SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

June 2017, Vol. 60 062102:1–062102:12
doi: 10.1007/s11432-015-5450-3

Adaptive multiple video sensors fusion based on decentralized Kalman filter and sensor confidence

Qingping LI, Junping DU*, Suguo ZHU & Liang XU

Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876*, China*

Received April 25, 2016; accepted May 18, 2016; published online February 8, 2017

Abstract The fusion of multiple video sensors provides an effective way to improve the robustness and accuracy of video surveillance systems. In this paper, an adaptive fusion method based on a decentralized Kalman filter (DKF) and sensor confidence is presented for the fusion of multiple video sensors. The adaptive scheme is one of the approaches used for preventing the divergence problem of the filter when statistical values of the measurement noises of the system models are not available. By introducing the sensor confidence, we can adaptively adjust the measurement noise covariance matrix of the local DKFs and thus, determine the weight of each sensor more correctly in the fusion procedure. Also, the DKF applied here can make full use of redundant tracking data from multiple video sensors and give more accurate fusion results in an efficient manner. Finally, the fusion result with improved accuracy is obtained. Experimental results show that the proposed adaptive decentralized Kalman filter fusion (ADKFF) method works well in the case of real-world video sequences and exhibits more promising performance than single sensors and comparative fusion methods.

Keywords video sensors fusion, decentralized Kalman filter, target tracking, sensor confidence, video surveillance

Citation Li Q P, Du J P, Zhu S G, et al. Adaptive multiple video sensors fusion based on decentralized Kalman filter and sensor confidence. Sci China Inf Sci, 2017, 60(6): 062102, doi: 10.1007/s11432-015-5450-3

1 Introduction

With the significant progress in sensor technology and its applications, an increased requirement to improve the ability to remotely monitor a complex environment has made automatic surveillance systems a promising research field over the last few decades [1, 2]. One of the most important goals of these recent surveillance systems is the automatic assessment of what is happening in the monitored scene and presenting suspect cases to the system. These tasks can be carried out by getting the targets' trajectories in the monitored place and analyzing them by comparing them with the existing well-studied patterns [3].

To obtain trajectories with better robustness and accuracy of the moving targets in the monitored place, multiple sensors are placed in the same area to produce redundant information and used to improve the monitoring accuracy of the surveillance [4, 5]. Then, a data fusion technique can be used for improving the performance of the systems by taking advantage of the redundant trajectory data from the source

* Corresponding author (email: junpingdu@126.com)

sensors [6, 7]. With the development of multisensor systems and communication techniques, data fusion has been increasingly used in surveillance systems [8].

Sensor fusion can be applied in different ways and at different levels. Signal-level fusion is the lowest level of sensor fusion; it involves the use of fusion techniques to generate a composite image. However, images need to be spatially registered correctly before fusion can actually be performed [9, 10]. Feature-level sensor fusion is a higher-level fusion method; here, the sensors to be fused can have very different fields of view, and thus, signal-level fusion cannot be applied to them [11]. Accordingly, it is necessary to extract features, such as position, velocity, trajectory, and posture, from source sensors that are to be fused in a shared representation format.

As mentioned above, we focus on the fusion of the trajectory feature in this paper, such a fusion can be achieved by extracting data regarding the position of a target and projecting them to a shared map of the monitored environment. There are a host of multisensor fusion methods that can accomplish this task. One traditional approach is called centralized Kalman filter (CKF) fusion. It sends all the observation source data from the local sensors to a fusion center to generate the fused estimates and involves minimal information loss [12]. However, in this method, all the observation data are treated as one observation matrix and the fusion center bears most of the computational burden. Also, the CKF may be relatively less accurate and stable when it encounters serious data errors or heavy noise [13]. Another classic method is federated Kalman filter fusion, which can generate a more accurate fused estimate using information sharing factors (ISF) [14]. However, ISF are calculated using a covariance matrix, which invariably contains some estimation errors [15].

Also, researchers have proposed many other approaches, such as fusion techniques using steady-state Kalman filters [16], covariance intersection (CI) Kalman fuser [17], standard DKF fusion [18], and sequential CI fusion [19]. Among them, the DKF fusion approach is more robust, flexible, and efficient, and it is widely used in multisensory data fusion. Also, some adaptive fusion methods have been proposed to obtain more robust and accurate fusion results; these adaptive strategies include weighted strategies [15, 20] and fuzzy logic-based strategies [21, 22]. However, these fusion methods are barely applied in real-world video sensor fusion and do not consider the confidence of the source sensors; the data from a malfunctioning sensor will affect the fusion process and lead to an inaccurate result. The existing confidence function of a source sensor [23, 24] cannot always be employed in most of the practical video surveillance situations [25].

Aimed at solving these problems, a feature-level approach of adaptively fusing video sensors in a surveillance system is proposed in this work. The main contribution of this paper is the introduction of the video sensor confidence to adaptively adjust the measurement noise covariance of local DKFs in the fusion procedure and then, apply the adaptive fusion method to real-world video sensor fusion. This adaptive approach can give a more accurate weight of each video sensor in the fusion procedure and then, perform adaptive fusion on the basis of these weights. Also, the DKF processes the sensor measurements locally to produce local estimates. This can reduce the burden of the fusion center and thus, improve the robustness and efficiency of the system. The proposed method is also compared with other fusion techniques, and a more thorough experimental section is presented.

The remainder of the paper is organized as follows. Section 2 gives a brief review of the video sensor confidence function called appearance ratio (AR). Section 3 describes the proposed adaptive video sensor fusion method in detail. Experimental results and analysis are discussed in Section 4, and the conclusion is presented in Section 5.

