

Research Article

A Hierarchical Estimator for Object Tracking

Chin-Wen Wu,¹ Yi-Nung Chung,² and Pau-Choo Chung¹

¹Department of Electrical Engineering, Institute of Computer and Communication Engineering, National Cheng Kung University, Tainan 701, Taiwan

²Department of Electrical Engineering, National Changhua University of Education, Changhua 500, Taiwan

Correspondence should be addressed to Yi-Nung Chung, ynchung@cc.ncue.edu.tw

Received 17 November 2009; Revised 27 March 2010; Accepted 14 May 2010

Academic Editor: Hsu-Yung Cheng

Copyright © 2010 Chin-Wen Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A closed-loop local-global integrated hierarchical estimator (CLGIHE) approach for object tracking using multiple cameras is proposed. The Kalman filter is used in both the local and global estimates. In contrast to existing approaches where the local and global estimations are performed independently, the proposed approach combines local and global estimates into one for mutual compensation. Consequently, the Kalman-filter-based data fusion optimally adjusts the fusion gain based on environment conditions derived from each local estimator. The global estimation outputs are included in the local estimation process. Closed-loop mutual compensation between the local and global estimations is thus achieved to obtain higher tracking accuracy. A set of image sequences from multiple views are applied to evaluate performance. Computer simulation and experimental results indicate that the proposed approach successfully tracks objects.

1. Introduction

Visual object tracking is an important issue in computer vision. It has applications in many fields, including visual surveillance, human behavior analysis, maneuvering target tracking, and traffic monitoring. The two main types of visual tracking algorithms are target representation and localization algorithms and filtering and data association algorithms [1]. For target representation and localization algorithms, tracking a moving object typically involves matching objects in consecutive frames using features such as edge, region, shape, texture, position, and color. Comaniciu et al. [1] presented a kernel-based framework for tracking nonrigid objects. The mean shift algorithm [2] uses the repeated movement of data points to the sample means. The mean shift algorithm is shown to have effective computation and good tracking performance, but it tends to converge to a local maximum. For filtering and data association algorithms, the state estimation method is used for modeling the dynamic system of visual tracking. The state space approach recursively estimates the state vector in two consecutive stages: prediction and updating. In the prediction step, the prior estimate of the current state is derived using a dynamic equation. In the updating step, the posterior estimate of

the state is updated based on measurements. A state space approach which incorporates measurements into existing object tracks within the framework of Kalman filtering was developed in [3]. Cui et al. [4] presented a laser-based dense crowd tracking method. Particle filters, which are based on the Monte Carlo integration method for implementing a recursive Bayesian filter, have also been proposed [5, 6]. The key idea is to represent the required posterior estimate by a set of random samples with associated weights. A particle filter can effectively deal with clutter and ambiguous situations. However, if the dimension of the state vector is high, a particle filter has a very large computational cost [7–9]. Cheng and Hwang [10] combined a Kalman filter with particle sampling for multiple-object video tracking.

In the tracking procedure, once measurements are received, data association must be applied to determine the exact relationship between measurements and predicted objects. Several algorithms have been developed for data association, such as probabilistic data association (PDA) and joint probabilistic data association (JPDA) [11]. The PDA approach for multitarget tracking, presented by Kershaw and Evans [12], reduces the complexity associated with more sophisticated algorithms by focusing on a few most likely hypotheses.

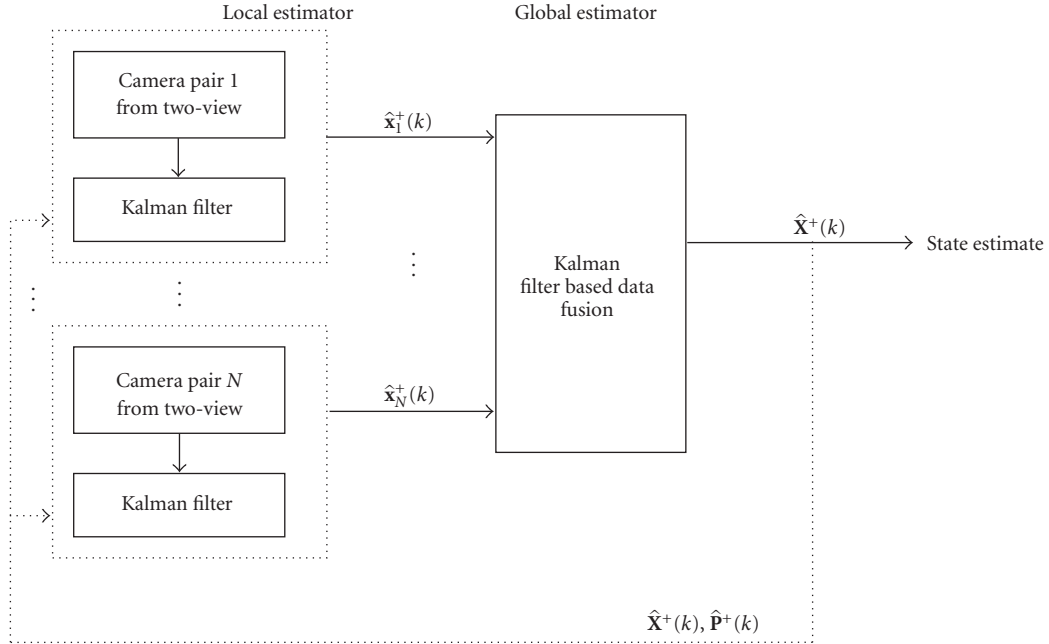


FIGURE 1: Proposed Kalman-filter-based hierarchical estimator for object tracking.

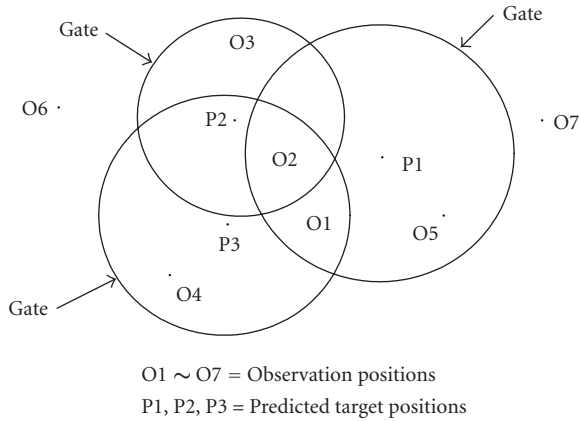


FIGURE 2: Relationship between predicted objects and measurements based on the gating technique.

