Visual Vehicle Tracking Using An Improved EKF*

Jianguang Lou, Hao Yang, Weiming Hu, Tieniu Tan
National Laboratory of Pattern Recognition (NLPR),
Institute of Automation, Chinese Academy of Sciences,
Beijing, 100080 P. R. China
{jglou, hyang, wmhu, tnt}@nlpr.ia.ac.cn

Abstract

In this paper, a dynamic model of car motion is proposed in which the turn of the steering wheel and the distance between the front and rear wheel are taken into account. Extended Kalman Filter (EKF) is widely used in visual tracking systems. However, because there is no direct link between the behaviour of the driver who controls the motion of the car and the assumed dynamic model, the traditional EKF does not perform well when the car carries out a complicated manoeuvre. In order to reduce the sensitivity of the filter to the model uncertainty, we use a modified EKF by adding a new objective function. Experimental results show that the performance of our approach is much better than the traditional EKF and its recent variant.

1. Introduction

For an autonomous visual traffic surveillance system, the ability to track and predict the vehicle motion is important. First, because of the presence of noise and inaccuracies in image data and object models, the final pose determined is often noisy. A filter is needed to obtain a smooth estimation of the tracked vehicle's motion parameters for semantic interpretation in high-level vision. Second, the predictive properties of the filter can be used to get an estimation of pose for the next frame based on the measurement of the preceding frames. An accurate prediction can simplify the measurement process and reduce the computational cost of searching in the object localization modules. In general, the accuracy of tracking and prediction depends on the structure of tracking filters that contain the dynamic model for the car motion.

Extended Kalman filter (EKF) is widely used in visual tracking systems [9, 7, 6, 5, 8, 10,11, 12], because it is an optimal linear recursive filter under certain conditions and can be implemented on-line. Blake etc. [5] uses Kalman filter and a stochastic differential equation model to track visual contours. In [6, 8, 10], the Kalman filter is used to track moving human in video sequences. In [7], a visual vehicle tracking system using Kalman filter is discussed, and in Koller etc. [11] an IEKF is used. Koller etc. [11] adopt a simple model that assumes that the car carries out a circular motion with a constant magnitude of the translational velocity and a constant angular velocity. However, the EKF needs a precise linear dynamic model and prior knowledge about the statistic characteristics of measurement noise. In reality, it is impossible to exactly model the behaviour of a driver who controls the motion of a car, and the dynamic car model is usually time-varying and nonlinear. Therefore, the traditional EKF does not perform well when the car carries out a complicated manoeuvre.

Maybank etc. [1] have proposed a Covariance-Updating filter (CUF) in which the mean and the covariance of the system states are propagated with errors of order $O(t^3)$. In the CUF, the change of steering wheel angle and power of engine are modeled by Brownian motion, and it is assumed that the measurement errors are Gaussian with zero mean and a known covariance matrix. It is reported that the performance of CUF is better than the traditional EKF [1]. However, the CUF is still sensitive to the uncertainty of motion model and fails when the car carries out a complicated motion. In fact, the dynamic model changes over time, so Gaussian process or Brownian motion cannot simply model the change of steering and the power of engine. Furthermore, the initial statistic characteristic of noise is unknown.

In this paper, we modify the EKF by adding a new optimising objective function in order to reduce the sensitivity of the filter to the model uncertainty. Experimental results show that this in general improves tracking performance. In the following, we first outline the overall system for vehicle tracking. The details of the tracking filter are given in Section 3 and experimental results are presented in Section 4.

2. System overview

* This work is funded by the Chinese NSFC.

We have developed a visual vehicle tracking system using a static and pre-calibrated camera. The system consists of the following major modules:

Motion Detection: The result of the motion detection module is regions of interest (ROI) indicating significant motion. Each region that represents a dynamic event is detected by background subtraction [13]. The background image describes the stationary portion of the scene, and stopped cars are considered as parts of the background image. In our system, we use Adaptive Kalman Filter to update the background image over time.