2 Confidence function: AR

In a surveillance system, different quality of sensors, view angles, distances between sensor and target, and light conditions can differently affect the ability of the sensors to detect a moving target. The fusion process generally employs a weighting strategy to adjust different source measurements. Making no distinction between the measurements from the source sensors can affect the stability of the filter

Figure 1 (Color online) Frames of detection results from underpass video sequences. The first row is from camera 1 sequence, and the second row is from camera 2 sequence.

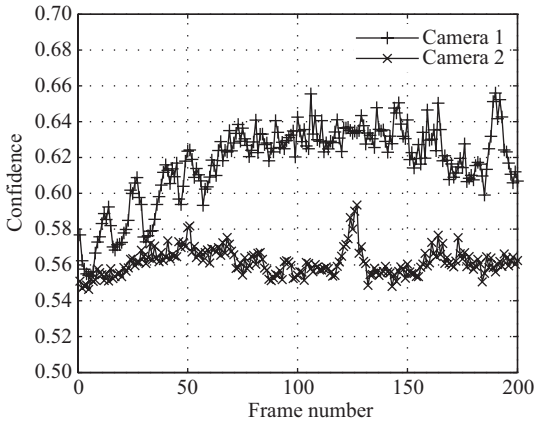

Figure 2 AR values of the detection results according to camera 1 and camera 2 for the underpass video sequences.

and yield an inaccurate estimate [3]. Therefore, to obtain a fused result with better performance than the result achieved from an individual sensor, we cannot equally weight the measurements of source sensors [25].

The AR measurement described in [26] establishes a model that estimates the discernibility of an extracted block from the referenced background. It indicates the confidence level of the detected blocks $b_{j,t}^m$, where j denotes the number of blocks, t represents the time, and m indicates the number of sensors. Then, the AR is given as follows:

$$\mathrm{AR}(b_{j,t}^m) = \frac{\sum_{\bar{c}} \sum_{x,y \in b_{j,t}^m} D_{\bar{c}}(x,y)}{\sum_{\bar{c}} \sum_{x,y \in b_{j,t}^m} \delta(R_{\bar{c}}(x,y))}, \tag{1}$$

where D denotes the absolute difference map between the current frame and the referenced background R, $\bar{c}$ represents the number of color bands of the associated frames, and δ refers to the spread of each pixel, which is defined as (2). In this paper, we use the RGB (red, green and blue) model; therefore, $\bar{c}$ represents one of the three color bands (red, green or blue). Therefore, $D_{\bar{c}}(x,y)$ denotes the value of band $\bar{c}$ of the difference map at position (x,y), and $R_{\bar{c}}(x,y)$ represents the value of band $\bar{c}$ of the referenced background image at pixel (x,y),

$$\delta(R_{\bar{c}}(x,y)) = \max(R_{\bar{c}}(x,y), 255 - R_{\bar{c}}(x,y)). \tag{2}$$

As can be seen from (2), δ computes the maximum difference that pixel (x,y) of the current frame image can show against the corresponding pixel of the referenced background. Therefore, D denotes the real absolute difference between the current frame and the background while δ represents the maximum difference. Then, AR is generated using (1); it ranges in the interval [0,1], which indicates the confidence level of the detected results from the source sensors.

Figure 1 shows some frames of the detection results of video sequences called underpass. In these videos, a person is moving through an underpass with poor illumination conditions. Figure 2 gives the confidence values (AR) of the detected results associated with the moving people in Figure 1. From Figure 2, we can see that the confidence curve of the detected blocks from camera 1 is higher than that from camera 2. This coincides with the fact that the person is more discernible in the illumination condition of camera 1 as he moves through the underpass.

3 Adaptive multiple video sensors fusion architecture

A schematic representation of the proposed adaptive fusion method for video surveillance is presented in Figure 3. A tracking technique is applied for detecting moving targets for each sensor, and it provides the

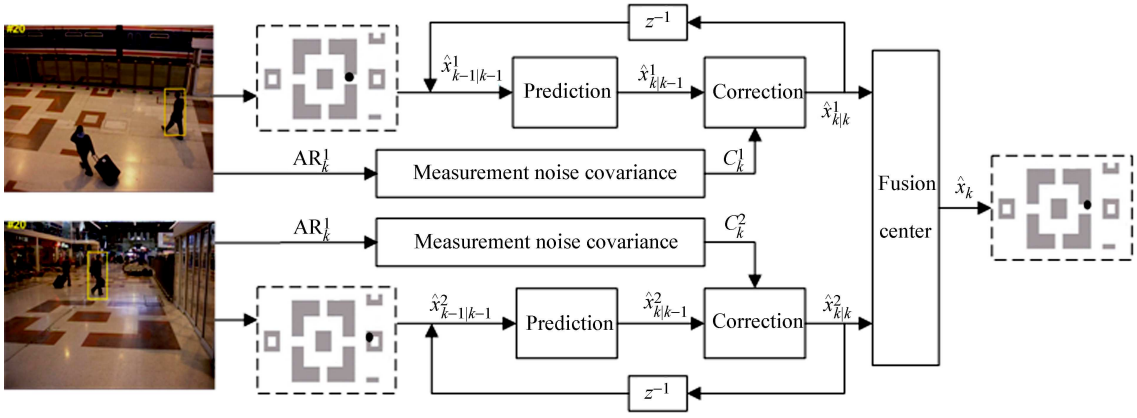

Figure 3 (Color online) Schematic representation of the proposed adaptive multiple video sensors fusion framework.

