# Network Consistent Data Association

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

**Abstract**—Existing data association techniques mostly focus on sequentially 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 lead 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 have tested NCDA in two application areas, viz., person re-identification and spatio-temporal cell tracking and observed consistent and highly accurate data association results in both 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 are sequential in nature, *i.e.* they 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. We explain the problem more precisely through two 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.

Spatio-temporal cell tracking is another application area where consistent data association is important. 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 four dimensional (3D + t) stack of images. Each cell, therefore, may have projections on different spatio-temporal planes. The spatio-temporal cell tracking aims to find correspondences between cell image slices along both 'z' (depth of the tissue) and time. 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 automatically. 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

• *A. Chakraborty, 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*

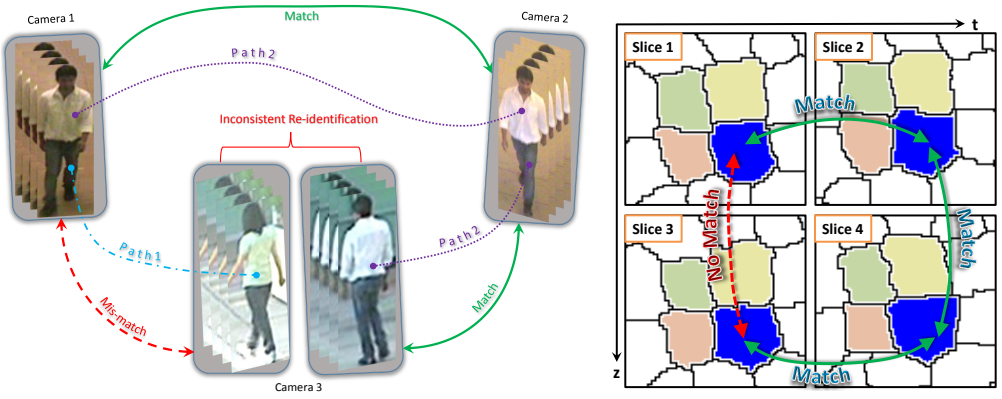



Fig. 1. Example of network inconsistency in data association. (a) A person re-identification case. Among the 3 possible re-identification results, 2 are correct. The match of the target from camera 1 to camera 3 can be found in two ways. The first one is the direct pairwise re-identification result between 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) A similar case of network inconsistency in spatio-temporal cell tracking problem. 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 to restore network consistency.

network-level inconsistent data association results in the spatio-temporal cell tracking problem is presented in Fig. 1(b).

The network inconsistency problem in data-association is not only observed in the person re-identification or the cell tracking tasks, but is also visible in other correspondence problems of similar nature, *e.g.* optical flow computation, feature tracking etc. In fact, a data-association approach that ensures network level consistency will be valuable for any problem where sets of data points observed at multiple spatio-temporal locations need to be associated with each other.

Contribution of the present work: Motivated by such scenarios, we propose a novel consistent data association scheme over a network. When the same set of data points is observed by different sensors (such as different camera FoVs) or observed by the same sensor repeatedly at different spatial and/or temporal locations (such as spatio-temporal imaging planes in confocal microscopy stacks), the possible associations between such projections of the same set of observations construct a network with the observed data points as nodes. To achieve a consistent and optimal data association, we pose the problem as an optimization problem that minimizes the global cost of associating pairs of targets on the entire network constrained by a set of consistency criteria (as mentioned before). Since consistency across the network is the motivation as well as the building block of the proposed method, we term this as the *Network Consistent Data Association (NCDA)* strategy.

We start by computing the pairwise similarity scores between sets of targets which are the input to the proposed method. Unlike assigning a match for which the similarity score is maximum among a set of probable candidates, our formulation picks the assignments for which the total similarity of all matches is the maximum, as well as 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 resulting optimization problem is translated into a binary integer program that can be solved using standard branch and cut, branch and bound or dynamic search algorithms based methods available in [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 as we image deeper into a tissue and hence the number of 2D cell slices widely vary across imaging planes. The objective function and and the constraints are modified to incorporate probable one-to-none mappings without jeopardizing the network consistency. We also provide a proof showing that the one-to-one NCDA can be directly derived as a special case of this more generalized formulation.

We show the general applicability of the proposed method by testing it in two previously mentioned computer vision application domain, *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

Network level consistency, as described before, has not been dealt with in popular data association problems such as person re-identification and multi dimensional tracking (spatio-temporal cell tracking) until very recently in [2], though a consistent association result in both these application areas is of utmost importance for attaining unambiguity and better accuracy. In this section, we shall briefly discuss about some of the most relevant and latest work in each of the two application areas. Later, we shall discuss about some related work in other optimization or flow based data-association problems such as multiple frame point correspondence, multi target tracking problem etc. and highlight their differences with the proposed method. Finally, we shall point out the differences between the current submission with our earlier paper [2] that introduced network consistency in person re-identification problems.

