Adaptive Diffusion-Based Track Assisted Multi-Object Labeling in Distributed Camera Networks

Freweyni K. Teklehaymanot*†, Michael Muma*, Abdelhak M. Zoubir*†

* Signal Processing Group
Technische Universität Darmstadt
Merckstr. 25, 64283 Darmstadt, Germany
{muma, zoubir}@spg.tu-darmstadt.de

† Graduate School CE
Technische Universität Darmstadt
Dolivostr. 15, D-64293 Darmstadt, Germany
teklehaymanot@gsc.tu-darmstadt.de

***Abstract*—The tracking and labeling of multiple objects in multiple cameras is a fundamental task in applications such as video surveillance, autonomous driving, and sports analysis. In an ad-hoc multi-camera network without a fusion center nodes can benefit from local cooperation to solve signal processing tasks, such as distributed image enhancement. A crucial first step for the successful cooperation of neighboring nodes is to answer the question: *Who observes what?*. In this paper, an adaptive algorithm is proposed that enables cameras with different view points to assign the same identity to the same object across time frames without assuming the availability of camera calibration information or requiring the registration of camera views. Information which is extracted directly from the videos and is shared in the network via a diffusion algorithm is exploited to jointly solve multi-object tracking and labeling problems in a multi-camera network. A real-data use case of pedestrian labeling is provided, which demonstrates that a high labeling accuracy can be achieved in a multi-object multi-camera setup with low video resolution and frequent object occlusions.**

I. Introduction

Distributed and adaptive signal processing is a rapidly growing field of research, which enables the development of novel paradigms for signal and parameter estimation. One such paradigm is the multiple devices multiple tasks (MDMT) paradigm where multiple devices cooperate to solve multiple complex signal processing tasks [1]–[3]. A distributed camera network containing multiple heterogeneous devices, such as smart phones, tablets and/or handheld cameras, which neither has a pre-defined network structure nor a centralized computing unit can make use of the MDMT paradigm. Nodes in a distributed camera network can be interested in, e.g., image enhancement, object detection, pose analysis, and object tracking. In most real-world applications, the signal received by these nodes is contaminated by noise, contains frequent object occlusions, and lacks visibility in densely crowded scenes. Under the MDMT paradigm, such nodes can benefit from cooperation with their immediate neighbors to solve their signal processing task of interest. The answer to the question "*who observes what?*" holds the key to the successful cooperation of neighboring nodes in a distributed network. Recently, some generic clustering and classification algorithms have been proposed for distributed sensor networks [4]–[6]. The research question addressed in this paper is to develop a distributed and adaptive multi-object labeling algorithm for a multi-camera network without assuming any form of camera calibration or utilizing a centralized computing unit that fuses all information collected from different cameras. The advantage of such an approach is that it is applicable to ad-hoc networks of mobile cameras. Furthermore, the approach is robust against a single node failure and since it is based on the diffusion principle [7] it is adaptive and scalable.

The work of F. K. Teklehaymanot is supported by the 'Excellence Initiative' of the German Federal and State Governments and the Graduate School of Computational Engineering at Technische Universität Darmstadt. The work of M. Muma has been supported by the project HANDiCAMS which acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 323944.

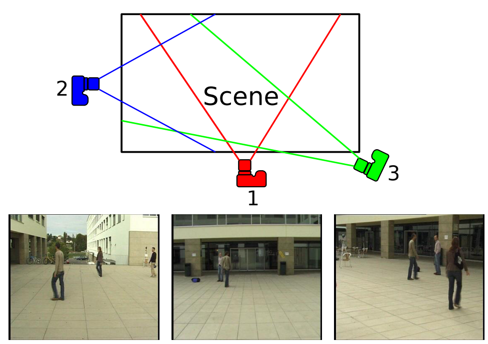

Fig. 1. A wireless camera network [8], [9] observing a scene of interest. The top image shows a camera network with $J = 3$ nodes continuously monitoring a scene of interest from different observation angles. The bottom images show frames captured at the same time instant by cameras $2, 1$, and 3, respectively.