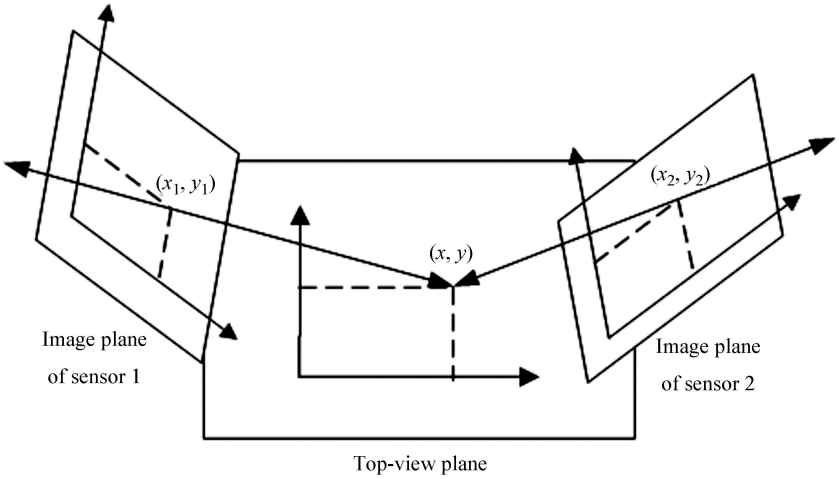

Figure 4 Homographic transformation between camera planes and top-view plane.

position of the detected target. Then, the positions of the targets are projected onto a shared top-view map of the observed area by homographic transformation and filtered by the local adaptive DKFs to get the optimal estimates. Finally, a fused estimate with more robustness is obtained by fusing all local estimates in the fusion center. For the sake of simplicity, here, we consider two sensors, but this method can be extended to more than two sensors.

3.1 Target's position projection

In order to fuse the position data from different video sensors, a shared coordinate frame is required. Here, the position data from a source sensor are projected onto a shared top-view map by homographic transformation, as illustrated in Figure 4. The shared top-view map usually covers the monitored area of the surveillance system.

The homographic transformations can be easily obtained from the connection between the salient points of the camera planes and the corresponding points of the top-view plane [27]. It will be set up in the surveillance system and give the projection between camera planes and the top-view plane.

When homographic transforms are found, the projection can be performed. The position (x_c, y_c) is the center of the target detected by the tracker. In this study, we project this point onto the bottom of the bounding box enclosing the target and suppose that point (x_g, y_g) touches the ground. Then, point (x_g, y_g) from each sensor is projected onto the top-view map by homographic transformation.

3.2 Adaptive decentralized Kalman filter fusion

In this work, an adaptive decentralized Kalman filter fusion (ADKFF) method is proposed to carry out the fusion task. This approach can adaptively adjust the measurement noise covariance matrix of the local DKFs and thus, determine the weight of each sensor more correctly in the fusion procedure. Then, the fused result will benefit from the source sensors with higher confidence.

When performing this fusion method, the predicted states $\hat{x}^m_{k|k-1}$ are first obtained from the estimated state $\hat{x}^m_{k-1|k-1}$ at time $k-1$, where $m = 1, 2, \ldots, M$ denotes the number of source sensor. Then, with the observations z^m_k and sensor confidence AR^m_k, these predicted states are corrected and the corrected estimates $\hat{x}^m_{k|k}$ at time k are obtained. Meanwhile, the corrected estimates are fed back to the prediction step for the next iteration. Finally, a fused state $\hat{x}_k$ at time k is generated with the local estimates $\hat{x}^m_{k|k}$.

3.2.1 *Local DKFs*

In a situation that mainly focuses on the fusion method, we assume that the local DKF process has a state vector $x \in \Re^n$ and the process is governed by

$$x_k = A_{k-1}x_{k-1} + w_{k-1}, \tag{3}$$

with a measurement $z \in \Re^r$ that is defined by

$$z_k = H_k x_k + v_k. \tag{4}$$

In the above equations, the random variables w_k and v_k represent the process and measurement noise with the covariance Q_k and C_k, respectively; they are zero mean Gaussian white noise having zero cross-correlation with each other [28]. The state transition matrix A_{k-1} relates the state at the previous time step $k-1$ to the state at the current time step k with process noise w_k. H_k represents the observation transition matrix and relates the state x_k to the measurement z_k.

In practical object tracking, $\hat{x}^m_{k-1|k-1}$ denotes the state of sensor m $(m = 1, 2, \ldots, M)$ at time $k-1$. Then, the prediction $\hat{x}^m_{k|k-1}$ is generated by (5) and the corresponding prior estimate error covariance $P^m_{k|k-1}$ is given by (6):

$$\hat{x}^m_{k|k-1} = A_{k-1}\hat{x}^m_{k-1|k-1}, \tag{5}$$

$$P^m_{k|k-1} = A_{k-1}P^m_{k-1|k-1}A^{\mathrm{T}}_{k-1} + Q_{k-1}. \tag{6}$$

With these predictions, the estimate of the next time $\hat{x}^m_{k|k}$ is obtained as follows:

$$\hat{x}^m_{k|k} = \hat{x}^m_{k|k-1} + K^m_k(z^m_k - H^m_k\hat{x}^m_{k|k-1}), \tag{7}$$

$$K^m_k = P^m_{k|k-1}(H_k)^{\mathrm{T}}[H_kP^m_{k|k-1}(H_k)^{\mathrm{T}} + C^m_k]^{-1}, \tag{8}$$

$$P^m_{k|k} = [I - K^m_k H_k]P^m_{k|k-1}, \tag{9}$$

where K^m_k denotes the local DKF gain matrix for sensor m at time k and $P^m_{k|k}$ represents the corresponding posteriori estimate error covariance.

