Network Consistent Data Association

Anirban Chakraborty, *Member, IEEE,* Abir Das, *Student Member, IEEE,* and Amit Roy-Chowdhury, *Senior Member, IEEE*

Abstract—Existing data association techniques mostly focus on matching pairs of data-point sets and then repeating this process along space-time to achieve long term correspondences. However, in many problems such as person re-identification, a set of data-points may be observed at multiple spatio-temporal locations and/or by multiple agents in a network and simply combining the local pairwise association results between sets of data-points often leads to inconsistencies over the global space-time horizons. In this paper, we propose a novel Network Consistent Data Association (NCDA) framework formulated as an optimization problem that not only maintains consistency in association results across the network, but also improves the pairwise data association accuracies. The proposed NCDA can be solved as a binary integer program leading to a globally optimal solution and is capable of handling the challenging data-association scenario where the number of data-points varies across different sets of instances in the network. We also present an online implementation of NCDA method that can dynamically associate new observations to already observed data-points in an iterative fashion, while maintaining network consistency. We have tested both the batch and the online NCDA in two application areas - person re-identification and spatio-temporal cell tracking and observed consistent and highly accurate data association results in all the cases.

Index Terms—Data association, Network consistency, Integer program, Person Re-identification, Spatio-temporal cell tracking.

1 Introduction

In many computer vision problems such as tracking, re-identification *etc.*, associating detected targets across space and/or time is of utmost importance. Most data association approaches try to find correspondences between pairs of instances of a set of datapoints and repeat this process along space/time to obtain long term correspondences. However, this local approach for finding correspondences may lead to inconsistencies over the global space-time horizons. The goal of this paper is to show how *globally consistent* correspondence results can be obtained by enforcing suitable network-level constraints over the entire set of observation data points.

The notion of *Network Consistency* is applicable to a wide class of data association problems, especially where the set of all observations can be partitioned into multiple non-overlapping subsets such that no two observations from within the same subset may be associated with one another. The target of such a problem is to estimate pairwise associations between observations belonging to different subsets, over all possible pairs of such subsets. The probable *links* associating pairs of observations form a network of associated observations and thus, the study of analyzing the feasibility of pairwise associations in context to the overall network can be termed as *Network Consistent Data Association*. The class of data association problems where such consistency is of utmost importance includes person re-identification, feature matching across multiple sensors (such as multi-robot localization, mapping, exploration etc.) and/or across multiple time points (tracking), feature matching in high dimensional biological datasets (such as 3D+t cell tracking) and many more. We explain the problem more precisely through two of such examples below.

Consider the well studied person re-identification problem where the objective is to associate targets across cameras with non overlapping field-of-views (FoVs). Most widely used approaches focus on pairwise re-identification, *i.e.*, association between two camera FoVs. Even if the re-identification accuracy for each camera pair is high, it might contain many global association inconsistencies over the entire network if three or more cameras are considered. Matches between targets given independently by every pair of cameras might not conform to one another and, in turn, may lead to inconsistent mappings. Thus, in person re-identification across a camera network, multiple paths of correspondences may exist between targets from any two cameras, but ultimately all these paths must point to the same correspondence maps for each target in each camera. An example scenario is shown in Fig. 1(a). Even though camera pairs 1-2 and 2-3 have correct re-identification of the target, the false match between the targets in camera pair 1-3 makes the overall re-identification across the triplet inconsistent. It can be noted that the error in re-identification manifests itself through inconsistency across the network, and hence by enforcing consistency the pairwise accuracies can be improved as well.

Multi-view feature tracking is another application area where consistent data association is important.

• *A. Chakraborty is with the Department of Diagnostic Radiology, National University of Singapore, Singapore 119077. A. Das and A. Roy-Chowdhury are with the Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521. E-mail: amitrc@ee.ucr.edu*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2491922, IEEE Transactions on Pattern Analysis and Machine Intelligence

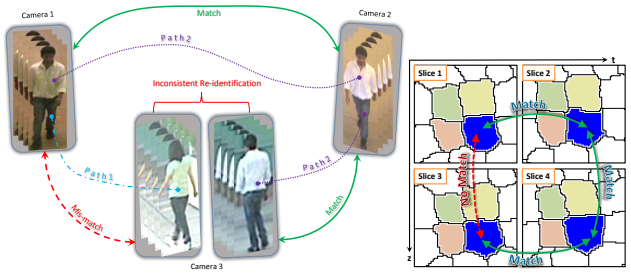


Fig. 1. Example of network inconsistency in data association. (a) Person re-identification: Among the 3 possible re-identification results, 2 are correct. Match of the target from camera 1 to camera 3 can be found in 2 ways. The first one is the direct pairwise re-identification result between cameras 1 and 3 (shown as 'Path 1'), and the second one is the indirect re-identification result in camera 3 given via the matched person in camera 2 (shown as 'Path 2'). The two outcomes do not match and thus the overall associations of the target across 3 cameras is not consistent. (b) Network inconsistency in spatio-temporal cell tracking: In this schematic, association results between 2D projections of the same 3D cell on four spatio-temporal image planes are analyzed. The pairwise associations need to be consistent across the loop over the four image slices. This consistency can be used to obtain correspondences when there are no direct pairwise matches or to correct wrong ones. For example, the correspondence between the same cell in image slice 1 and slice 3 (broken arrow) is established via an indirect path (solid arrows) through slices 2 and 4.

One such feature tracking problem is the spatio-temporal cell tracking. Using confocal microscopes, multicellular biological tissues are often imaged at multiple time points to observe the growth of hundreds of individual cells in the tissue. At each time point, cells within the tissue are imaged at various confocal planes, thus resulting in a (3D+t) stack of images. Each cell, therefore, may have projections on different spatio-temporal planes. A cell tracker aims to find correspondences between cell image slices along both 'z' (depth of the tissue) and time and hence the problem is same as any multi-view feature tracking problem. Because of the multi-dimensional nature of this tracking problem, spatial and temporal correspondences obtained by choosing the most similar candidate for each cell independently do not guarantee consistent results. Note that, as in the case of re-identification, a 2D cell segment in any spatio-temporal image slice must not have more than one match in any other spatio-temporal image and if at least one spatio-temporal path exists in the network that associates two cell slices, they must be projections of the same cell onto two image planes. Example of network-level association inconsistencies in the (3D+t) cell tracking problem is shown in Fig. 1(b).

The examples above represent the class of relevant data association problems where network consistency has to be enforced and where all the data points are available before the associations between them are estimated. Besides these, for many other *dynamic data association* problems where more observations become available with time, the consistency needs to be applied in an online fashion. Dynamic feature tracking (across space and time), tracklet association, online spatio-temporal cell tracking etc. are some of such problems and the proposed data association method has to be equipped with both online and offline mode of operations.

Contribution of the present work: We propose a novel consistent data association scheme over a network with individual observations as nodes. We pose the problem as an optimization problem that minimizes the global cost for associating pairs of targets on the entire network constrained by a set of consistency criteria and dub this as *Network Consistent Data Association (NCDA)*.

The inputs to NCDA are pairwise similarity scores between targets. Unlike assigning a match for which the similarity score is maximum, our formulation picks the assignments for which the total similarity of all matches is the maximum, under the constraint that there is no inconsistency in the assignment among any two sets of targets given the assignments between all other sets of targets across the network. The problem is translated into a binary integer program that can be solved using standard methods [1].

The proposed NCDA method is further generalized to a more challenging scenario in data association where the number of targets may vary across different sets of instances in the network. For example, in person re-identification problems all persons may not appear in all the cameras or in case of cell tracking, 2D projections of new cells appear deeper into a tissue. The objective function and the constraints are modified to incorporate probable one-to-none mappings without jeopardizing the network consistency.

For dynamic data association problems, we propose an online version of the generalized NCDA method. We show that the online NCDA uses the same core ideas as the batch version, but the size of the optimization problem solved at each time point is substantially smaller than that of the batch version - in terms of both the number of variables optimized and the number of constraints. The online NCDA method is able to associate newer observations available over time with those from the past while strictly maintaining the network consistency.

We show the general applicability of the proposed method by testing it in two previously mentioned computer vision application domains, *viz.* person re-identification and spatio-temporal cell tracking. We describe how each of these challenges can be mapped to the exact same NCDA problem, which can then be solved to generate unambiguous and more accurate data-association results.

2 Related Work

After discussing about the related work in each of the application areas , we shall highlight the differences

between the current submission with our earlier paper [2] that introduced network consistency in person re-identification.

Person Re-Identification: The existing person re-identification approaches are camera pairwise and they can be roughly divided into 3 categories - (i) discriminative signature based methods [3], [4], [5], [6], (ii) metric learning based methods [7], [8], [9], and (iii) transformation learning based methods [10], [11]. Person specific discriminative signatures are computed using multiple local features (color, shape and texture) [4], [5], [6], [12], [13]. Metric learning based methods try to improve the re-identification performance by learning optimal non-Euclidean metric defined on pairs of true and wrong matches [14], [15], [16]. Works exploring transformation of features between cameras learn a brightness transfer function (BTF) between appearance features [11], a subspace of the computed BTFs [10], linear color variations model [17], or a Cumulative BTF [18] between cameras. As the above methods do not take consistency into account, applying them to a camera network does not give consistent re-identification. Since the proposed method is built upon the pairwise similarity scores, any of the above methods can be the building block to generate the camera pairwise similarity between the targets.