Occlusion is considered an essential challenge in tracking moving objects. Consequently, a number of recent studies have used multiple views to handle occlusion [13–19]. In [13], a recursive algorithm for stereo was developed. The scheme uses an extended Kalman filter to recursively estimate 3D motion and the depth of moving objects. In [14], a discrete relaxation approach for reducing the intrinsic combinatorial complexity was introduced. The algorithm uses prior knowledge from 2D tracking of each view to obtain real-time 3D tracking. Hu et al. [15] proposed a framework for tracking multiple people about uncalibrated occlusion reasoning. Khan and Shah [16] presented a tracking system based on the field of views (FOVs) of multiple cameras. Another 3D object tracking method that uses multiple views was presented in [17]. Ercan et al. [18] proposed

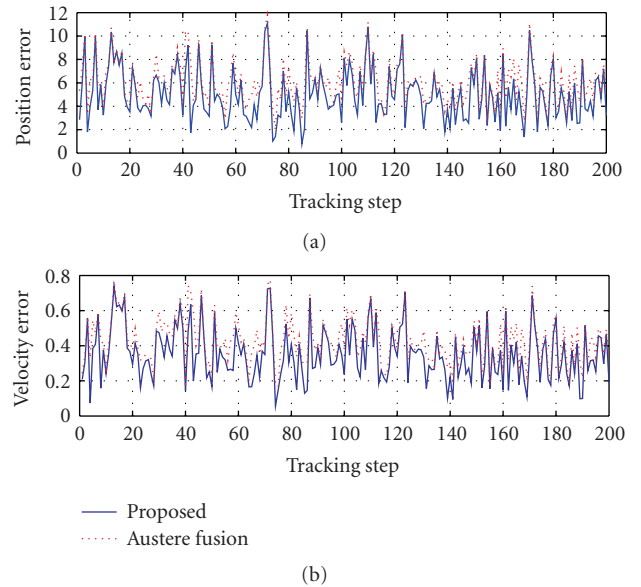


FIGURE 3: Comparison of the position and velocity errors between proposed method and Austere fusion method.

a particle-based framework for single-object tracking with occlusions in a camera network. This approach requires prior knowledge of the environment and the FOV of each camera for estimating the likelihood of whether the object will be occluded from the view of a camera. Furthermore, they did not address the issue of data fusion. Multiple-view data fusion systems have been investigated in several studies [20, 21].

Several studies on hierarchical data fusion [22–26] have also been conducted. Majji et al. [22] presented an algorithm

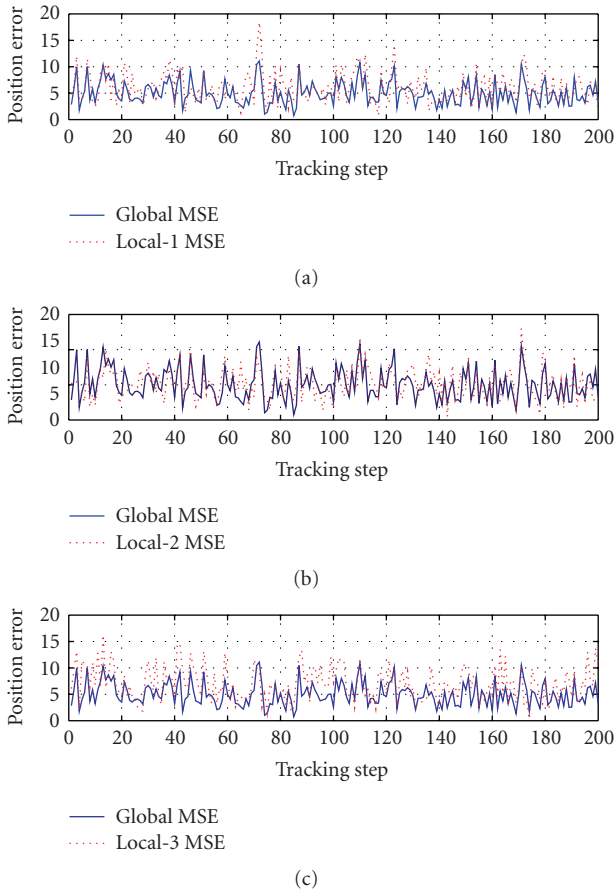


FIGURE 4: Comparison of global and local 3D estimation errors.

using centralized hierarchical fusion. However, the system does not provide feedback to the local filters for modifying their estimate. As such, their approach cannot achieve truly local-global integration to obtain highly accurate estimate. Ajgl et al. [23] discussed various fusion approaches and showed that hierarchical fusion with Millman’s formula has the best performance. Wang et al. [24] developed a two-stage hierarchical framework with partial feedback and applied it to compressed video. Local estimators consist of motion, color, and face detectors. However, the measurements of some local estimates in this scheme are not always available due to intracoded frame prediction. Strobel et al. [25] presented a joint audio-video object tracking method based on decentralized Kalman filters. The front end local estimation uses two Kalman filters, one to track objects based on video and the other to track objects based on audio. The results are then passed through two inverse Kalman filters to obtain measurements, which are applied to another Kalman filter for global fusion to obtain the final tracking result. Due to the use of both Kalman filtering and inverse Kalman filtering, the method is relatively time consuming. Furthermore, it is designed as an open-loop mechanism and thus mutual compensation between the global and local estimates cannot be achieved. Medeiros et al. [26] proposed a cluster-based Kalman filter algorithm for a wireless camera

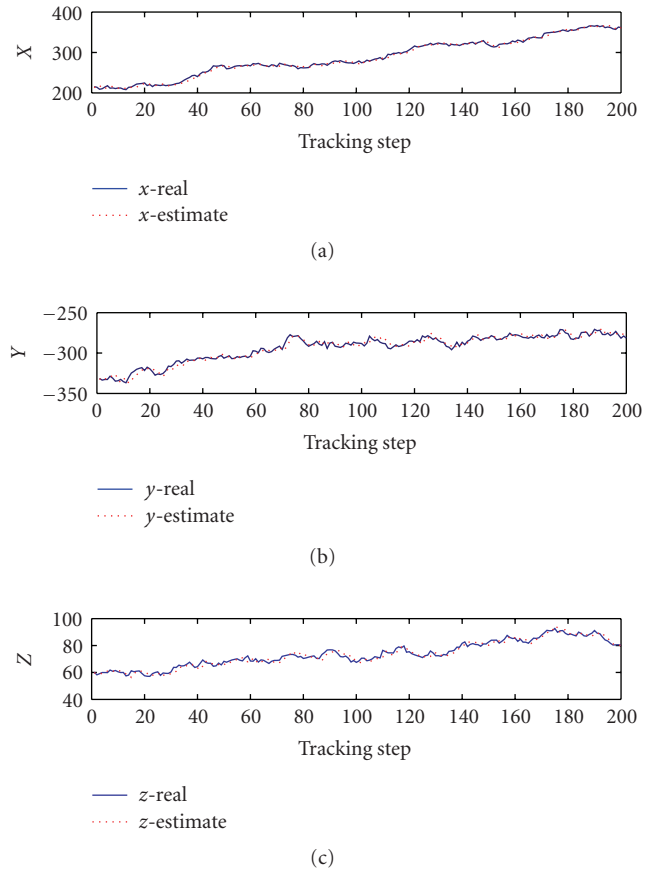


FIGURE 5: Simulation results of tracking.

(sensor) network for object tracking. In their approach, sensors that detect the same object are grouped into a cluster and the information sensed from each individual sensor in the cluster is sent to the cluster head for aggregation by a Kalman filter. The Kalman filter is divided into blocks to improve the computation efficiency. An innovative protocol procedure between individual sensors and the cluster head was developed. However, how to improve the tracking efficiency through a local-global hierarchical fusion mechanism was not discussed.