3.2.2 *Adjustment of measurement error covariance matrix*

In the procedure described above, the measurement error covariance matrix C models the uncertain and inaccurate information of the filter. It reflects the precision of the source sensors and plays an important role in the state estimate. In the traditional method, this matrix is usually set a fixed value, which implies that the corresponding sensor is also set a fixed confidence. This will significantly affect the fusion result. For solving this problem, in this study, the sensor confidence is applied to adaptively adjust the measurement error covariance matrix. It is defined by

$$C^m_k = \begin{pmatrix} c^{xx}_{m,k} & 0 \\ 0 & c^{yy}_{m,k} \end{pmatrix}. \tag{10}$$

It is assumed that the measurement error is not cross-correlated. Thus, we set c^{xy} and c^{yx} to 0. The function for the c^{xx} and c^{yy} is defined as follows:

$$c^{xx}_{m,k} = c^{yy}_{m,k} = GD \times [1 - \mathrm{AR}(b^m_{j,k})], \tag{11}$$

where GD denotes the maximum value of the measurement error variance and is usually assigned an experimental value to determine the measurement error covariance matrix according to a specific situation. The measurement error covariance is then adjusted using this function and the position of the detected targets with a relatively high AR values is assigned more trust. Otherwise, the detected targets with relatively small AR values are assigned less trust.

The measurement with a relatively small eigenvalue of matrix C^m_k for the corresponding sensor m has a greater precision and vice versa. Then, AR values from the source sensors can be used for adjusting the measurement error covariance matrix through which source estimates will be given a different weight in the fusion procedure. The matrix C^m_k will finally affect the fused estimate by influencing the corrected estimate $\hat{x}^m_{k|k}$, the local Kalman filter gain matrix K^m_k and their corresponding estimate error covariance matrix $P^m_{k|k}$, as shown in (7)–(9).

3.2.3 *Fusion center*

When the local prediction $\hat{x}^m_{k|k-1}$, corrected estimate $\hat{x}^m_{k|k}$, the corresponding error covariance $P^m_{k|k-1}$ and $P^m_{k|k}$ are ready, a fused estimate $\hat{x}_k$ can be generated from the fusion center. The fusion center also has two steps, namely prediction step and correction step. The prediction step is performed on the basis of the previous corrected estimate as follows:

$$\hat{x}^-_k = A_{k-1}\hat{x}_{k-1}, \tag{12}$$

$$P^-_k = A_{k-1}P_{k-1}A^{\mathrm{T}}_{k-1} + Q_{k-1}. \tag{13}$$

The final fusion result $\hat{x}_k$ is obtained from the correction step, which is calculated on the basis of the local estimates as in (14) and (15). It will be fed back to the next prediction step [18].

$$\hat{x}_k = P_k\left[(P^-_k)^{-1}\hat{x}^-_k + \sum_{m=1}^{M}(P^m_{k|k})^{-1}\hat{x}^m_{k|k} - \sum_{m=1}^{M}(P^m_{k|k-1})^{-1}\hat{x}^m_{k|k-1}\right], \tag{14}$$

$$(P_k)^{-1} = (P^-_k)^{-1} + \sum_{m=1}^{M}(P^m_{k|k})^{-1} - \sum_{m=1}^{M}(P^m_{k|k-1})^{-1}. \tag{15}$$

In the fusion procedure described above, the posteriori estimate error covariance $P^m_{k|k}$ is affected by sensor confidence and thus, adaptively adjusts the weight of each source sensor. Finally, the fusion result will contain more information from the sensors with greater confidence.

4 Experimental results

In this section, two experiments are implemented on real-world video sequences to demonstrate the effectiveness of the proposed fusion method. In the experiments, the trajectory result from the proposed ADKFF method is compared with those from single sensors, centralized Kalman filter fusion (CKFF), federated Kalman filter fusion (FKFF), and decentralized Kalman filter fusion (DKFF) methods without considering the sensor confidence against the referenced ground truth trajectory. Also, we compare the ADKFF method with the adaptive federated Kalman Filter fusion (AFKFF) [15] and fuzzy logic-based adaptive Kalman filter fusion (FL-AKFF) [22] methods. These two latter methods are adaptive methods that use different strategies to adaptively adjust the filter's measurement covariance in their fusion procedure. All the Kalman filters use the random shift model, which is a classic model usually used in video tracking. The maximum value of the measurement error variance GD in (11) is usually set to 6 in a practical video surveillance situation. The ground truth is obtained manually by slowly tracing the mouse position as the user tracks a target in a video sequence and then, smoothing the trajectory to get a more accurate ground truth.

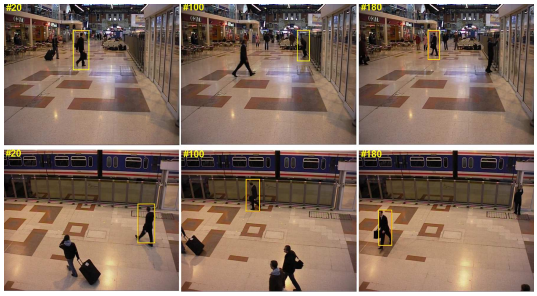

Figure 5 (Color online) Frames of detection results from subway sequences. The first row shows frames from camera 1 and the second row shows frames from camera 3.

