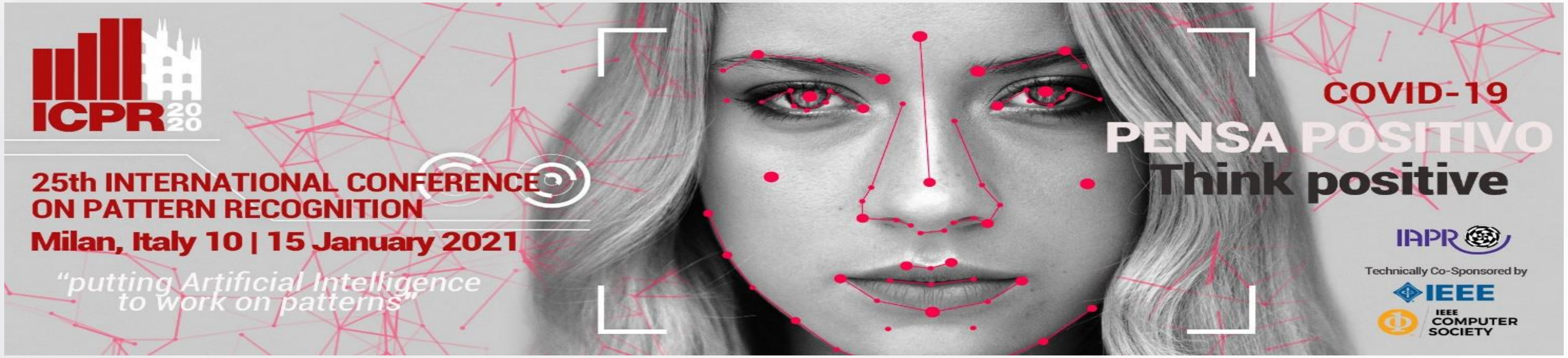


An Adaptive Fusion Model Based on Kalman Filtering and LSTM for Fast Tracking of Road Signs

Chengliang Wang

Xin Xie

Chao Liao

Contents

1 background

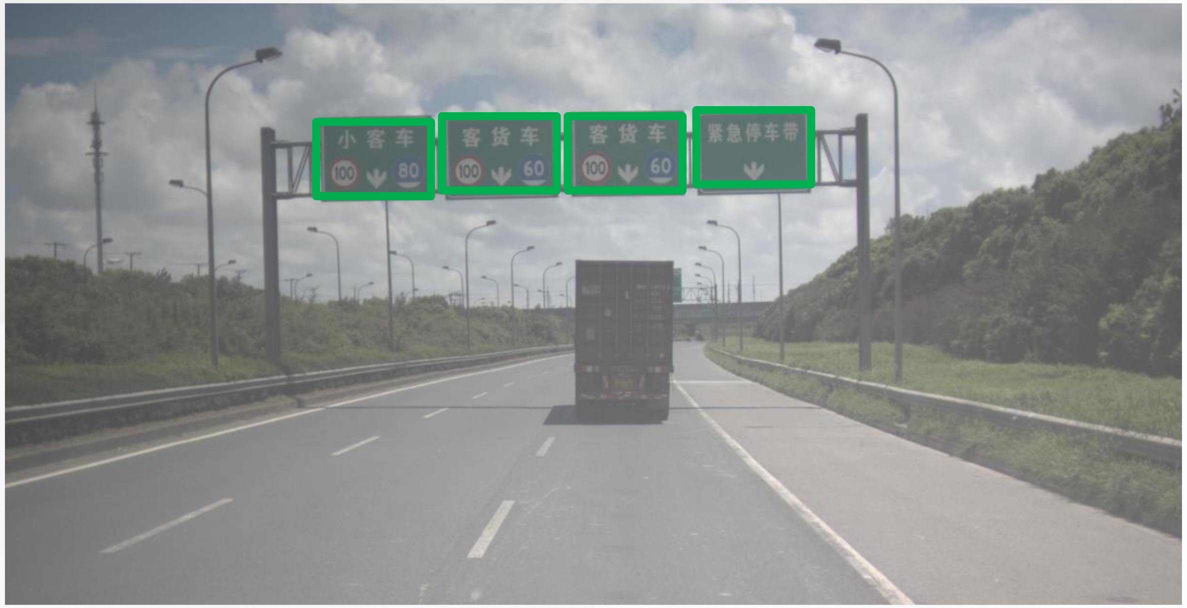
Multi-Object Tracking

The main task of Multiple Object Tracking or Multiple Target Tracking (MOT) is **to locate multiple targets of interest at the same time in a given video, maintain their IDs and record their tracks**. These targets can be pedestrians on the road, vehicles on the road athletes on the playground , or multiple groups of animals (birds , bats, ants, fish, cell, etc.)