Spatio-temporal Cell Tracking: There has been some work on automated tracking and segmentation of cells in time-lapse images, for both plants and animals. Some of the well-known approaches for segmenting and tracking cells are active contours based methods [19], [20], [21], [22], [23], Softassign method [24], [25], tracking based on association between detections[26], [27], [28], multiple hypotheses based tracking[29], joint detection and tracking[30], [31]. In [32], [33], a spatio-temporal tracking algorithm for Arabidopsis SAM was proposed, where relative positional information of neighboring cells were used to generate unique features for each cell. In [34], the spatio-temporal cell tracking problem is posed as an inference problem on a conditional random field. However, most of these methods have focused on slice to slice/pairwise cell tracking. The method in [32] utilizes indirect *paths* between any two slices to improve the pairwise tracking accuracy. However, this method does not ensure spatio-temporally consistent association results. The proposed NCDA method yields globally optimal and consistent correspondences between 2D cell slices when built on any method that can generate similarity scores between cells, such as [34].

Other relevant work: Some recent work aims to find point correspondences in monocular image sequences [35] or links detections in a tracking scenario by solving a constrained flow optimization [36] or using sparse appearance preserving tracklets [37]. Another flow based method for multi target tracking was presented in [38], which allows for one-to-many/many-to-one matchings and therefore can keep track of targets even when they merge into groups. With known flow direction, a flow formulation of a data-association problem will yield consistent results. But in data-association problems with no temporal or spatial layout information (e.g. person re-identification), the flow directions are not natural and thus the performance may widely vary with different choices of temporal or spatial flow. Using the transitivity of correspondence, point correspondence problem was addressed in a distributed as well as computationally efficient manner [39]. However, 'Consistency' and 'transitivity' being complementary to each other, less computation comes at the cost of local conflicts and mismatch cycles in absence of any consistency constraints, requiring a heuristics based approach to correct the conflicts subsequently. The proposed NCDA approach, on the other hand, uses maximal information by enforcing consistency and produces a globally optimal solution without needing to correct the correspondences at later stages.

In a very recent paper [2], we have introduced the network consistency in solving the person re-identification problem. In [2], the method and the constraints used in the integer program are specific to that particular problem (re-identification). However, in this paper, we provide a generalized problem formulation for solving any network level data association problem first, and describe the way the generalized constraints can be simplified further for problems in specific application areas. Moreover, in this paper we introduce a novel online implementation of the NCDA for solving dynamic/online data association problems. We show how the design of the optimization problem can reduce the size of the problem per iteration than that of the batch generalized NCDA method. Besides the re-identification problem, we also show applications of this data association method in the (3D+t) cell tracking problem and how the generalized constraints can be translated into their problem specific forms.

3 The Network Consistent Data Association Problem

3.1 Notations and Terminologies

1. Node: A node is a datapoint/target that needs to be associated with other datapoints via NCDA. For person re-identification problems, a node represents a target in the FoV of a camera, whereas, in cell tracking problem a node is a 2D segmented cell (at any given spatio-temporal location).

2. Group: A 'group' is a collection of nodes. A node can never be associated with any other node from the same group it belongs to. For example, in a typical person re-identification problem, the set of all targets appearing in the FoV of the same camera is a group and for spatio-temporal tracking, the collection of 2D cell segmentations in one image slice can be assumed

a group. Thus, a node is a member of a group. Let the i^{th} node in the group g be denoted as $\mathcal{P}_i^g$.

3. Similarity score matrix: This is a matrix data structure containing feature similarity scores between nodes belonging to two different groups. Therefore, for each pair of groups in a network there is one such matrix. Let $\mathbf{C}^{(p,q)}$ denote the similarity matrix between groups p and q. Then $(i,j)^{th}$ element in $\mathbf{C}^{(p,q)}$ is the similarity score between the nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$.

4. Assignment matrix: We need to know whether the nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ are associated or not, $\forall i,j = \{1,\cdots n\}$ and $\forall p,q = \{1,\cdots m\}$. The associations between targets across groups can be represented using 'Assignment matrices', one for each pair of groups. Each element $x_{i,j}^{p,q}$ of the assignment matrix $\mathbf{X}^{(p,q)}$ between the group pair (p,q) is defined as follows,

$$x_{i,j}^{p,q} = \begin{cases} 1 & \text{if } \mathcal{P}_i^p \text{ and } \mathcal{P}_j^q \text{ are the same targets} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

If the number of nodes is the same in all groups, then $\mathbf{X}^{(p,q)}$ is a permutation matrix, *i.e.*, only one element per row and per column is 1, all the others are 0. Mathematically, $\forall\, x_{i,j}^{p,q} \in \{0,1\}$

$$\sum_{j=1}^{n} x_{i,j}^{p,q} = 1 \;\forall i = 1 \text{ to } n,\; \sum_{i=1}^{n} x_{i,j}^{p,q} = 1 \;\forall j = 1 \text{ to } n \quad (2)$$

5. Edge: An 'edge' between two nodes $\mathcal{P}_i^p$, and $\mathcal{P}_j^q$ from two different groups of nodes is constructed between the i^{th} node in group p and the j^{th} node in group q. It should be noted that there will be no edge between the nodes of the same group. There are two attributes connected to each edge. They are the similarity score $c_{i,j}^{p,q}$ and the association value $x_{i,j}^{p,q}$.

6. Path: A 'path' between two nodes $(\mathcal{P}_i^p, \mathcal{P}_j^q)$ is a set of edges that connect the nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ without traveling through a node twice. Moreover, each node on a path belongs to a different group. A path between $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ can be represented as the set of edges $e(\mathcal{P}_i^p, \mathcal{P}_j^q) = \{(\mathcal{P}_i^p, \mathcal{P}_a^r), (\mathcal{P}_a^r, \mathcal{P}_b^s), \cdots\ (\mathcal{P}_c^t, \mathcal{P}_j^q)\}$, where $\{\mathcal{P}_a^r, \mathcal{P}_b^s, \cdots \mathcal{P}_c^t\}$ are the set of intermediate nodes on the path between $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$. The set of association values on all the edges between the nodes is denoted as $\mathcal{L}$, *i.e.* $x_{i,j}^{p,q} \in \mathcal{L}$, $\forall i,j = [1,\cdots,n]$, $\forall p,q = [1,\cdots,m]$ and $p < q$. Finally, the set of all paths between any two nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ is represented as $\mathcal{E}(\mathcal{P}_i^p, \mathcal{P}_j^q)$ and the z^{th} path is $e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q) \in \mathcal{E}(\mathcal{P}_i^p, \mathcal{P}_j^q)$.

3.2 NCDA Objective Function and Constraints

Let us first discuss the problem of one-to-one data association where the number of datapoints per group is constant and each datapoint from one group would have exactly one match in another group. This type of data association problem is often relevant to the person re-identification datasets, where the same set of persons appears across the FoVs of all the cameras. Later, we shall present a more generalized version of NCDA where number of datapoints belonging to different groups may vary and therefore a datapoint may or may not have a match in another group.

For the pair of groups (p,q), the sum of the similarity scores of association is given by $\sum_{i,j=1}^{n} c_{i,j}^{p,q} x_{i,j}^{p,q}$. Summing over all possible pairs of groups, the global similarity score can be written as

$$\mathbf{C} = \sum_{\substack{p,q=1 \\ p<q}}^{m} \sum_{i,j=1}^{n} c_{i,j}^{p,q} x_{i,j}^{p,q} \quad (3)$$

The set of constraints are as follows.

1. Pairwise association constraint: For the one-to-one association scenario, a datapoint from the group p can have only one match from another group q. This is mathematically expressed by the set of equations (2). This is true for all possible pairs of data groups and can be expressed as,

$$\begin{aligned} &\sum_{j=1}^{n} x_{i,j}^{p,q} = 1 \;\forall i = 1 \text{ to } n \;\forall p,q = 1 \text{ to } m,\; p < q \\ &\sum_{i=1}^{n} x_{i,j}^{p,q} = 1 \;\forall j = 1 \text{ to } n \;\forall p,q = 1 \text{ to } m,\; p < q \end{aligned} \quad (4)$$

2. Loop constraint: This constraint comes from the consistency requirement. If two nodes are indirectly associated via nodes in other groups, then these two nodes must also be directly associated. Therefore, given two nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$, it can be noted that for consistency, a logical 'AND' relationship between the association value $x_{i,j}^{p,q}$ and the set of association values $\{x_{i,a}^{p,r}, x_{a,b}^{r,s}, \cdots x_{c,j}^{t,q}\}$ of any possible path between the nodes has to be maintained. The association value between the two nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ has to be 1 if the association values corresponding to all the edges of any possible path between these two nodes are 1. Keeping the binary nature of the association variables and the pairwise association constraint in mind the relationship can be compactly expressed as,

$$x_{i,j}^{p,q} \geq \left(\sum_{(\mathcal{P}_k^r, \mathcal{P}_l^s) \in e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)} x_{k,l}^{r,s} \right) - |e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)| + 1 \quad (5)$$

$\forall$ paths $e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q) \in \mathcal{E}(\mathcal{P}_i^p, \mathcal{P}_j^q)$, where $|e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)|$ denotes the cardinality of the path $|e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)|$, *i.e.* the number of edges in the path. The relationship holds true for all i and all j. For the case of a triplet of cameras the constraint in Eqn. (5) simplifies to,

$$x_{i,j}^{p,q} \geq x_{i,k}^{p,r} + x_{k,j}^{r,q} - 2 + 1 = x_{i,k}^{p,r} + x_{k,j}^{r,q} - 1 \quad (6)$$

An example from person re-identification involving 3 cameras and 2 persons is illustrated with the help of Fig. 2. Say, the raw similarity score between pairs of targets across cameras suggests associations between $(\mathcal{P}_1^1, \mathcal{P}_1^2)$, $(\mathcal{P}_1^2, \mathcal{P}_1^3)$ and $(\mathcal{P}_2^1, \mathcal{P}_1^3)$ independently. However, when these associations are combined together over the entire network, it leads to an infeasible scenario - $\mathcal{P}_1^1$ and $\mathcal{P}_1^2$ are the same person. This infeasibility is also correctly captured through the constraint in Eqn. (6), *i.e.*, $x_{1,1}^{1,3} = 0$ but $x_{1,1}^{1,2} + x_{1,1}^{2,3} - 1 = 1$, thus violating the loop constraint.