The tracking and labeling of multiple objects in multiple cameras is fundamental, e.g. for applications such as video surveillance, autonomous driving, and sports analysis. Multi-object multi-camera tracking systems must maintain consistent labels of objects of interest across camera views and over time to take advantage of the information available from different

ISBN 978-0-9928626-7-1 © EURASIP 2017

camera views in the network. Previous approaches to the labeling of multiple objects across views include principal axis-based integration of multi-camera information [10], nonlinear manifold learning and system dynamics identification [11], and approaches that either use homography or camera calibration information to register stationary camera views on top of a known ground plane [8], [9], [12]–[14]. These state-of-the-art methods are centralized approaches in the sense that camera views are aggregated into a ground plane to make sure that unique and consistent labels are assigned to the objects in the scene of interest.

In this paper, we propose a new fully distributed algorithm which does not require camera view registration to ensure that the same object is provided with the same identity in a multi-camera network. The information available in a neighborhood of cameras is exploited to provide unique and consistent labels to multiple objects across multiple camera views and time frames. Each node solves a regularized cost function during the assignment of a label to a particular object to exploit both the information obtained from a local (single node) Kalman filter-based tracker and a diffusion-based labeling algorithm.

The paper is organized as follows. Section II formulates the distributed and adaptive multi-object multi-camera labeling problem and Section III presents the proposed algorithm in detail. A numerical evaluation of the proposed algorithm using a real multi-camera network example is provided in Section IV. Finally, conclusions are drawn and future work is briefly discussed in Section V.

II. Problem Formulation

Consider a wireless camera network with J nodes distributed over some geographic region as the one shown in Fig. 1. The set of nodes that communicates directly with node $j \in \{1, \dots, J\} \triangleq \mathcal{J}$ is called the neighborhood of node j and is denoted by $\mathcal{B}_j \subseteq \mathcal{J}$. Let $\boldsymbol{X}_{jn} \in \mathbb{R}^{q \times m_{jn}}$ represent the q-dimensional feature vectors extracted at the jth node from the m_{jn} objects that are observed by the camera of node j at time instant n. Each feature vector belongs to a certain cluster $\mathcal{C}_k, k \in \{1, \dots, K_n\}$, where k is the cluster label. The total number of objects (clusters) K_n at time instant n is assumed to be known or estimated a priori, e.g. using [15]. Due to the different view points of the cameras, even at the same time instant, the number of objects observed by different cameras differs. Our research goal is to adaptively estimate cluster centroids and enable cameras with different view points to assign the same identity to the same object in the scene of interest.

III. Proposed Adaptive Diffusion-Based Track Assisted Multi-Object Labeling Algorithm

A distributed and adaptive track assisted multi-object labeling algorithm for multi-camera networks, which is based on the Adapt Then Combine (ATC) diffusion principle [7], is proposed and depicted in Fig. 2. The general procedure involved in the proposed framework for node $j \in \mathcal{J}$ is summarized as follows.

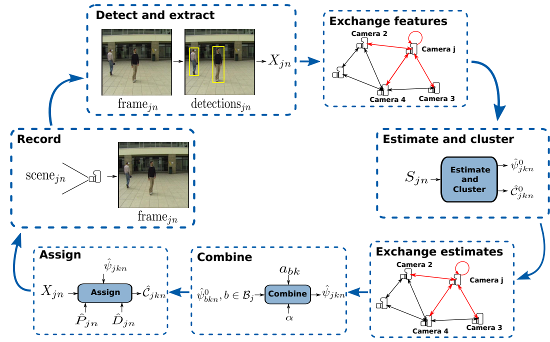

Fig. 2. An overview of the distributed and adaptive diffusion-based track assisted multi-object labeling algorithm.