1 background

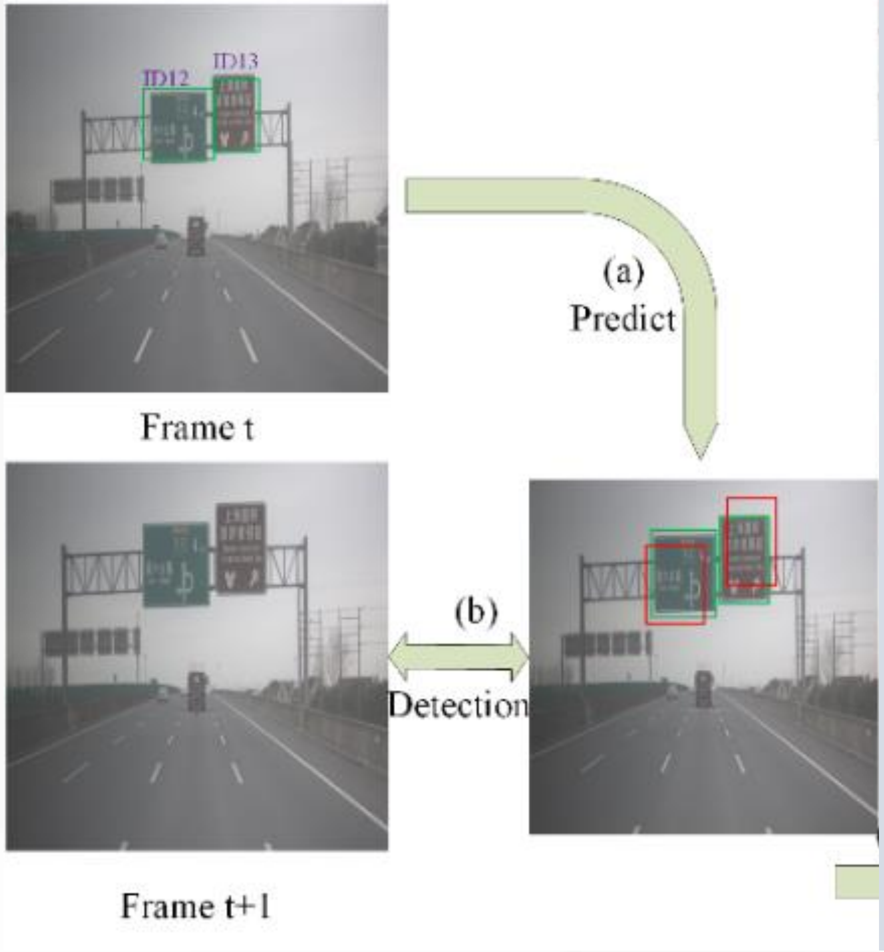
Tracking of Road Signs

Detection and tracking of road signs have recently gained more attention from the computer vision research and industrial community. The main objective of these algorithms is **to detect and track road signs with different maintenance conditions and variable light sources from videos acquired from a moving vehicle**.



2 algorithm：Fast Road Sign Tracking algorithm

- **An Illustration for the tracking of road signs**

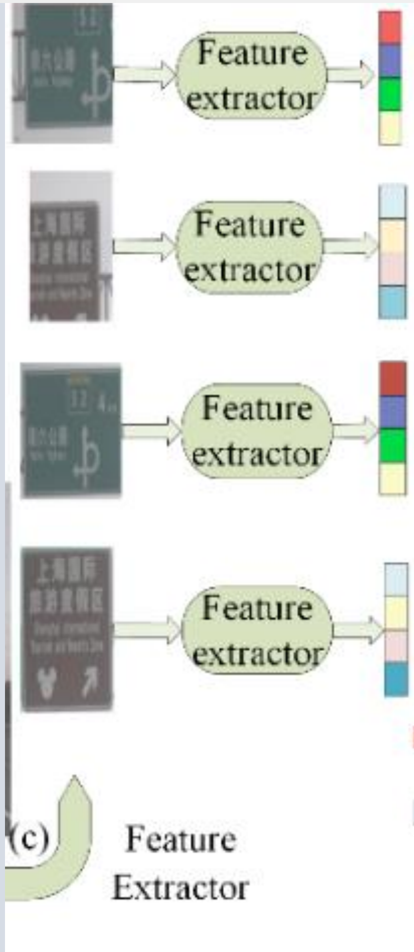


In our model, firstly, predict the state in frame t+1. Then, read the frame t+1 to detect the road signs.

2 algorithm：Fast Road Sign Tracking algorithm

- **An Illustration for the tracking of road signs**

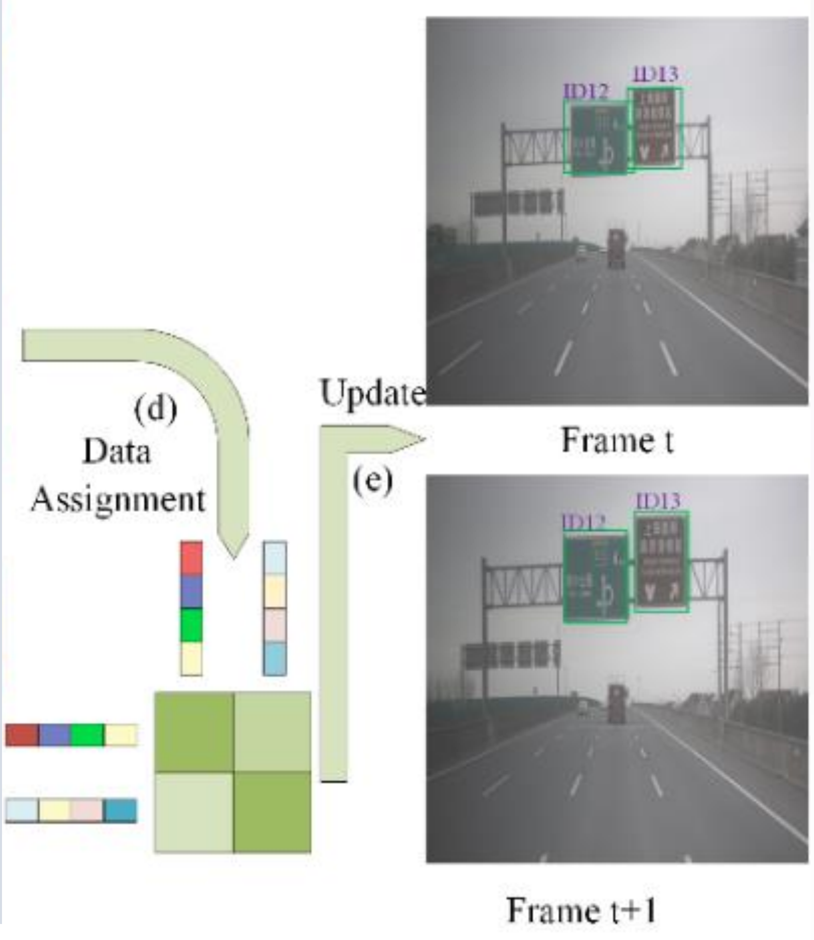
Extract visual features and motion features according to the road signs.