For a generic scenario involving a large number of groups of nodes where similarity scores between every pair of groups may not be available the loop

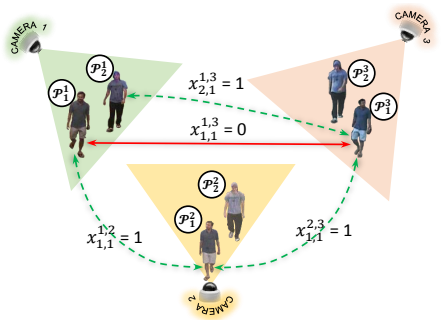


Fig. 2. An illustrative example showing the importance of the loop constraint in a data-association problem. It presents a simple person re-identification scenario in a camera network involving 2 persons (data points) in 3 cameras (groups).

constraint equations (*i.e.* Eqn. (5)) have to hold for every possible triplet, quartet (and so on) of groups. On the other hand, if the similarity scores between all nodes for every possible pair of groups are available, the loop constraints on quartets and higher order loops are not necessary. If loop constraint is satisfied for every triplet of groups then it automatically ensures consistency for every possible combination of groups taking 3 or more of them. So, in such a case, the loop constraint for the network can be written as,

$$x_{i,j}^{p,q} \geq x_{i,k}^{p,r} + x_{k,j}^{r,q} - 1 \quad (7)$$
$$\forall\, i,j,k = [1,\cdots n],\ \forall\, p,q,r = [1,\cdots m],\ \text{and}\ p < r < q$$

In spatio-temporal cell tracking problem, 3D structures of tightly packed multilayer tissues are imaged at various depths and at multiple observational time points. This yields a (3D+t) data structure of images, each of which contains 2D spatio-temporal projections of numerous 3D cells. More details on the structure of the data can be found at Sec. 6.2. Now, the target is to estimate correspondences between these 2D cellular projections along both space and time, i.e., between images at various depths at the same time point or at the same depth across consecutive time points. Thus, the entire spatio-temporal cell tracking network can be exhaustively partitioned into quartets of cell slices. One such quartet is shown (by arrows) in Fig. 1(b). Unlike triplets as in person re-identification case, the number of edges in an indirect path between any two nodes in an quartet is 3 and hence the value of $|e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)|$ in Eqn. (5) is 3. The loop constraints for the cell tracking problem can, therefore, be derived from Eqn. (5) as

$$\begin{aligned} x_{i,j}^{p,q} &\geq x_{i,k}^{p,r} + x_{k,l}^{r,s} + x_{l,j}^{s,q} - 3 + 1 \\ &= x_{i,k}^{p,r} + x_{k,l}^{r,s} + x_{l,j}^{s,q} - 2 \\ \forall\, i,j,k,l = [1,\cdots n],\ &\forall\, p,q,r,s = [1,\cdots m], \\ &\text{and}\ p < r < s < q \end{aligned} \quad (8)$$

Note that the above expressions are valid when there are equal number of nodes (cells/persons) in all groups and an exact one-to-one correspondence between each pair of groups is expected.

3.3 Overall Problem For One-to-One Associations

By combining the objective function in Eqn. (3) with the constraints in Eqn. (4) and Eqn. (7), we pose the overall optimization problem for the case of one-to-one mapping between groups as,

$$\begin{aligned} &\underset{\substack{x_{i,j}^{p,q} \\ i,j=[1,\cdots,n] \\ p,q=[1,\cdots,m]}}{\text{argmax}} \left(\sum_{\substack{p,q=1 \\ p<q}}^{m} \sum_{i,j=1}^{n} c_{i,j}^{p,q} x_{i,j}^{p,q} \right) \\ &\text{subject to} \sum_{j=1}^{n} x_{i,j}^{p,q} = 1\ \forall i = [1,\cdots,n] \\ &\forall p,q = [1,\cdots,m],\ p < q \\ &\sum_{i=1}^{n} x_{i,j}^{p,q} = 1\ \forall j = [1,\cdots,n]\ \forall p,q = [1,\cdots,m],\ p < q \\ &x_{i,j}^{p,q} \geq \left(\sum_{(\mathcal{P}_k^r, \mathcal{P}_l^s) \in e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)} x_{k,l}^{r,s} \right) - |e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)| + 1 \\ &\forall\, i,j = [1,\cdots n],\ \forall\, p,q = [1,\cdots m],\ \text{and}\ p < q \\ &\forall\ \text{paths}\ e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q) \in \mathcal{E}(\mathcal{P}_i^p, \mathcal{P}_j^q) \\ &x_{i,j}^{p,q} \in \{0,1\}\ \forall i,j = [1,\cdots,n],\ \forall p,q = [1,\cdots,m],\ p < q \end{aligned} \quad (9)$$

The above optimization problem for optimal and consistent data association is a binary integer program. The exact and simplified form of the integer program is discussed in the supplementary material.

4 NCDA for Variable Number of Data-points in Each Group

As explained in the previous sub-section, the NCDA can be achieved by solving the binary IP formulated in Eqn. (9). However, the assumption of one-to-one association between targets across groups may not be valid in many practical scenarios, especially when there are unequal numbers of datapoints in different groups. For re-identification in a camera network, there may be situations when every person does not go through the FoV of every camera. For the spatio-temporal cell tracking problems, there could be variable number of segmented cell slices on the images at different spatio-temporal locations. In such cases, a datapoint may not have association with any datapoint from another group and hence the values of assignment variables in every row or column of the assignment matrix can all be 0. However, a one-to-many association is still infeasible as before. For re-identification, a person from any camera p can have *at most* one match from another camera q. As a result, the pairwise association constraints now change from equalities to inequalities as follows,

$$\begin{aligned} &\sum_{j=1}^{n_q} x_{i,j}^{p,q} \leq 1\ \forall i = [1,\cdots,n_p]\ \forall p,q = [1,\cdots,m],\ p < q \\ &\sum_{i=1}^{n_p} x_{i,j}^{p,q} \leq 1\ \forall j = [1,\cdots,n_q]\ \forall p,q = [1,\cdots,m],\ p < q \end{aligned} \quad (10)$$

where, n_p snd n_p are the number of nodes (datapoints) in groups p and q respectively.

However, with this generalization, it is easy to see that the objective function (ref. Eqn. (9)) is no longer valid. Even though the provision of 'no match' is now available, the optimal solution will try to get as many associations as possible across the network. This is due to the fact that the current objective function assigns reward to both true positive (correctly associating a datapoint across groups) and false positive associations. Thus the optimal solution may contain many false positive associations. This situation can be avoided by incorporating a modification in the objective function as follows,

$$\sum_{\substack{p,q=1\\p<q}}^{m}\sum_{i,j=1}^{n_p,n_q}(c_{i,j}^{p,q}-k)x_{i,j}^{p,q} \tag{11}$$

where k is any value in the range of the similarity scores. This modification leverages upon the idea that, typically, similarity scores for most of the true positive matches in the data would be much larger than majority of the false positive matches. In the new cost function, instead of rewarding all positive associations we give reward to most of the true positives, but impose penalties on the false positives. As the rewards for all true positive (TP) matches are discounted by the same amount k and as there is penalty for false positive (FP) associations, the new cost function gives us optimal results for both 'match' and 'no-match' cases. The choice of the parameter k depends on the similarity scores generated by the chosen method, and thus can vary from one pairwise similarity score generating method to another. Ideally, the distributions of similarity scores of the TPs and FPs are non-overlapping and k can be any real number from the region separating these two distributions. However, for practical scenarios where TP and FP scores overlap, an optimal k can be learned from training data. A simple method to choose k could be running NCDA for different values of k over the training data and choosing the one giving the maximum accuracy on the cross validation data. So, for this more generalized case, the NCDA problem can be formulated as follows,

$$\begin{aligned}
&\underset{\substack{x_{i,j}^{p,q}\\ i=[1,\cdots,n_p]\\ j=[1,\cdots,n_q]\\ p,q=[1,\cdots,m]}}{\operatorname{argmax}}\left(\sum_{\substack{p,q=1\\p<q}}^{m}\sum_{i,j=1}^{n_p,n_q}(c_{i,j}^{p,q}-k)x_{i,j}^{p,q}\right)\\
&\text{subject to }\sum_{j=1}^{n_q}x_{i,j}^{p,q}\leq 1\ \forall i=[1,\cdots,n_p]\\
&\forall p,q=[1,\cdots,m],\ p<q\\
&\sum_{i=1}^{n_p}x_{i,j}^{p,q}\leq 1\ \forall j=[1,\cdots,n_q]\ \forall p,q=[1,\cdots,m],\ p<q\\
&x_{i,j}^{p,q}\geq\left(\sum_{(\mathcal{P}_k^r,\mathcal{P}_l^s)\in e^{(z)}(\mathcal{P}_i^p,\mathcal{P}_j^q)}x_{k,l}^{r,s}\right)-|e^{(z)}(\mathcal{P}_i^p,\mathcal{P}_j^q)|+1\\
&\forall\, i=[1,\cdots,n_p],\ j=[1,\cdots,n_q],\\
&\forall\, p,q=[1,\cdots m],\ \text{and } p<q\\
&\forall\ \text{paths } e^{(z)}(\mathcal{P}_i^p,\mathcal{P}_j^q)\in\mathcal{E}(\mathcal{P}_i^p,\mathcal{P}_j^q)\\
&x_{i,j}^{p,q}\in\{0,1\}\ \forall i=[1,\cdots,n_p],\ j=[1,\cdots,n_q],\\
&\forall p,q=[1,\cdots,m],\ p<q
\end{aligned} \tag{12}$$

5 Online Implementation of NCDA

5.1 Motivation

The generalized NCDA implementation, as presented in Sec. 4, is aimed at solving data association problems where all the observations are available and the target is to establish an optimal set of correspondences between them while maintaining network consistency. It is implemented as a batch method and the size of the problem increases quite dramatically with increase in the number of observations.