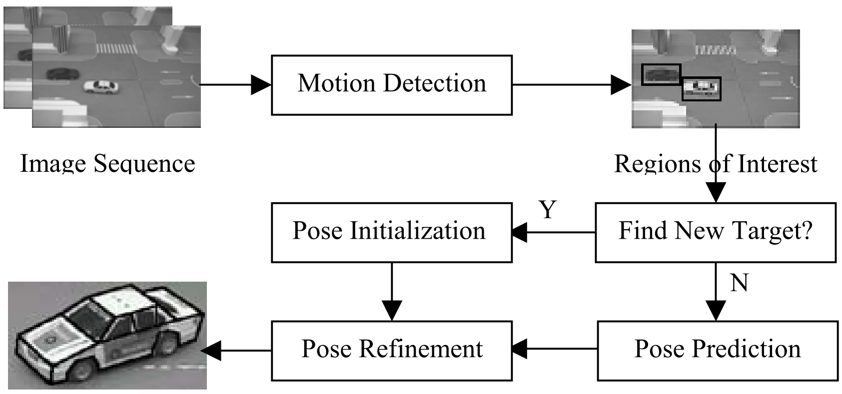


Figure 1. Overview of our tracking system (The camera is static and calibrated, and 3D wireframe models are used.)

Pose Initialization: Issues about pose initialization are discussed in our previous paper [2].

Pose Prediction: It is the main subject of this paper.

Pose Refinement: Ground-Plane Constraint (GPC) [2] is applied in our pose determination, which means that under normal conditions, vehicles are constrained to be in contact with the ground-plane. Furthermore, the pose of a vehicle can be decomposed into two independent 3D components: translation on the Ground-Plane (GP) and rotation around its axis that is perpendicular to the GP. The translation parameters are obtained based on Point-to-Line-Segment Distance (PLS Distance) between an image region and the 3D wire frame model's projection under certain orientation, while rotation parameters are determined by exploiting the geometric relationship among a set of imaginary planes. The details of this pose refinement algorithm are described in another paper [3].

3. The filter

3.1. The kinematics model

In general, the more precise the dynamic model we could obtain, the better performance the filter would conduct. Koller etc. [11] introduce a simple circular motion model. A more precise dynamic model is studied in Maybank etc. [1]. We extend the model. The car is controlled by the driver through varying the steering angle φ and changing the velocity v. Most vehicles are driven by their rear wheels, while the front wheels are only used to turn the direction of the vehicle. In this paper, we use a Two-Point-Bicycle model to simulate the motion of the car. As shown in Figure 2, $[\ x,\ y\ ]^T$ is the position of the car on the ground plane, v is the velocity of the rear wheel, θ is the orientation of the whole car, φ is the orientation of the front wheel. In this model, it is assumed that the car is rigid, in other words, the distance between the front wheel and the rear wheel does not change over time. It is also assumed that the wheels of the car cannot slip sideways.

According to these constraints, we can obtain the speed of the center point O:

Magnitude $\quad \dfrac{v}{2\cos\phi}\sqrt{1+3\cos^2\phi}$

Direction $\quad \theta + arctg(\dfrac{tg\phi}{2})$

Therefore, the motion can be modeled by a fifth-order dynamic process $X=[\ x,\ y,\ v,\ \theta,\ \varphi\ ]^T$, and X is called state vector. Let l be the wheelbase of the car (typically $l \approx$ 2.5m), and we assume that the measurement noise can be approximated by White Noise e. The dynamic model of the motion can then be described as follows:

$$\begin{cases} \dot{x} = \dfrac{v}{2\cos\phi}\sqrt{1+3\cos^2\phi}\cos[\theta + arctg(\dfrac{tg\phi}{2})] \\ \dot{y} = \dfrac{v}{2\cos\phi}\sqrt{1+3\cos^2\phi}\sin[\theta + arctg(\dfrac{tg\phi}{2})] \\ \dot{v} = a \\ \dot{\theta} = \dfrac{v\cdot tg\phi}{l} \\ \dot{\phi} = b \end{cases} \tag{1}$$