In contrast to existing approaches, the present study proposes a closed-loop local-global integrated hierarchical estimator (CLGIHE) for object tracking using multiple cameras. The Kalman filter is used to combine the local and global estimates into one estimate for mutual compensation since it can be efficiently integrated into a hierarchical fusion algorithm. The local estimate is input into the global fusion and the obtained global estimate is fed back to the local estimator to achieve iterative optimization-based improvement in both local and global estimates. The local and global estimates are combined into one estimate using the derived equations. The global estimate includes the covariance (environment conditions) from all the local estimators in the derived global fusion equations in the adjustment of fusion gain for dynamically adjusting

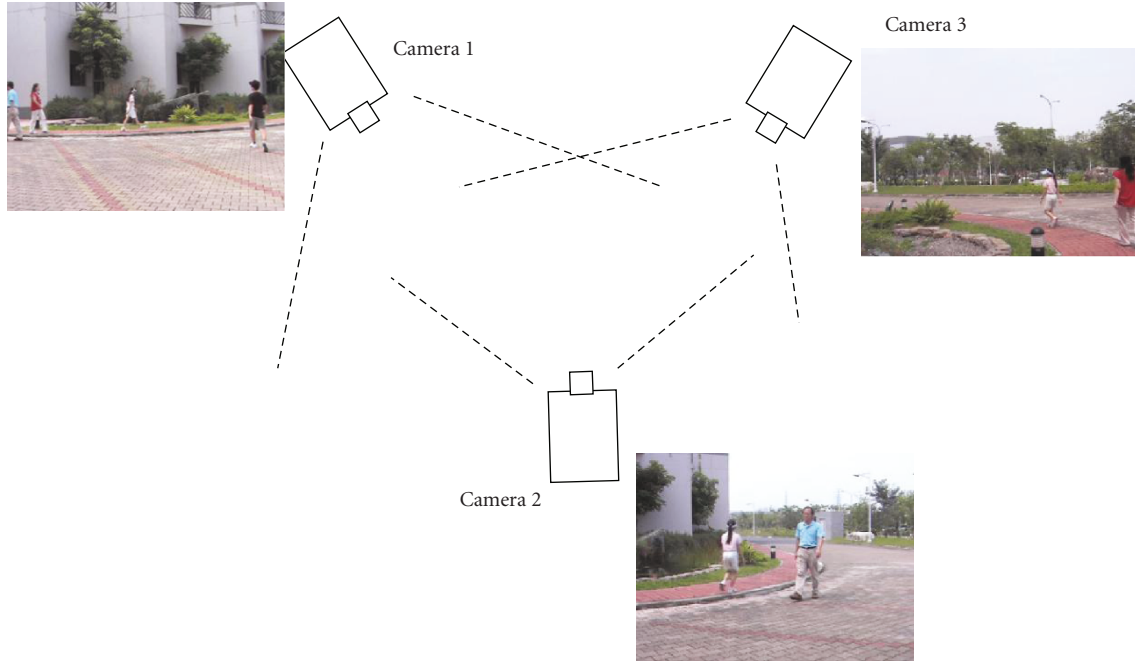


FIGURE 6: Configuration of the tracking system in the experiment.

the tracking in the optimal estimate. Mutual compensation between the local and global estimates is thus achieved to obtain more accurate position estimation.

The rest of this paper is organized as follows. Section 2 provides a brief overview of the proposed system. The proposed object tracking with hierarchical estimation is described in Section 3. The simulation and experimental results of the proposed approach are described in Section 4. Finally, the conclusions are given in Section 5.

2. System Overview

The proposed hierarchical tracking system (CLGIHE) consists of Local Estimator and Global Estimator, as shown in Figure 1. Local Estimator uses data association and the Kalman filter for estimating the 3D position of the object using a camera pair. It should be noted that in the system, every two cameras are considered to be a camera pair. Given 2D images and the camera matrices, the positions of 3D points are computed using the triangulation method presented in [27].

Global Estimator performs data fusion of the estimates obtained by Local Estimator to obtain more accurate 3D global estimates. Object tracking is achieved primarily using measurements received from local estimates that are integrated using a data fusion algorithm to form the global estimate. The fusion algorithm concludes the result considering that different local estimators have different reliability to achieve the best estimation result. Therefore, it can provide increased robustness and accurate estimates. After the global estimate is produced, the estimated 3D position of the tracking object is fed back to local filters for modifying the estimated states.

Suppose that there are N camera pairs and a total of L objects in the system. In the system, the local and global estimates are modeled in world coordinates, whereas 3D measurements are reconstructed by each camera pair. The motion segmentation approach is used in each image plane, for example, background subtraction is used to detect a moving object to obtain a measurement for the local estimate. After the measurement has been reconstructed and assigned to the local estimator, the state estimate is performed for the local filter with the measurement.

The following nomenclature is used throughout this study: \mathbf{x}_i denotes local estimate, “ $\hat{\cdot}$ ” denotes estimate, “ $\text{super } T$ ” denote transpose, “ $-$ ” denotes the a priori estimate, “ $+$ ” denotes the a posteriori estimate, \mathbf{p}_i denotes the local covariance matrix, \mathbf{k}_i denotes the Kalman gain of the local estimate, \mathbf{X} , \mathbf{P} , and \mathbf{K} denote the global estimate, the covariance matrix, and the Kalman gain, respectively, \mathbf{I}_n denotes an $n \times n$ identity matrix, and n denotes the dimension of state vector \mathbf{x}_i .

3. Proposed Hierarchical Estimator for Object Tracking

The algorithm for CLGIHE is described in this section. The basic idea of the proposed fusion algorithm with a hierarchical estimation approach is to combine local and global estimates for object tracking. The local predictor produces a 3D position estimate based on the local information perceived by a camera pair. The local estimate results are then sent to the global estimator to generate a global estimate of the object.

