Supplementary Materials
for
"Network Consistent Data Association"

Anirban Chakraborty, Abir Das and Amit K. Roy-Chowdhury

Overview

This supplementary material contains the following:

– Normalized Area Under Curves (nAUC) values of the CMC curves on the WARD and RAiD datasets.
– Formulation of the optimization problems as standard Binary Integer Programs (BIP).
– Equivalence of one-to-one and generalized NCDA problems (*i.e.*, NCDA with same and variable set of persons across cameras) is proved. One-to-one NCDA can be derived from the generalized NCDA if the condition that the same set of persons appear in all the cameras is imposed on the latter problem.
– Settings for Person Re-Identification Experiments and Similarity Score Generation.
– Methodology for Pairwise Similarity Score Generation for Spatio-temporal Cell Tracking.

1 Comparison of nAUC Values

Table 1: Comparison of NCDA with state-of-the-art methods on the WARD dataset in terms of the nAUC values.

Camera pair	SDALF	WACN	ICT	FT	NCDA on ICT	NCDA on FT
1-2	0.6487	0.7328	0.8780	0.9136	0.8835	**0.9317**
1-3	0.6825	0.7496	0.8240	0.8905	0.8299	**0.8981**
2-3	0.7206	0.7966	0.8881	0.9278	0.8910	**0.9330**

For all 3 camera pairs, 'NCDA on FT' gives the best nAUC values.

Table 2: Comparison of NCDA with state-of-the-art methods on the RAiD dataset in terms of the nAUC values.

Camera pair	SDALF	WACN	ICT	FT	NCDA on ICT	NCDA on FT
1-2	0.7987	0.9072	0.9138	0.9220	**0.9373**	0.9345
1-3	0.6576	0.6979	0.8145	0.8110	**0.8660**	0.8618
1-4	0.7274	0.7674	0.8413	0.8523	0.8790	**0.8885**
2-3	0.7802	0.8057	0.8328	0.8648	0.8700	**0.9008**
2-4	0.7956	0.8441	0.8615	0.9010	0.9008	**0.9210**
3-4	0.8014	0.8256	0.8813	0.8943	0.8990	**0.9138**

For all 6 camera pairs, either 'NCDA on ICT' or 'NCDA on FT' give the best nAUC values. Also for the 2 cases where 'NCDA on ICT' gives better nAUC values than 'NCDA on FT', the difference is in the 3^{rd} place of decimal.

2 Standard BIP Formulation of the Optimization Problems for Network Consistent Person Re-identification

For a matrix $\mathbf{A} = [\mathbf{a}_1\, \mathbf{a}_2\, \cdots\, \mathbf{a}_n] \in \mathbb{R}^{a \times b}$, let us define the vectorization operator $vec\,(.)$ as, $vec\,(\mathbf{A}) = [\mathbf{a}_1^T\, \mathbf{a}_2^T\, \cdots\, \mathbf{a}_n^T]^T = \underline{\mathbf{A}}$ (say) $\in \mathbb{R}^{ab}$. Now, the global similarity score in Eqn. (3) of the main paper can be written as,

$$\begin{aligned} \mathbf{C} &= \sum_{\substack{p,q=1 \\ p<q}}^{m} (vec\,(\mathbf{C}^{(p,q)}))^T vec\,(\mathbf{X}^{(p,q)}) \\ &= \sum_{\substack{p,q=1 \\ p<q}}^{m} (\underline{\mathbf{C}}^{(p,q)})^T \underline{\mathbf{X}}^{(p,q)} \end{aligned} \tag{1}$$

Clubbing all the vectorized $\mathbf{C}^{(p,q)}$s in a single vector we can write,

$$\underline{\mathbf{C}} = [(\underline{\mathbf{C}}^{(1,2)})^T (\underline{\mathbf{C}}^{(1,3)})^T \cdots (\underline{\mathbf{C}}^{(1,m)})^T (\underline{\mathbf{C}}^{(2,3)})^T (\underline{\mathbf{C}}^{(2,4)})^T \cdots (\underline{\mathbf{C}}^{(2,m)})^T \cdots\cdots (\underline{\mathbf{C}}^{(m-1,m)})^T]^T \tag{2}$$

Similarly, all the vectorized $\mathbf{X}^{(p,q)}$s can be clubbed as,

$$\underline{\mathbf{X}} = [(\underline{\mathbf{X}}^{(1,2)})^T (\underline{\mathbf{X}}^{(1,3)})^T \cdots (\underline{\mathbf{X}}^{(1,m)})^T (\underline{\mathbf{X}}^{(2,3)})^T (\underline{\mathbf{X}}^{(2,4)})^T \cdots (\underline{\mathbf{X}}^{(2,m)})^T \cdots\cdots (\underline{\mathbf{X}}^{(m-1,m)})^T]^T \tag{3}$$

Using Eqn. (2) and (3), Eqn. (1) can be written compactly as,

$$\mathbf{C} = \underline{\mathbf{C}}^T \underline{\mathbf{X}} \tag{4}$$

Let us express Eqn. (2) of the main paper in terms of $\underline{\mathbf{X}}^{(p,q)}$.