For example, a typical person re-id system is designed to run on datasets where all the observations across multiple camera FoVs are given. However, in a realistic setting, new observations are obtained with time and the data association method must be capable of assigning ids to observations as and when they become available.

In this section, we present an online formulation of the NCDA method. The online formulation is a direct theoretical extension of the batch problem (Eqn. (12)), as all the constraints (pairwise/loop) from the batch NCDA are preserved. Additionally, the online implementation of NCDA is capable of handling another very important and realistic scenario that the batch NCDA is not designed to. In a person re-id or a multi-view tracking problem, the same target may reappear in the same camera FoV after passing through the FoVs of some other camera(s) in the network. Unlike the batch method, the online NCDA can correctly re-id the target while maintaining global consistency.

5.2 Method

The definitions of *nodes*, *groups*, *edges* and *paths* follow from Sec. 3.1. Let us assume that there are m groups of observations upto time point t and the number of unique observations in group k is $n_k^{(t)}, k=1,2,\cdots m$. Thus, until time t, the total number of unique observations is $N^{(t)}=\sum_1^m n_k^{(t)}$. Let us also assume that the $N^{(t)}$ observations are already associated and the association is represented using a set of estimated labels $x_{i,j}^{p,q}={}^{(t)}x_{i,j}^{p,q},\ \forall i=[1,\cdots n_p^{(t)}],\forall j=[1,\cdots n_q^{(t)}],p,q=[1,\cdots m],\ p<q$.

In the next time window $[t,t+w]$, say there are l^w new observations across different groups and the objective is to associate these new observations to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2491922, IEEE Transactions on Pattern Analysis and Machine Intelligence

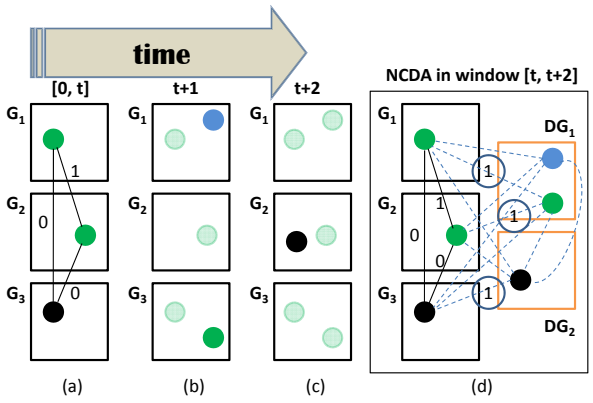


Fig. 3. A schematic showing time evolution of the online NCDA implementation. (a) The green target appeared in both group 1 (G_1) and group 2 (G_2) until time t and the black target only appeared in group 3 (G_3). Data association problem is solved until time t using NCDA and the labels are shown. (b) Two new observations (from the blue and the green target) are observed simultaneously at time point $t+1$ in groups 1 and 3 respectively. The past observations are blurred out. (c) At time $t+2$, no observation in G_1 or G_3, but an observation (from the black target) is obtained in G_2. All past observations are blurred including the ones from time $t+1$. (d) NCDA is run to associate the 3 new observations obtained in the time window $[t, t+2]$ to the ones obtained until t. Two dummy groups (DG_1 and DG_2) are created for the new observations - DG_1 holds the observations from the blue and the green target as they have time overlap and DG_2 holds the observation from the black target as it has no space/time overlap with the other two and hence can be legally associated with either of them. Nodes fron the dummy groups are connected via edges (dashed lines) to one another and to the nodes corresponding to all past observations. Online NCDA (as in Eqn. (18)) is run with past associations (solid lines) as constraints and the new labels are estimated (only label 1s are shown).

already observed targets and among each other. Now, some of these $l^{(w)}$ observations may have temporal or spatial overlap with some other new observation and therefore may not be associated with each other. As example, in person re-identification/multi-camera tracking problem, multiple observations may have temporal overlap between them. Likewise, in spatio-temporal cell tracking, cells in a 2D image slice have spatio-temporal overlap and hence cannot have associations among one another. The $l^{(w)}$ new observations can therefore be partitioned into s subsets where no two observations within a subset may have come from the same target. Thus, $n_{m+1}+n_{m+2}+\cdots n_{m+s} = l^{(w)}$, where n_p is the number of unique observations in the p^{th} subset.

Following the definitions in Sec. 3.1, each of these s subsets can be called a *virtual/dummy* 'group'. Thus in the aforementioned time window, the data-association problem can be solved using NCDA with a total of $N^{(t)} + l^{(w)}$ nodes (observations) and $m+s$ groups.

Time evolution of the online NCDA and the set of unlabeled edges are shown in Fig. 3. Edges are constructed between each node in a dummy group and the nodes in the other $m+s-1$ groups. Now the target is to assign labels (0/1) on each of these unlabeled edges maintaining optimality and network consistency. As the data-association between all the past observations ($N^{(t)}$ in m groups) is already solved, all the labeled edges can be treated as a set of additional constraints in NCDA. As these additional constraints use the already assigned labels for edges present till time t, the resulting IP only solves for the edge labels each of which involves (at least) one node from the new $l^{(w)}$ nodes. For re-identification problems, moving new observations into dummy groups also enables us to associate observations of the same target re-appearing in the same camera FoV.

Objective function: The objective function is similar to that of generalized NCDA (Eqn. (11)), though it is defined only on the set of unlabeled edges for the online implementation, *i.e.*,

$$\sum_{\substack{p,q=m+1\\p<q}}^{m+s}\sum_{i,j=1}^{n_p,n_q}(c_{i,j}^{p,q}-k)x_{i,j}^{p,q} + \sum_{\substack{p=m+1,\\q=1}}^{m+s,m}\sum_{i,j=1}^{n_p,n_q}(c_{i,j}^{p,q}-k)x_{i,j}^{p,q} \tag{13}$$

The first part of the objective function is defined over all unlabeled edges between nodes of the dummy groups ($m+1,\cdots m+s$), whereas, the second part constitutes of unlabeled edges between the nodes in the dummy groups and the past observed nodes (in groups $1,\cdots m$). Let, the set containing all unlabeled edges at any iteration is represented as E_u.

Pairwise association constraints: The set of pairwise association constraints between pairs of groups of observations is defined as in Eqn. (10), except the fact that at least one of the groups must be a dummy group. Mathematically,

$$\begin{aligned}\sum_{j=1}^{n_q} x_{i,j}^{p,q} \leq 1\ \forall i=[1,\cdots,n_p],\ \forall(p,q)\in\mathcal{E}^{(w)}\\ \sum_{i=1}^{n_p} x_{i,j}^{p,q} \leq 1\ \forall j=[1,\cdots,n_q],\ \forall(p,q)\in\mathcal{E}^{(w)}\end{aligned} \tag{14}$$

where the pairs of groups ($\mathcal{E}^{(w)}$) are given as

$$\begin{aligned}\mathcal{E}^{(w)} \ = \ &\{(p,q): p,q\in[1,\cdots m+s], p<q\}\\ &\backslash\{(p,q): p\leq m, q\leq m\}\end{aligned} \tag{15}$$

Loop constraints: The loop constraints remain the same as in Eqn. (5). However, we only need a much smaller subset in the online NCDA implementation as each of these inequality constraints must involve at least one unlabeled edge, *i.e.*, at least one edge from the set $\{(\mathcal{P}_i^p,\mathcal{P}_j^q)\cup\ e^{(z)}(\mathcal{P}_i^p,\mathcal{P}_j^q)\}$ must belong to the set of unlabeled edges E_u. Thus, mathematically, $\{(\mathcal{P}_i^p,\mathcal{P}_j^q)\cup\ e^{(z)}(\mathcal{P}_i^p,\mathcal{P}_j^q)\}\ \cap E_u \neq \emptyset$ and, the loop

constraints are, therefore,

$$x_{i,j}^{p,q} \geq \left(\sum_{(\mathcal{P}_k^r, \mathcal{P}_l^s) \in e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)} x_{k,l}^{r,s} \right) - |e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)| + 1$$
$$\text{and, } \{(\mathcal{P}_i^p, \mathcal{P}_j^q) \cup e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)\} \cap E_u \neq \emptyset$$
$$\forall\, i = [1, \cdots, n_p],\ j = [1, \cdots, n_q],$$
$$\forall\, p, q = [1, \cdots m+s],\ \text{and } p < q$$
$$\forall \text{ paths } e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q) \in \mathcal{E}(\mathcal{P}_i^p, \mathcal{P}_j^q) \tag{16}$$

Associations between past observations: All observations ($N^{(t)}$) upto time t are associated with one another and these set of already estimated labels are imposed in the online NCDA as linear constraints, *i.e.*,

$$x_{i,j}^{p,q} = {}^{(t)}x_{i,j}^{p,q},\ \forall i = [1, \cdots n_p], \forall j = [1, \cdots n_q], \quad p, q = [1, \cdots m],\ p < q \tag{17}$$