Person Re-Identification: In the last few years there has been increasing attention in the field of person re-identification across camera networks. The proposed approaches addressing the pairwise re-identification problem across non-overlapping cameras 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] or salient features learned in an unsupervised framework [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] or by maintaining redundancy in colorspace using a local Fisher discriminant analysis based metric [16]. Works exploring transformation of features between cameras tried to 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. Some of these works [17], [10] learned space-time probabilities of moving targets between cameras which may be unreliable if camera FoVs are significantly non-overlapping. 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 where the relative positional information of a cell with respect to its neighbors are utilized to generate robust associations between cells in two spatially/temporally consecutive image slices. 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. Also, the tracking in the (3D+t) stack is done in a sequential manner and a globally optimal solution is not achieved. The proposed NCDA method yields globally optimal and consistent correspondences between 2D cell slices when built on top of any method that can generate similarity scores between cells, such as [34].

Other relevant work: There have been a few correspondence methods proposed in recent years in other aspects of computer vision, *e.g.*, point correspondence in multiple frames and multi target tracking that are relevant to the proposed method. In one of the early works [35], finding point correspondences in monocular image sequences is formulated as finding a graph cover and solved using a greedy method. A suboptimal greedy solution strategy was used in [36] to track multiple targets by finding a maximum cover path of a graph of detections where multiple features like color, position, direction and size determined the edge weights. In [37], the authors linked detections in a tracking scenario across frames by solving a constrained flow optimization. The resulting convex formulation of finding k-shortest node-disjoint paths guaranteed the global optima. However, this method does not actively use appearance features into the data association process which might lead to ID switches among different pairs of cameras resulting in inconsistency. An extension of the work using sparse appearance preserving tracklets was proposed in [38]. 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.

In a very recent paper [2], we have introduced the network consistency in solving the person re-identification problem. In [2], the presentation of 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, as evidenced by the generalized expression for the 'loop constraint' (see Eqn. 5). We further show how the generalized constraints can be simplified for problems in specific application areas. Besides the person re-identification problem, we also show applications of this data association method in the spatio-temporal (3D+t) cell tracking problem and how the generalized constraints can be translated into their cell tracking problem specific form.

## 3 The Network Consistent Data Association Problem

In this section we describe the proposed approach in details. The Network Consistent Data Association (NCDA) method starts with the pairwise similarity scores between the targets. First we describe the notation and define the terminologies associated to this problem that would be used throughout the rest of the paper before delving deeper into the problem formulation.

### 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 score matrix between groups $p$ and $q$. Then $(i,j)^{th}$ element in $\mathbf{C}^{(p,q)}$ denotes 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,

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

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, $\forall\, x_{i,j}^{p,q} \in \{0,1\}$

$$\begin{aligned} \sum_{j=1}^{n} x_{i,j}^{p,q} &= 1 \;\forall i = 1 \text{ to } n \\ \sum_{i=1}^{n} x_{i,j}^{p,q} &= 1 \;\forall j = 1 \text{ to } n \end{aligned} \tag{2}$$

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 The NCDA Objective Function

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

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

### 3.3 Identification of 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 in the network. 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.

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,

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

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,

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

$\forall$ 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,

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

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

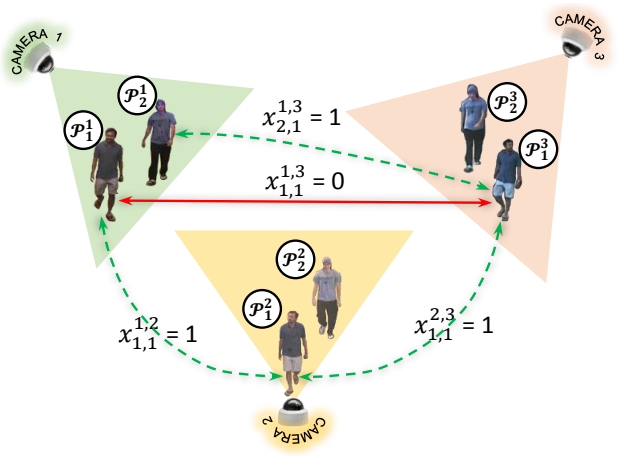



Fig. 2. An illustrative example showing the importance of the loop constraint in a data-association problem. It presents a simple person re-identification scenario in a camera network involving 2 persons (data points) in 3 cameras (groups).