2 algorithm： Fast Road Sign Tracking algorithm

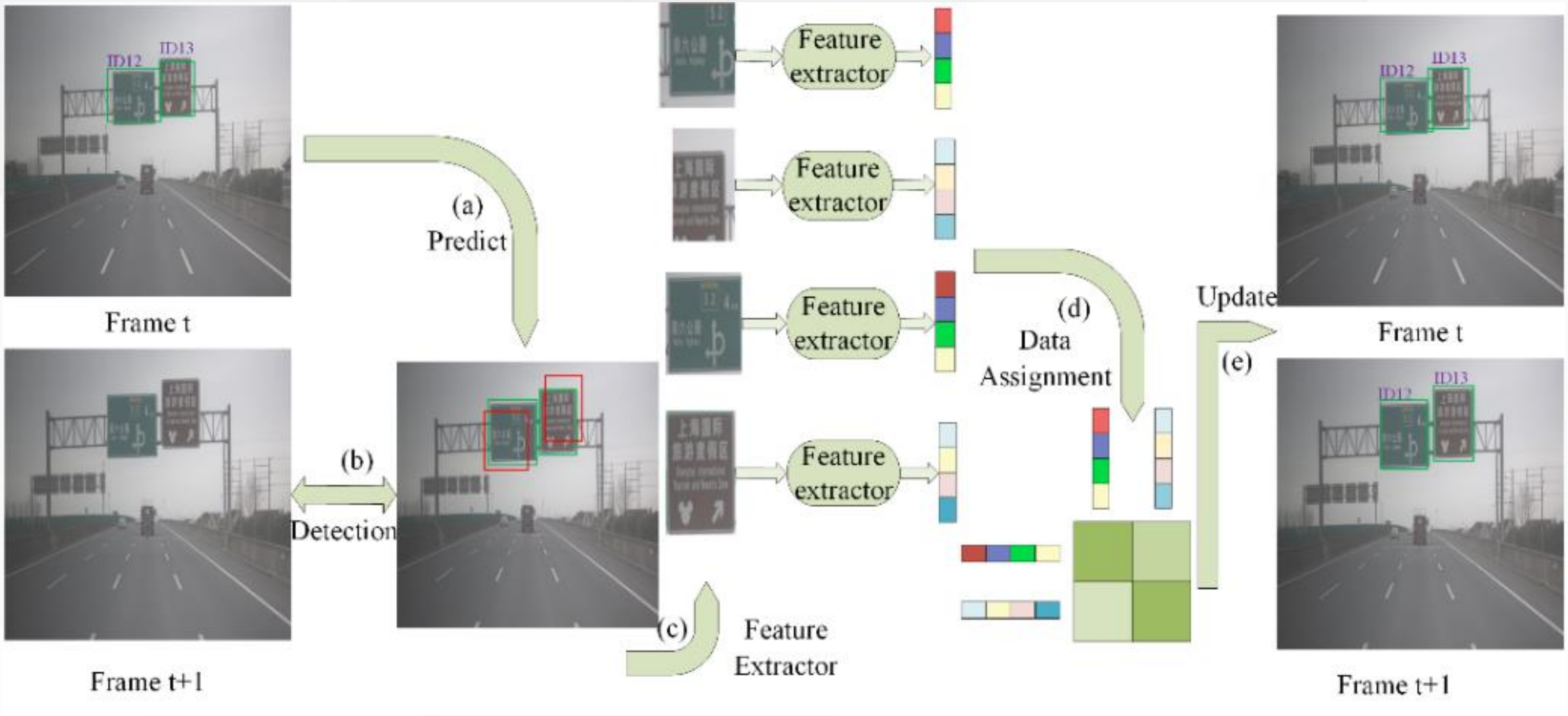
- **An Illustration for the tracking of road signs**

If the road signs assign the existing trackers by assignment algorithm, update the estimation model to estimate the state of the next frame, otherwise create a new tracker.



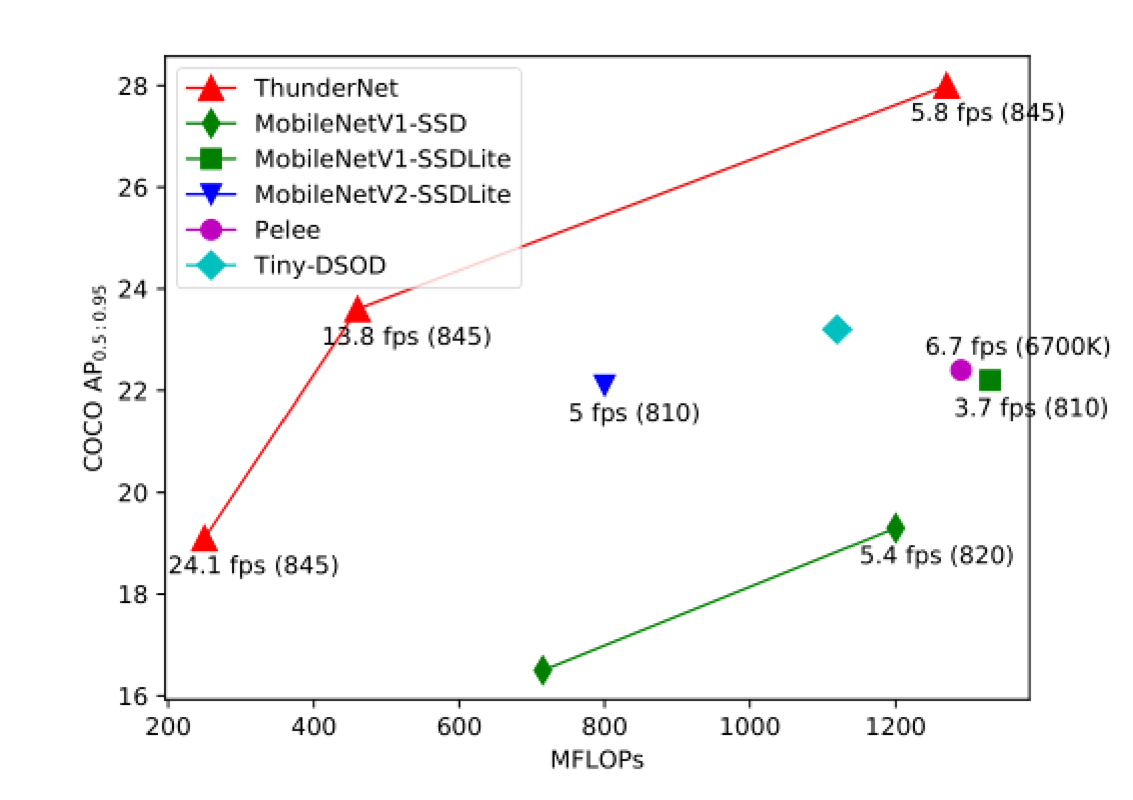
2 algorithm：Fast Road Sign Tracking algorithm

