Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Distributed Human 3D Pose Estimation and Action Recognition[*]

Guoliang Liu, Tiantian Liu and Guohui Tian
School of Control Science and Engineering
Shandong University
Jinan, Shandong, China
{*liuguoliang*}*@sdu.edu.cn*

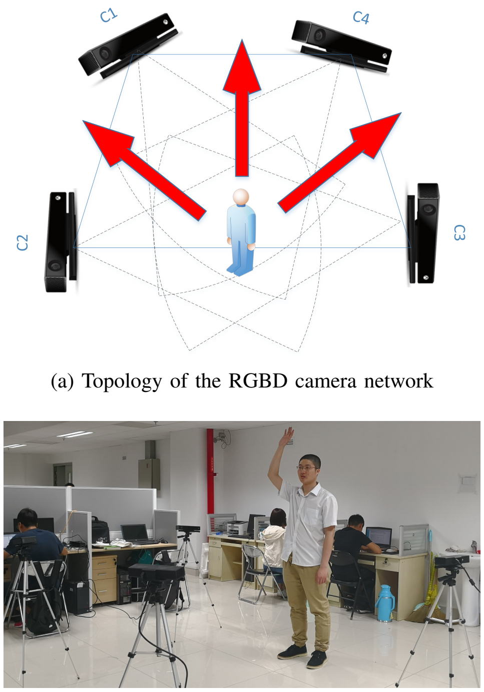
Ze Ji
School of Engineering
Cardiff University
Cardiff, CF24 3AA, UK
jiz1@cardiff.ac.uk

***Abstract*— In this paper, we propose a distributed solution for 3D human pose estimation using a RGBD camera network. The key feature of our method is a dynamic hybrid consensus filter (DHCF) is introduced to fuse the multiple view information of cameras. In contrast to the centralized fusion solution, the DHCF algorithm can be used in a distributed network, which requires no central information fusion center. Therefore, the DHCF based fusion algorithm can benefit from many advantages of distributed network. We also show that the proposed fusion algorithm can handle the occlusion problems effectively, and achieve higher action recognition rate compared to the ones using only single view information.**

***Index Terms*— Distributed information fusion, RGBD camera network, consensus filtering, action recognition, human pose estimation.**

I. Introduction

DUE to the widespread of the low-cost RGBD sensors, the real time 3D human pose estimation is available on the commercial products, such as Microsoft Xbox using Kinect for human computer interaction (HCI) games [1], [2]. More potential applications have been studied in many research fields. For instance, Morato et al. developed a N-Kinect system to build an explicit model of the human body, which is used to detect an imminent collision between the robot and the human [3]. Pathirana et al. proposed a Kinect based bio-kinematic measurement system for rehabilitation and physiotherapy applications, i.e., monitoring of home-based prescribed exercise routines which can significantly reduce the need for patients to travel to regional centers [4]. Geiselhart et al. estimated the workers movements and interaction with digital object models using multi-depth camera system [5], such that a low cost camera setup to facilitate interaction with virtual environments for planning experts is possible.



(a) Topology of the RGBD camera network

(b) Real experimental scenario

Fig. 1: The experimental setup uses four Kinect V2 cameras, which have a loop topology and construct a distributed sensor network.

To accurately estimate the human pose is still a challenge problem due to the occlusion problem [6]. Many researchers propose to use multiple 3D cameras to handle occlusion problem by fusing multiple view information. The key problem is how to fuse. Current works mostly employ a centralized network topology, and use a central computer node to fuse information from all sensor nodes. The fusion algorithm can be a simple weighted summation method using skeleton joint tracking status [7], [8], [9], or additional bone length constrain conditions [10]. In order to use the temporal information and get smooth trajectory of skeleton joints, a

[*] This work is partially supported by the National Key R&D Program of China (#2018YFB1306504), National Natural Science Foundation of China (#61603213, #91748115), Young Scholars Program of Shandong University (#2018WLJH71), the Fundamental Research Funds of Shandong University, and the Taishan Scholars Program of Shandong Province. (Corresponding author: Guoliang Liu).

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

Kalman filter or particle filter can be employed [11], [12], [13].