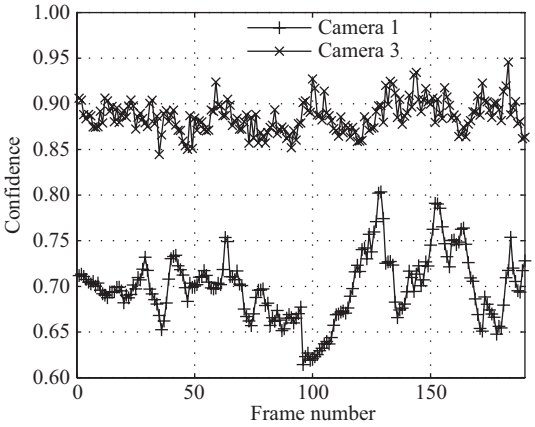

Figure 6 Target detection confidence according to camera 1 and camera 3 for subway sequences.

4.1 Comparison with CKFF, FKFF, and DKFF methods

In the first experiment, the PETS2006 datasets[1)] is used. Here, we employ a segment of Scenario 3 (Take 7-A) and call it subway sequences. One person enters the scene, walks through the passageway, and makes a curved trajectory. We consider only camera 1 and camera 3 because camera 2 contains no informative data for our purpose (the target is barely detectable and far away from the camera's point of view). Because of the challenge from different camera views, background ambiguities, and illumination conditions, this segment becomes a valid sequence for testing multisensor fusion. Some frames of detected target from the different cameras are shown in Figure 5. Frames in the first row are from the camera 1 and the corresponding frames from camera 3 are shown in the second row.

As we will see, the background in camera 1 can affect the detector, as from its angle the illumination condition is poor and the background is ambiguous. On the other hand, as can be seen from the figure, camera 3 gives a brighter and clearer scene; the target is also more discernable from its view angle. The results from camera 3 are therefore more reliable than those from camera 1 with respect to target detection. The confidence values associated with the detection results also confirm the conclusion that the results from camera 3 are clearly better than those from camera 1, as shown in Figure 6.

The trajectory results from the single sensors and four fusion methods are shown in Figure 7. As shown in the figure, by comparing the trajectories results from single sensors and the CKFF, FKFF, and DKFF methods, we find that the fused trajectories obtained from the FKFF and DKFF methods give more accurate results than that obtained using the CKFF method. Also, the fusion result obtained from DKFF is slightly better than that obtained using FKFF. As can be seen from Figure 7(d), the ADKFF method gives the best fusion result, which is more coincident with the ground truth trajectory. Also, the fusion result is mainly benefited from camera 3 that provides higher confidence in the target detection. This excellent performance is attributed to the use of the confidence-based adaptive adjusting model in the fusion procedure.

For further comparison, the position distance error (the distance between the estimated position and the corresponding ground truth position) of the fusion results obtained using the four aforementioned fusion methods against the ground truth is shown in Figure 8. As can be seen from the figure, considering the closeness of fused trajectories with respect to the ground truth, we find that the proposed ADKFF method gives the lowest distance error and therefore has the best performance.

A more objective account of the performance for the single sensors and the fusion methods is reported in Table 1, where the root mean square error (RMSE) and the average position error (APE) of the

1) It is available on http://www.cvg.rdg.ac.uk/PETS2006/.

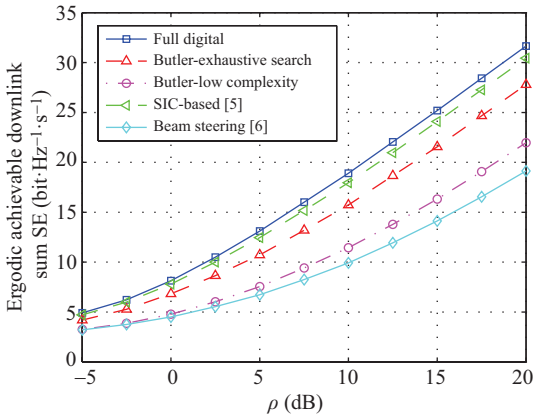

Figure 7 Trajectory results associated with the two cameras, ground truth, and the (a) CKFF, (b) FKFF, (c) DKFF, and (d) ADKFF methods for subway sequences.

Table 1 RMSE and APE for subway sequences

	X-error	Y-error	APE
Camera 1	14.0750	25.4559	29.0879
Camera 3	5.2310	4.9507	7.2023
CKFF	8.5688	12.9289	15.5107
FKFF	6.0083	12.6347	13.9906
DKFF	6.7301	11.3297	13.1779
ADKFF	4.6595	7.2846	8.6474

trajectories with respect to the ground truth are calculated. A lower metric denotes a better performance. As can be seen from Table 1, the objective results coincide with those shown in Figure 7; the FKFF and DKFF methods perform better than the CKFF method, and the proposed ADKFF method exhibits the best performance. This is different from the known conclusion that the CKFF should be optimal theoretically.

In a traditional Kalman filter, the covariance matrix of process noise and observation noise is assumed to be available. However, in most practical applications, we cannot get an accurate covariance matrix of noise and usually approximate it. In Kalman filter, the quality of approximation is closely related to the quality of the statistical property of priori noise. An inaccurate priori statistical property of the input noise is known to considerably deteriorate the performance of the Kalman filter and even leads to a divergence of the filter [22]. In the case of a simulation, the observation data are generated using a linear model with a known covariance matrix of noise and are filtered by the same linear model; therefore, the fusion results can be theoretically optimized. However, the observation data used in this study are obtained from target tracking (not generated by a known model) in a practical real-world situation, but they are filtered with the given linear model, which did not perfectly fit the theoretical conditions of optimal CKFF. Also, the CKFF method cannot isolate the source data with considerable error, while the FKFF and DKFF methods provide improved error correction and enhanced redundancy management. Therefore, when we use real-world noisy data from a tracker, the CKFF method cannot provide an optimal result because of a lack of perfect theoretical conditions, and the advantages of the FKFF and DKFF methods are then reflected.