- **An Illustration for the tracking of road signs**



2 algorithm：Fast Road Sign Tracking algorithm

- **A. Detection Model**

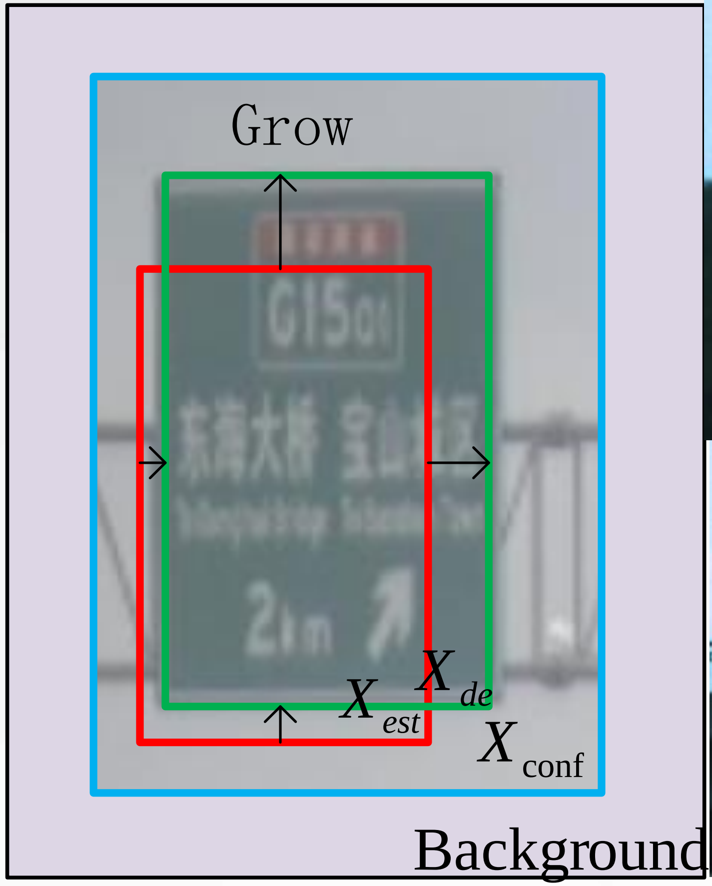


- ThunderNet is a CNN-based detector with real-time object detection on mobile platforms.
- ThunderNet achieves improvements in both accuracy and efficiency.

2 algorithm： Fast Road Sign Tracking algorithm

- **A. Detection Model**

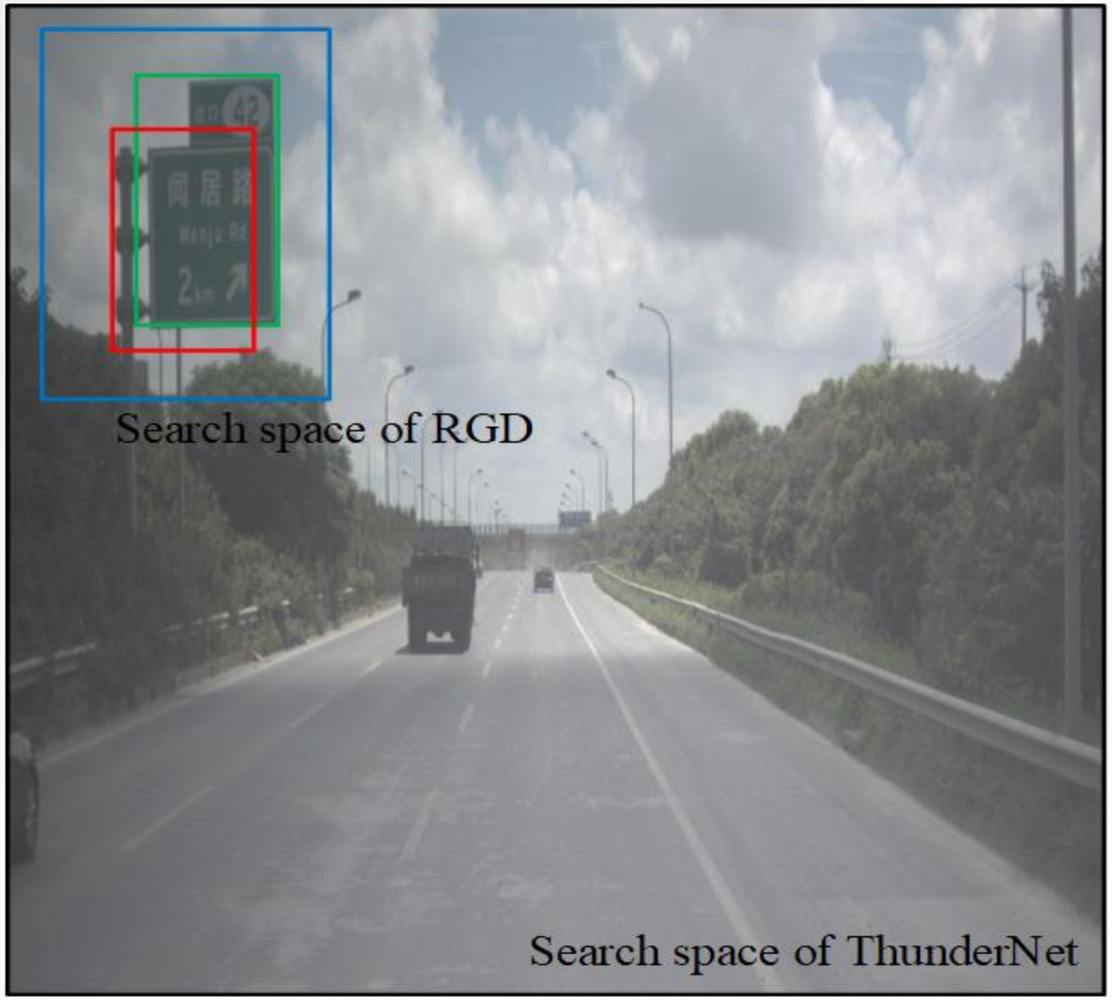
  - Region growing detector (RGD) use the information about the video and the color of road signs
  - For each initialized candidate X_{est} , color detection is performed on each edge.
  - If the ratio of color classified as a road sign **exceeds the threshold, it grows outward**.
  - **Otherwise**, **it is deleted inward**.
  - Get a detection candidate X_{de}.



