992 occludees, and 188 trajectories, are labelled frame by frame. The `Dynamic` sequences contain scenarios with dynamic occlusions, recorded with regular driving style. To evaluated

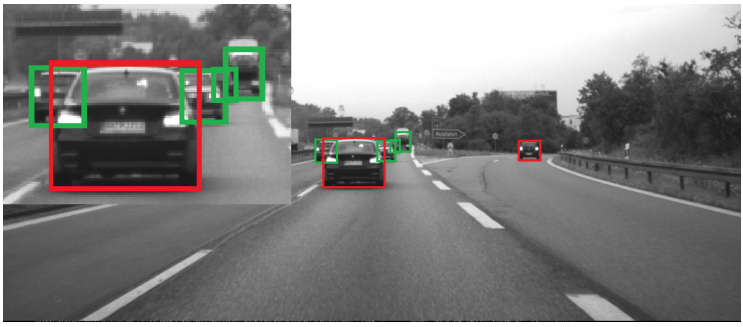

Fig. 5. A frame with labelled vehicles. Red rectangles denote the occluders; green rectangles denote occludees.

the proposed occluder-occludee integrative system, there are scenarios occluders changing lane with their occludees becoming fully visible, or with occluders on fast lane driving pass the ego-vehicle becoming occludees. The `Dense` sequence contains 8300 frames with every 100 frames labelled (approx. 5.5 minutes). 343 vehicles and 67 occludees are labelled. The number of objects at the first glance seems limited, but those 83 frames are randomly sampled from thousands of frames. This sequence is adopted to estimate the proposed system on dense highway traffic, a more general evaluation. Both `Dynamic` and `Dense` sequences are recorded at 25 fps, from stereo cameras mounted behind the windscreen of an ego-vehicle.

One example frame with object labels shown in Fig. 5. The occluders and occludees are explicitly labelled respectively, denoted with different color rectangles. The occludee labels (green rectangles) are overlapped with their occluders. The proposed integrated system and occludee detector output the visible part exclusive to its occluder. Thus, the overlap ratio for measuring is set relatively low 0.25. A zoom-in region shown on top left corner illustrates what the perfect system is expected to detect. There are occluded vehicles appear further away from occludees. We will focus on those situations in future work.

6.1 Integrated System *vs* Occluder System

We begin the evaluation with comparing the proposed integrated system with the baseline system (occluders detection and tracking system). The frame-wise detection rate and precision are adopted. Since tracking is involved, the recall

Table 1. The test sequences.

Sequences	Frames	Objects	Occludees	Trajectories
Dynamic	1,650	6,259	992	188
Dense	8,300	343	67	-

curve is not applicable. The trajectory detection rate is used to evaluate tracking performance. A trajectory is counted as detected if 50% frames over the trajectory length are detected.

For `Dynamic` sequences, both frame-wise detection and trajectory measures are shown in Tab.2, left. 'Integrated' denotes the proposed integrated system. 'Occluder' denotes the occluder detection and tracking system. The proposed system fires more false positives, but improves both the detection rate and trajectory detection rate by significant margins 11% and 27.9% respectively. The proposed system detects and tracks occludees with occlusion portion up to 80%. The detection rate is improved due to the occludee detection and tracking system, and the integration between occluder and occludee trackers. The evaluation results on `Dense` sequence are illustrated in Tab. 2, right. Similar performance is observed. 'Integrated' outperforms 'Occluder' by a large margin 17.8%.

6.2 Application Scenario

Different levels of autonomous driving on highways are available in serial production cars, e.g auto-brake, distance keeping, lane keeping etc.. In order to enable more advanced autonomous driving, better understanding the environment offers better foundation for that purpose. Driving environment in real world is dynamic. Vehicles changing from one lane to the other, because of the vehicle in front (their occludees) driving slow, or even, suddenly broken down. The occludees affects the behavior of their occluders. If the ego-vehicle observes a bit further away (the occludees), a more advanced reaction could be made, instead of just braking abruptly.

Using occluder detection and tracking system, the occludees are not picked up due to occlusion, although they are visible, partially. The proposed integration system detects and tracks the occludees with the occlusion portion up to 80%. Four `Dynamic` sequences are adopted to measure the time (frame) difference between the proposed system and the occluder system picking up the previously heavy occluded then fully-visible vehicles. The evaluation results are shown in Tab. 3.

In the first three `Dynamic` sequences, the proposed system picks up the occludees $30-40$ frames ahead of the occluder. With recording frame rate 25 fps, the proposed system 'sees' the occludee $1.2-1.6$ seconds before the occluder system. With high speed, even a few milliseconds make a difference. In 'sequence 4', the occludee is partially visible for 7 s before appearing fully visible. This

Table 2. Performance measured on the `Dynamic` sequences (*left*) and on the `Dense` sequence (*right*).