or collectively as $dX/dt=f(X)$, and the measurement model as

$$\begin{bmatrix} x \\ y \\ \theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ v \\ \theta \\ \phi \end{bmatrix} + e \tag{2}$$

or $Y=h(X)$

In this dynamic model, a and b are used to describe the behaviour of the driver. a reflects the driver's press on the accelerator or the brake, or changing the position of the gear; b reflects the turn of the steering wheel. a

and b are the parameters which depend on the driver's activity and can not be modeled easily. The problem we have is to estimate the state vector of a time-varying nonlinear system.

3.2. A new objective function

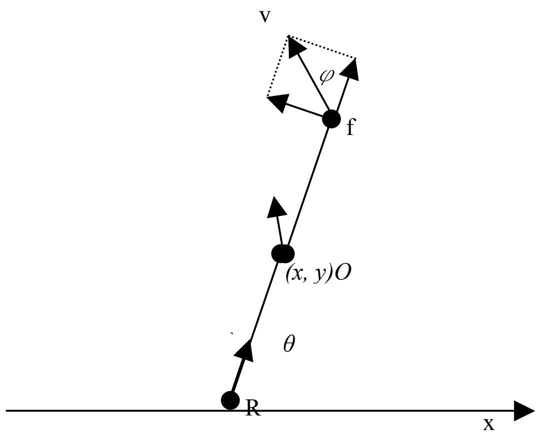


Figure 2. The bicycle model of car motion

Our filter is based on EKF. In order to improve the robustness of the filter against the model's inaccuracy, we modify the EKF by adding a new objective function. The same objective function has been developed elsewhere by Zhou etc. [4] in the field of system's diagnosis. We extend the filter to be applied in visual vehicle tracking, and also add a weighted matrix for handling the situation that x, y and θ have different influences on the efficiency of the pose refinement algorithm [3].

Extended Kalman Filter is an optimal filter that minimizes the square error between system's state vector X and the predicted state vector $\hat{X}$:

$$E[(X_{k+1}-\hat{X}_{k+1|k+1})^T(X_{k+1}-\hat{X}_{k+1|k+1})]=minimum \quad (3)$$

When the system model is time-invariant and linear, the traditional Extended Kalman Filter can work well. But when the system model is time-varying, the EKF cannot estimate the system's states accurately.

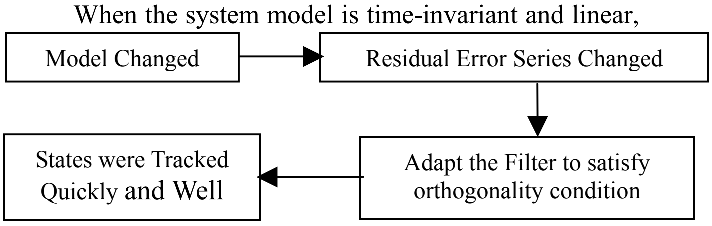


Figure 3. The idea of the filter

In general, the measurement noise e is White Noise. To adapt the EKF to the change of the system model, a new criterion is added which forces the residual error series γ to satisfy an orthogonality condition:

$$E(\gamma_{k+1}^T\alpha\gamma_{k+1+j})=0, \qquad k=0,1,2,\cdots, j=1,2,3,\cdots \quad (4)$$

where α is a weight diagonal matrix. Thus, each residual error has a different weight.

The idea is that once the model has changed, the residual error series would change immediately, and then we adapt the filter to satisfy the orthogonality condition (just like White Noise) in order that the filter's estimated states can track the system's real states quickly and accurately (see Figure 3). If the model's parameters match the real system, the orthogonality condition will be self- satisfying for the EKF. But if the model changes over time, the traditional EKF's residual errors do not satisfy the orthogonality condition, and they reflect the instability of the model parameter. Because the measurement noise is assumed as White Noise, the residual error process should be White Noise. We adapt the filter to make sure that the residual error series has the similar characteristic with White Noise in order that the estimated states of the filter can track the system's states as quickly as the system's parameters change. For a real nonlinear system, the orthogonality condition (4) can only be satisfied approximately.