2 algorithm: Fast Road Sign Tracking algorithm

- **A. Detection Model**

  - The search space of RGD is much smaller than the search space of ThunderNet.
  - In most cases, RGD is faster than ThunderNet.
  - However, the applicable conditions of RGD are limited.



2 algorithm：Fast Road Sign Tracking algorithm

- **A. Detection Model**

Detection Model	Advantage	Disadvantage
ThunderNet	Suitable for various conditions, does not depend on estimation model	Slow
RGD	Fast	depends on estimation model

We combine these detectors to get better performance.

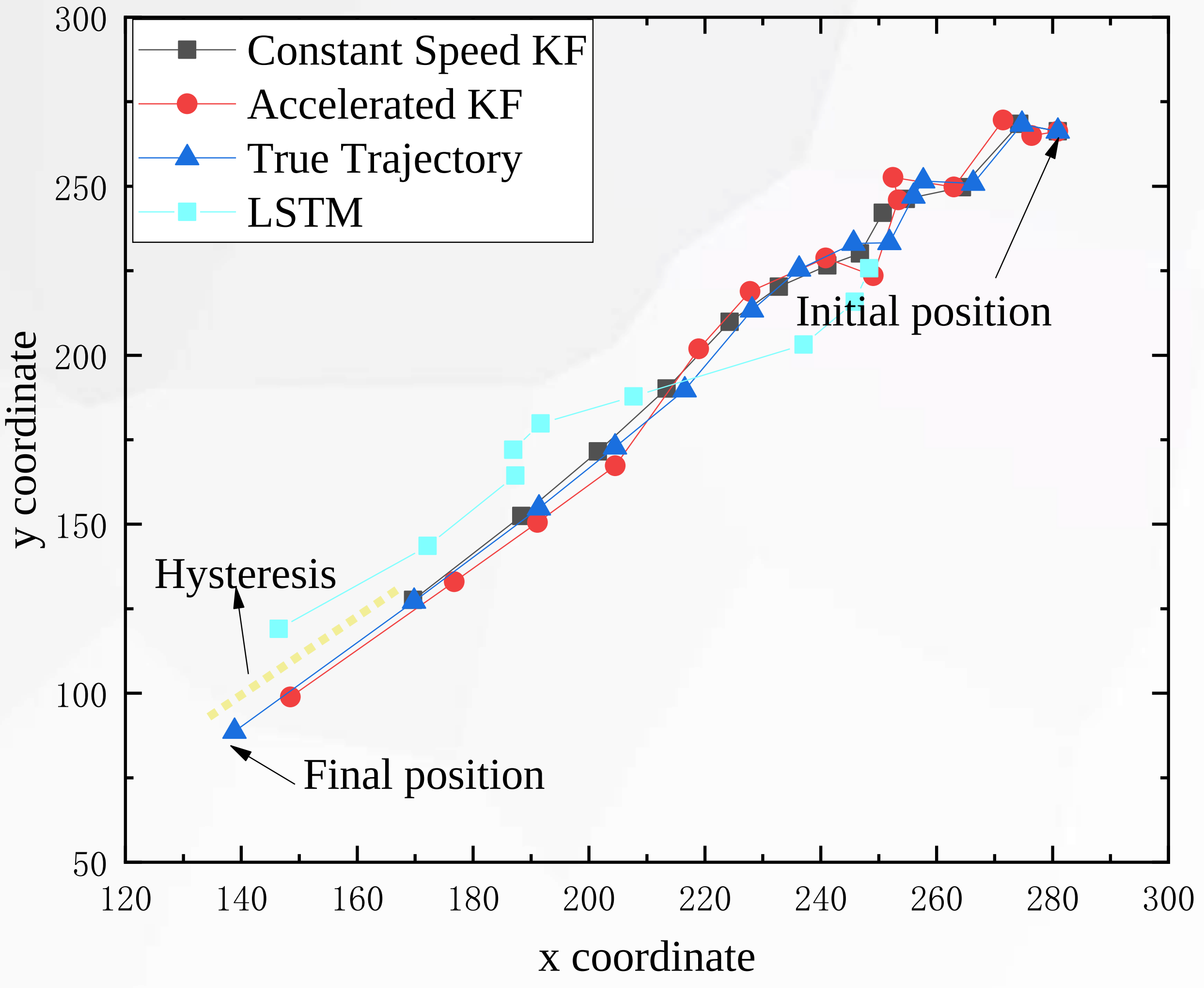
2 algorithm：Fast Road Sign Tracking algorithm

- **B. Estimation Model**

- The process from the appearance to the disappearance of road signs can be regarded as an accelerated motion model.
- To take advantage of the characteristics of road signs, adopt a linear uniformly accelerated KF(Kalman filter) to estimate the motion state.