Compared to the centralized network topology, the distributed network topology is more desirable in many application scenarios due to its scalability to a large number of sensors, ease of installation and high tolerance to node failure [14]. Furthermore, each camera can process information and make decision locally, such that the camera nodes become distributed smart agents, which can achieve consensus by exchanging information with neighbor nodes after a number of iterations [15]. Song et al. propose a distributed 2D camera network to track human trajectories and recognize actions using a Kalman consensus filter (KCF) [16], [15]. The KCF was introduced by Olfati-Saber and has been applied in many fields [17], [18]. Kamal et al. proved that the KCF can not handle information redundant and naive node problems for the camera network, and proposed an information weighted consensus filter (IWCF) to replace the KCF, which shows improved performance for 2D human tracking using a 2D camera network [14]. Wang et al. employ the IWCF to track 3D human skeleton joints using multiple Kinects [19]. Li et al. further prove that the IWCF based human skeleton fusion algorithm can achieve higher action recognition [20], [21].

In this paper, we propose a dynamic hybrid consensus filter (DHCF) based human pose estimation and action recognition algorithm using a 3D RGB-D camera network, which can handle occlusion problem and achieve higher recognition rate than single views. The proposed method is a proof-of-concept study in using consensus for 3D human pose estimation, which can not only been used for small network in the indoor environment using RGB-D cameras, e.g., Microsoft Kinect, Asus Xtion and Intel Realsense, but also can been used for large and scalable network in the outdoor environment using advanced 3D cameras.

II. Distributed Human Pose Estimation

In this section, we first introduce the DHCF algorithm and show how the skeletons from multiple views can be fused, and then discuss the human action recognition method used for demonstrating the performance of the proposed idea compared to the single views.

A. Dynamic Hybrid Consensus Filter for Skeleton Fusion

The DHCF is a distributed fusion algorithm, whose core idea is the iterative information exchanges between the sensor node and its neighbors [22], [23]. After a number of iterations, the sensor nodes in the whole network can achieve consensus state about the target. Compared to the IWCF proposed in [14], DHCF can achieve faster convergence rate and have no requirement about the total number of sensor nodes. The distributed camera network used in this paper construct an undirected graph $G = (C, E)$ where $C = \{1, 2, 3, , N\}$ means vertex set that has N nodes, and $E \subset \{\{i, j\}|i, j \subset C\}$ denotes the edge set. We define the