3.3. Incorporate the orthogonality condition into EKF

Because our approach is based on EKF, propagation equations of the traditional EKF are listed below for self-completeness.

State estimation equation: $\hat{X}_{k+1|k+1}=\hat{X}_{k+1|k}+K_{k+1}\gamma_{k+1}$ (5)

State prediction equation: $\hat{X}_{k+1|k}=f(\hat{X}_{k|k})$ (6)

Gain matrix:

$$K_{k+1}=P_{k+1|k}H_{k+1}^T(\hat{X}_{k+1|k})[H_{k+1}(\hat{X}_{k+1|k})P_{k+1|k}H_{k+1}^T(\hat{X}_{k+1|k})+Q_{k+1}]^{-1} \quad (7)$$

Covariance matrix prediction:

$$P_{k+1|k}=F_k(\hat{X}_{k|k})P_{k|k}F_k^T(\hat{X}_{k|k})+V_k \quad (8)$$

Covariance matrix:

$$P_{k+1|k+1}=[I-K_{k+1}H_{k+1}(\hat{X}_{k+1|k})]P_{k+1|k} \quad (9)$$

Residual error: $\gamma_{k+1}=Y_{k+1}-h_{k+1}(\hat{X}_{k+1|k})$ (10)

where F is the discrete form of f, H is the discrete form of h, K is the gain matrix, P is the covariance matrix of state vector error, Q is the covariance of the process noise, and V is the covariance of the measurement error.

We can adapt the filter to satisfy the orthogonality condition by adjusting the EKF's gain matrix on-line. This is achieved by using a fading parameter, that is, the covariance matrix prediction equation is updated to:

$$P_{k+1|k}=\lambda_{k+1}F_k(\hat{X}_{k|k})P_{k|k}F_k^T(\hat{X}_{k|k})+V_k \quad (11)$$

where λ is the fading parameter.

3.4. Determine the fading parameter

Now, we need to determine the fading parameter λ that can make the filter satisfy the orthogonality condition. First, we give two lemmas.

Lemma 1:

$$E(\gamma_{k+1}^T \alpha \gamma_{k+1+j}) = 0 \quad \Leftrightarrow \quad tr(E(\gamma_{k+1+j}\gamma_{k+1}^T \alpha)) = 0$$

Proof: $E(\gamma_{k+1}^T \alpha \gamma_{k+1+j}) = 0 \quad \Leftrightarrow tr[E(\gamma_{k+1}^T \alpha \gamma_{k+1+j})] = 0$

$\Leftrightarrow \quad E[tr(\gamma_{k+1}^T \alpha \gamma_{k+1+j})] = 0 \quad \Leftrightarrow E[tr(\gamma_{k+1+j}\gamma_{k+1}^T \alpha)] = 0$

$\Leftrightarrow \quad$ tr[$E(\gamma_{k+1+j}\gamma_{k+1}^T \alpha)$] = 0

By using the similar method as Zhou etc. [4], we can get a characteristic of the EKF as follows.

Lemma 2:

If $\left|\Delta X_k\right| \underline{\underline{\Delta}} \left|X_k - \hat{X}_{k|k}\right| << \left|X_k\right|$ then

$$E(\gamma_{k+1+j}\gamma_{k+1}^T \alpha) \approx G \times (P_{k+1|k} H_{k+1}{}^T - K_{k+1} U_{k+1}) \alpha$$

where $U_{k+1} = H_{k+1} P_{k+1|k} H_{k+1}{}^T + V_{k+1}$ and G is a nonzero function.