See from Figure, accelerated KF has the best performance.

Comparing Constant Speed KF, Accelerated KF and LSTM

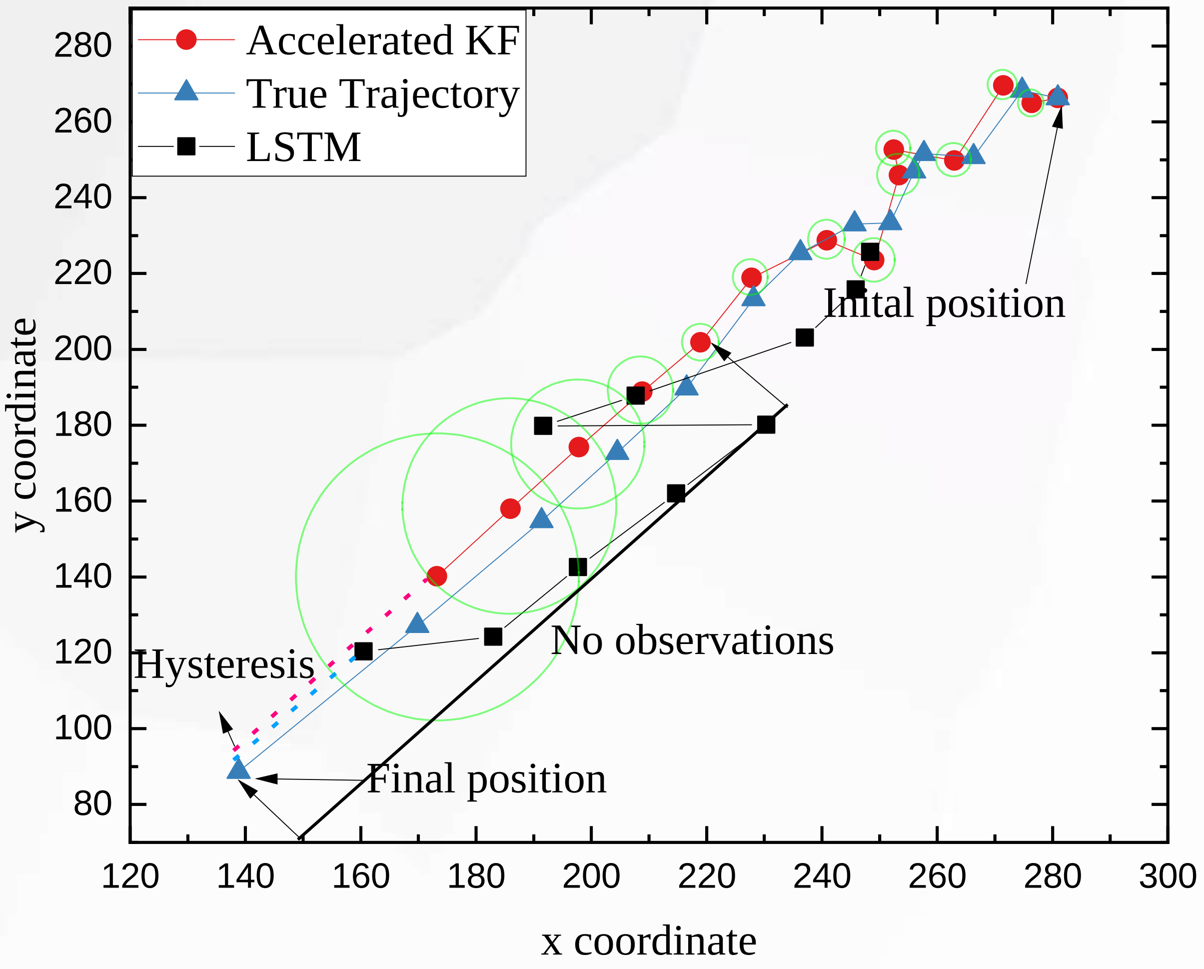


2 algorithm: Fast Road Sign Tracking algorithm

- **B. Estimation Model**
- In the state estimation of road signs, we cannot measure y each frame because there may be occlusion problems or missed detection, and these problems usually occur in consecutive frames.
- Utilize the estimation of the LSTM to correct the estimation of the KF.

See from the Figure, the prediction trajectory of Accelerated KF has a significant hysteresis to the true trajectory, and the prediction trajectory of LSTM is closer to true trajectory than the prediction trajectory of Accelerated KF when there are no observations

Visualization of covariance matrix of prediction.



2 algorithm：Fast Road Sign Tracking algorithm

- **B. Estimation Model**

$$y_{t+1}^{best} = \alpha \cdot f_{KF}(t+1) + (1-\alpha) \cdot f_{LSTM}(t+1) \quad (1)$$

$$S_{pre} = \begin{bmatrix} \hat{\Sigma}_{(x,x)} & \hat{\Sigma}_{(x,y)} & \hat{\Sigma}_{(x,w)} & \hat{\Sigma}_{(x,h)} \\ \hat{\Sigma}_{(y,x)} & \hat{\Sigma}_{(y,y)} & \hat{\Sigma}_{(y,w)} & \hat{\Sigma}_{(y,h)} \\ \hat{\Sigma}_{(w,x)} & \hat{\Sigma}_{(w,y)} & \hat{\Sigma}_{(w,w)} & \hat{\Sigma}_{(w,h)} \\ \hat{\Sigma}_{(h,x)} & \hat{\Sigma}_{(h,y)} & \hat{\Sigma}_{(h,w)} & \hat{\Sigma}_{(h,h)} \end{bmatrix} \quad (2)$$