	Detection rate	Precision	Trajectory detection rate
Occluder	59.9	88.9	44.2
Integrated	76.9	79.7	72.1

	Detection rate	Precision	Trajectory detection rate
Occluder	55.7	89.2	-
Integrated	73.5	80.3	-

Table 3. Frames ahead of occluder tracker by the integrated system picked up the occluded car in front of a leading car.

	Frames	Time(s)
Sequence 1	40	1.6
Sequence 2	35	1.4
Sequence 3	29(27,32)	1.2
Sequence 4	191	7.64

information can be used for higher level decision making, e.g regarding changing lane for the ego-vehicle.

7 Conclusions

We proposed a vehicle detection and tracking system for handling dynamic occlusions. The proposed method integrates detection and tracking of occludees and occluders. We proposed a context-based multiple-cue method for occludee detection. The applied classifiers for occluders and occludees may be replaced by other bounding-box-based classifiers. A tracking-by-detection method is used for tracking occludee and occluder respectively. The proposed integrated occluder-occludee tracking system shows promising results on handling dynamic occlusions. The proposed system improves detection rate and trajectory detection rate by significant margins, compared with the occluder-only system. The proposed context-based multiple-cue occludee detector detects the immediate occludees for left and right sides of an occluder. It detects slightly to heavily occluded vehicles, occlusion portion up to 80%. The proposed system contributes to handle emergency situations in highway autonomous driving. Generally, instead of focusing on the target-object, e.g occludees in isolation, adopting the contextual information improves the performance.

References

1. Brendel, W., Amer, M.R., Todorovic, S.: Multiobject tracking as maximum weight independent set. In: CVPR (2011)
2. Chen, G., Ding, Y., Xiao, J., Han, T.X.: Detection evolution with multi-order contextual co-occurrence. In: CVPR (2013)
3. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. TPAMI **32**(9), 1627–1645 (2010)
4. Gao, T., Packer, B., Koller, D.: A segmentation-aware object detection model with occlusion handling. In: CVPR (2011)
5. Girshick, R.B., Felzenszwalb, P.F., Mcallester, D.A.: Object detection with grammar models. In: NIPS (2011)
6. Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: CVPR (2005)

7. Huang, C., Li, Y., Nevatia, R.: Multiple target tracking by learning-based hierarchical association of detection responses. TPAMI **35**(4), 898–910 (2013)
8. Karlinsky, L., Dinerstein, M., Harari, D., Ullman, S.: The chains model for detecting parts by their context. In: CVPR (2010)
9. Leal-Taix, L., Fenzi, M., Kuznetsova, A., Rosenhahn, B., Savarese, S.: Learning an image-based motion context for multiple people tracking. In: CVPR (2014)
10. Li, B., Hu, W., Wu, T., Zhu, S.C.: Modeling occlusion by discriminative and-or structures. In: ICCV (2013)
11. Li, B., Wu, T., Zhu, S.-C.: Integrating context and occlusion for car detection by hierarchical and-or model. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part VI. LNCS, vol. 8694, pp. 652–667. Springer, Heidelberg (2014)
12. Li, B., Song, X., Wu, T., Hu, W., Pei, M.: Coupling-and-decoupling: A hierarchical model for occlusion-free object detection. PR **47**(10), 3254–3264 (2014)
13. Ouyang, W., Wang, X.: Single-pedestrian detection aided by multi-pedestrian detection. In: CVPR (2013)
14. Pepikj, B., Stark, M., Gehler, P., Schiele, B.: Occlusion patterns for object class detection. In: CVPR (2013)
15. Pirsiavash, H., Ramanan, D., Fowlkes, C.C.: Globally-optimal greedy algorithms for tracking a variable number of objects. In: CVPR (2011)
16. Rezaei, M., Klette, R.: Look at the driver, look at the road: no distraction! no accident! In: CVPR (2014)
17. Tang, S., Andriluka, M., Schiele, B.: Detection and tracking of occluded people. IJCV **110**(1), 58–69 (2009)
18. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR (2001)
19. Klette, R.: Concise Computer Vision. Springer, London (2014)
20. Wang, X., Han, T.X., Yan, S.: An HOG-LBP human detector with partial occlusion handling. In: ICCV (2009)
21. Wen, L., Li, W., Yan, J., Lei, Z., Yi, D., Li, S.Z.: Multiple target tracking based on undirected hierarchical relation hypergraph. In CVPR, 2014
22. Wöhler, C., Joachim, K.A.: An adaptable time-delay neural-network algorithm for image sequence analysis. IEEE Trans. Neural Networks **10**(6), 1531–1536 (1999)
23. Yang, Y., Baker, S., Kannan, A., Ramanan, D.: Recognizing proxemics in personal photos. In: CVPR (2012)