According to the above two lemmas, $(P_{k+1|k} H_{k+1}{}^T - K_{k+1} U_{k+1}) \alpha \equiv 0$ is the sufficient precondition for the orthogonality objective function. Thus, the parameter λ can be obtained by minimizing the expression $(P_{k+1|k} H_{k+1}{}^T - K_{k+1} U_{k+1}) \alpha$ using some optimization algorithms. But, in general, optimization algorithms are always time-consuming. We only calculate the suboptimal parameter for λ similarly as Zhou etc. [4] have done. The suboptimal estimation of λ is calculated using the following expressions.

$$\lambda_{k+1} = \begin{cases} \lambda_0, & \lambda_0 \geq 1, \\ 1, & \lambda_0 < 1, \end{cases}$$

where $\lambda_0 = \dfrac{tr(N_{k+1})}{tr(M_{k+1})}$

$$N = \alpha U_{k+1} - H_{k+1}(\hat{X}_{k+1|k}) Q_k H_{k+1}^T(\hat{X}_{k+1|k}) - V_{k+1}$$

$$M = H_{k+1}(\hat{X}_{k+1|k}) F_k(\hat{X}_{k|k}) P_{k|k} F_k^T(\hat{X}_{k|k}) H_{k+1}^T(\hat{X}_{k+1|k})$$

3.5. Adapt the noise covariance

The traditional EKF needs some prior knowledge about the statistic characteristics of the measurement noise and the process noise. But, in reality, such information is hard to obtain. Furthermore, the statistic characteristic may change slowly over time. To handle this situation, we should update the noise covariance as time goes on. The covariance matrices are updated using the following equations.

$$Q_{k+1} = (1 - \frac{1-\beta}{1-\beta^{k+1}}) Q_k + \frac{1-\beta}{1-\beta^{k+1}} [\gamma_{k+1}\gamma_{k+1}^T$$
$$- H_{k+1}(\hat{X}_{k+1|k}) P_{k+1|k+1} H_{k+1}^T(\hat{X}_{k+1|k})]$$

$$V_{k+1} = (1 - \frac{1-\beta}{1-\beta^{k+1}}) V_k + \frac{1-\beta}{1-\beta^{k+1}} [K_{k+1}\gamma_{k+1}\gamma_{k+1}^T K_{k+1}^T$$
$$+ P_{k+1|k} - F_k(\hat{X}_{k|k}) P_{k|k} F_k^T(\hat{X}_{k|k})]$$

where β is the vanishing factor, with a typical value between 0.8 and 1.0.

When k is a small number, the measurements play an important role in the estimation of the noise characteristic. As the number k becomes larger, the influence of the measurements to the noise covariance reduces.

4. Experimental Results

4.1. Predictive Tracking

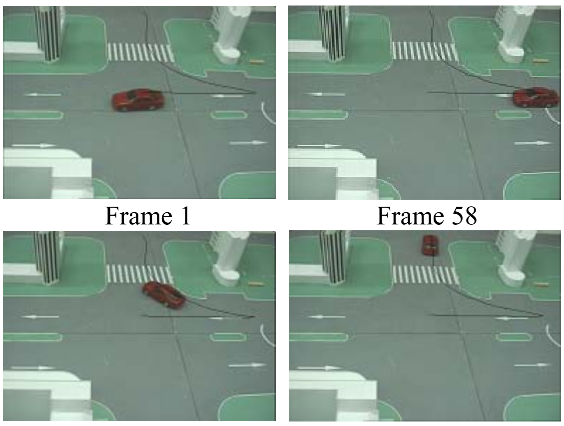
The filter for vehicle tracking has been implemented using C++ under Microsoft Windows. The performance of this filter is then assessed in comparison with a traditional EKF and the CUF proposed by Maybank etc. [1]. Typical vehicle behaviour has been carried out for experimental assessment of the filter. In the test video sequence, a car reverses and rapidly slows down, stops, then, turn right round the intersection (see Figure 4).