Algorithm 1 DHCF based skeleton fusion

```
• Initialization: measurement noise R, process noise Q and total consensus iteration steps L.
• For $k = 1, \cdots, \infty$:
    1) Predicted human pose for the next time step:
           $\hat{x}_{i,k} = F_k x_{i,k-1}$                                        (1)
           $\hat{Y}_{i,k} = (F_k Y_{i,k-1}^{-1} F_k^T + Q_k)^{-1}$                 (2)
           $\hat{y}_{i,k} = \hat{Y}_{i,k} \hat{x}_{i,k}$                           (3)
    2) Computation of consensus proposals:
       if $i \in S$ then
           $u_{i,k} = H_{i,k}^T R_{i,k}^{-1} z_{i,k},\ U_{i,k} = H_{i,k}^T R_{i,k}^{-1} H_{i,k}$   (4)
           $b_{i,k} = 1$                                                         (5)
       else
           $u_{i,k} = 0,\ U_{i,k} = 0,\ b_{i,k} = 0$                             (6)
       end if
    3) Perform consensus iteratively:
       Initialization: $b_{i,k}^0 = b_{i,k}$, $(\hat{y}_{i,k}^0 = \hat{y}_{i,k},\ \hat{Y}_{i,k}^0 = \hat{Y}_{i,k})$, $(u_{i,k}^0 = u_{i,k},\ U_{i,k}^0 = U_{i,k})$
       for $l = 1$ to $L$ do
           (a) Send $b_{i,k}^{l-1}$, $(\hat{y}_{i,k}^{l-1},\ \hat{Y}_{i,k}^{l-1})$, and $(u_{i,k}^{l-1},\ U_{i,k}^{l-1})$ to all neighbors $j \in \mathcal{N}_i$
           (b) Receive $b_{j,k}^{l-1}$, $(\hat{y}_{j,k}^{l-1},\ \hat{Y}_{j,k}^{l-1})$, and $(u_{j,k}^{l-1},\ U_{j,k}^{l-1})$ from all neighbors $j \in \mathcal{N}_i$
           (c) Update consensus terms
           $\hat{y}_{i,k}^l = \epsilon_{i,j,k} \sum_{j \in \mathcal{N}_i} \hat{y}_{j,k}^{l-1}, \quad \hat{Y}_{i,k}^l = \epsilon_{i,j,k} \sum_{j \in \mathcal{N}_i} \hat{Y}_{j,k}^{l-1}$   (7)
           $u_{i,k}^l = \epsilon_{i,j,k} \sum_{j \in \mathcal{N}_i} u_{j,k}^{l-1}, \quad U_{i,k}^l = \epsilon_{i,j,k} \sum_{j \in \mathcal{N}_i} U_{j,k}^{l-1}$   (8)
           $b_{i,k}^l = \epsilon_{i,j,k} \sum_{j \in \mathcal{N}_i} b_{j,k}^{l-1}$                                 (9)
       end for
    4) Compute the posterior estimation at $k$ time step:
           $w_{i,k}^L = \begin{cases} 1/b_{i,k}^L & if\ b_{i,k}^L \neq 0 \\ 1 & otherwise \end{cases}$   (10)
           $y_{i,k} = \hat{y}_{i,k}^L + w_{i,k}^L u_{i,k}^L,\ Y_{i,k} = \hat{Y}_{i,k}^L + w_{i,k}^L U_{i,k}^L$   (11)
           $x_{i,k} = Y_{i,k}^{-1} y_{i,k}$                                       (12)
```

neighbor nodes of i_{th} node as $\mathcal{N}_i = \{j \in C | i, j \in E\}$. To model the human motion, we use a linear dynamic model, which is

$$x(t+1) = Fx(t) + w \tag{13}$$

where the state vector $x(t) = (p_x(t), p_y(t), p_z(t), v_x(t), v_y(t), v_z(t))^T$ includes the position and velocity of the skeleton joints, F is the state transition matrix and $w \sim N(0, Q)$ is the Gaussian noise with mean 0 and covariance Q. The skeleton measurement from each camera node can be used to update the system state of the filter. The measurement model for each joint using the 3D camera is

$$z(t) = Hx(t) + v \tag{14}$$

where H is the linear observation matrix, z is the predicted measurement of the joint, which has a Gaussian noise $v \sim N(0, R)$ with zero mean and covariance R. The DHCF filter can use the dynamic model for prediction and the measurement model for state updating.

The dynamic hybrid consensus filter (DHCF) which handles information fusion problem of multiple RGBD cameras can be seen in Algorithm 1. First, the main parameters of the algorithm are initialized, i.e., number of consensus iterations L, measurement noise R and process noise Q. Second, the DHCF is conducted in four steps for the time step $k = 1, \cdots, \infty$. The first step is the state prediction based on the human skeleton motion model, where the prior mean, information matrix and information vector are calculated as shown in (1), (2) and (3) respectively. The second step is to compute the information contributions $U_{i,k}$ and $u_{i,k}$ according to skeleton measurements as shown in (4). To weight the information contributions, the DHCF uses a distributed estimation of ratio factor S/N to solve the naive node and overweighting problems, where S is the number of cameras that have valid measurements of human target. We introduce a quantity $b_{i,k}$ to indicate whether the camera is a naive node, i.e., $b_{i,k} = 0$ for the naive sensor and $b_{i,k} = 1$ for the valid sensor ($i \in S$). Since the naive nodes have no valid measurement about current human target, the information contributions for these naive sensor nodes are zero. The third step is to perform the hybrid consensus iteration for L steps. For each iteration l, the i_{th} node sends its prior informations ($y_{i,k}^{l-1}$, $Y_{i,k}^{l-1}$), information contributions ($u_{i,k}^{l-1}$, $U_{i,k}^{l-1}$) and $b_{i,k}^{l-1}$ to neighbor nodes and receives these consensus quantities from neighbors in parallel, then the consensus on this sensor node is performed according to (7), (8) and (9).