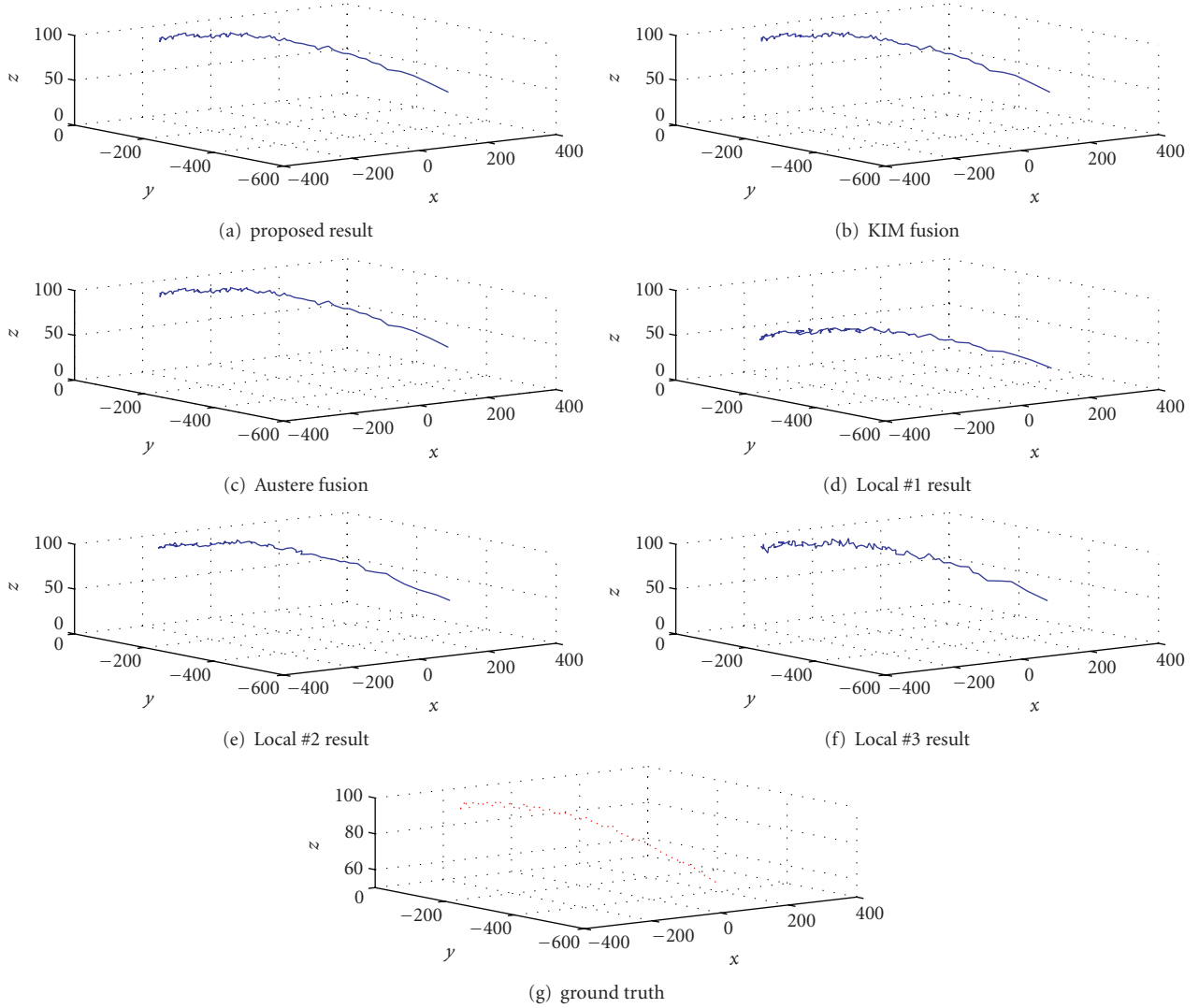


FIGURE 7: 3D tracking results for sequence 1. (a) Global estimate, (b) Kim's fusion, (c) Austere's fusion, (d)–(f) local estimates. (g) Ground truth.

3.1. Local Estimate. The local estimate is computed by the Kalman filters from measurements obtained by a camera pair. Let \mathbf{x}_i be the estimated state vector in the i th Kalman filter at step k given by:

$$\mathbf{x}_i(k) = [x_i(k) \quad \dot{x}_i(k) \quad y_i(k) \quad \dot{y}_i(k) \quad z_i(k) \quad \dot{z}_i(k)]^T, \quad (1)$$

where $[x_i(k) \quad y_i(k) \quad z_i(k)]$ and $[\dot{x}_i(k) \quad \dot{y}_i(k) \quad \dot{z}_i(k)]$ represent the position and velocity of the tracked object, respectively. The Kalman-filter-based local estimate is modeled as

$$\mathbf{x}_i(k+1) = \mathbf{f}_i(k)\mathbf{x}_i(k) + \mathbf{g}_i(k)\mathbf{w}_i(k), \quad (2)$$

where $\mathbf{x}_i(k+1)$ and $\mathbf{x}_i(k)$ are the state vectors at time $k+1$ and k , respectively, which is the number of camera pairs since one Kalman filter is used for each local estimate from two camera views, and $\mathbf{f}_i(k)$ and $\mathbf{g}_i(k)$ are the state transition and noise coupling matrices, respectively. The system noise, $\mathbf{w}_i(k)$, associated with the moving object at frame k is assumed

to be white Gaussian noise distributed with zero mean and covariance matrix $\mathbf{q}_i(k)$.

The measurement equation can be expressed as

$$\mathbf{y}_i(k) = \mathbf{h}_i(k)\mathbf{x}_i(k) + \mathbf{v}_i(k), \quad (3)$$

where the measurement $\mathbf{y}_i(k)$ is formed by a pair of image positions of the i th local estimator at time k , $\mathbf{h}_i(k)$ is the observation matrix of the filter i , and $\mathbf{v}_i(k)$ is the measurement error, which is assumed to be white Gaussian noise with zero mean and covariance matrix $\mathbf{r}_i(k)$.

According to the dynamic system defined in (2) and (3), the solution of the Kalman filter for this model for each camera pair i is given by the state prediction in [3].

The updating step is expressed as

$$\hat{\mathbf{x}}_i^+(k) = \hat{\mathbf{x}}_i^-(k) + \mathbf{k}_i(k)[\mathbf{y}_i(k) - \mathbf{h}_i(k)\hat{\mathbf{x}}_i^-(k)] \quad (4)$$

with error covariance

$$\mathbf{p}_i^+(k) = [\mathbf{I}_n - \mathbf{k}_i(k)\mathbf{h}_i(k)]\mathbf{p}_i^-(k), \quad (5)$$