Each measurement has different influence on the pose refinement algorithm. In our pose refinement algorithm, we firstly obtain the translation parameters using the PLS Distance (Point-Line-Segment Distance) between the ROI and the 3D wire frame model's projection under the predicted orientation θ, and then, we use VNC (Vertex Neighbourhood Constraint) to determine the rotation parameters[3]. Efficiency analysis of the refinement algorithm shows that the influence of the prediction error of θ is much larger than that of x or y. In our filter, the proportion among the errors can be changed by tuning the weight matrix α as mentioned above. Matrix α is a diagonal matrix and has three parameters which are noted as α_x , α_y and α_θ. There is a constraint *(* $\alpha_x + \alpha_y + \alpha_\theta = 3$*)* to maintain the convergence of the filter. We can select different values for α_x , α_y and α_θ and have a trade-off among the influences of the prediction errors. In our system, $\alpha_x = \alpha_y = 0.1$ and $\alpha_\theta = 2.8$, other similar values may be used.

4.2 Performance comparison

Figure 5 shows the tracking performance of three filters (the traditional EKF, the CUF and our new filter). The diagrams in the left column are obtained when the car carries out a circular motion with a constant velocity, and the diagrams in the right column are obtained when the car carries the motion described in Figure 4. In the video sequence, the car began to slow down at frame 50, began to turn right at about frame 70 and to speed up. In other words, the velocity v and the orientation θ changed dramatically between frame 50 and frame 100. In the diagrams, the CUF has a performance better than the traditional EKF, and they both have relatively large errors from frame 50 to 120, while our new filter can work well between frame 50 and frame 100.The right column shows that when the driver of the car carries out some dramatic behaviors, the CUF and EKF cannot predict the states of the car's motion very well. However, our filter can track the change of the driver's behavior fairly accurately. The traditional EKF is a degenerated form of our filter where the fading parameter $\lambda=1$. When the car carries a smooth motion, the fading parameter will converge to $\lambda=1$, and the new filter has a performance almost the same as traditional EKF, as it is demonstrated by the left column in Figure 5. The left column in Figure 5 also shows that the CUF performs best when the car carries a smooth motion as described in [1]. If we could combine the CUF filter with our filter, we would probably obtain much better result, so we intend to explore this in our future work.

Another sequence in Figure 6 is also demonstrated which is captured by a PanasonicTM NV-DX100 digital video camera at 5 frames per second. In this sequence, the car moves forward slowly after it enters the field of view, then it accelerates suddenly and turns right, and finally, it stops near the intersection. Figure 7 shows the tracking result of this sequence.



Frame 1 Frame 58

Frame 96 Frame 148

Figure 4. The test sequence I

From Figure 7 (a), we can see that the prediction error of our approach is less than that of the traditional EKF. By comparing (b) and (c), we can note the relationship between the fading parameter and the change of the car's speed. When the speed changed dramatically, a fading parameter $\lambda>1$ is applied to adapt the filter, however, in the traditional EKF, λ is always 1.

The results show that our filter is much more robust when the car's motion behavior changes suddenly.

5. Conclusion

In this paper, we have described a modified Extended Kalman Filter incorporated with a precise kinematics model for visual vehicle tracking. A Two-Point-Bicycle model discussed above is used to model the motion of the car, which is accurate enough for the pose prediction filter. By adding an additional orthogonality condition, the filter has less sensitivity to the time varying model of system. Because the filter is based on EKF, the computational cost is low. Experiments show that the filter has a good performance when the tracked car carries a complex motion. Therefore, this filter is well suited for application in visual traffic surveillance.

6. References

[1] S. J. Maybank, A. D. Worrall, and G. D. Sullivan, "Filter for car tracking based on acceleration and steering angle", *Proc. of British Machine Vision Conference*, Sep. 1996, pp. 615--624.

[2] T.N. Tan and K.D. Baker, "Efficient Image Gradient Based Vehicle Localization", *IEEE Tran. on Image Processing*, Vol.9, No.8, Aug. 2000, pp.1343-1356.