$(\mathcal{P}_1^1, \mathcal{P}_1^2), (\mathcal{P}_1^2, \mathcal{P}_1^3)$ 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 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 constraint equations (*i.e.* eqn. (5)) have to hold for every possible triplet, quartet, quintet (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,

$$\begin{aligned}x_{i,j}^{p,q} \geq x_{i,k}^{p,r} + x_{k,j}^{r,q} - 1 \\ \forall\, i, j, k = [1, \cdots n],\ \forall\, p, q, r = [1, \cdots m],\ \text{and } p \leq r \leq q\end{aligned} \tag{7}$$

Unlike the person re-identification case, formation of triplets is not possible for the cell tracking problem because of the structure of the data (see Sec. 6.2). However, it can be shown that the entire spatio-temporal cell tracking network can be exhaustively partitioned into quartets of cell slices. The loop constraints for this problem are, therefore,

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

### 3.4 Overall Optimization 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,

$$\underset{\substack{x_{i,j}^{p,q} \\ i,j=[1,\cdots,n] \\ p,q=[1,\cdots,m]}}{\text{argmax}} \left( \sum_{\substack{p,q=1 \\ p<q}}^{m} \sum_{i,j=1}^{n} c_{i,j}^{p,q} x_{i,j}^{p,q} \right)$$

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

The above optimization problem for optimal and consistent re-identification 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 Datapoints In Each Group

As explained in the previous sub-section, the NCDA problem can be solved 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 example, 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,

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

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,

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

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 methods 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,

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

## 5 Equivalence Between One-to-One NCDA (Eqn. (9)) and The Generalized NCDA (Eqn. (12))

If the similarity score matrix and the assignment matrix are vectorized, one can rewrite the problems in eqn. (9) and eqn. (12) in standard binary integer program form. The one-to-one NCDA problem in eqn. (9) can be rewritten as

$$\begin{gathered}\underset{\underline{\mathbf{X}}}{\operatorname{argmax}}\, \underline{\mathbf{C}}^T \underline{\mathbf{X}} \\ \text{subject to } \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}},\ \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}} \\ \underline{\mathbf{X}} \text{ is composed of binary variables.} \end{gathered} \tag{13}$$

where $\mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}$ is the pairwise association constraint (same as eqn. (4)) and $\mathbf{B}\underline{\mathbf{X}} \leq \underline{\mathbf{1}}$ is the rewritten loop constraint (same as eqn. (7)). The construction of the vectors $\underline{\mathbf{C}}$, $\underline{\mathbf{X}}$ and the matrices $\mathbf{A}$, $\mathbf{B}$ is detailed in the supplementary materials. The value of $d$ is 1 in case of person re-identification problems where the loop constraints are expressed on triplets of groups. However, for the cell tracking problem, $d = 2$ as here the IP is written using quartet based loop constraints (see eqn. (8) and supplementary for more details).

The generalized form of NCDA (eqn. (12)) can, similarly be rewritten as,

$$\begin{gathered}\underset{\underline{\mathbf{X}}}{\operatorname{argmax}}\, (\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}} \\ \text{subject to } \mathbf{A}\underline{\mathbf{X}} \ \leq \underline{\mathbf{1}},\ \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}} \\ \underline{\mathbf{X}} \text{ is composed of binary variables.} \end{gathered} \tag{14}$$

Now let us prove that the problem expressed by eqn. (14) is equivalent to the problem expressed by eqn. (13) under the condition that the number of datapoints/targets is constant and there exists a one-to-one mapping between targets across groups. Let $\underline{\mathbf{X}}^*$ be the optimal solution to the problem expressed by eqn. (14). To prove the equivalence, we have to show that $\underline{\mathbf{X}}^*$ also maximizes the problem expressed by eqn. (13).

Since $\underline{\mathbf{X}}^*$ maximizes the objective function under the constraints as expressed by eqn. (14), we can write,

$$\begin{gathered}(\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}}^* \geq (\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}} \\ \text{for } \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} \leq \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \end{gathered} \tag{15}$$

where both $\underline{\mathbf{X}}^*$ and $\underline{\mathbf{X}}$ are composed of binary variables.