So, by combining Eqns. (13),(14), (15), (16) and (17), the online NCDA problem for the time window $[t, t+w]$ is given as follows.

$$\underset{\substack{x_{i,j}^{p,q} \\ i=[1,\cdots,n_p],\ j=[1,\cdots,n_q] \\ (p,q)\in\mathcal{E}^{(w)}}}{\text{argmax}} \left(\sum_{\substack{p,q=m+1 \\ p<q}}^{m+s} \sum_{i,j=1}^{n_p,n_q} (c_{i,j}^{p,q} - k) x_{i,j}^{p,q} + \sum_{\substack{p=m+1, \\ q=1}}^{m+s,m} \sum_{i,j=1}^{n_p,n_q} (c_{i,j}^{p,q} - k) x_{i,j}^{p,q} \right)$$

$$\text{subject to } \sum_{j=1}^{n_q} x_{i,j}^{p,q} \leq 1\ \forall i = [1, \cdots, n_p],\ \forall (p,q) \in \mathcal{E}^{(w)}$$
$$\sum_{i=1}^{n_p} x_{i,j}^{p,q} \leq 1\ \forall j = [1, \cdots, n_q],\ \forall (p,q) \in \mathcal{E}^{(w)}$$
$$x_{i,j}^{p,q} \geq \left(\sum_{(\mathcal{P}_k^r, \mathcal{P}_l^s) \in e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)} x_{k,l}^{r,s} \right) - |e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)| + 1$$
$$\text{and, } \{(\mathcal{P}_i^p, \mathcal{P}_j^q) \cup e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)\} \cap E_u \neq \emptyset$$
$$\forall\, i = [1, \cdots, n_p],\ j = [1, \cdots, n_q],$$
$$\forall\, p, q = [1, \cdots m+s],\ \text{and } p < q$$
$$\forall \text{ paths } e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q) \in \mathcal{E}(\mathcal{P}_i^p, \mathcal{P}_j^q)$$
$$x_{i,j}^{p,q} = {}^{(t)}x_{i,j}^{p,q},\ \forall i = [1, \cdots n_p], \forall j = [1, \cdots n_q],$$
$$p, q = [1, \cdots m],\ p < q, \quad \text{and,}$$
$$x_{i,j}^{p,q} \in \{0, 1\}\ \forall i = [1, \cdots n_p],\ j = [1, \cdots n_q],\ (p,q) \in \mathcal{E}^{(w)} \tag{18}$$

Once the association labels are obtained by solving Eqn. (18), the dummy groups are dissolved and the new observations, labeled according to the association results, are put back to the original groups they belong to. If, according to the newly estimated labels, multiple observations come from same target, they are clubbed together into one node using any suitable fusion strategy. The association results are further used to re-estimate edge labels after this observation re-assignment step. This set of labels, denoted as ${}^{(t+w)}x_{i,j}^{p,q}$, is used in the next iteration of the online method once the new set of observations are available. Please note that information on timestamps of the 'past' observations is not remembered once the associations are done and is not used anywhere in the online NCDA in the subsequent time steps. Also, as observable from Eqn. (18), the number of unknown labels being optimized as well as the total number of constraints used in online NCDA at each time window is substantially less than that in batch NCDA based implementations, thereby reducing both the time complexity and the memory requirements.

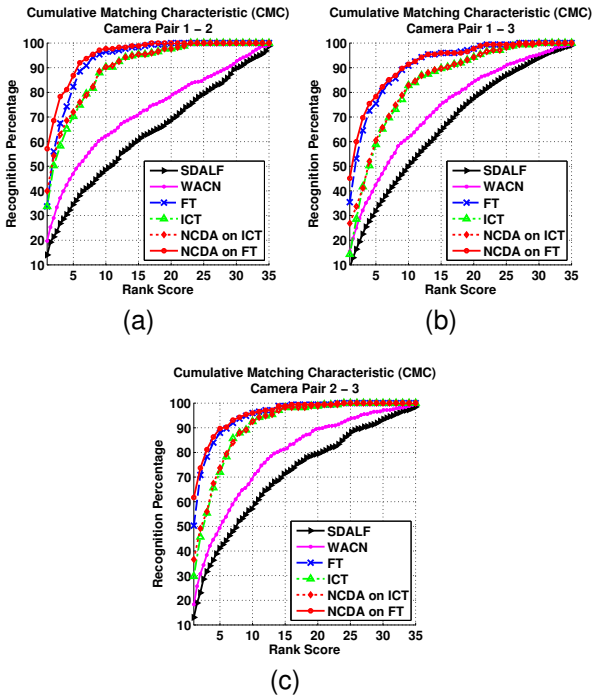


Fig. 4. CMC curves for the WARD dataset. Results and comparisons in (a), (b) and (c) are shown for the camera pairs 1-2, 1-3, and 2-3 respectively.

6 Experiments and Results

In this section, we evaluate the NCDA method on two different computer vision application areas, viz. 1. person re-identification and 2. spatio-temporal cell tracking. Analysis of the results in each application area is provided in the respective subsections.

6.1 Person Re-identification

Datasets and Performance Measures: To validate our approach, we performed experiments on two benchmark datasets - WARD [6] and RAiD [2]. Though state-of-the-art methods for person re-identification *e.g.*, [40], [4], [41] evaluate their performances using other datasets too (*e.g.*, ETHZ, CAVIAR4REID, CUHK) these do not fit our purposes since these are either two camera datasets or several sequences

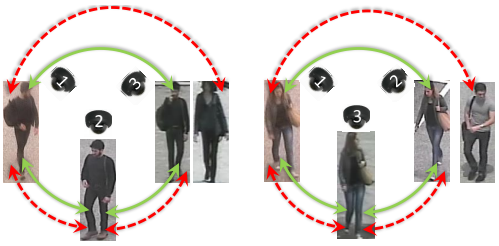


Fig. 5. 2 examples of correction of inconsistent re-identification from WARD dataset. The red dashed lines denote re-identifications performed on 3 camera pairs independently by FT. The green solid lines show the re-identification results on application of NCDA on FT. The NCDA algorithm exploits the consistency requirement and makes the resultant re-identification across 3 cameras correct.

of different two camera datasets. Results are shown in terms of recognition rate as Cumulative Matching Characteristic (CMC) curves and normalized Area Under Curve (nAUC) values (provided in the supplementary), as is the common practice in the literature. The CMC curve is a plot of the recognition percentage versus the ranking score and represents the expectation of finding the correct match inside top t matches. nAUC gives an overall score of how well a re-identification method performs irrespective of the dataset size. In the case where every person is not present in all cameras, we show the accuracy as total number of true positives (true matches) and true negatives (true non matches) divided by the total number of unique people present. All the results used for comparison were either taken from the corresponding works or by running codes which are publicly available or obtained from the authors on datasets for which reported results could not be obtained.

Pairwise Similarity Score Generation: The camera pairwise similarity score generation starts with extracting appearance features in the form of HSV color histogram from the images of the targets. The low level features are extracted as proposed in [4]. Given the extracted features, we generate the similarity scores by learning the way features get transformed as proposed in [42]. To keep the procedure simple, we used only HSV appearance features unlike [42] where several other appearance and texture features are studied. Due to the use of HSV features only, we differentiate this method of similarity score generation from [42] by calling our similarity score generation method as Feature Transform (FT). In addition to the feature transformation based method, similarity scores are also generated using the publicly available code of a recent work - ICT [3] where pairwise re-identification was posed as a classification problem in the feature space formed of concatenated features of persons viewed in two different cameras. More details about the similarity score generation is provided in the supplementary.

6.1.1 WARD Dataset

The WARD dataset [6] has 4786 images of 70 different people acquired in a real surveillance scenario in 3 non-overlapping cameras. This dataset has a huge illumination variation apart from resolution and pose changes. The cameras here are denoted as camera $1, 2$ and 3. Fig. 4(a), (b) and (c) compare the performance for camera pairs $1-2, 1-3$, and $2-3$ respectively. The 70 people in this dataset are equally divided into training and test sets of 35 persons each. The proposed approach is compared with the methods SDALF [4], ICT [3] and WACN [6]. The legends 'NCDA on FT' and 'NCDA on ICT' imply that the NCDA algorithm is applied on similarity scores generated by learning the feature transformation and by ICT respectively. For all 3 camera pairs the proposed method outperforms the rest with rank 1 recognition percentage as high as 61.71% for the camera pair 2-3. The next runner up is the method applying only FT which has the recognition percentage of 50.29% for rank 1.

To show how NCDA yields consistent re-identification where pairwise method fails, two example cases are provided in Fig. 5. At first, re-identification is performed on 3 camera pairs independently on the WARD data by FT method. In the first example, though the camera pairs 1-2 and 2-3 gave correct association (red dashed lines) for both the targets, the incorrect associations between camera pair 1-3 (red dashed line) make the re-identification across the 3 cameras inconsistent. Similarly, in the second example, incorrect associations between targets across camera pair 1-2 make the overall re-identification results inconsistent. However, in both the cases, NCDA exploits the consistency requirement and makes the resultant re-identification across 3 cameras correct (shown using green arrows).

6.1.2 RAiD Dataset

This dataset [2] was collected using 2 indoor (camera 1 and 2) and 2 outdoor (camera 3 and 4) cameras It has large illumination variation that is not present in most of the publicly available benchmark datasets. 41 subjects were asked to walk through these 4 cameras and 6920 images of 41 persons are present in it.