The Metropolis weights are used with the DHCF due to its fast convergence rate as shown in [24], which is defined as

$$\epsilon_{i,j,k} = \begin{cases} \frac{1}{1+max\{d_{i,k}, d_{j,k}\}} & if\ j \in \mathcal{N}_i \\ 1 - \sum_{j \in \mathcal{N}_i} \epsilon_{i,j,k} & if\ i = j \\ 0 & otherwise \end{cases} \tag{15}$$

where $d_{i,k}$ and $d_{j,k}$ are the degrees of the camera node i and camera node j at the time step k respectively. Finally, the estimated human skeleton results of the algorithm at the discrete time step k are derived by (12). The DHCF algorithm requires no information of the number of sensor nodes of the camera network, which is suitable for scalable networks.

B. Skeleton Based Human Action Recognition

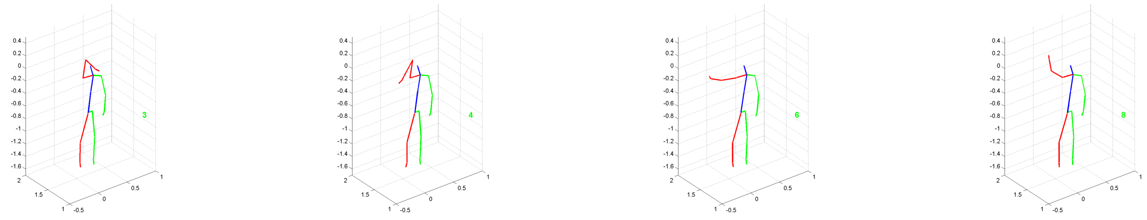
To demonstrate the performance of the fusion algorithm, and to show potential application, we use a skeleton based human action recognition algorithm proposed by [25] to compare the recognition rate. Each skeleton joint can be treated as a point in the Lie group, and the 3D geometric relationships between skeleton joints can be modeled as rotations and translations. The human actions are then modeled as curves in the Lie group, which are mapped to vectors in their Lie algebra. The action classification can be further processed by a combination of dynamic time warping, Fourier temporal pyramid representation and linear SVM.

III. Experimental Setup and Results

We construct a RGBD camera network using four Kinect V2 sensors (C1, C2, C3 and C4), which has a loop topology as shown in Fig. 1. The cameras are calibrated using a chessboard, followed by an iterative closest point optimization method on the point cloud captured by four Kinect V2 sensors to achieve accurate calibration result. Each camera is connected to a computer to record and process the skeleton sequences.

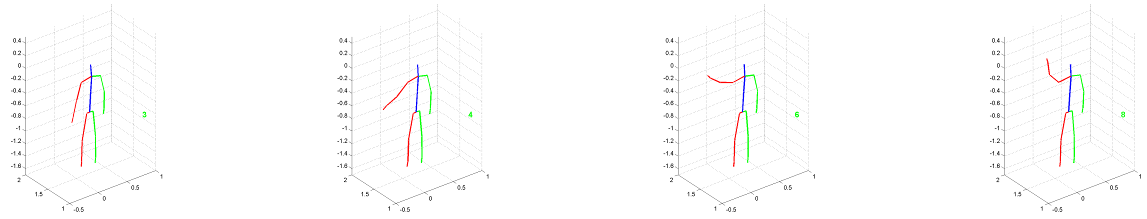
Eight actors are invited to stand in the central position surrounding by cameras as shown in Fig. 1, and do 20 action classes according to the MSRAction3D dataset [23], i.e., two hand wave, bend, side-boxing, forward kick, side kick, jogging, tennis serve, tennis swing, golf swing, pickup and throw, high arm wave, horizontal arm wave, hammer, hand catch, high throw, forward punch, draw x, draw circle, draw tick, hand clap. To show the occlusion problems, we ask each actor to do the same action class five times in three directions, i.e., one in middle, twice in left and twice in right, which are shown as red arrows in Fig. 1.