Since $\{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \subset \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} \leq \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\}$, the relation (15) holds true for the feasible set of eqn. (13), *i.e.*,

$$\begin{gathered}(\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}}^* \geq (\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}} \\ \text{for } \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \\ \implies \underline{\mathbf{C}}^T\underline{\mathbf{X}}^* - k\underline{\mathbf{1}}^T\underline{\mathbf{X}}^* \geq \underline{\mathbf{C}}^T\underline{\mathbf{X}} - k\underline{\mathbf{1}}^T\underline{\mathbf{X}} \\ \text{for } \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \\ \text{with both } \underline{\mathbf{X}}^* \text{ and } \underline{\mathbf{X}} \text{ composed of binary variables.} \end{gathered} \tag{16}$$

Now for all $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}}^*$ that satisfy $\mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}$ (*i.e.*, for the case when the same set of $n$ targets appear in all $m$ groups),

$$\begin{gathered}\underline{\mathbf{1}}^T\underline{\mathbf{X}}^* = \underline{\mathbf{1}}^T\underline{\mathbf{X}} \\ = \text{Num. of group pairs } \times \text{ Num. of targets} \end{gathered}$$

This is because, each row and column of the assignment matrix for pair of groups contains exactly one 1, resulting in the sum of all elements of the assignment matrices being $n$.

Using the above relation in eqn. (16) we get,

$$\begin{gathered}\underline{\mathbf{C}}^T\underline{\mathbf{X}}^* \geq \underline{\mathbf{C}}^T\underline{\mathbf{X}} \\ \text{for } \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \\ \text{with both } \underline{\mathbf{X}}^* \text{ and } \underline{\mathbf{X}} \text{ composed of binary variables.} \end{gathered} \tag{17}$$

Therefore, $\underline{\mathbf{X}}^*$ also maximizes the problem (13), thus proving the equivalence.

## 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 individual subsections.

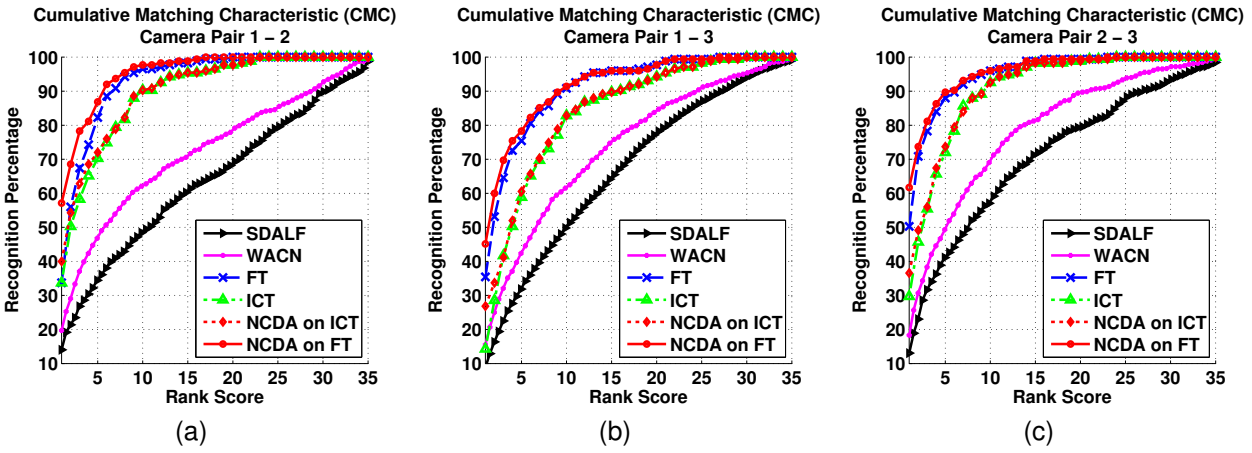



Fig. 3. CMC curves for the WARD dataset. Results and comparisons in (a), (b) and (c) are shown for the camera pairs 1-2, 1-3, and 2-3 respectively.

### 6.1 Person Re-identification

Datasets and Performance Measures: To validate our approach, we performed experiments on two benchmark datasets - WARD [6] and one new dataset RAiD introduced in [2]. Though state-of-the-art methods for person re-identification *e.g.*, [39], [4], [40] 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 of different two camera datasets. WARD is a 3 camera dataset with 70 people while RAiD has been collected across 4 cameras with 43 persons walking through them. 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. We did not re-implement other methods as it is very difficult to exactly emulate all the implementation details.

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. Before computing these features, the foreground is segmented out to extract the silhouette. Three salient regions (head, torso and legs) are extracted from the silhouette as proposed in [4]. The head region $S^H$ is discarded, since it often consists of a few and less informative pixels. We additionally divide both body and torso into two horizontal sub-regions based on the intuition that people can wear shorts or long pants, and short or long sleeves tops.