The proposed approach is compared with the same methods as for the WARD dataset. 21 persons were used for training while the rest 20 were used in training. Figs. 6(a) - (f) compare the performance for camera pairs 1-2, 1-3, 1-4, 2-3, 2-4 and 3-4 respectively. We see that the proposed method performs better than all the rest for both the cases when there is not much appearance variation (for camera pair 1-2 where both cameras are indoor and for camera pair 3-4 where both cameras are outdoor) and when there is significant lighting variation (for the rest 4 camera pairs). Expectedly, for camera pairs 1-2 and 3-4 the performance of the proposed method is the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2491922, IEEE Transactions on Pattern Analysis and Machine Intelligence

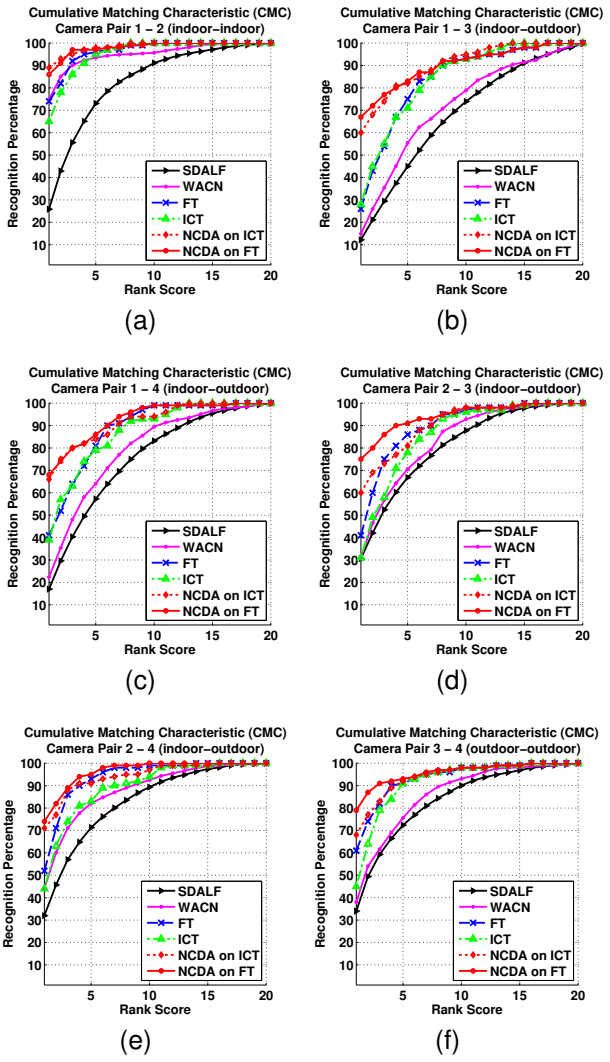


Fig. 6. CMC curves for RAiD dataset. In (a), (b), (c), (d), (e), (f) comparisons are shown for the camera pairs 1-2 (both indoor), 1-3 (indoor-outdoor), 1-4 (indoor-outdoor), 2-3 (indoor-outdoor), 2-4 (indoor-outdoor) and 3-4 (both outdoor) respectively.

best. For the indoor camera pair 1-2 the proposed method applied on similarity scores generated by feature transformation (NCDA on FT) and on the similarity scores by ICT (NCDA on ICT) achieve 86% and 89% rank 1 performance respectively. For the outdoor camera pair 3-4 the same two methods achieve 79% and 68% rank 1 performance respectively. For the rest of the cases where there is significant illumination variation the proposed method is superior to all.

In all the camera pairs, the top two performances come from the NCDA method applied on two different camera pairwise similarity scores generating methods. It can further be seen that for camera pairs with large illumination variation (*i.e.* 1-3, 1-4, 2-3 and 2-4) the performance improvement is significantly large. For camera pair 1-3 the rank 1 performance shoots up to 67% and 60% on application of NCDA algorithm to FT and ICT compared to their original rank 1 performance of 26% and 28% respectively. Clearly, imposing consistency improves the overall performance with the best absolute accuracy achieved for camera pairs consisiting of only indoor or only outdoor cameras. On the other hand, the relative improvement is significantly large in case of large illumination variation between the two cameras.

6.1.3 Re-identification with Variable Number of Persons

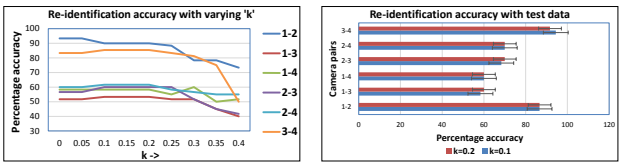
Next we evaluate the performance of the proposed method for the generalized setting when all the people may not be present in all cameras. For this purpose, from the RAiD dataset we chose two cameras (namely camera 3 and 4) and removed 8 (40% out of the test set containing 20 people) randomly chosen people keeping all the persons intact in camera 1 and 2. For this experiment the accuracy of the proposed method is shown with similarity scores as obtained by learning the feature transformation between the camera pairs. The accuracy is calculated by taking both true positive and true negative matches into account and it is expressed as $\frac{(\#\text{ true positive}+\#\text{ true negative})}{\#\text{ of unique people in the testset}}$.

Since the existing methods do not report re-identification results on variable number of persons nor is the code available which we can modify easily to incorporate such a scenario, we can not provide a comparison of performance here. However we show the performance of the proposed method for different values of k. The value of k is learnt using 2 random partitions of the training data in the same scenario (*i.e.*, removing 40% of the people from camera 3 and 4). The average accuracy over these two random partitions for varying k for all the 6 cameras are shown in Fig. 7(a). As shown, the accuracy remains more or less constant till $k = 0.25$. After that, the accuracy for camera pairs having the same people falls rapidly, but for the rest of the cameras where the number of people are variable remains significantly constant. This is due to the fact that the reward for 'no match' increases with the value of k and for camera pair 1-2 and 3-4 there is no 'no match' case. So, for these two camera pairs, the optimization problem (in Eqn. (12)) reaches the global maxima at the cost of assigning 0 label to some of the true associations (for which the similarity scores are on the lower side). So any value of k in the range $(0 - 0.25)$ will be a reasonable choice. The accuracy of all the 6 pairs of cameras for $k = 0.1$ and 0.2 is shown in Fig. 7(b), where it can be seen that the performance is significantly high and does not vary much with different values of k.

6.1.4 Online Re-identification

In this section, we present results on application of the online NCDA in an online person re-identification problem. We use the RAiD dataset again for this purpose. We randomly choose 20 subjects for training while the remaining 21 form the test set. The results are averaged over 5 such random partitions.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2491922, IEEE Transactions on Pattern Analysis and Machine Intelligence



(a) (b)

Fig. 7. Performance of the NCDA algorithm after removing 40% of the people from both camera 3 and 4 in the RAiD dataset. In (a) re-identification accuracy on the training data is shown for camera pairs by varying k after removing 40% of the training persons. (b) shows the re-identification accuracy on the test data for the chosen values of $k = 0.1$ and 0.2 when 40% of the test people were not present.

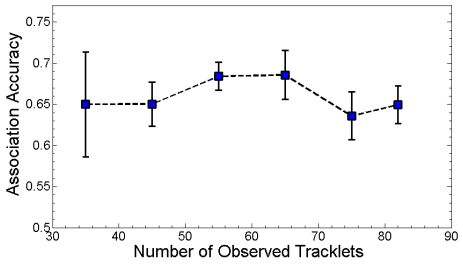


Fig. 8. Application of online NCDA for the online re-id problem. Mean association accuracies with standard deviation are plotted for increasing number of observed tracklets in the online setup. The online NCDA maintains a high accuracy ($\sim 65\%$) even as the number of observed tracklets increases.

In a classical person re-identification problem setup, all observations are assumed available at runtime and the association problem is solved in batch. This makes the time information for the observations redundant. However, the RAiD dataset contains timestamps associated with observations and we utilize the timestamps to generate the temporal data stream in the online experimental setup. We assume that the input to the NCDA method is a set of tracklets (a temporally consecutive series of observations/images from the same target obtained from within the same camera FoV), which were made available to the NCDA at their respective times of appearance. Moreover, as explained in Sec. 5.2, the tracklets which are temporally overlapping (even at different camera FoVs) may not be associated with one another. Hence, during runtime of the online NCDA, tracklets having temporal overlaps were clustered into the same dummy group. The pairwise similarity scores between two tracklets were computed using the method described in Sec. 6.1.

In the test set, there were a total of 84 tracklets for 21 targets. Based on time overlap, the 84 tracklets were clustered into 35 groups on an average. These groups of observations were time ordered and at each iteration of the online NCDA, one such group is introduced as input, Each group may contain more than one tracklet(s). At each iteration, the new tracklets are associated with the previously observed ones and labeled accordingly. The association accuracy at each iteration is estimated as $\frac{(\#\text{ true positive}+\#\text{ true negative})}{\#\text{ of unique people in the testset}}$. The variation of estimated accuracy with increasing number of observations is plotted in Fig. 8. It can be observed that the association accuracy is more than 65% and remains constant on average over time. Therefore, as new data is being observed, the data association results do not deteriorate on account of the larger and larger datasets - thus indicating robustness of the proposed online NCDA method.