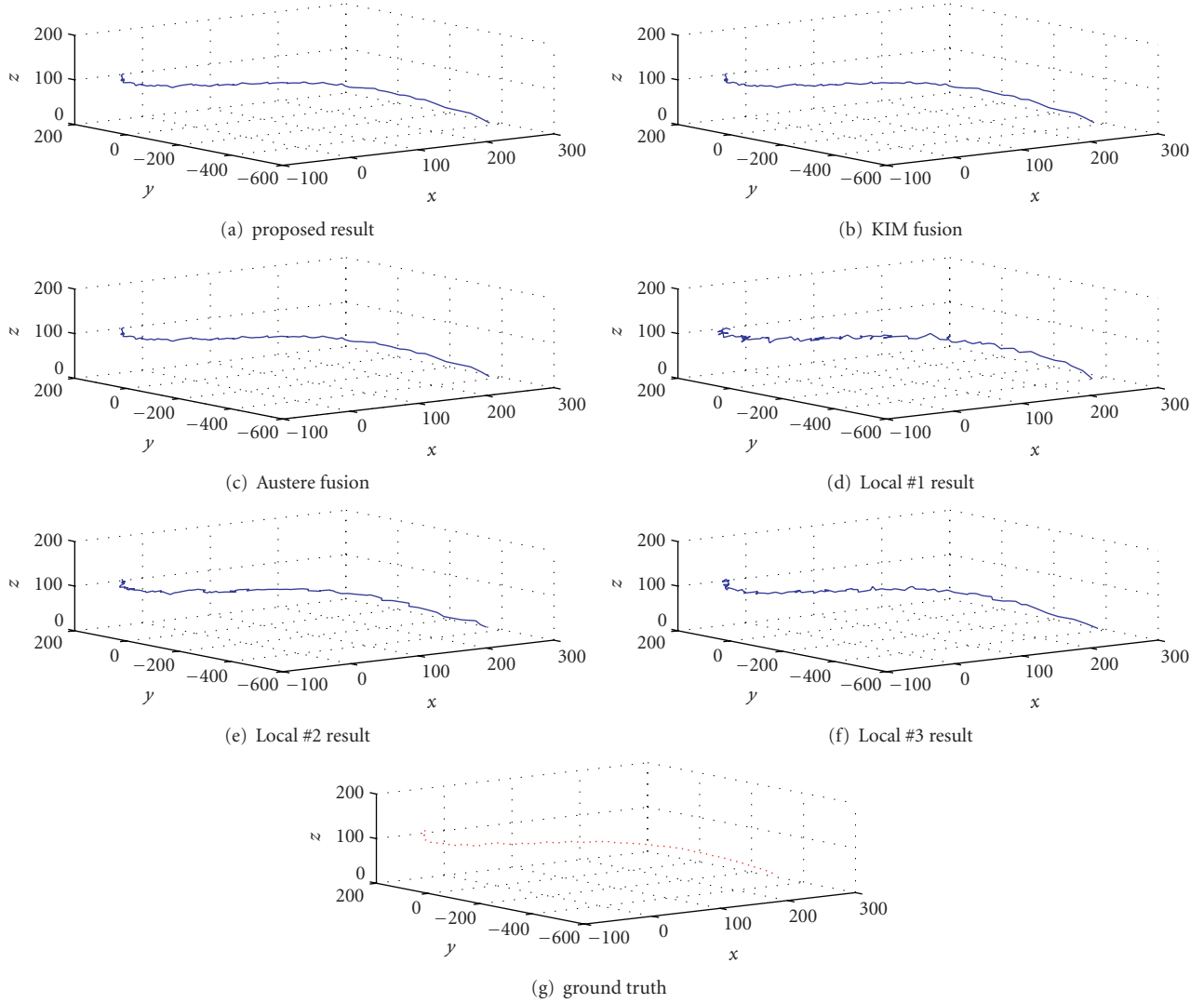


FIGURE 8: 3D tracking results for sequence 2. (a) Global estimate (b) Kim's fusion (c) Austere's fusion (d)–(f) local estimates (g) Ground truth.

where $\hat{\mathbf{x}}_i^-(k) = \mathbf{f}_i(k)\hat{\mathbf{x}}_i^+(k-1)$ and

$$\mathbf{p}_i^-(k) = \mathbf{f}_i(k)\mathbf{p}_i^+(k-1)\mathbf{f}_i^T(k) + \mathbf{g}_i(k)\mathbf{q}_i(k)\mathbf{g}_i^T(k). \quad (6)$$

This process is repeated iteratively at each time instant in all the local tracking processes. The iteration generates one instant-time estimate and the system iteratively updates the estimate.

3.2. Data Association. In a state estimation algorithm, one important procedure is data association, which can be used to determine the relationship between measurements and existing objects. Data association usually consists of two procedural steps: gating and correlation computation logic. If more than one measurement exists, the data association technique can be used to reduce the number of measurements. Figure 2 shows a typical gate diagram, which consists of three objects, P1, P2, and P3. In this figure, there are three objects and seven observations. The gating technique is

applied to eliminate the least probable observations, such as \mathbf{O}_6 and \mathbf{O}_7 . Then, \mathbf{O}_1 , \mathbf{O}_2 , \mathbf{O}_3 , \mathbf{O}_4 , and \mathbf{O}_5 measurements, whose association with the objects has to be determined, remain. A suboptimal Bayesian approach, denoted as 1-step conditional maximum likelihood, is applied to determine the association between the remaining measurements and the objects. For the above equations, let $\mathbf{s}_i(k) = \mathbf{h}_i(k)\mathbf{p}_i^-(k)\mathbf{h}_i^T(k) + \mathbf{r}_i(k)$ be the residual covariance matrix, and $\tilde{\mathbf{y}}_i(k) = \mathbf{y}_i(k) - \mathbf{h}_i(k)\hat{\mathbf{x}}_i^-(k)$ the measurement residual vector at time k . In each local estimator, 1-step conditional maximum likelihood is used to obtain the state estimate $\hat{\mathbf{x}}_i^+(k)$ from all the valid measurements. The Gaussian likelihood $\beta_{j,i}$ of associated measurement i with object j is

$$\beta_{j,i} = \frac{1}{\sqrt{(2\pi)^n |s_i(k)|}} \exp\left\{-\frac{1}{2}\tilde{\mathbf{y}}_i^T(k)s_i^{-1}(k)\tilde{\mathbf{y}}_i(k)\right\}, \quad (7)$$

where $|s_i(k)|$ is the determinant of $s_i(k)$. Since one object may be observed by several local filters, generating multiple