The skeletons are recorded using the Microsoft Kinect SDK library, which have poor quality when the camera has a side view or back view of the target human. Each skeleton has 20 joints, and each joint has a confidence score, which is corresponding to the probability of the joint detection. However, the skeleton detection algorithm of the Kinect only considers that the human is facing the sensor, such that the algorithm can make mistake if the human shows backside to the Kinect, i.e., the skeleton joint has a high confidence



(a) 3 (b) 4 (c) 6 (d) 8

Fig. 2: This sequences of the skeleton data is the high arm wave action captured by Kinect C1 which is on the left side of the actor, We can see that the estimation of the right arm (red color) is incorrect due to the occlusion by the torso. The blue color, red color and green color corresponding to torso part, right body part and left body part respectively. Since C1 is on the left side of the actor, the right arm of the actor is occluded by the torso, which result in incorrect estimation in the first two frames.



(a) 3 (b) 4 (c) 6 (d) 8

Fig. 3: The DHCF based skeleton fusion result of sensor node C1, which is converged after 9 iterations, where the joints with poor quality are corrected. Other nodes C2, C3 and C4 have the same estimation results as node C1, since the DHCF algorithm is converged.

score for the wrong detection (right elbow labeled as left elbow, etc.). To solve these problems, the joint angle and joint position constraints are used to correct the confidence score, such that a rear view and a front view about the human can be discriminated.

We introduce the DHCF algorithm to fuse the skeleton from multiple views as shown in Algorithm 1, where the transition matrix F and the observation matrix H are defined as

$$F = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{16}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{17}$$

The Gaussian process noise Q and measurement noise R are defined as

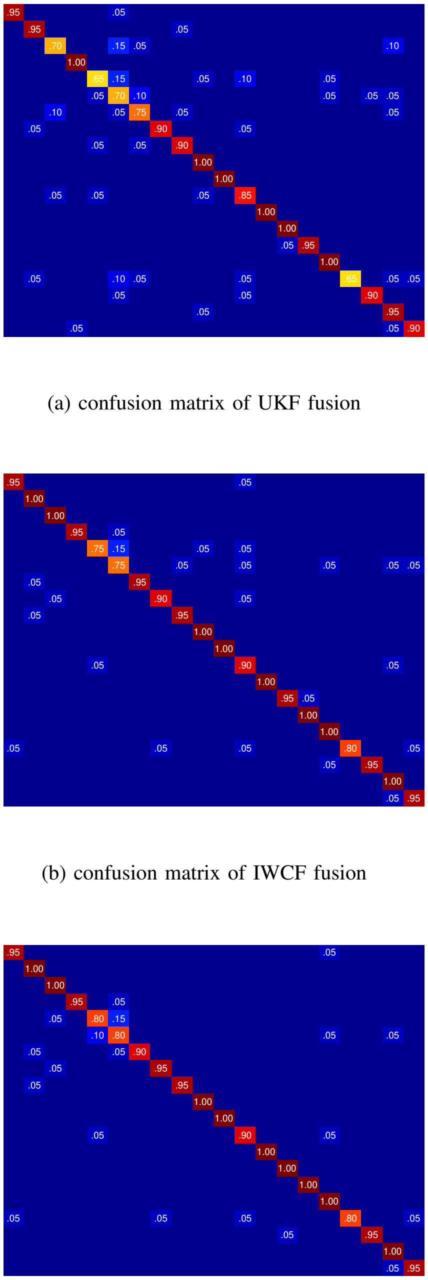
$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 50 & 0 & 0 \\ 0 & 0 & 0 & 0 & 50 & 0 \\ 0 & 0 & 0 & 0 & 0 & 50 \end{bmatrix}, \tag{18}$$

$$R = \begin{bmatrix} 20 & 0 & 0 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 & 0 & 0 \\ 0 & 0 & 20 & 0 & 0 & 0 \end{bmatrix}. \tag{19}$$

The iteration number L can be set as different values, such that we can see the convergence rate of the algorithm.