Given the extracted features, we generate the similarity scores by learning the way features get transformed between cameras in a similar approach as [11], [10]. Instead of using feature correlation matrix or the feature histogram values directly, we capture the feature transformation by warping the feature space in a nonlinear fashion inspired by the principle of Dynamic Time Warping (DTW). The feature bin number axis is warped to reduce the mismatch between feature values of two feature histograms from two cameras. Considering two non-overlapping cameras, a pair of images of the same target is a feasible pair, while a pair of images between two different targets is an infeasible pair. Given the feasible and infeasible transformation functions from the training examples, a Random Forest (RF) [41] classifier is trained on these two sets. The camera pair wise similarity score between targets are obtained from the probability given by the trained classifier of a test transformation function as belonging to either the set of feasible or infeasible transformation functions. 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.

Experimental Setup: In our implementation we used the following settings:

- To be consistent with the evaluations carried out by state-of-the-art methods, images were normal-

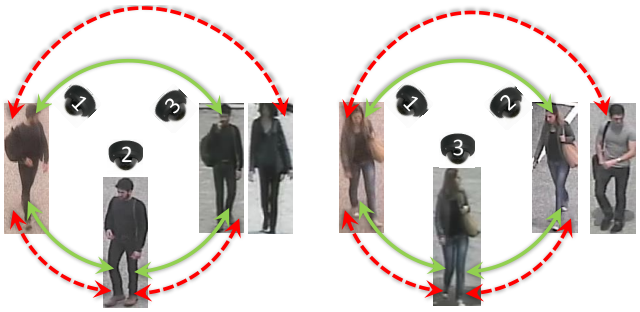


Fig. 4. Two examples of correction of inconsistent re-identification from WARD dataset. The red dashed lines denote re-identifications performed on 3 camera pairs independently by FT method. The green solid lines show the re-identification results on application of NCDA on FT. The NCDA algorithm exploits the consistency requirement and makes the resultant re-identification across 3 cameras correct.

ized to $128 \times 64$. The H, S and V color histograms extracted from the body parts were quantized using 10 bins each.

- Image pairs of the same or different person(s) in different cameras were randomly picked to compute the feasible and infeasible transformation functions respectively.
- All the experiments are conducted using a multi-shot strategy where 10 images per person is taken for both training and testing
- The RF parameters such as the number of trees, the number of features to consider when looking for the best split, etc. were selected using 4-fold cross validation.
- For each test we ran 5 independent trials and report the average results.

#### 6.1.1 WARD Dataset

The WARD dataset [6] has 4786 images of 70 different people acquired in a real surveillance scenario in three 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. 3(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 feature transformation which has the recognition percentage of 50.29% for rank 1.

To show how the proposed method yields consistent re-identification results where pairwise method fails, two example cases are provided in Fig. 4. 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 case, the NCDA exploits the consistency requirement and makes the resultant re-identification across 3 cameras correct, which are shown using green arrows.

#### 6.1.2 RAiD Dataset

This dataset [2] was collected to test the proposed method on a larger network. The dataset 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. Sample images from this dataset showing the variation of illumination between the cameras are shown in the supplementary material.

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. 5(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 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 the rest.