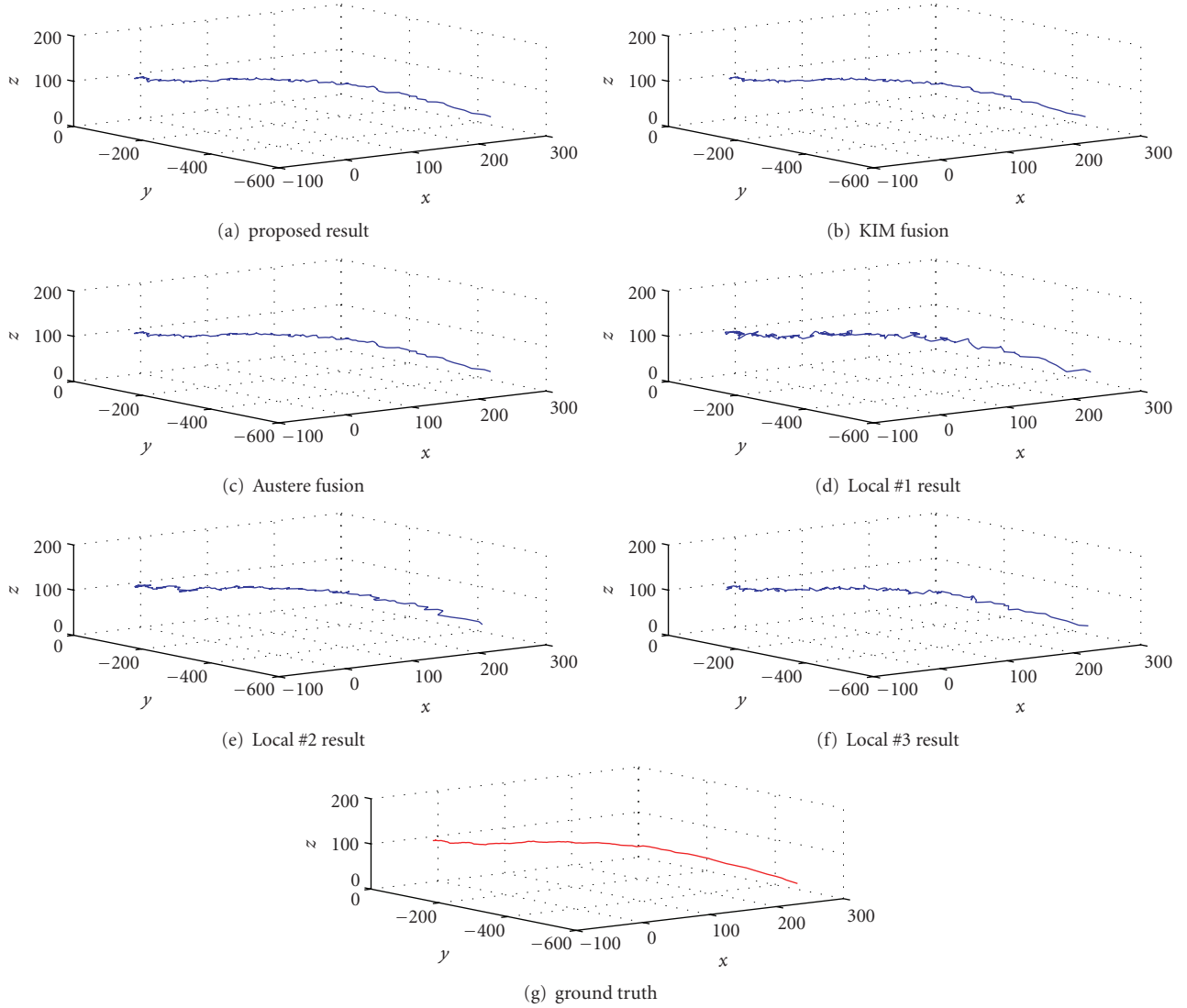


FIGURE 9: 3D tracking results for sequence 3. (a) Global estimate, (b) Kim's fusion, (c) Austere's fusion, (d)–(f) local estimates, (g) Ground truth.

estimates, the local estimates are sent to the global estimate to obtain a fused final result.

3.3. Global Estimate with Data Fusion. The global estimate is composed of the information integrated from the local estimators for tracking and identification. The estimated state of the object at time k is $\mathbf{X}(k)$, where $\mathbf{X}(k) = [X(k) \dot{X}(k) Y(k) \dot{Y}(k) Z(k) \dot{Z}(k)]^T$ contains the object position $[X(k) Y(k) Z(k)]$ and velocity $[\dot{X}(k) \dot{Y}(k) \dot{Z}(k)]$. The discrete-time global dynamic and measurement models of the tracking object are, respectively, defined as

$$\begin{aligned} \mathbf{X}(k+1) &= \mathbf{F}(k)\mathbf{X}(k) + \mathbf{G}(k)\mathbf{W}(k), \\ \mathbf{Y}(k) &= \mathbf{C}(k)\mathbf{X}(k) + \mathbf{V}(k). \end{aligned} \quad (8)$$










Assume that

$$\mathbf{Y}(k) = \begin{bmatrix} \mathbf{y}_1(k) \\ \mathbf{y}_2(k) \\ \vdots \\ \mathbf{y}_N(k) \end{bmatrix}, \quad \mathbf{C}(k) = \begin{bmatrix} \mathbf{C}_1(k) \\ \mathbf{C}_2(k) \\ \vdots \\ \mathbf{C}_N(k) \end{bmatrix}, \quad i = 1, 2, \dots, N, \quad (9)$$

where N is the total number of measurements obtained from N local estimators for the tracked object. The system noise, $\mathbf{W}(k)$, associated with the moving object at step k is assumed to be white Gaussian noise distributed with zero mean and covariance matrix $\mathbf{Q}(k)$. $\mathbf{C}(k)$ is the global observation matrix, and $\mathbf{V}(k)$ is the measurement error, which is assumed to be white Gaussian noise with zero mean and covariance matrix $\mathbf{R}(k)$.

In order to determine the relationship between the local estimate and global estimate, mapping matrix $\mathbf{M}_{j,i}(k)$ is

TABLE 1: Average geometric error for test sequences.

	Sequence			Average geometric error (pixels)	
	View-1	View-2	View-3	Position x	Position y
1				5.4674	6.8053
2				5.8902	7.2142
3				6.5584	8.6403

defined. Let $\mathbf{M}_{j,i}(k)$ be the mapping matrix of the object j seen by camera pair i at time k . It is defined as follows:

$$\mathbf{M}_{j,i}(k) = \begin{cases} \mathbf{I}_6, & j\text{th object is seen by camera pair } i, \\ \mathbf{0}_6, & \text{otherwise,} \end{cases} \quad (10)$$

where \mathbf{I}_6 is a 6-by-6 identity matrix, $\mathbf{0}_6$ is a 6-by-6 matrix of zeros, $j = 1, 2, \dots, L$, and $i = 1, 2, \dots, N$.

The proposed data fusion algorithm in the tracking system is applied to combine the local estimate with mapping matrix $\mathbf{M}_{j,i}(k)$. Thus, recording the output of the local estimators and $\mathbf{M}_{j,i}(k)$ to form global measurement, matrix $\mathbf{C}(k)$ is expressed as

$$\mathbf{C}_i(k) = \mathbf{h}_i(k)\mathbf{M}_{j,i}(k), \quad i = 1, 2, \dots, N. \quad (11)$$

If object j is seen by camera pair i in the local estimate, the output of the local estimate is fed into the global estimate with global estimate matrix $\mathbf{C}_i(k) = \mathbf{h}_i(k)$. Otherwise, there is no need to be updated for none measurement provided.

In order to derive the estimation algorithm, local estimates are combined to form the global estimate. The goal is to compute $\hat{\mathbf{X}}(k)$ in each time step. The global estimate, $\hat{\mathbf{X}}(k)$, for a tracking object can be computed with the Kalman filter as

$$\hat{\mathbf{X}}^+(k) = \hat{\mathbf{X}}^-(k) + \mathbf{B}(k) \sum_{i=1}^N \mathbf{k}_i(k) [\mathbf{y}_i(k) - \mathbf{C}_i(k)\hat{\mathbf{X}}^-(k)], \quad (12)$$

TABLE 2: Average MSE for test sequences.

Method	Sequence		
	Sequence 1	Sequence 2	Sequence 3
Proposed	12.9327	13.2828	13.5486
Kim	13.0883	13.3073	13.5314
Austere	13.1239	13.5741	13.8357
Local 1	14.4163	14.6282	15.9262
Local 2	14.4579	15.9337	16.7499
Local 3	13.9209	13.4702	13.8708

where $\mathbf{B}(k)$ is a normalizing matrix for a tracking object, defined as

$$\mathbf{B}(k) = \left[\sum_{i=1}^N \mathbf{M}_{j,i}(k) \right]^{-1}. \quad (13)$$

The global error covariance is update by

$$\mathbf{P}^+(k) = \left[\mathbf{I} - \mathbf{B}(k) \sum_{i=1}^N \mathbf{k}_i(k) \mathbf{C}_i(k) \right] \mathbf{P}^-(k). \quad (14)$$