4.2 Comparison with adaptive methods AFKFF and FL-AKFF

In the second experiment, the AFKFF [15] and FL-AKFF [22] methods which also use adaptive strategies are compared with the proposed ADKFF method to demonstrate the effectiveness of the proposed method. The robot sequences from two cameras that observe a robot running across a floor are used in this experiment.

Figure 9 illustrates some detection results of the robot sequences. As shown in Figure 9, the video

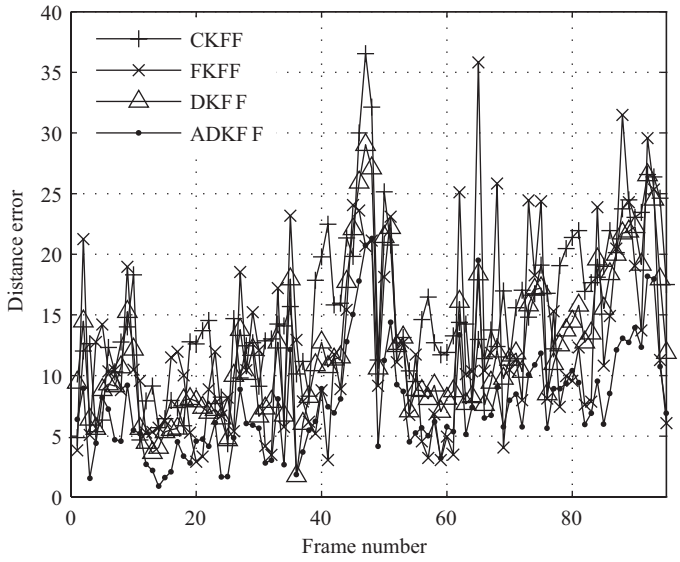

Figure 8 The position distance error of the fused trajectory result estimated using the CKFF, FKFF, DKFF and the ADKFF methods for subway sequences.

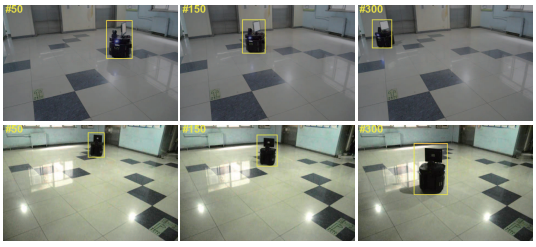

Figure 9 (Color online) Frames of detection results from robot sequences. The first row shows frames from camera 1, and the second row shows frames from camera 2.

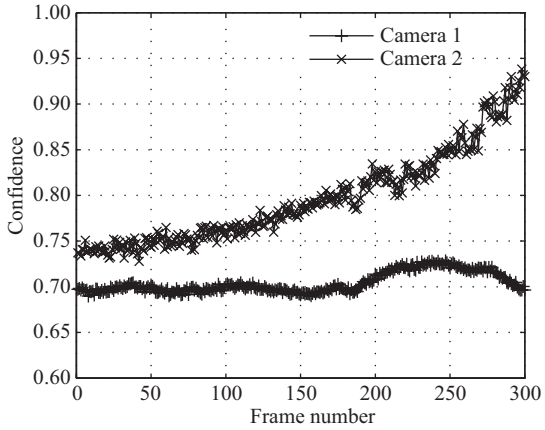

Figure 10 Target detection confidence according to camera 1 and camera 2 for robot sequences.

sequence obtained from camera 1 is underexposed. The poor illumination condition can clearly influence the tracker and eventually affect the detection results, which is common in practical surveillance situations. However, camera 2 provides more abundant and clear information for target detecting. The confidence values associated with the detection results from camera 1 and camera 2, as shown in Figure 10, also confirm the conclusion that the results from camera 2 are obviously better than those from camera 1.

The trajectory results associated with the single sensors and three fusion methods are presented in Figure 11. Obviously, by comparing the results shown in the figure, we conclude that the trajectory result obtained from the proposed ADKFF method is considerably closer to the ground truth than that from the AFKFF and FL-AKFF methods and the FL-AKFF performs better than AFKFF. Also, as can be seen, the fusion result of the proposed method is mainly benefited from camera 2, which provides higher confidence in the target detection. This behavior coincides with the confidence result shown in Figure 10. Also, the excellent performance discussed above can be attributed to the use of the sensor confidence and DKF in the fusion procedure.

The position distance errors of the fusion results according to the three fusion methods against the