1) **Record:** the camera at node j captures frame$_{jn}$ from scene$_{jn}$ at time instant n.
2) **Detect and extract:** if there are objects of interest in frame$_{jn}$, then each node j extracts feature vectors from the detected bounding boxes of the objects and collects them in $\boldsymbol{X}_{jn}$. Otherwise, the camera at node j continues to record at time instant $n+1$.
3) **Exchange features:** each node j exchanges its feature vectors $\boldsymbol{X}_{jn}$ within its neighborhood $\mathcal{B}_j$. The own and received feature vectors are stored in the matrix $\tilde{\boldsymbol{X}}_{jn} \in \mathbb{R}^{q \times \sum_{b \in \mathcal{B}_j} m_{bn}}$, where m_{bn} represents the number of objects detected by node $b \in \mathcal{B}_j$ at time instant n. Next, each node j accumulates $\tilde{\boldsymbol{X}}_{ji}, i = 1, \dots, n$, inside the matrix $\boldsymbol{S}_{jn} \in \mathbb{R}^{q \times N_{jn}}$, where $N_{jn} = N_{j(n-1)} + \sum_{b \in \mathcal{B}_j} m_{bn}$ is the total number of feature vectors at node j at time instant n.
4) **Estimate and cluster:** each node j performs K-means to minimize the ℓ_2 distance between the feature vectors in $\boldsymbol{S}_{jn}$ and the cluster centroids $\boldsymbol{w}^0_{jkn} \in \mathbb{R}^{q \times 1}$
$$\arg\min_{\boldsymbol{w}^0_{jkn}} \sum_{k=1}^{K_n} \sum_{i=1}^{N_{jn}} \| \boldsymbol{S}_{jn}(:, i) - \boldsymbol{w}^0_{jkn} \|_2, \tag{1}$$
where $\boldsymbol{S}_{jn}(:, i)$ denotes the ith column of $\boldsymbol{S}_{jn}$ and the initial cluster centroids are obtained using the K-means++ initialization algorithm [16]. This results in a unique cluster label $\hat{\mathcal{C}}^0_{jkn}$ for each feature vector in $\boldsymbol{S}_{jn}$. The feature vectors that belong to the same cluster $k \in \{1, \dots, K_n\}$ are saved in $\boldsymbol{V}_{jkn} \in \mathbb{R}^{q \times N_{jkn}}$, where $\sum_{k=1}^{K_n} N_{jkn} = N_{jn}$. Then, the row-wise mean of $\boldsymbol{V}_{jkn}$ is computed as
$$\hat{\boldsymbol{\psi}}^0_{jkn} = \mathrm{mean}\left(\boldsymbol{V}_{jkn} \right), \tag{2}$$
where $\hat{\boldsymbol{\psi}}^0_{jkn} \in \mathbb{R}^{q \times 1}$. The minimization of the ℓ_2 distance using $\boldsymbol{w}^0_{jkn}$ in Eq. (1) is performed only if $K_n > K_{n-1}$. Otherwise, $\boldsymbol{w}^0_{jkn}$ is replaced with $\hat{\boldsymbol{\psi}}_{jk(n-1)}$ in Eq. (1).
5) **Exchange estimates:** each node j exchanges its intermediate centroid estimates $\hat{\boldsymbol{\psi}}^0_{jkn}$ within its neighborhood $\mathcal{B}_j$. Synchronization of $\hat{\boldsymbol{\psi}}^0_{bkn}, b \in \mathcal{B}_j$, is necessary because the order of $\hat{\boldsymbol{\psi}}^0_{jkn}$ is random at different nodes.

ISBN 978-0-9928626-7-1 © EURASIP 2017

The reordering of the intermediate centroid estimates is performed by computing the Euclidean distance relative to an arbitrarily chosen neighborhood head in $\mathcal{B}_j$.

6) **Combine estimates:** each node j adapts its centroid estimates using

$$\hat{\boldsymbol{\psi}}_{jkn} = \alpha \cdot \hat{\boldsymbol{\psi}}^0_{jkn} + (1-\alpha) \cdot \sum_{b \in \mathcal{B}_j \setminus \{j\}} a_{bj} \cdot \hat{\boldsymbol{\psi}}^0_{bkn}, \tag{3}$$

where α controls the tradeoff between the weight given to the own and neighborhood estimates. In this paper, uniform combination weights are used which are given by

$$a_{bj} = 1/|\mathcal{B}_j \setminus \{j\}|, \tag{4}$$

where $|\mathcal{B}_j \setminus \{j\}|$ represents the cardinality of the set $\mathcal{B}_j$ without node j.