6.2 Spatio-temporal Cell Tracking - A Multi-View Feature Tracking Problem

Dataset: For the experiments performed in the present study, the 3D structure of the tissues are imaged using single-photon confocal laser scanning microscope and we have specially dealt with the 'Shoot Apical Meristem' (SAM) of the plants that showcase all the challenges associated with any spatio-temporal cell tracking problem in a tightly packed multilayer, multicellular tissue. By changing the depth of the focal plane, CLSM can provide in-focus images from various depths of the specimen. To make the cells visible under laser, fluorescent dyes are used. The set of images, thus obtained at each time point, constitute a 3-D stack, also known as the 'Z-stack'. Each Z-stack is imaged at a time interval of 3 hours and it is comprised of a series of optical cross sections of SAMs that are separated by 1.5 μm. Thus, in this 4D image stack, every cell can have 2D projections on various 'z-planes' and the same cell can be imaged at multiple time points. The problem of cell tracking is to associate these spatio-temporal projections of the individual cells in the tissue along with detection of cell division events.

Pairwise Similarity Score Generation: Each 2D image slice in the 4D confocal image stack is segmented into individual cell slices using an adaptive Watershed segmentation method [43] and are temporally registered using a landmark-based registration scheme [44]. The similarity scores between 2D cell slices in spatio-temporally neighboring images are obtained using the method described in [34]. First, cell division events are detected between every pair of temporally neighboring images by utilizing the fact that combined shape of children cells is typically similar to that of the parent. Then, for every spatially/temporally neighboring pairs of images the a spatial graph is built on one of the images, where the 2D cells are nodes and any two neighboring cells share a link. Note that this graph does not contain the parent/children cells detected in the previous step. For each cell in first image, a set of candidate cells (one probable match in this set) is obtained from the second image. Now, a Conditional Random Field (CRF) is defined on this graph, and the node and edge potentials are computed based on shape similarity between cells and their candidates and by utilizing tight spatial topology of the tissue respectively. A loopy belief propagation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2491922, IEEE Transactions on Pattern Analysis and Machine Intelligence

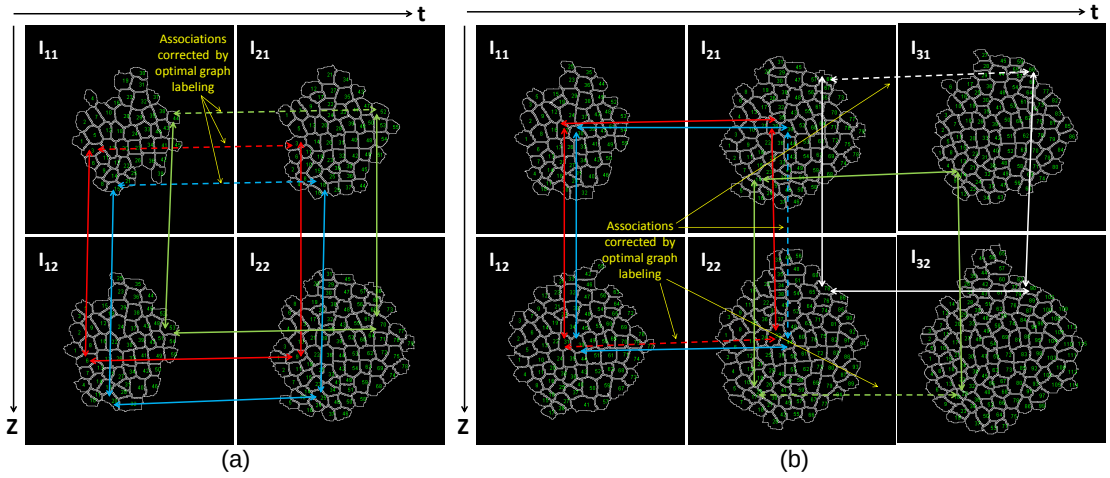


Fig. 9. Effect of NCDA towards improvement of spatio-temporal tracking results. (a) The figure shows a spatio-temporal 2X2 block of confocal images. Pairwise assignments between cells in spatial or temporal pairs of images are obtained by performing MAP inference on graphs formed on every image slice. Infeasible 4D assignments are observed when these pairwise associations are combined over the stack. The solid arrows represent correct associations between cells and the broken arrows depict no association which is incorrect and cause the infeasibility. Our proposed data association approach establishes consistency in association and corrects these errors. (b) Similar results are observed in a 2X3 confocal stack.

is run on this CRF and the marginal posteriors, thus obtained, are treated as similarity scores between the cells and their candidates. This method is repeated for all spatially/temporally consecutive pairs of image slices. More details on this can be found in [34] and also in the suppl. materials.

Establishing Network Consistency: Now, the objective is to obtain consistent associations between the 2D cell slices in the entire spatio-temporal image stack using the similarity scores generated via previous method. Now, each 2D image slice (containing a cluster of tightly packed cell slices) is treated as a 'group' and individual 2D cells on these slices are the nodes, as before. Also, for any given image slice, similarity scores are computed only to its immediate spatio-temporal neighboring slices (i.e. slice above, slice below, slice at same 'z' at previous time point and the same at next time point). This architecture yields a network of image slices (groups) that can be exhaustively covered using quartets of groups. Fig. 9(a) shows one such quartet in a large network. Please note that, unlike the person re-identification problem, the loop constraints for the cell tracking problem cannot be expressed as triplets, as similarity scores are not generated between temporally neighboring image slices that lie on different 'z-planes'. Using the marginal posteriors as similarity scores between a cell slice to its spatial/temporal candidates, we run NCDA for generating complete optimal 4D spatio-temporal correspondences between 2D cell slices.

Analysis of Results: The effect of NCDA towards improvement of spatio-temporal tracking results is shown in Fig. 9. In Fig. 9(a), a sample 2X2 block of images of Arabidopsis SAM are shown, which contains two spatially neighboring image slices at each of two consecutive time points of observation. Pairs of image slices are chosen and CRFs are formed for each of the pairs ($I_{11} - I_{12}, I_{12} - I_{22}, I_{21} - I_{22}$ and $I_{11} - I_{21}$). Now, marginal posteriors are estimated using LBP and MAP inferences are drawn to generate pairwise correspondences. When these pairwise associations are combined together, spatio-temporally infeasible associations are observed for a number of cells. For example, correct associations are found between cell 15 in I_{11} and cell 20 in I_{12}, cell 20 in I_{12} and cell 25 in I_{22}, cell 25 in I_{22} and cell 18 in I_{21}. Therefore, for spatio-temporal feasibility, cell 15 in I_{11} and cell 18 in I_{21} must also be associated. However, according to the aforementioned MAP inference, no associations for cell cell 15 from I_{11} is found in I_{21}. Similar infeasibilities are observed for cells 3 and 44 in I_{11}. The network consistent data association technique, when applied on the previously computed marginal posteriors for pairs of images, corrects these infeasibilities and establishes the associations. Fig. 9(b) shows similar results on a 2X3 confocal image stack, where the false negative associations (broken arrows) are corrected using the generalized NCDA for both spatial/temporal tracking.

Although the number of network inconsistencies may seem a small (3 in the 2X2 stack and 4 in 2X3 stack) percentage of the total number of cells, it is of utmost importance that each such error is rectified. A few inconsistencies per slice may add up to a large number of errors in a typical confocal stack consisting of thousands of 2D cell slices. Moreover, a tracking error not only affects the corresponding cell lineage, but may also affect the tracking accuracies for a number of its neighbors in the tightly packed multilayer tissue.

6.2.1 Online Spatio-temporal Cell Tracking

Typically in CLSM based live cell imaging, the observations are obtained every 3-6 hours. As explained in Sec. 6.2, each of these observations is a 3D stack of

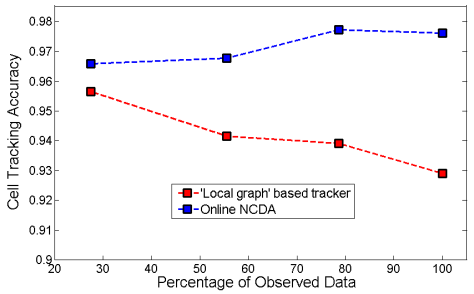


Fig. 10. Application of online NCDA in Online spatio-temporal cell tracking problem and comparison with *Local graph* based cell tracker [32]. Cell tracking accuracies with increasing amount of observations (expressed as percentage of total number of observations) are plotted for both [32] and the proposed online NCDA. The online NCDA outperforms [32] and the margin increases with more observations.

confocal images containing 2D projections of cells. Instead of waiting for the entire 4D stack to be collected, especially when it often takes 3-4 days, the online NCDA can generate correspondence results whenever a new 3D observation is obtained. In the cell tracking problem, the graph is built only between temporally/spatially neighboring pairs of images. Hence, when a new 3D stack of images is available at the observational time point t, the online NCDA is used to find correspondences between spatially neighboring 2D cell slices in the 3D stack at t and between these cells and that in the stack at $t-1$. Note that the spatial correspondences between cells at time $t-1$ are already established in previous iteration of the online NCDA.

In a 4D confocal image stack, total of 1170 2D unique cellular projections are observed across 4 time points and 3 z-slices. With each time point, more cells are observed and the cell tracking accuracies with number of observations, as obtained by using the online NCDA, are plotted in Fig. 10. As in the case for re-identification, the accuracies are estimated as a function of both true positives and true negatives. As observed, the accuracy increases and stabilizes as more observations become available, thereby establishing robustness of the online NCDA in case of biological cell tracking problem, too.

To make a quantitative comparison of the proposed NCDA over the state-of-the-art, we compared against the *local graph* based cell tracking method [32] on the same image stack. The output of this method are association results between pairs of 2D image slices and association accuracies are plotted in the same figure (Fig. 10). It can be observed that NCDA outperforms [32] substantially over time as the latter does not enforce additional consistency constraints over the pairwise association results.