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

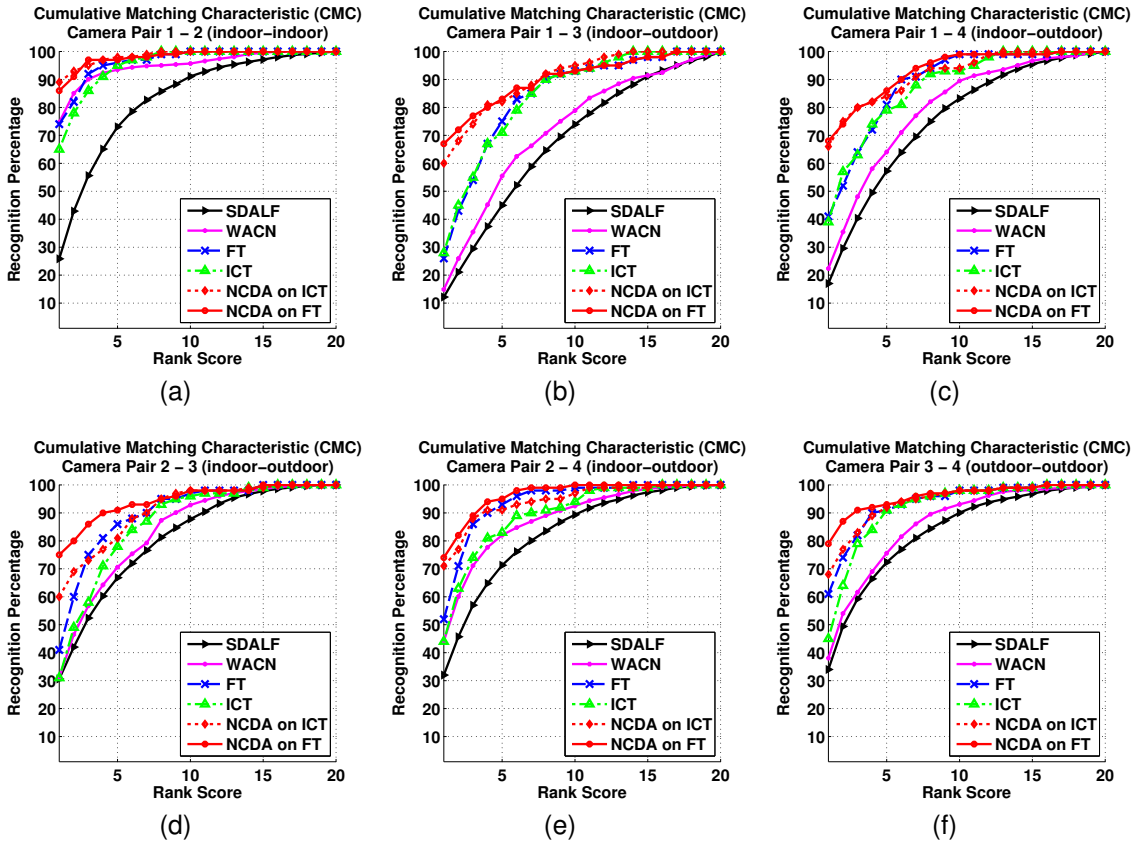


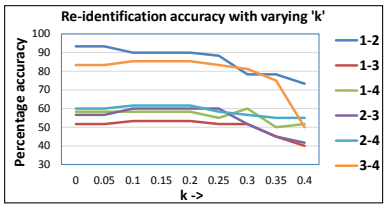

Fig. 5. CMC curves for RAiD dataset. In (a), (b), (c), (d), (e), (f) comparisons are shown for the camera pairs 1-2 (both indoor), 1-3 (indoor-outdoor), 1-4 (indoor-outdoor), 2-3 (indoor-outdoor), 2-4 (indoor-outdoor) and 3-4 (both outdoor) respectively.

large. For camera pair 1-3 the rank 1 performance shoots up to 67% and 60% on application of NCDA algorithm to TF 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.

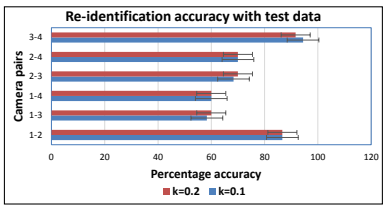
### *6.1.3 Re-identification with Variable Number of Persons*

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 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. 6(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 (namely camera pairs 1-2 and 3-4) 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




(a)




(b)

Fig. 6. Performance of the NCDA algorithm after removing 40% of the people from both camera 3 and 4. In (a) re-identification accuracy on the training data is shown for every camera pair by varying the parameter $k$ 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.

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. 6(b), where it can be seen that the performance is significantly high and does not vary much with different values of $k$.

### 6.2 Spatio-temporal Cell Tracking

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 tissue. The SAM of *Arabidopsis Thaliana* consists of approximately 500 cells and they are organized into multiple cell layers that are clonally distinct from one another. 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 $\mu$m. 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. Some sample confocal images from a typical 4D live imaging stack are presented in Fig. 7. The problem of cell tracking is to associate these spatio-temporal

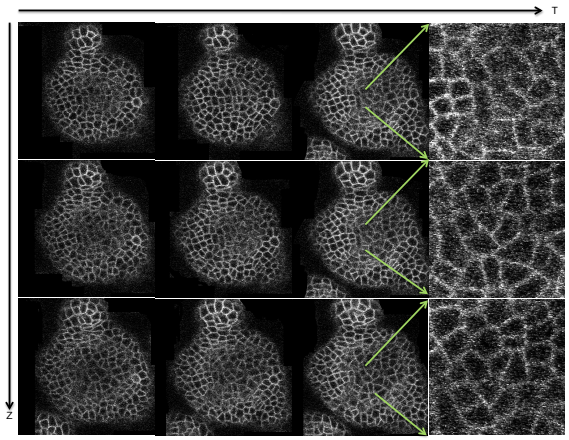



Fig. 7. A typical 4D (X-Y-Z-T) live-imaging data. A live Arabidopsis shoot meristem tissue is imaged using a confocal laser scanning microscope at multiple time points. The plasma membranes of the cells are stained with fluorescent proteins and that is why the cell walls are the only visible parts. Each of the first three columns of images presents Z stack of image slices, i.e the cross sections of the tissue imaged at various depths of it. When such images are collected over time to capture the growth of the tissue along with that of individual cells in it, it forms a 4D image stack. For a better visualization of the cellular structure, we have zoomed into a cluster of cells from the $4^{th}$ column of the figure.

projections of the individual cells in the tissue along with detection of cell division events.