7) **Assign:** at this step, each node j assigns unique labels $\hat{\mathcal{C}}_{jkn}$ to objects of interest in frame$_{jn}$. We propose a regularized cost function that aggregates the information obtained from a local Kalman filter-based tracker and a diffusion-based labeling algorithm. In particular,

$$\begin{aligned} \boldsymbol{Z}_{jn}(t,m) = \lambda \cdot \|\hat{\boldsymbol{d}}_{jnm} - \hat{\boldsymbol{p}}_{jnt}\|_2 + \\ (1-\lambda) \cdot \|\boldsymbol{X}_{jn}(:,m) - \hat{\boldsymbol{\psi}}_{jtn}\|_2, \end{aligned} \tag{5}$$

where λ is the regularization parameter, $\hat{\boldsymbol{d}}_{jnm}$ and $\hat{\boldsymbol{p}}_{jnt}$ are detected and predicted bounding box center positions for $m = 1, \ldots, m_{jn}$ and $t = 1, \ldots, t_{jn}$, respectively. The total number of open tracks in the jth node at time instant n is denoted by t_{jn} and $\hat{\boldsymbol{\psi}}_{jtn}$ represents the cluster centroid that belongs to the tth track. The two ℓ_2 distances in Eq. (5) are normalized by their respective maximum to make sure that they are comparable. Then, the Hungarian algorithm [17] is applied on $\boldsymbol{Z}_{jn}$ to assign unique labels $\hat{\mathcal{C}}_{jkn}$ to feature vectors in $\boldsymbol{X}_{jn}$.

Algorithm 1 summarizes the adaptive diffusion-based track assisted multi-object labeling algorithm.

IV. Real-Data Application: Multi-Camera Network for Pedestrian Labeling

In this section, we first describe the multi-camera network setup and the pedestrian detection algorithm with the associated feature extraction method. Then, the network-wide performance measures used to evaluate the labeling performance of the proposed algorithm are explained. Finally, real-data results of the diffusion-based track assisted multi-object labeling algorithm are provided.

A. Multi-Camera Video Sequence

We use an outdoor video sequence that was recorded by three unsynchronized digital video cameras on the campus of École Polytechnique Fédérale de Lausanne in Switzerland [8], [9]. The cameras were mounted at head level ($\simeq$ 1.80m), observing the scene of interest from different angles, and the captured videos were synchronized by hand. Each video sequence contains 2000 frames and up to five people are seen entering and exiting the scene of interest at different times. The multi-camera video sequence is challenging in the sense that the videos have low resolution, the cameras monitor pedestrians from different angles, and there are frequent pedestrian occlusions.

Algorithm 1 Distributed and adaptive diffusion-based track assisted multi-object labeling in multi-camera networks

for $n = 1, 2, \ldots$ **do**
    for $j = 1, 2, \ldots, J$ **do**
        record frame$_{jn}$
        if objects are detected **then**
            extract feature vectors $\boldsymbol{X}_{jn}$
        else
            proceed with record step at $n+1$
        end if
    end for
    for $j = 1, 2, \ldots, J$ **do**
        exchange $\boldsymbol{X}_{jn}$ within $\mathcal{B}_j$
        store own and received feature vectors in $\tilde{\boldsymbol{X}}_{jn}$
        accumulate $\tilde{\boldsymbol{X}}_{ji}, i = 1, \ldots, n$, in $\boldsymbol{S}_{jn}$
    end for
    for $j = 1, 2, \ldots, J$ **do**
        perform K-means according to Eq. (1)
        calculate $\hat{\boldsymbol{\psi}}^0_{jkn}$ via Eq. (2)
    end for
    for $j = 1, 2, \ldots, J$ **do**
        exchange $\hat{\boldsymbol{\psi}}^0_{jkn}$ within $\mathcal{B}_j$
        synchronize $\hat{\boldsymbol{\psi}}^0_{bkn}, b \in \mathcal{B}_j$
    end for
    for $j = 1, 2, \ldots, J$ **do**
        combine $\hat{\boldsymbol{\psi}}^0_{bkn}, b \in \mathcal{B}_j$ via Eq. (3)
    end for
    for $j = 1, 2, \ldots, J$ **do**
        solve Eq. (5) using Hungarian algorithm
    end for
end for

B. Pedestrian Detection and Feature Extraction

To detect pedestrians in the scene of interest, we use an already trained MATLAB© implementation of the Aggregated Channel Features (ACF) pedestrian detector [18]. The ACF pedestrian detector uses boosting to train decision trees over features and a multiscale sliding window approach to distinguish objects of interest from background.