7 Conclusion

When sets of data-points are observed at multiple spatio-temporal locations, pairwise data-association may often lead to infeasible scenarios over the global space-time horizon. To overcome this, we have proposed a generalized data-association method as a binary integer program on a graph. This proposed NCDA method not only maintains consistency across the global *network* of data-point sets, but also improves the pairwise data-association accuracy, even when the number of data-points varies across different sets of instances in the network. We have also proposed a novel mathematical framework for establishing network consistency in online data association problems. The online NCDA builds on similar ideas used for the generalized batch method, but solves a much smaller optimization problem at each iteration. Two applications of the batch and online version of proposed NCDA are shown in person re-identification and (3D+t) cell tracking. Analysis of the results indicates robustness of the method as well as significant improvements in accuracy over the state-of-the-arts.

References

[1] A. Schrijver, *Theory of linear and integer programming*. John Wiley and Sons, 1998.
[2] A. Das, A. Chakraborty, and A. Roy-Chowdhury, "Consistent re-identification in a camera network," in *European Conference on Computer vision*, 2014.
[3] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning implicit transfer for person re-identification," in *European Conference on Computer Vision, Workshops and Demonstrations*, 2012, pp. 381–390.
[4] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 130–144, Nov. 2013.
[5] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification : What features are important ?" in *European Conference on Computer Vision, Workshops and Demonstrations*. Florence, Italy: Springer Berlin Heidelberg, 2012, pp. 391–401.
[6] N. Martinel and C. Micheloni, "Re-identify people in wide area camera network," in *International Conference on Computer Vision and Pattern Recognition Workshops*. Providence, RI: IEEE, Jun. 2012, pp. 31–36.
[7] A. Bellet, A. Habrard, and M. Sebban, "A survey on metric learning for feature vectors and structured data," *ArXiv e-prints*, 2013.
[8] A. Alavi, Y. Yang, M. Harandi, and C. Sanderson, "Multi-shot person re-identification via relational stein divergence," in *Image Processing, IEEE International Conference on*, 2013.
[9] L. Yang and R. Jin, "Distance metric learning : A comprehensive survey," Michigan State University, Tech. Rep., 2006.
[10] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera spacetime and appearance relationships for tracking across non-overlapping views," *Computer Vision and Image Understanding*, vol. 109, no. 2, pp. 146–162, Feb. 2008.
[11] F. Porikli and M. Hill, "Inter-camera color calibration using cross-correlation model function," in *IEEE International Conference on Image Processing (ICIP)*, 2003, pp. 133–136.
[12] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person re-identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1622–1634, 2013.
[13] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *International Conference on Computer Vision and Pattern Recognition*, 2013.
[14] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Asian conference on Computer vision*, 2010, pp. 501–512.
[15] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Asian Conference on Computer Vision*, 2012, pp. 31–44.
[16] S. Pedagadi, J. Orwell, and S. Velastin, "Local fisher discriminant analysis for pedestrian re-identification," in *International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3318–3325.

[17] A. Gilbert and R. Bowden, "Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity," in *European Conference Computer Vision*, 2006.
[18] B. Prosser, S. Gong, and T. Xiang, "Multi-camera matching using bi-directional cumulative brightness transfer functions," in *British Machine Vision Conference*, Sep. 2008.
[19] O. Dzyubachyk, W. Van Cappellen, J. Essers, W. Niessen, and E. Meijering, "Advanced level-set-based cell tracking in time-lapse fluorescence microscopy," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 3, pp. 852–867, 2010.
[20] K. Li and T. Kanade, "Cell population tracking and lineage construction using multiple-model dynamics filters and spatiotemporal optimization," in *2nd International Workshop on Microscopic Image Analysis with Applications in Biology*, 2007.
[21] K. Li, M. Chen, T. Kanade, E. Miller, L. Weiss, and P. Campbell, "Cell population tracking and lineage construction with spatiotemporal context," *Medical Image Analysis*, vol. 12, no. 5, pp. 546 – 566, 2008.
[22] D. R. Padfield, J. Rittscher, N. Thomas, and B. Roysam, "Spatio-temporal cell cycle phase analysis using level sets and fast marching methods." *Medical Image Analysis*, vol. 13, no. 1, pp. 143–155, 2009.
[23] A. Dufour, V. Shinin, S. Tajbakhsh, N. Guillen-Aghion, J. C. Olivo-Marin, and C. Zimmer, "Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1396–1410, 2005.
[24] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching," in *in CVPR*, 2000, pp. 44–51.
[25] V. Gor, M. Elowitz, T. Bacarian, and E. Mjolsness, "Tracking cell signals in fluorescent images," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 0, p. 142, 2005.
[26] N. N. Kachouie, P. Fieguth, J. Ramunas, and E. Jervis, "Probabilistic model-based cell tracking," *International Journal of Biomedical Imaging*, 2006.
[27] T. Kirubarajan, Y. Bar-Shalom, and K. R. Pattipati, "Multiassignment for tracking a large number of overlapping objects," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 37, no. 1, pp. 2–21, 2001.
[28] R. Bise, Z. Yin, and T. Kanade, "Reliable cell tracking by global data association," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 1004–1010.
[29] L. Liang, H. Shen, P. Rompolas, V. Greco, P. D. Camilli, and J. S. Duncan, "A multiple hypothesis based method for particle tracking and its extension for cell segmentation," in *Information Processing in Medical Imaging*, 2013, pp. 98–109.
[30] D. Delibaltov, S. Karthikeyan, V. Jagadeesh, and B. S. Manjunath, "Robust biological image sequence analysis using graph based approaches," in *Asilomar Conference On Signals, Systems and Computers (ACSSC)*, 2012.
[31] S. Karthikeyan, D. Delibaltov, U. Gaur, M. Jiang, D. Williams, and B. Manjunath, "Unified probabilistic framework for simultaneous detection and tracking of multiple objects with application to bio-image sequences," in *International Conference on Image Processing*, 2012.
[32] M. Liu, R. K. Yadav, A. Roy-Chowdhury, and G. V. Reddy, "Automated tracking of stem cell lineages of arabidopsis shoot apex using local graph matching," *Plant journal, Oxford, UK*, vol. 62, pp. 135–147, 2010.
[33] M. Liu, A. Chakraborty, D. Singh, R. K. Yadav, G. Meenakshisundaram, G. V. Reddy, and A. Roy-Chowdhury, "Adaptive cell segmentation and tracking for volumetric confocal microscopy images of a developing plant meristem," *Molecular Plant*, vol. 4, no. 5, pp. 922–31, 2011.
[34] A. Chakraborty and A. Roy-Chowdhury, "A conditional random field model for tracking in densely packed cell structures," in *International Conference on Image Processing*, 2014.
[35] K. Shafique and M. Shah, "A noniterative greedy algorithm for multiframe point correspondence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 1, pp. 51–65, 2005.
[36] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 9, pp. 1806–1819, 2011.
[37] H. Ben Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Tracking multiple people under global appearance constraints," in *Computer Vision, IEEE International Conference on*, 2011, pp. 137–144.
[38] J. F. Henriques, R. Caseiro, and J. Batista, "Globally optimal solution to multi-object tracking with merged measurements," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2470–2477.
[39] S. Avidan, Y. Moses, and Y. Moses, "Centralized and distributed multi-view correspondence," *International Journal of Computer Vision*, vol. 71, no. 1, pp. 49–69, 2007.
[40] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," 2011.
[41] W. Li and X. Wang, "Locally aligned feature transforms across views," in *International Conference on Computer Vision and Pattern Recognition*, 2013.
[42] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Re-identification in the function space of feature warps," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, In press.
[43] K. Mkrtchyan, D. Singh, M. Liu, G. V. Reddy, A. K. R. Chowdhury, and M. Gopi, "Efficient cell segmentation and tracking of developing plant meristem," in *IEEE International Conference on Image Processing*, 2011, pp. 2165–2168.
[44] K. Mkrtchyan, A. Chakraborty, and A. K. Roy-Chowdhury, "Automated registration of live imaging stacks of arabidopsis," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013, pp. 672–675.

Anirban Chakraborty Anirban Chakraborty received his B.E. in Electrical Engineering from Jadavpur University, India in 2007 and M.S. and Ph.D. in the same subject from University of California, Riverside in 2010 and 2014 respectively. He is currently a research fellow at the Department of Diagnostic Radiology, National University of Singapore. His research interests include computer vision, pattern recognition, bio-medical image analysis, applications of probabilistic graphical models, stochastic processes and optimization in vision problems.

Abir Das Abir Das received his B.E. and M.S. degrees in Electrical Engineering from Jadavpur University, India and University of California, Riverside, USA in 2007 and 2013 respectively. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering at University of California, Riverside. His main research interests include computer vision, person re-identification, multi-camera multi-target tracking and video processing using machine learning based methods.

Amit K. Roy-Chowdhury Amit K. Roy-Chowdhury received the Bachelors degree in Electrical Engineering from Jadavpur University, Calcutta, India, the Masters degree in systems science and automation from the Indian Institute of Science, Bangalore, India, and the Ph.D. degree in Electrical Engineering from the University of Maryland, College Park. He is a Professor of Electrical Engineering at U. of California, Riverside. His research interests include image processing and analysis, computer vision, and video communications and statistical methods for signal analysis. His current research projects include intelligent camera networks, wide-area scene analysis, motion analysis in video, activity recognition and search, video-based biometrics (face and gait), biological video analysis, and distributed video compression. He is coauthor of The Acquisition and Analysis of Videos over Wide Areas. He is the editor of the book Distributed Video Sensor Networks. He has been on the organizing and program committees of multiple conferences and serves on the editorial boards of a number of journal.