The global priori estimate and its prediction error covariance are computed as

$$\hat{\mathbf{X}}^-(k) = \mathbf{F}(k)\hat{\mathbf{X}}^+(k-1). \quad (15)$$

Then, combining (12) and (15), the global estimate for the object becomes

$$\hat{\mathbf{X}}^+(k) = \mathbf{A}(k)\hat{\mathbf{X}}^+(k-1) + \mathbf{B}(k)\sum_{i=1}^N \mathbf{k}_i(k)\mathbf{y}_i(k), \quad (16)$$

where

$$\mathbf{A}(k) = \left[\mathbf{I} - \mathbf{B}(k)\sum_{i=1}^N \mathbf{k}_i(k)\mathbf{C}_i(k) \right] \mathbf{F}(k). \quad (17)$$

The local estimates are combined to produce the global estimate, $\hat{\mathbf{X}}^+(k)$. The local estimates are computed by the Kalman filters and rearranged as

$$\hat{\mathbf{x}}_i^+(k) = \mathbf{a}_i(k)\hat{\mathbf{x}}_i^+(k-1) + \mathbf{k}_i(k)\mathbf{y}_i(k), \quad (18)$$

where

$$\mathbf{a}_i(k) = [\mathbf{I} - \mathbf{k}_i(k)\mathbf{h}_i(k)]\mathbf{f}_i(k). \quad (19)$$

By rewriting (18), one can obtain

$$\mathbf{k}_i(k)\mathbf{y}_i(k) = \hat{\mathbf{x}}_i^+(k) - \mathbf{a}_i(k)\hat{\mathbf{x}}_i^+(k-1). \quad (20)$$

Let $\Phi_i(k) = \mathbf{P}^+(k)^{-1}\mathbf{M}_{j,i}^T(k)\mathbf{p}_i^+(k)$, where $\Phi_i(k)$ can be considered as the adjusting factor between the local and global error covariance. Then, in the global estimate,

$$\mathbf{K}_i(k)\mathbf{Y}_i(k) = \Phi_i(k)\{\hat{\mathbf{x}}_i^+(k) - \mathbf{a}_i(k)\hat{\mathbf{x}}_i^+(k-1)\} \quad (21)$$

Therefore, the final result of the global estimate for the tracking object, $\hat{\mathbf{X}}^+(k)$, in (16) is

$$\begin{aligned} \hat{\mathbf{X}}^+(k) &= \mathbf{A}(k)\hat{\mathbf{X}}^+(k-1) \\ &+ \mathbf{B}(k)\sum_{i=1}^N \Phi_i(k)[\hat{\mathbf{x}}_i^+(k) - \mathbf{a}_i(k)\hat{\mathbf{x}}_i^+(k-1)] \\ &= \mathbf{S}\mathbf{I}^+(k) + \mathbf{B}(k)\sum_{i=1}^N \Phi_i(k)\hat{\mathbf{x}}_i^+(k), \end{aligned} \quad (22)$$

where

$$\begin{aligned} \mathbf{S}\mathbf{I}^+(k) &= \mathbf{A}(k)\hat{\mathbf{X}}^+(k-1) - \mathbf{B}(k)\sum_{i=1}^N \mathbf{t}_i(k)\hat{\mathbf{x}}_i^+(k-1), \\ \mathbf{t}_i(k) &= \Phi_i(k)\mathbf{a}_i(k). \end{aligned} \quad (23)$$

The global estimate $\hat{\mathbf{X}}^+(k)$ is fed back to local filters for improving the local estimates using

$$\hat{\mathbf{x}}_i^+(k) = \hat{\mathbf{X}}^+(k), \quad i = 1, 2, \dots, N. \quad (24)$$

In summary, each local estimate, $\hat{\mathbf{x}}_i^+(k)$, is computed by each local estimator using (4) and then all local estimates are sent to the global estimator. The global estimate, $\hat{\mathbf{X}}^+(k)$ in (22), is obtained after performing the data fusion process in the global estimator. The global estimate $\hat{\mathbf{X}}^+(k)$ is then sent to each local estimator to update the estimate of the local state vector. When the global estimate is fed back, $\mathbf{M}_{j,i}(k)$ can be determined.

4. Experimental Results

To evaluate performance, the proposed CLGIHE algorithm was compared with Austere's method and Kim's method [28] using computer simulation and real image sequences. Since Austere's method and Kim's method use the fusion method without specifying the local filters, to provide an accurate comparison, the Kalman filter was used as the local filter for Austere's fusion and Kim's fusion algorithms.

In the simulation, the state noise, measurement noise, and 3D object positions were created using synthetic data generators. The measurement data were obtained via a homogeneous transformation of the two-camera model in addition to measurement errors. Kalman filters were used to estimate the local state vectors. Once the measurement data was received, the corresponding probability was calculated based on each hypothesis. The conditional estimate of the object states was evaluated and combined with the individual estimate for each hypothesis, weighted by the corresponding probability function. The performance of multiple-view tracking was simulated under epipolar geometry.

After several Monte Carlo runs, the results of position and velocity errors for the proposed method and Austere's method were obtained. A comparison is shown in Figure 3. The horizontal axis indicates the tracking steps, and the vertical axis indicates the position or velocity errors. The position and velocity errors are defined as the mean squared errors. The results indicate that the proposed method has lower MSE values than those of Austere's method. The average MSE values for the proposed method and Austere's fusion method are 4.4750 and 4.8893, respectively.

In order to determine the effect of global fusion, the proposed system's performance was measured with and without global fusion. Figures 4(a)–4(c) show the 3D estimation error comparisons between the global estimator and three local estimators (local 1, local 2, and local 3). The global estimator has lower MSE values than those of each of the three local estimators. The average MSE values obtained for local 1, local 2, and local 3 are 4.7982, 5.0101, and 4.8596, respectively, whereas that for the global estimator is 4.4750. The performance in terms of measured positions of the object compared with the ground truth is shown in Figure 5. The results show that the estimates of x , y , and z coordinates are close to those of the object trajectory.

The performance of the proposed algorithm was also evaluated using real image sequences. In order to show the performance in real situations, three fixed calibrated digital cameras were set up to track people who were moving outdoors. Figure 6 shows the configuration of the tracking system in the experiment. The test image sequences have an image size of 640×480 pixels. All the image sequences were taken with calibrated cameras. At each local estimator, a 3D state vector is determined based on the reconstruction of the camera pair. Every two views form a camera pair and are applied to a local estimator for observations. The direct linear transform (DLT) [27] is adopted as the reconstruction method for each camera pair. To evaluate the accuracy of reconstruction, the geometric error [29] is used for measuring the results. The geometric error is

the sum of the projection error in each camera view for a pair of correspondence points. Before the experiment, a self-made calibrated board was used for camera calibration. The calibration uses a set of control points whose coordinates are already known. Then, several reconstructions and re-projections are used to tune the camera matrices by adjusting geometric error.

When the objects are occluded, observations are unavailable. If there is no measurement to obtain, the object is seen by neither camera. In this situation, the local predicted state is not updated until new observations are generated and the global estimate is updated using only available camera pairs.

In the initial step of the experiment, the local and global estimators were initialized, and background subtraction [30] was used to separate the moving foreground objects. The measurement of the local estimator was obtained from two camera views, that is, a camera pair. The local estimate performed its Kalman filter with the estimated state and the Kalman gain was updated. Each output of the local estimator was sent to the global estimator. The global estimator and estimated 3D positions of the tracked object were computed using (22).