We consider the extraction of two types of color features for the purpose of unsupervised labeling of detected pedestrians across the camera network. For the first color feature, the image patch (detected pedestrian) is subdivided into three concentric rings and a 10-bin histogram per color channel is computed for every region in a cumulative manner. The resulting feature vector is the concatenation of the three color channels, which correspond to red, green, and blue (RGB), resulting in a 90 dimensional feature vector per detected pedestrian. For the second color feature, the detected image patch is cut horizontally into four equal parts and the average of the pixel values inside each part is computed for each color channel. The concatenation of the features for the three color channels results in a vector of dimension 12. Finally, these two color features are concatenated to create a feature vector of dimension 102.

C. Network-Wide Performance Measures

We define two network-wide performance measures, i.e., the average labeling rate $\mathcal{L}_{k,k}$ and the average mislabeling rate $\mathcal{M}_{k,h}$ as follows

$$\mathcal{L}_{k,k} = \frac{1}{J}\sum_{j=1}^{J}\frac{(\hat{\mathcal{C}}_{jk} == \mathcal{C}_k)}{N_{jk}} \tag{6}$$

$$\mathcal{M}_{k,h} = \frac{1}{J}\sum_{j=1}^{J}\frac{(\hat{\mathcal{C}}_{jh} \neq \mathcal{C}_k)}{N_{jk}}, \tag{7}$$

where $\mathcal{L}_{k,k}$ indicates if object k is provided with the correct label k, $\mathcal{M}_{k,h}$ indicates if object k is provided with a wrong label h, $\mathcal{C}_k$ is the set of ground truth labels, and N_{jk} is the number of times the kth object was detected. To evaluate the performance of the proposed algorithm, $\mathcal{L}_{k,k}$ and $\mathcal{M}_{k,h}$ are placed in the diagonal and off-diagonal, respectively, of a confusion matrix. In the confusion matrix, both $\mathcal{L}_{k,k}$ and $\mathcal{M}_{k,h}$ are given in percentage.

D. Real-Data Results

The multi-camera video sequence contains $J = 3$ stationary cameras and a neighborhood size of $|\mathcal{B}_j| = 3$ is considered. The weight parameter is set to $\alpha = 0.5$ and the regularization parameter is $\lambda = 0.4$. The number of pedestrians seen until the nth time instant, K_n, is assumed to be known.

The labeling performance of camera 3 on pedestrian 1 is depicted in Fig 3. A zero label indicates that either the pedestrian is not detected in the current frame or he/she is no longer in the scene of interest. For this particular multi-camera video sequence, the ACF pedestrian detector has a high misdetection rate and the position of the bounding boxes is unstable. In some frames, multiple bounding boxes are detected for a single pedestrian in the scene. These problems affect the performance of the Kalman filter-based tracker and the diffusion-based labeling algorithm. The red spikes in Fig. 3 indicate mislabels and result due to multiple detections for a single pedestrian. However, even under such conditions, the proposed diffusion-based multi-object labeling algorithm performs reasonably well.

Fig. 4 shows an example of the network-wide labeling results of the proposed algorithm using ACF pedestrian detector. The bounding box of each pedestrian is provided with a unique color and number. We define the multi-object labeling algorithm to be performing well if the same pedestrian is provided with the same color of bounding box and number across different camera views and time frames. The proposed algorithm is able to provide unique and consistent labels to pedestrians in the scene even when there are partial occlusions.

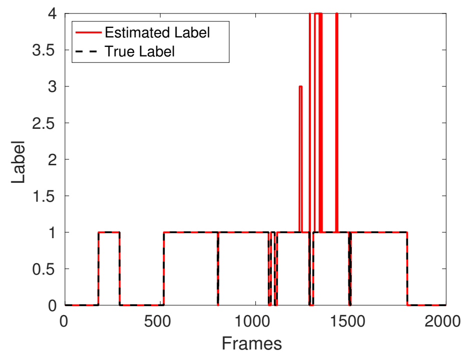

Fig. 3. A comparison of the estimated and true labels for pedestrian 1 in the video sequence captured by camera 3.