Methodology: Each 2D image slice in the 4D confocal image stack is segmented into individual cell slices using an adaptive Watershed segmentation method [42] that learns the 'h-minima' threshold directly from the image data so that a uniformity in cell sizes is maintained as a result of the segmentation. Further the 3D image stacks are temporally registered using a landmark-based registration scheme [43].

The similarity scores between 2D cell slices in spatio-temporally neighboring images are obtained using the method described in [34], which is briefly given below. However, please note that any other method that estimates the similarities between the cell slices could also be used in conjunction with the proposed NCDA method.

First, cell division events are detected between every pair of temporally neighboring images. If a cell has divided into two children cells in the next temporal image slice, then ideally the shape of the parent cell should be very similar to the combined shape of the children taken together and each of the children cells would have approximately half the size of the parent cell. This prior knowledge is utilized in detecting cells undergoing division and both the parent and the children cells are removed from the set

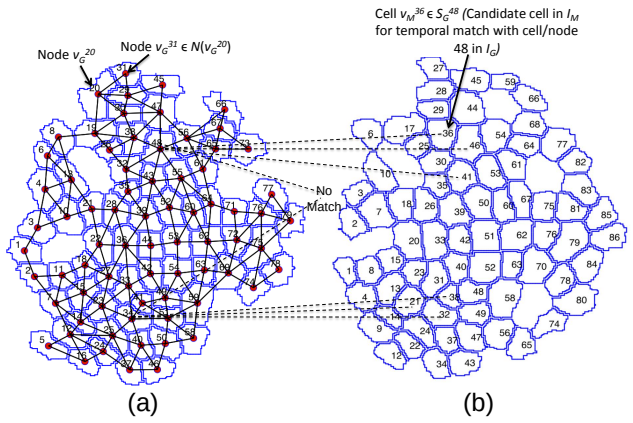



Fig. 8. Graph Structure. (a) For tracking cells between two spatially and temporally consecutive image slices, a graph is built on one of the images, where the nodes of the graph are the segmented cells and two neighboring cells share an edge between them. For temporal tracking, the cells undergoing division are set aside before constructing the graph. (b) From the next image slice, the candidate matches for each cell in A are estimated. Again, for temporal tracking, the children cells after division are also removed from the image and the candidate set of best 'K' states for each node in A is estimated through a search in B in a spatial window around the location of each of the nodes in A. An additional state is added to each of the candidate sets corresponding to the 'no-match' case.

if cells needed to be tracked (for temporal association only).

For every spatially/temporally neighboring pairs of images, a spatial graph is built on one of the images. Each 2D segmented cell slice is considered a node and any two cells that share a boundary have an undirected link between them. Please note that these graphs do not include the cells undergoing division and the resulting children cells. As every image pair is registered in the dataset under study, the set of candidate cells for matching from the second image (other than the one on which the graph is built) is further reduced via spatial windowing. These candidate cells constitute the set of probable states for each node. To account for the case that a cell may or may not have a match in the neighboring image slice, another 'no match' state is added each node's candidates. The graph formation and the candidate states for each node is presented through Fig. 8.

We further define a Conditional Random Field (CRF) on the graph constructed for each pair of images. A distance defined on the physical features extracted from a cell and that of each of its candidate matches is used to constitute the node potential. The spatial context is modeled on each of the edges based on the relative location of the cell and its neighbors by utilizing the tight spatial topology of the cell clusters. Details on the computation of node and edge potentials can be found in [34].

Loopy belief propagation based on 'sum-product' algorithm and message passing scheme is run on every CRF thus formed and the marginal posteriors for each node in pairs of images are computed. The posterior for a node is treated as similarity scores between the corresponding cell on one image and each of its candidate matches from the other image in the pair.