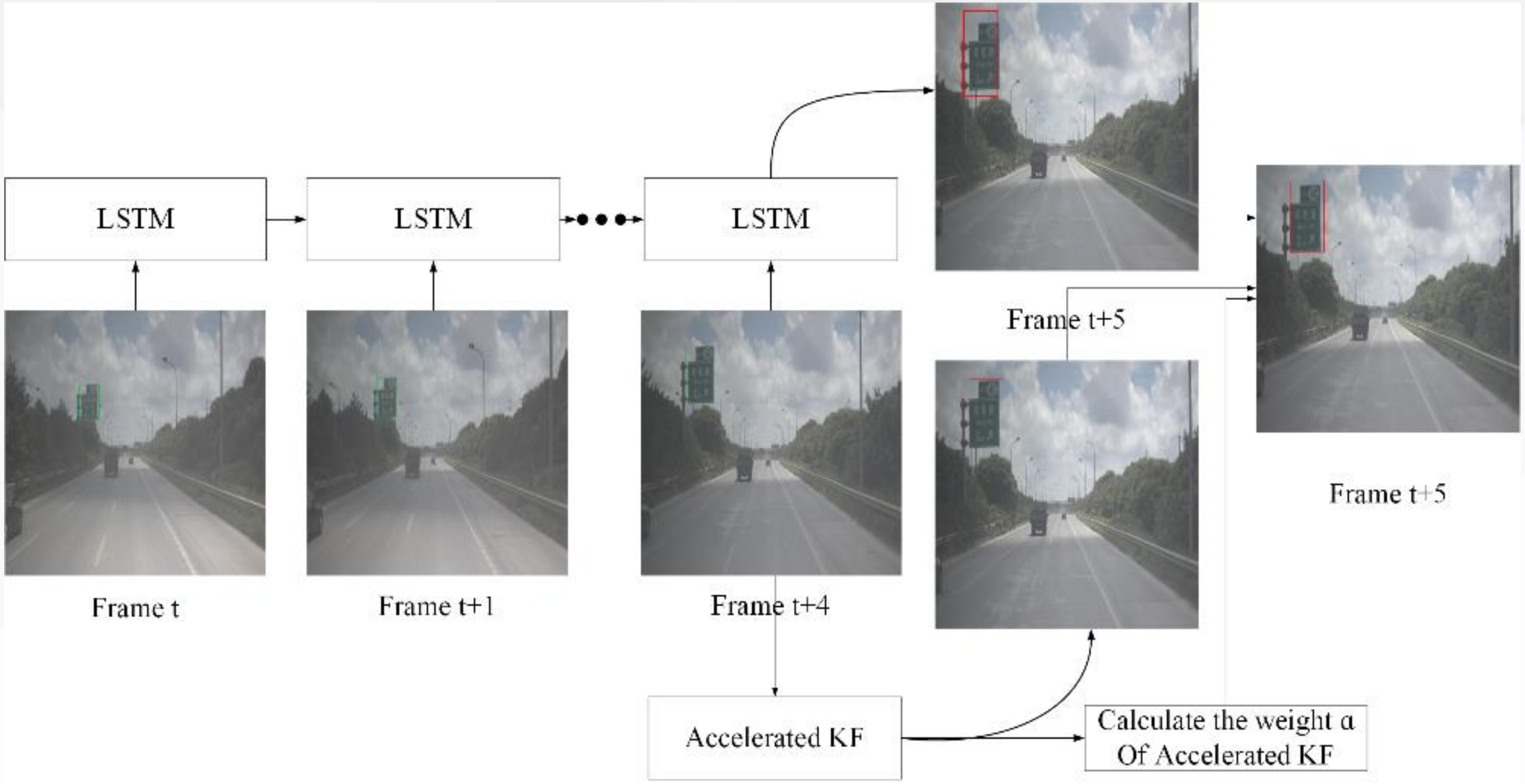
$$\begin{cases} \alpha = \alpha_{max}, |S_{pre}| \le th_{min} \\ \alpha = 1 - \left(\dfrac{|S_{pre}| - th_{min}}{th_{max} - th_{min}} \right), th_{min} < |S_{pre}| < th_{max} \\ \alpha = \alpha_{min}, |S_{pre}| \ge th_{max} \end{cases} \quad (3)$$

- Adaptive fusion model based on KF and LSTM (AKL) is proposed. See from Eq. 1,we combine these predictions. How to get α?
- $|S_{pre}|$ represents the determinant of the covariance matrix related to observations to determine the α (see from Eq. 3).