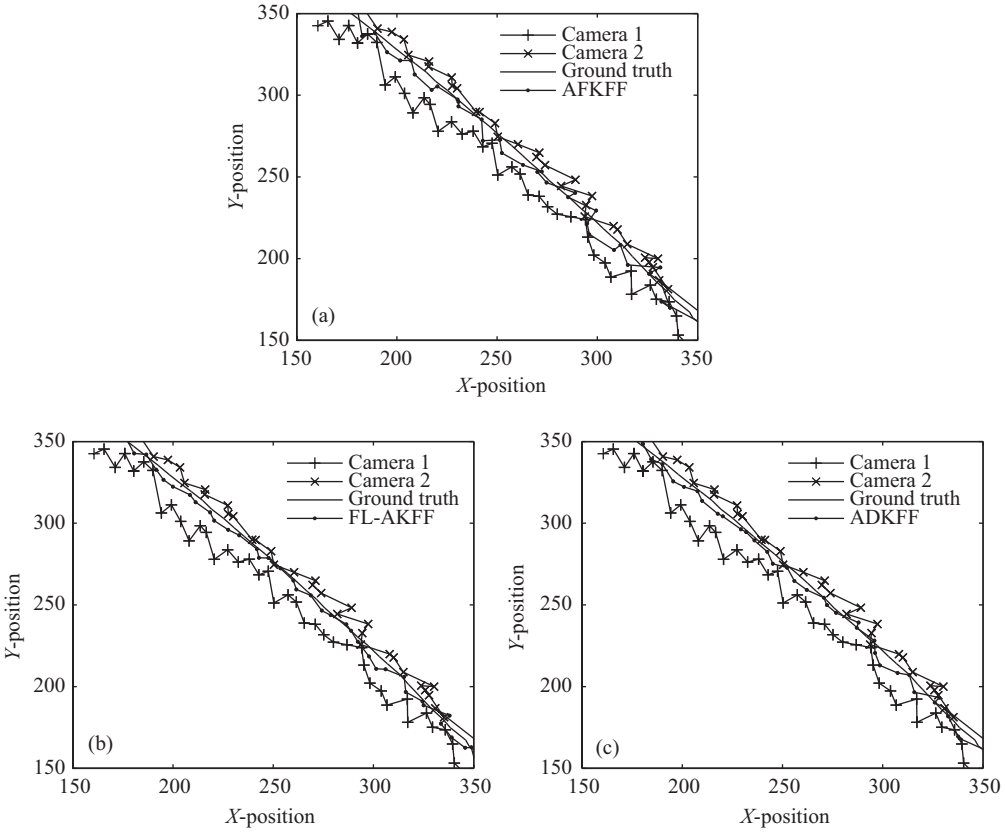

Figure 11 Trajectory results associated with the two cameras, ground truth, and (a) AFKFF, (b) FL-AKFF, and (c) ADKFF methods for robot sequences.

ground truth are shown in Figure 12. As can be seen from the figure, the fusion result from the AFKFF method had the greatest position error, and the proposed ADKFF method yielded the fusion result with the least position error, and therefore, this method exhibited the best performance.

For a further comparison, the objective metrics, namely RMSE and APE, are used for evaluating the fusion performance. The result is reported in Table 2. From the objective results, we can see that the result obtained from the proposed ADKFF method with the least error and these quantitative results are coincident with the visual effect, as shown in Figures 11 and 12, respectively. Therefore, it can be concluded that the proposed ADKFF method exhibits better performance than the AFKFF and FL-AKFF methods.

The comparison results of the three adaptive fusion methods show that the proposed ADKFF method outperforms the other two methods and that the FL-AKFF method performs better than the AFKFF method. These results can be attributed to the fact that the AFKFF and FL-AKFF methods essentially use a strategy that enhances the consistency between the actual covariance value of the residual and the theoretical value. Therefore, these two methods can adaptively adjust the measurement error covariance in the inner system. However, the ADKFF method uses the sensor confidence from the outer system and the sensor confidence is independent of the fusion system and has a more accurate ability to adjust the measurement error covariance adaptively. Also, FL-AKFF performs better than AFKFF because of the application of the fuzzy logic inference system to the adjustment of the measurement error covariance; the fuzzy logic inference system has a relatively good ability to adjust the measurement error covariance.

5 Conclusion

In this paper, we proposed an adaptive video sensor fusion method based on DKF and sensor confidence. The proposed method can effectively improve the robustness and accuracy of tracking results by auto-

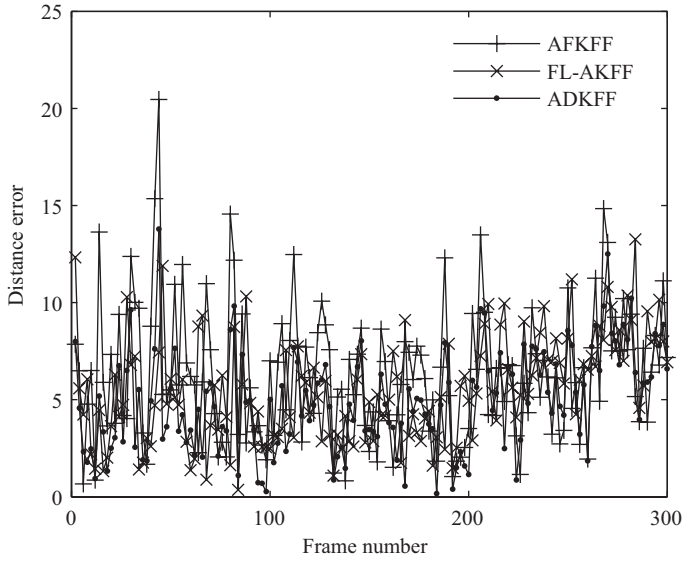

Figure 12 The position distance error of the fused trajectory result achieved using the AFKFF, FL-AKFF and proposed ADKFF methods for robot sequences.

Table 2 RMSE and APE for robot sequences

	X-error	Y-error	APE
Camera 1	20.3655	7.1487	21.5837
Camera 2	7.5412	4.0542	8.5619
AFKFF	5.8346	4.4208	7.3203
FL-AKFF	5.5758	3.8297	6.7644
ADKFF	4.9831	3.4663	6.0702

matically fusing the position data of the tracking targets from the source video sensors. In the fusion procedure, a sensor confidence function is taken into account to evaluate the performance of the sensor in target detection, and then, the confidence value is used for automatically adjusting the measurement noise covariance matrix of the local filters and adaptively determines the weight of each video sensor more correctly in the fusion process. Also, the application of DKF can make full use of the redundant tracking data from multiple video sensors and give more accurate fusion results in an efficient manner. Thus, the position errors due to inaccurate target tracking and position projection can be reduced. Experimental results of visual and objective evaluations also demonstrate that the proposed ADKFF method has better performance than the other comparative fusion methods. In future, we intend to address the development of a more efficient confidence function and fusion model for nonlinear situations.