For evaluation, three sequences, for which sample images are shown in the three rows of Table 1, were used for the test. The 3D tracking results obtained for the person wearing blue clothes (sequence 1) are shown in Figure 7(a). For comparison, the results obtained with Kim's fusion and Austere's fusion algorithms are shown in Figures 7(b) and 7(c), respectively. The average MSE values for the proposed method, Kim's fusion, and Austere's fusion are 12.9327, 13.0883, and 13.1239, respectively, (see Table 2). Results show that the proposed method has lower MSE values than those of the other fusion methods. To show the fusion effect, the results obtained from local estimates are shown in Figures 7(d)–7(f). The average MSE values for the three local estimates are 14.4163, 14.4579, and 13.9209, respectively, (see Table 2). The results were also evaluated by mapping the obtained 3D positions onto 2D image planes for comparison. The average errors in the x - and y -directions are listed in the last column of the first row in Table 1.

Similarly, the obtained 3D tracking results for sequence 2 and sequence 3 are shown in Figures 8 and 9, respectively. The average errors in the x - and y -directions of the projected 2D images are shown in the last column of the second and the third row in Table 1, respectively. The MSE values of 3D positions obtained using the proposed approach, Austere's method, Kim's method, and the three local estimators for sequence 2 and sequence 3 are listed in Table 2.

5. Conclusion

A closed-loop local-global integrated hierarchical estimator (CLGIHE) approach was proposed for object tracking using multiple cameras. CLGIHE adopts the Kalman filter to build an integrated hierarchical fusion estimator because it allows the local and global estimates to be combined into one estimate for mutual compensation. Compared to existing multiple-camera Kalman-filter-based object tracking approaches, CLGIHE has the following advantages.

Firstly, it is implemented with a feedback loop to achieve iterative optimization-based improvement from both the local and global mutual compensation. Secondly, local and global estimates are integrated into one estimate to allow the optimal adjustment of the fusion gain based on environment conditions from each local estimator to obtain accurate and smooth tracking results. The simulation and experimental results show that the proposed algorithm is capable of tracking objects in various situations. Moreover, the data fusion algorithm applied to the multiple-view images reduces the probability of misdetection.

Acknowledgment

This work was supported in part by National Science Council, Taiwan, under Grant NSC 98-2218-E-006-004.

References

- [1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.
- [2] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, pp. 142–149, June 2000.
- [3] M. S. Gerwal and A. P. Andrews, *Kalman Filtering Theory and Practice*, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.
- [4] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Laser-based detection and tracking of multiple people in crowds," *Computer Vision and Image Understanding*, vol. 106, no. 2-3, pp. 300–312, 2007.
- [5] J. Czyz, B. Ristic, and B. Macq, "A particle filter for joint detection and tracking of color objects," *Image and Vision Computing*, vol. 25, no. 8, pp. 1271–1281, 2007.
- [6] C. Hue, J.-P. Le Cadre, and P. Pérez, "Sequential Monte Carlo methods for multiple target tracking and data fusion," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 309–325, 2002.
- [7] C. Chang and R. Ansari, "Kernel particle filter: iterative sampling for efficient visual tracking," in *Proceedings of the International Conference on Image Processing (ICIP '03)*, pp. 977–980, September 2003.
- [8] N. Bouaynaya, W. Qu, and D. Schonfeld, "An online motion-based particle filter for head tracking applications," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, pp. 225–228, March 2005.
- [9] C. Shan, Y. Wei, T. Tan, and F. Ojardias, "Real time hand tracking by combining particle filtering and mean shift," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04)*, pp. 669–674, May 2004.
- [10] H.-Y. Cheng and J.-N. Hwang, "Adaptive particle sampling and adaptive appearance for multiple video object tracking," *Signal Processing*, vol. 89, no. 9, pp. 1844–1849, 2009.
- [11] Y. Bar-shalom and T. Fortmann, *Tracking and Data Association*, Academic Press, New York, NY, USA, 1988.
- [12] D. J. Kershaw and R. J. Evans, "Waveform selective probabilistic data association," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1180–1188, 1997.

- [13] J.-W. Yi and J.-H. Oh, "Recursive resolving algorithm for multiple stereo and motion matches," *Image and Vision Computing*, vol. 15, no. 3, pp. 181–196, 1997.
- [14] Y. Li, A. Hilton, and J. Illingworth, "A relaxation algorithm for real-time multiple view 3D-tracking," *Image and Vision Computing*, vol. 20, no. 12, pp. 841–859, 2002.
- [15] W. Hu, X. Zhou, M. Hu, and S. Maybank, "Occlusion reasoning for tracking multiple people," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 1, pp. 114–121, 2009.
- [16] S. Khan and M. Shah, "Consistent labeling of tracked objects in multiple cameras with overlapping fields of view," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1355–1360, 2003.
- [17] J. Black and T. Ellis, "Multi camera image tracking," *Image and Vision Computing*, vol. 24, no. 11, pp. 1256–1267, 2006.
- [18] A. O. Ercan, A. El Gamal, and L. J. Guibas, "Object tracking in the presence of occlusions via a camera network," in *Proceedings of the 6th International Symposium on Information Processing in Sensor Networks (IPSN '07)*, pp. 509–518, April 2007.
- [19] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance models for occlusion handling," *Image and Vision Computing*, vol. 24, no. 11, pp. 1233–1243, 2006.
- [20] S. L. Dockstader and A. M. Tekalp, "Multiple camera fusion for multi-object tracking," in *Proceedings of IEEE Workshop on Multi-Object Tracking*, pp. 95–102, July 2001.
- [21] Q. Zhou and J. K. Aggarwal, "Object tracking in an outdoor environment using fusion of features and cameras," *Image and Vision Computing*, vol. 24, no. 11, pp. 1244–1255, 2006.
- [22] M. Majji, J. J. Davis, and J. L. Junkins, "Hierarchical multi-rate measurement fusion for estimation of dynamical systems," in *AIAA Guidance, Navigation, and Control Conference 2007*, pp. 3967–3978, usa, August 2007.
- [23] J. Ajgl, et al., "Millman's formula in data fusion," in *Proceedings of the 10th International PhD Workshop on Systems and Control*, pp. 1–6, Prague, Czech Republic, 2009.
- [24] J. Wang, R. Achanta, M. Kankanhalli, and P. Mulhem, "A hierarchical framework for face tracking using state vector fusion for compressed video," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 209–212, April 2003.
- [25] N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video object localization and tracking: a presentation general methodology," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 22–31, 2001.
- [26] H. Medeiros, J. Park, and A. C. Kak, "Distributed object tracking using a cluster-based Kalman filter in wireless camera networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 4, pp. 448–463, 2008.
- [27] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, Mass, USA, 2nd edition, 2003.
- [28] K. H. Kim, "Development of track to track fusion algorithms," in *Proceedings of the American Control Conference*, pp. 1037–1041, July 1994.
- [29] D. J. Bardsley and L. Bai, "3D surface reconstruction and recognition," in *Biometric Technology for Human Identification IV*, vol. 6539 of *Proceedings of SPIE*, Orlando, Fla, USA, April 2007.
- [30] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, 2003.