2 algorithm：Fast Road Sign Tracking algorithm

- **B. Estimation Model**

An Illustration for AKL model

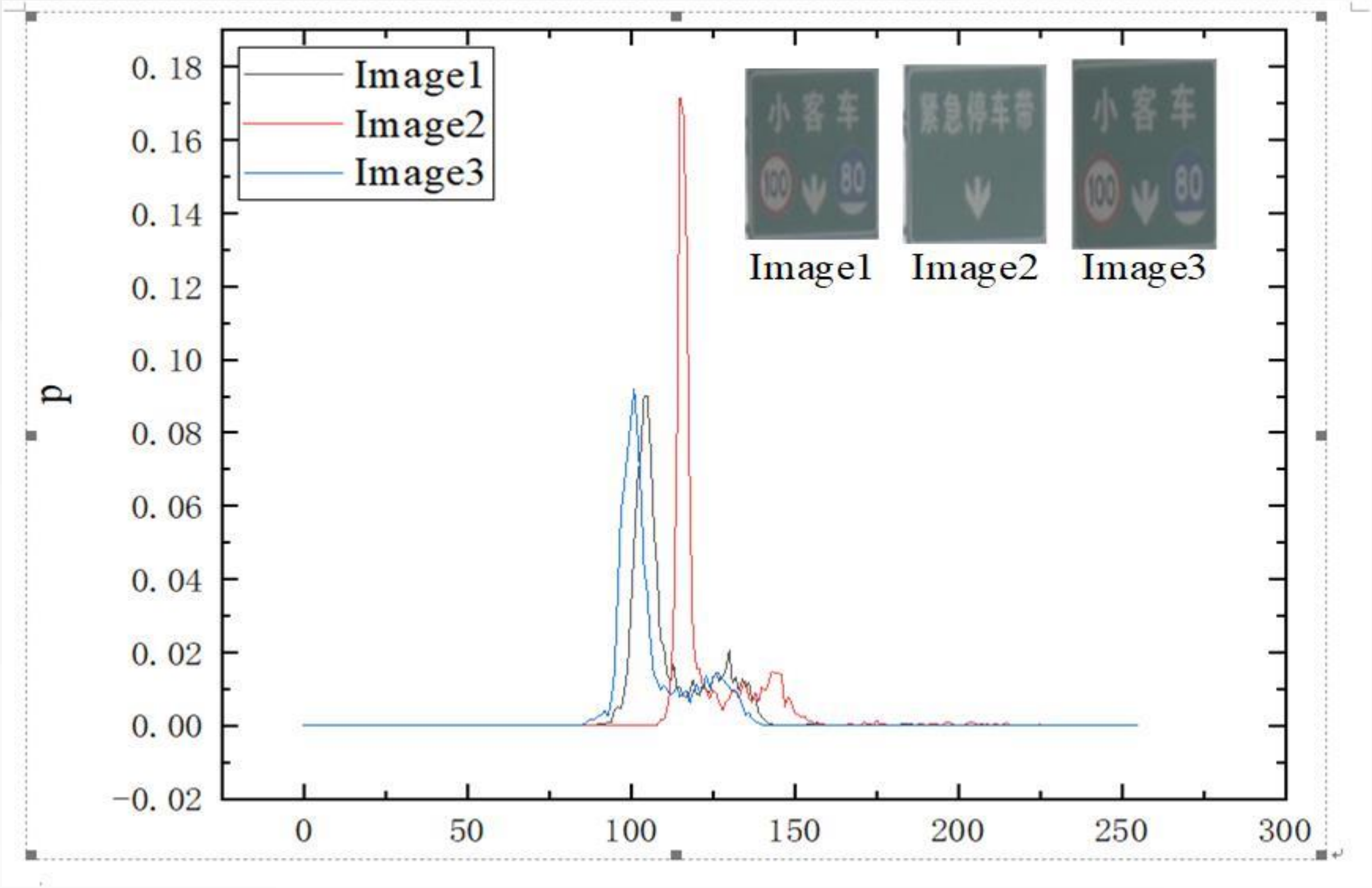


KF model for prediction. When the confidence of the prediction is low, merge the prediction of the LSTM model.

2 algorithm：Fast Road Sign Tracking algorithm

- **C. Appearance Descriptor**

- CNN algorithm to extract the appearance will spend too much time and need offline training
- Adopt a simple color histogram algorithm, which takes advantage of the characteristics of different road signs having different color distribution without external training.



2 algorithm：Fast Road Sign Tracking algorithm

- **D.Assignment Problem**

- Hungarian algorithm is an algorithm for solving general bipartite graph matching to solve the association between the predicted states and newly arrived measurements. Extract motion features and visual features,
- Calculate the distance matrix according to Eq. 1 and Eq. 2
- Use the Hungarian algorithm for matching.

$$d^m(i,j) = \{IoU(d_i, d_j)\} \quad (1)$$

d_j represents the state of the j-th track, and r_i represents the state of the i-th detection.

$$d^a(i,j) = \{r_j^T r_i\} \quad (2)$$

r_j represents the feature of the j-th track, and r_i is the feature of the i-th detection.

3 Experimental results

- **Dataset**

Dataset	Images number	Training Set Numbers	Testing Set Numbers
Dataset1	1013	721	292
Dataset2	239	0	239
Dataset3	239	0	239
Dataset4	238	238	0
Dataset5	239	239	0
Dataset6	239	239	0
Dataset7	238	238	0
Dataset8	1000	1000	0
Dataset9	1000	0	1000
total	4445	2675	1770

- Nine video sequences for 4445 images collected on the highway with different lighting conditions at 30FPS.
- Images are taken on a Millet recorder, including data from nine different highways in Shanghai, China. Dataset-8 and Dataset-9 are collected at night.

3 Experimental results

- **Experiment**

Road sign tracking performance of different models

Detection Models	Estimation Models	Appearance Models	MOTA	Mostly tracked	False Alarm	Runtime(ms)
Yolov3	KF	ResNet50	91.25	90.68	5.21	127
Thunder Net	KF	ResNet50	90.17	89.72	5.74	62
Thunder Net	Accelerated KF	Color	93.54	92.35	3.58	38
Thunder Net	AKL	Color	95.12	94.56	2.40	43
Fusion Detectors	AKL	Color	95.63	94.98	2.31	24

The performance has been significantly improved after replacing it with our algorithm.

3 Experimental results

- **Experiment**

Comparing FRST and the state of the art algorithms

Tracker	MOTA	Mostly tracked	False Alarm	Runtime(ms)
MOTDT	92.37	91.49	4.76	97
JDE	94.98	94.63	2.56	109
FairMot	96.25	95.72	2.49	78
Ours	95.63	94.98	2.31	24

3 Experimental results

- **Experiment**

Comparing performance of different models at night

Tracker	MOTA	Mostly tracked	False Alarm	Runtime(ms)
MOTDT	89.22.	88.36	6.26	99
JDE	90.54	90.73	4.12	112
FairMot	92.98	91.64	3.15	77
Ours	91.29	91.11	3.40	29

4 Conclusion

This paper describes an algorithm for fast detecting and tracking road signs from a sequence of video images. We utilized the ThunderNet and RGD to detect road signs. Furthermore, we combined the KF model with LSTM model to estimate the state of road signs as a new estimation model. Experimental results showed that our algorithm is both accurate and efficient

THANK YOU