[3] H. Yang, J.G. Lou, H.Z. Sun, W.M. Hu and T.N. Tan, "Efficient and Robust Vehicle Localization", *IEEE Int. Conf. on Image Processing (ICIP 2001)*, Sep. 2001 (accepted).

[4] D.H. Zhou, Y.G. Xi and Z.J. Zhang, "A Nonlinear Adaptive Fault Detection Filter", *Int. Journal of System Science*, Vol22 (12) , 1991, pp. 2563-2571.

[5] A. Blake, M. Isanc and D. Reynard, "Learning to track the visual motion of contours", *Artificial Intelligence*, Vol. 78, 1995, pp. 101-133.

[6] I. Haritaoglu, D. Harwood, and L.S. Davis, "W4: Who? When? Where? What? A real time system for detecting and tracking people", *3rd Int. Conf. on Face and Gesture Recognition*, Nara, Japan, 1998.

[7] D. Koller, J. Weber, J. Malik, "Robust Multiple Car Tracking with Occlusion Reasoning", *Proc. of Europe Conference on Computer Vision*, 1994, vol. A, pp. 189-196.

[8] R. Rosales and S. Sclaroff. "Improved tracking of multiple humans with trajectory prediction and occlusion modeling", *CVPR Workshop on the Interpretation of Visual Motion*, 1998.

[9] I.J. Cox and S. L. Hingorani, "An Efficient Implementation and Evaluation of Reid's Multiple Hypothesis Tracking Algorithm for Visual Tracking", *Proc. of IAPR Int. Conf. on Pattern Recognition*,1994, pp.437-442.

[10] C. Bregler, "Learning and Recognizing Human Dynamics in Video Sequences", *Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition*, 1997, pp. 568-574.

[11] D. Koller, K. Daniilidis and H.-H. Nagel, "Model-Based Object Tracking in Monocular Image Sequences of Road

Traffic Scenes", *Int.l Journal of Computer Vision*, vol. 10(3) , 1993, pp. 257-281.
[12] Robert T. Collins, Alan J. Lipton and Takeo Kanade, "A System for Video Surveillance and Monitoring", *Proc. of the American Nuclear Society (ANS) Eighth Int. Topical Meeting on Robotics and Remote Systems*, Apr. 1999.
[13] H. Z. Sun, T. Feng and T. N. Tan, "Robust extraction of moving objects from video sequences", *Proc. of the Fourth Asian Conference on Computer Vision*, 2000, pp. 961-963.