Table I shows the confusion matrix of the proposed algorithm averaged over all cameras and all time frames. The confusion matrix shows the average labeling rate $\mathcal{L}_{k,k}$ in the diagonal and the average mislabeling rate $\mathcal{M}_{k,h}$ in the off-diagonal as defined in Section IV-C. The proposed algorithm performs well in providing unique and consistent labels to the first four pedestrians in the scene. Lower labeling performance is exhibited for the fifth pedestrian because he is only visible for a short time and is not well detected by the ACF pedestrian detector. Table II shows the confusion matrix of the proposed algorithm when we replace the ACF pedestrian detector with the ground truth pedestrian detections which provides the best achievable performance of our algorithm. If a good pedestrian detector is used, the proposed algorithm achieves a very high labeling performance even when the color features have strong similarities, which is the case when pedestrians are dressed similarly.

TABLE I
CONFUSION MATRIX IN PERCENTAGES WITH $\mathcal{L}_{k,k}$ IN THE DIAGONAL AND $\mathcal{M}_{k,h}$ IN THE OFF-DIAGONAL USING ACF PEDESTRIAN DETECTOR.

		Estimated Labels				
		1	2	3	4	5
True Labels	1	88.75	3.56	0.93	0.96	5.80
	2	0	95.12	1.56	0.07	3.18
	3	0.79	3.36	83.66	4.57	8.12
	4	0	3.89	5.03	91.08	0
	5	4.66	8.65	48.59	18.33	19.77

TABLE II
CONFUSION MATRIX IN PERCENTAGES WITH $\mathcal{L}_{k,k}$ IN THE DIAGONAL AND $\mathcal{M}_{k,h}$ IN THE OFF-DIAGONAL WHEN ACF PEDESTRIAN DETECTOR IS REPLACED WITH GROUND TRUTH PEDESTRIAN DETECTIONS.

		Estimated Labels				
		1	2	3	4	5
True Labels	1	96.25	0.79	1.31	0.10	1.55
	2	0.54	97.26	1.11	0.85	0.24
	3	1.24	1.80	89.44	1.48	6.05
	4	0	0.21	1.88	97.92	0
	5	0	0	0.29	1.11	98.60