For action recognition, the cross subject test setting is employed, i.e., we use half of the subjects half $(2, 4, 6, 8)$ for testing and the other half $(1, 3, 5, 7)$ for training, since we have 8 individual persons to demonstrate the 20 action classes.

The average action recognition results are summarized in Table I. We can see that the DHCF based skeleton fusion algorithm has higher recognition rate than the results from single views (C1 to C4). The recognition rates of DHCF



(a) confusion matrix of UKF fusion

(b) confusion matrix of IWCF fusion

(c) confusion matrix of DHCF fusion

Fig. 4: This figure shows that the confusion matrix of UKF based centralized fusion (a), and IWCF and DHCF based distributed fusion (b) respectively. The values in the confusion matrix present the recognition accuracy. It is clear that the DHCF based method outperforms the others.

based algorithm for all cameras are converged to 94.50% after 9 iterations. The occlusion problem can be seen in Fig. 2, where the person is facing the direction between C3 and C4, such that the C1 and C2 have side view and rear view of the human body respectively. Therefore, the right part of the human body is occluded from the C1 and C2 views. However, the proposed fusion algorithm can correct this situation after few number of iterations as shown in Fig. 3. All cameras can converge to the same human pose result and achieve the same recognition rate, so each camera can make decision locally with correct human pose information through the distributed camera network using the proposed method.

TABLE I: Comparison of Recognition Rate

Data source	Method	Recognition result
Single view	Kinect C1	82.75%
	Kinect C2	67.00%
	Kinect C3	75.50%
	Kinect C4	87.50%
Multiple view	UKF [13]	88.50%
	IWCF (L=9) [19]	93.75%
	DHCF (L=9)	94.50%

The experiment result also shows that the DHCF based fusion algorithm outperforms the IWCF based one for skeleton fusion, since the DHCF has faster convergence rate and can preserve the consistency of the local filters, such that the novel information is never overestimated [22], [23]. Furthermore, we also compare our method with the state of the art work OpenPTrack [13], which is based on centralized unscented Kalman filters (UKF) to track multiple people with asynchronous data sources. For comparison, we use the same filter parameters and keep the same asynchronous mechanism as the original OpenPTrack algorithm. The data association part of the original OpenPTrack algorithm is removed, since our dataset only has one person for each image frame. The results in the Table I show that our method has better recognition accuracy than the OpenPTrack algorithm, i.e., 94.50% vs 88.50%. The confusion matrix of recognition results for all methods are shown in Fig. 4. which show that the DHCF has better performances than UKF and IWCF based methods.

IV. Conclusion and Future Works

In this work, we propose a dynamic hybrid consensus filter for skeleton fusion from multiple views, which can handle occlusion problem efficiently and further improve the action recognition accuracy. To demonstrate the idea, we use a distributed RGBD camera network and collect 20 action classes from 8 individual persons. Finally, a Lie group based action recognition method is used to show the improved performance of the proposed idea. In future, a distributed data association method can be combined with our current work to handle multiple human tracking and pose estimation problems.

References

[1] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *CVPR 2011*, June 2011, pp. 1297–1304.

[2] J. Han, L. Shao, S. Member, D. Xu, and J. Shotton, "Enhanced Computer Vision with Microsoft Kinect Sensor : A Review," *IEEE transactions on cybernetics*, vol. 43, no. 5, pp. 1318–1334, 2013.

[3] C. Morato, K. N. Kaipa, B. Zhao, and S. K. Gupta, "Toward safe human robot collaboration by using multiple kinects based real-time human tracking," *Journal of Computing and Information Science in Engineering*, vol. 14, no. 1, p. 011006, 2014.

[4] P. N. Pathirana, S. Li, H. M. Trinh, and A. Seneviratne, "Robust real-time bio-kinematic movement tracking using multiple kinects for tele-rehabilitation," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1–1, 2016.

[5] F. Geiselhart, M. Otto, and E. Rukzio, "On the use of multi-depth-camera based motion tracking systems in production planning environments," *Procedia Cirp*, vol. 41, pp. 759–764, 2016.

[6] J. K. Aggarwal and L. Xia, "Human activity recognition from 3d data: A review ," *Pattern Recognition Letters*, vol. 48, no. 1, pp. 70–80, 2014.