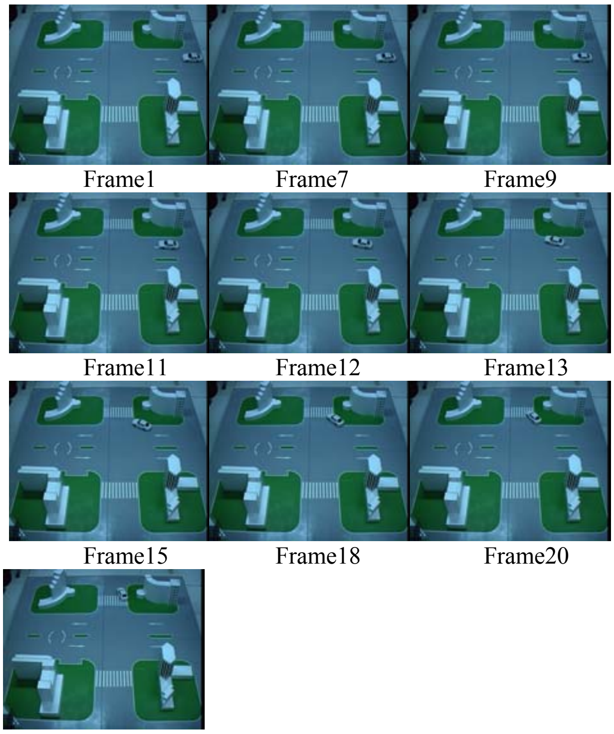


Figure 6. Test Sequence II

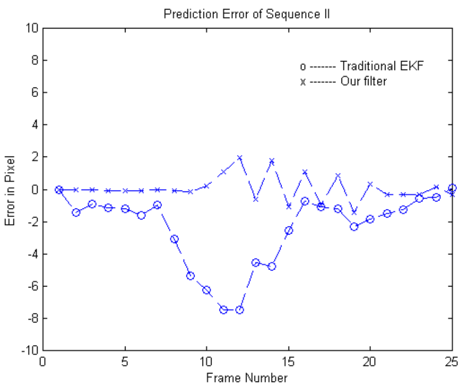


Fig7.(a)

Figure 7. The result of sequence II (a) Prediction Error, (b) The fading parameter,(c) The displacement of the car between two sequential frames over time, we use this to manifest the variation of car's speed.

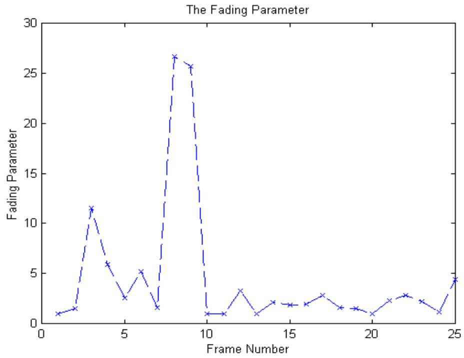


Fig7. (b)

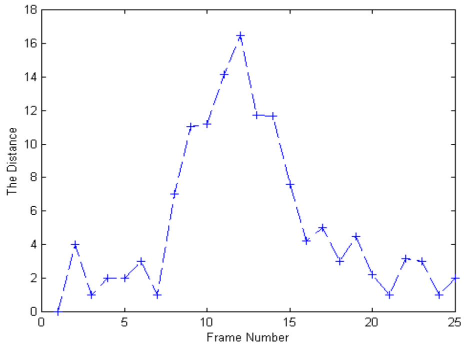


Fig7. (c)

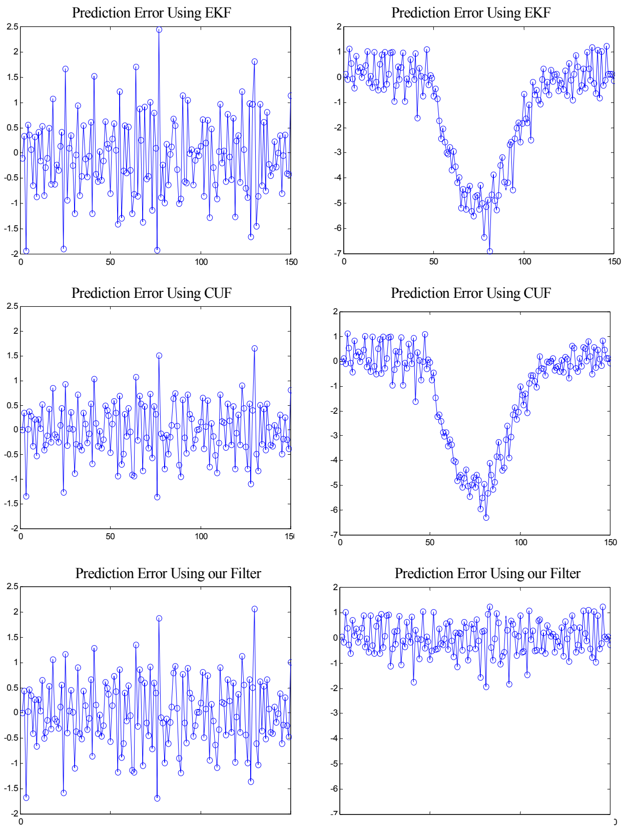


Figure 5. Prediction errors (between prediction and ground truth) of an EKF predictor,