Establishing Network Consistency: Now, the objective is to obtain network 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 further run the 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

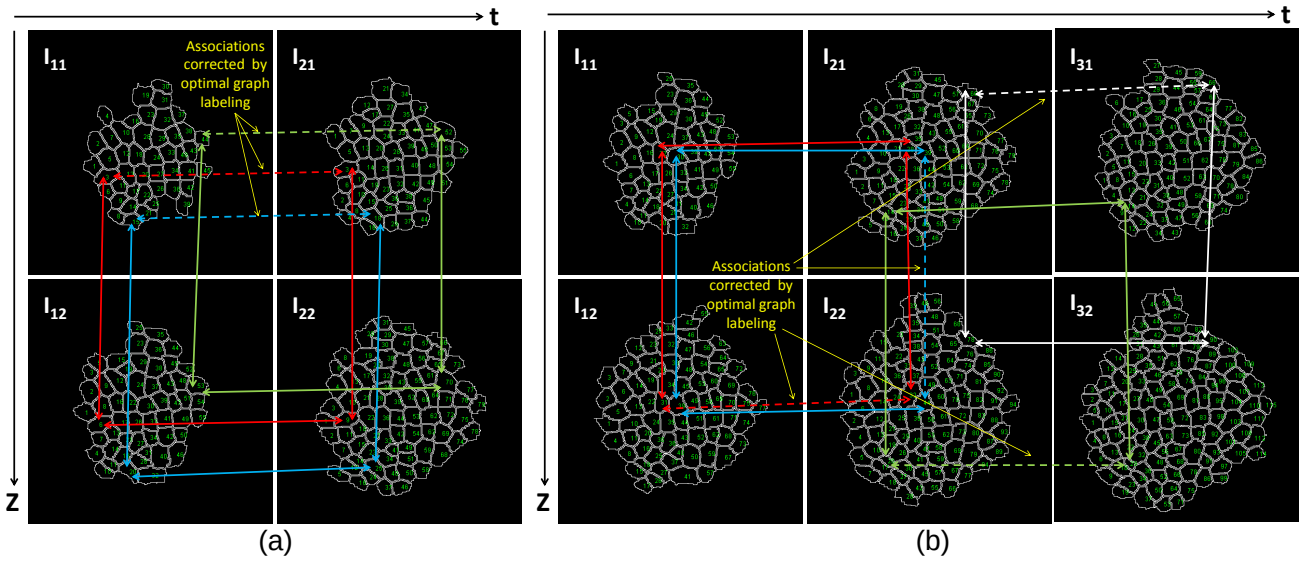



Fig. 9. Effect of the NCDA towards improvement of spatio-temporal tracking results. (a) The figure shows a spatio-temporal 2X2 block of confocal images. Pairwise assignments between cells in spatial or temporal pairs of images are obtained by performing MAP inference on graphs formed on every image slice. Infeasible 4D assignments are observed when these pairwise associations are combined over the stack. Examples of such infeasibilities are shown for three cell slices. The solid arrows represent correct associations between cell slices 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. False negatives in pairwise spatial or temporal tracking results are rectified using NCDA.

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. As before, correct associations obtained using MAP inference on the graphs are shown in solid arrows. The false negatives are shown using broken arrows, which are further corrected using the application of NCDA (by enforcing the loop constraints on the corresponding quartets). Note that the NCDA method can correct a false positive irrespective of its appearance in spatial or temporal tracking in the 4D stack.

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.

## 7 Conclusion

When the same set of data-points are observed by multiple agents and/or at multiple spatio-temporal locations, pairwise data-association may often lead to infeasible scenarios over the network of agents and the global space-time horizon. In this paper, we have proposed a generalized data-association method as a binary integer program on a graph. This proposed method, called NCDA, not only maintains consistency across the network of agents or amongst observations across spatio-temporal locations, but also improves the data-association accuracy. We have shown that the proposed NCDA method is also capable of handling the challenging data-association scenario where the number of data-points varies across different sets of instances in the network. Two applications of the proposed NCDA are shown - 1. Multi camera person re-identification and 2. Spatio-temporal (3D+t) cell tracking. Analysis of the experimental results indicate that the proposed method improves both network level consistency and pairwise association accuracy in these challenging problems. The future directions of our research will be not only to apply our approach to bigger and denser networks with large numbers of agents, and cope with wider space-time horizons but also to explore other data-association problems, (*e.g.*, consistent data association in multi-robot systems, feature point tracking, social network analysis etc.) where consistency is the key to robustness.

## 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 *Proceedings of the 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] M. Taj, E. Maggio, and A. Cavallaro, "Multi-feature graph-based object tracking," in *Multimodal Technologies for Perception of Humans*. Springer Berlin Heidelberg, 2007, vol. 4122, pp. 190–199.
[37] 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.
[38] 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.
[39] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," 2011.
[40] W. Li and X. Wang, "Locally aligned feature transforms across views," in *International Conference on Computer Vision and Pattern Recognition*, 2013.
[41] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[42] 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.
[43] 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.