Acknowledgements This work was supported by National Basic Research Program of China (973 Program) (Grant No. 2012CB821206) and National Natural Science Foundation of China (Grant Nos. 61320106006, 61532006, 61502042).

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Aghajan H, Cavallaro A. Multi-Camera Networks: Principles and Applications. Pittsburgh: Academic Press, 2009
2 Jia Y. Alternative proofs for improved LMI representations for the analysis and the design of continuous-time systems with polytopic type uncertainty: a predictive approach. IEEE Trans Automat Contr, 2003, 48: 1413–1416
3 Snidaro L, Visentini L, Foresti G L. Intelligent Video Surveillance: Systems and Technology. Boca Raton: CRC Press, 2009. 363–388
4 Li B, Yan W. A sensor fusion framework using multiple particle filters for video-based navigation. IEEE Trans Intell Trans Syst, 2010, 11: 348–358

5 Denman S, Lamb T, Fookes C, et al. Multi-spectral fusion for surveillance systems. Comput Electr Eng, 2010, 36: 643–663
6 Loreto S, Jose M M, Ander A, et al. RGB-D, laser and thermal sensor fusion for people following in a mobile robot. Int J Adv Robot Syst, 2013, 10: 271
7 Federico C. A review of data fusion techniques. Sci World J, 2013, 2013: 704504
8 Christoph S, Fernando P L, Marco K. Information fusion for automotive applications—an overview. Inform Fusion, 2011, 12: 244–252
9 Chan A L, Schnelle S R. Fusing concurrent visible and infrared videos for improved tracking performance. Opt Eng, 2013, 52: 017004
10 Jia Y. Robust control with decoupling performance for steering and traction of 4WS vehicles under velocity-arying motion. IEEE Trans Contr Syst Tech, 2000, 8: 554–569
11 Snidaro L, Visentini I, Foresti G L. Fusing multiple video sensors for surveillance. ACM Trans Multim Comput Commun Appl, 2012, 8: 7
12 Chong C Y, Mori S. Optimal fusion for non-zero process noise. In: Proceedings of the 16th International Conference on Information Fusion, Istanbul, 2013. 365–371
13 Xu J, Song E B, Luo Y T, et al. Optimal distributed Kalman filtering fusion algorithm without invertibility of estimation error and sensor noise covariances. IEEE Signal Process Lett, 2012, 19: 55–58
14 Li Z G, Tian X Y. The application of federated Kalman filtering in the information fusion technique. In: Cross Strait Quad-Regional Radio Science and Wireless Technology Conference, Harbin, 2011, 2: 1228–1230
15 Zhang H, Sang H S, Shen X B. Adaptive federated Kalman filtering attitude estimation algorithm for double-FOV star sensor. J Comput Inf Syst, 2010, 6: 3201–3208
16 Qi W J, Zhang P, Deng Z L. Weighted fusion robust steady-state Kalman filters for multisensor system with uncertain noise variances. J Appl Math, 2014, 2014: 369252
17 Julier S J, Uhlmann J K. General decentralized data fusion with covariance intersection. In: Handbook of Multisensor Data Fusion. Boca Raton: CRC Press, 2009. 319–342
18 Markus S S, Kristian K. Performance analysis of decentralized Kalman filters under communication constraints. J Adv Inf Fusion, 2007, 2: 65–75
19 Deng Z L, Zhang P, Qi W J, et al. Sequential covariance intersection fusion Kalman filter. Inform Sciences, 2012, 189: 293–309
20 Deng Z L, Zhang P, Qi W J, et al. The accuracy comparison of multisensor covariance intersection fuser and three weighting fusers. Inform Fusion, 2013, 14: 177–185
21 Ibarra-Bonilla M N, Escamilla-Ambrosio P J, Ramirez-Cortes J M, et al. Pedestrian dead reckoning with attitude estimation using a fuzzy logic tuned adaptive kalman filter. In: Proceedings of the IEEE 4th Latin American Symposium on Circuits and Systems, Cusco, 2013. 1–4
22 Li J, Lei Y H, Cai Y Z, et al. Multi-sensor data fusion algorithm based on fuzzy adaptive Kalman filter. In: Proceedings of the 32nd Chinese Control Conference, Xi'an, 2013. 4523–4527
23 Correia P L, Pereira F. Objective evaluation of video segmentation quality. IEEE Trans Image Process, 2003, 12: 186–200
24 Jia Y. General solution to diagonal model matching control of multi-output-delay systems and its applications in adaptive scheme. Progress Nat Sci, 2009, 19: 79–90
25 Xu T, Cui P. Data fusion of integrated navigation system based on confidence weighted. Acta Aeronaut Et Astronaut Sin, 2007, 28: 1389–1394
26 Snidaro L, Foresti G L, Niu R X, et al. Sensor fusion for video surveillance. In: Proceedings of the 7th International Conference on Information Fusion, Stockholm, 2004, 2: 739–746
27 Hartley R, Zisserman A. Multiple View Geometry in Computer Vision. 2nd ed. New York: Cambridge Univercity Press, 2004
28 Shen X J, Luo Y T, Zhu Y M, et al. Globally optimal distributed Kalman filtering fusion. Sci China Inf Sci, 2012, 55: 512–529