[7] M. Caon, Y. Yue, J. Tscherring, E. Mugellini, and O. Abou Khaled, "Context-Aware 3D Gesture Interaction Based on Multiple Kinects," in *The First International Conference on Ambient Computing, Applications, Services and Technologies*, 2011, pp. 7–12.

[8] N. A. Azis, Y.-S. Jeong, H.-J. Choi, and Y. Iraqi, "Weighted averaging fusion for multi-view skeletal data and its application in action recognition," *IET Computer Vision*, vol. 10, no. 2, pp. 134–142, 2016.

[9] J. J. Kim, I. Lee, J. J. Kim, and S. Lee, "Implementation of an omnidirectional human motion capture system using multiple Kinect sensors," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E98-A, no. 9, pp. 2004–2008, 2015.

[10] K.-Y. Yeung, T.-H. Kwok, and C. C. L. Wang, "Improved Skeleton Tracking by Duplex Kinects: A Practical Approach for Real-Time Applications," *Journal of Computing and Information Science in Engineering*, vol. 13, no. 4, p. 041007, 2013.

[11] R. Paper, S. Moon, Y. Park, D. W. Ko, and I. H. Suh, "Multiple Kinect Sensor Fusion for Human Skeleton Tracking Using Kalman Filtering Regular Paper," *International Journal of Advanced Robotic Systems*, vol. 13, no. 65, pp. 1–10, 2016.

[12] V. Stohne, "Real-time filtering for human pose estimation using multiple Kinects," Ph.D. dissertation, 2014.

[13] M. Carraro, M. Munaro, J. Burke, and E. Menegatti, "Real-time marker-less multi-person 3D pose estimation in RGB-Depth camera networks," in *ICRA*, 2018. [Online]. Available: http://arxiv.org/abs/1710.06235

[14] A. T. Kamal, J. H. Bappy, J. A. Farrell, and A. K. Roy-Chowdhury, "Distributed multi-target tracking and data association in vision networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1397–1410, July 2016.

[15] B. Song, C. Ding, A. T. Kamal, J. A. Farrell, and A. K. Roy-Chowdhury, "Distributed camera networks," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 20–31, 2011.

[16] B. Song, a. T. Kamal, C. Soto, C. Ding, J. a. Farrell, and a. K. Roy-Chowdhury, "Tracking and Activity Recognition Through Consensus in Distribution Camera Network," *IEEE Trans. on Image Processing*, vol. 19, no. 10, pp. 2564–2579, 2010.

[17] R. Olfati-Saber, "Kalman-Consensus filter: Optimality, stability, and performance," in *Proceedings of the IEEE Conference on Decision and Control*, 2009, pp. 7036–7042.

[18] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[19] Z. Wang, G. Liu, and G. Tian, "Human Skeleton Tracking Using Information Weighted Consensus Filter in Distributed Camera Networks," in *Chinese Automation Congress*, 2017, pp. 4640 – 4644.

[20] J. Li, G. Liu, G. Tian, X. Zhu, and Z. Wang, "Distributed rgbd camera network for 3d human pose estimation and action recognition," in *2018 21st International Conference on Information Fusion (FUSION)*, July 2018, pp. 1554–1558.

[21] G. Liu, G. Tian, J. Li, X. Zhu, and Z. Wang, "Human action recognition using a distributed rgb-depth camera network," *IEEE Sensors Journal*, vol. 18, no. 18, pp. 7570–7576, Sept 2018.

[22] G. Battistelli, L. Chisci, and C. Fantacci, "Parallel consensus on likelihoods and priors for networked nonlinear filtering," *IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 787–791, 2014.

[23] G. Liu and G. Tian, "Distributed Vision Network for Multiple Target Tracking Using a Dynamic Hybrid Consensus Filter," in *IEEE International Conference on Automation Science and Engineering*, 2016, pp. 805 – 808.

[24] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *International Symposium on Information Processing in Sensor Networks*, 2005, pp. 63–70.

[25] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3d skeletons as points in a lie group," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 588–595.