Visual Saliency Oriented Vehicle Scale Estimation

Jiali Ding, Tie Liu, Qixin Chen, Zejian Yuan and Yuanyuan Shang

Problem

For various traffic scenarios, the Vehicle Collision Warning system (VCW) can predict the risk of the coming back-end collision between the main vehicle and the leading vehicle, which has very important practical significance. The level of danger is quantified by time-to-collision (TTC) of them. In addition to judging the collision distance through expensive distance sensors, it can also be estimated through visual scale changes.

The existing methods have the following two problems.

1. The intensity-based methods cannot handle the challenging situations such as occlusion and rapid scale changes.
2. Descriptor-based methods can capture the changes, but lack of the accuracy of scale estimation.

Contributions

In this paper, we propose a novel robust regularized intensity matching method, which combines the merits of the two categories of methods, i.e., accuracy and robustness.

The contributions of this paper are mainly in four aspects:

1. We propose to integrate intensity-based method and descriptor-based method into our proposed method with a regularization term in Lie Algebra.
2. We introduce a method of salient object detection that enables the system to quickly focus on the most relevant target vehicle.
3. We introduce a two-stage attention mechanism to handle occlusion and noise corruption.
4. We propose to use an efficient second-order minimization scheme that computes the scale change by minimizing the objective function.

Funding

This work is supported by the National Natural Science Foundation of China under Grants 61603022 , 61876112 and 61601311, Support Project of High-level Teachers in Beijing Municipal Universities in the Period of 13th Five-year Plan (CIT&TCD20170322).

References

[1] Redmon, Joseph, and Ali Farhadi. Yolov3: An Incremental Improvement. In *CVPR '2018*

[2] Danelljan, Martin and Häger, Gustav and Khan, Fahad and Felsberg, Michael. Accurate Scale Estimation for Robust Visual Tracking. In *BMVC '2014*

Results

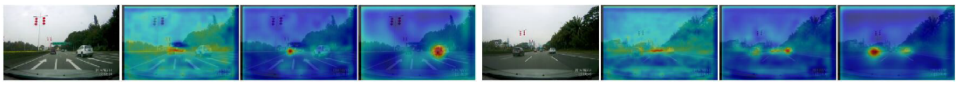
(a) (b)

(c) (d)

The pictures above present our tracking results on four representative videos. The video sequence shown in (a), contains large displacements caused by abrupt bumps of the lead vehicle. (b)(c) demonstrate our results at night. (d) shows the target is severely occluded by the wiper and blurred by raindrops.

Our approach not only tracks the lead vehicle successfully but also accurately estimate the scale.

Methods-PartA&B&C



(a) (b)

we have introduced a method of salient object detection, which matches the required characteristics. We constructed a new SOD network based on the structure of yolov3[1]. We extract the saliency maps from different depth layers in the network and visualize them, as shown above.They can accurately distinguish the scale difference of the targets, which enhances the robustness of the whole system.

As for partB, Descriptor Matching method, we formulate the descriptor matching problem as follows: $\mathbf{y} = \arg\min_{\mathbf{y}} \varphi_m(\mathbf{y})$, where $\mathbf{y} = [t_x, t_y, s]$, and t_x , t_y and s represent translation in x direction, translation in y direction and scaling. And $\varphi_m(\mathbf{y})$ provides a matching score. And we follow the method presented in [2] to calculate $\hat{\mathbf{y}}$.

As for partC, Intensity Matching method. The cornerstone of intensity matching is the intensity consistency assumption: corresponding pixels have the same intensity value in all frames. We penalize deviation with the following loss function: $\mathcal{L}_{gray}(\mathbf{x}) = \frac{1}{2}\sum_{\mathbf{p}\in\Omega}\left\|\delta(\mathbf{x},\mathbf{p})\right\|_2^2$. The original tracking problem is equivalent to estimating an unknown parameter vector **x** that minimizes the sum of $L2-norm$ of the difference between the intensity value of pixel **p** in the reference frame and the intensity value of **p**'s corresponding pixel in the current frame.

Methods-PartD&E&F

(a) (b) (c)

For abnormal conditions such as bad weather, low illumination, and obstruction of unrelated objects, we propose a two-stage attention mechanism that is composed of hard attention mechanism and soft attention mechanism.For each pixel **p**, hard attention mechanism generates a binary mask $M_t(\mathbf{p})$ which indicates whether pixel **p** should contribute to the loss function or not. Soft attention mechanism generates positive weight $W_t(\mathbf{p})$ which controls the amount of attention our method pays to pixel **p**.Pictures above show how mask and weight function.The small yellow and green circles in (a) represent the key points that have been selected using Harris Corner Selector. (b) shows the resulting template after applying the mask. (c) is a visualization of the weight matrix.

As for partE & partF, both of them is about regularization and optimization.