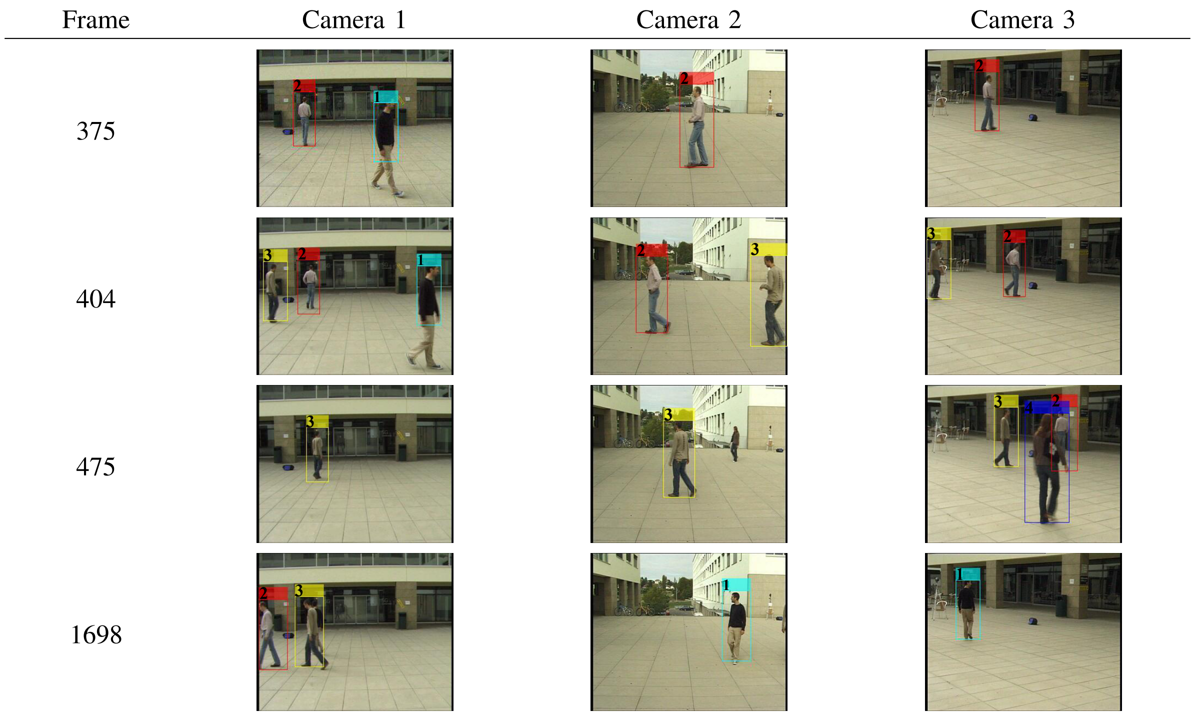

Fig. 4. An example of the network-wide labeling results of the proposed algorithm using ACF pedestrian detector. Each row displays several views coming from different cameras at the same time frame. The color of each bounding box and the number displayed on it show the identity that is given to the particular pedestrian.

V. Conclusion

We proposed a distributed and adaptive diffusion-based track assisted multi-object labeling algorithm for multi-camera networks. The presented algorithm is able to provide unique and consistent labels to multiple objects across camera views and time frames without camera view registration. The performance was tested on a real multi-camera network use-case. Good labeling performance was achieved given the number of pedestrians in the scene of interest. In future work, we will extend our multi-object multi-camera labeling framework by automatically estimating the number of objects in the scene of interest. The approach can also be extended to the case where information from heterogeneous sensor modalities is used to consistently label objects across the network.

References

[1] A. Bertrand and M. Moonen, "Distributed signal estimation in sensor networks where nodes have different interests," *Signal Process.*, vol. 92, pp. 1679–1690, Jul. 2012.

[2] J. Chen, C. Richard, and A. H. Sayed, "Diffusion LMS over multitask networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733–2748, Jun. 2015.

[3] J. Plata-Chaves, A. Bertrand, M. Moonen, S. Theodoridis, and A. M. Zoubir, "Heterogeneous and multi-task wireless sensor networks - algorithms, applications and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 3, pp. 450–465, Apr. 2017.

[4] P. Binder, M. Muma, and A. M. Zoubir, "Robust and adaptive diffusion-based classification in distributed networks," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 34, pp. 1–13, Mar. 2016.

[5] F. K. Teklehaymanot, M. Muma, B. Béjar-Haro, P. Binder, A. M. Zoubir, and M. Vetterli, "Robust diffusion-based unsupervised object labelling in distributed camera networks," in *Proc. 12th IEEE AFRICON*, Apr. 2015.

[6] J. Chen and I. D. Schizas, "Distributed information-based clustering of heterogeneous sensor data," *Elsevier Signal Process.*, vol. 126, pp. 35–51, Sep. 2016.

[7] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, pp. 460–497, Apr. 2014.

[8] J. Berclaz, F. Fleuret, E. Türetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1806–1819, Sep. 2011.

[9] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 267–282, Feb. 2008.

[10] W. Du and J. Piater, "Multi-camera people tracking by collaborative particle filters and principal axis-based integration," in *Asian Conf. Comput. Vis. (ACCV)*, Nov. 2007, pp. 365–374.

[11] V. I. Morariu and O. I. Camps, "Modeling correspondences for multi-camera tracking using nonlinear manifold learning and target dynamics," in *Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jul. 2006, pp. 545–552.

[12] J. Kang, I. Cohen, and G. Medioni, "Tracking people in crowded scenes across multiple cameras," in *Asian Conf. Comput. Vis. (ACCV)*, Jan. 2004.

[13] S. Khan and M. Shah, "A multiview approach to tracking people in crowded scenes using a planar homography constraint," in *Eur. Conf. Comput. Vis.: Lecture Notes Comput. Sci.*, vol. 3954, May 2006, pp. 133–146.

[14] M. Taj and A. Cavallaro, "Multi-camera track-before-detect," in *3rd ACM/IEEE Int. Conf. Distrib. Smart Cameras*, Oct. 2009.

[15] F. K. Teklehaymanot, M. Muma, J. Liu, and A. M. Zoubir, "In-network adaptive cluster enumeration for distributed classification/labeling," in *Eur. Signal Process. Conf. (EUSIPCO)*, Apr. 2016.

[16] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[17] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, Mar. 1957.

[18] P. Dollár, R. Appel, S. J. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1532–1545, Aug. 2014.