$$\begin{aligned} &\sum_{j=1}^{n} x_{i,j}^{p,q} = x_{i,1}^{p,q} + x_{i,2}^{p,q} \cdots + x_{i,n}^{p,q} = 1 \;\forall i = 1 \text{ to } n \\ \implies &[1, 1, \cdots, 1][x_{i,1}^{p,q}, x_{i,2}^{p,q}, \cdots, x_{i,n}^{p,q}]^T = 1 \;\forall i = 1 \text{ to } n \end{aligned} \tag{5}$$

and

$$\begin{aligned} &\sum_{i=1}^{n} x_{i,j}^{p,q} = x_{1,j}^{p,q} + x_{2,j}^{p,q} \cdots + x_{n,j}^{p,q} = 1 \;\forall j = 1 \text{ to } n \\ \implies &[1, 1, \cdots, 1][x_{1,j}^{p,q}, x_{2,j}^{p,q}, \cdots, x_{n,j}^{p,q}]^T = 1 \;\forall j = 1 \text{ to } n \end{aligned} \tag{6}$$

Writing Eqns. (5) and (6) respectively for all rows and columns of $\mathbf{X}^{(p,q)}$ we get,

$$\left[\begin{array}{ccccc}
\overbrace{1\,0\,0\cdots 0}^{n \text{ columns}} & \overbrace{1\,0\,0\cdots 0}^{n \text{ columns}} & \cdots & \overbrace{1\,0\,0\cdots 0}^{n \text{ columns}} \\
0\,1\,0\cdots 0 & 0\,1\,0\cdots 0 & \cdots & 0\,1\,0\cdots 0 \\
\vdots\,\vdots\,\vdots\,\ddots\,\vdots & \vdots\,\vdots\,\vdots\,\ddots\,\vdots & \ddots & \vdots\,\vdots\,\vdots\,\ddots\,\vdots \\
0\,0\,0\cdots 1 & 0\,0\,0\cdots 1 & \cdots & 0\,0\,0\cdots 1 \\
\hline
1\,1\,1\cdots 1 & 0\,0\,0\cdots 0 & \cdots & 0\,0\,0\cdots 0 \\
0\,0\,0\cdots 0 & 1\,1\,1\cdots 1 & \cdots & 0\,0\,0\cdots 0 \\
\vdots\,\vdots\,\vdots\,\ddots\,\vdots & \vdots\,\vdots\,\vdots\,\ddots\,\vdots & \ddots & \vdots\,\vdots\,\vdots\,\ddots\,\vdots \\
0\,0\,0\cdots 0 & 0\,0\,0\cdots 0 & \cdots & 1\,1\,1\cdots 1
\end{array}\right]
\begin{bmatrix} x^{p,q}_{1,1} \\ x^{p,q}_{2,1} \\ \vdots \\ x^{p,q}_{n,1} \\ x^{p,q}_{1,2} \\ x^{p,q}_{2,2} \\ \vdots \\ x^{p,q}_{n,2} \\ \vdots \\ x^{p,q}_{1,n} \\ x^{p,q}_{2,n} \\ \vdots \\ x^{p,q}_{n,n} \end{bmatrix}
= \begin{bmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix} \tag{7}$$

Writing the matrix as $\mathbf{A}$ and the vector as $\underline{\mathbf{X}}^{(p,q)}$ in Eqn. (7) the equation can be written as,

$$\mathbf{A}\underline{\mathbf{X}}^{(p,q)} = \underline{\mathbf{1}} \tag{8}$$

where, $\underline{\mathbf{1}}$ is a vector consisting of all 1's.

Relation (8) can be written for all the camera pairs, *i.e.*, $\forall p, q = [1, \cdots, m], p < q$. In a block matrix form these can be written as,

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A} \end{bmatrix}
\begin{bmatrix} \underline{\mathbf{X}}^{(1,2)} \\ \underline{\mathbf{X}}^{(1,3)} \\ \vdots \\ \underline{\mathbf{X}}^{(1,m)} \\ \underline{\mathbf{X}}^{(2,3)} \\ \underline{\mathbf{X}}^{(2,4)} \\ \vdots \\ \underline{\mathbf{X}}^{(2,m)} \\ \vdots \\ \underline{\mathbf{X}}^{(m-1,m)} \end{bmatrix} = \underline{\mathbf{1}} \tag{9}$$

Denoting the matrix as $\underline{\mathbf{A}}$ and the vector as $\underline{\mathbf{X}}$, Eqn. (9) can be written as,

$$\underline{\mathbf{A}}\underline{\mathbf{X}} = \underline{\mathbf{1}} \tag{10}$$

Let us express Eqn. (7) of the main paper in a similar way. The equation can be written as,

$$-x^{p,q}_{i,j} + x^{p,r}_{i,k} + x^{r,q}_{k,j} \leq 1 \tag{11}$$

Let us, first, write the set of these constraints for a particular triplet of cameras denoted by p, q, r with $p < r < q$. For this let us introduce some more notations. Let the j^{th} column of $\mathbf{X}^{(p,q)}$ be denoted as,

$$\mathbf{x}_j^{(p,q)} = [x_{1,j}^{p,q}, x_{2,j}^{p,q}, \cdots, x_{n,j}^{p,q}]^T$$

Let $\underline{\mathbf{1}}_i$ denote a vector of all 0's except a 1 at the i^{th} position. Let $\underline{\mathbf{0}}$ denote a vector of all 0's.

Keeping $i = 1; j = 1$ we, first, vary $k = 1, 2, \cdots, n$ in Eqn. (11) to get,

$$\begin{aligned}
-x_{1,1}^{p,q} + x_{1,1}^{p,r} + x_{1,1}^{r,q} &\leq 1 \quad \text{for } k = 1 \\
-x_{1,1}^{p,q} + x_{1,2}^{p,r} + x_{2,1}^{r,q} &\leq 1 \quad \text{for } k = 2 \\
-x_{1,1}^{p,q} + x_{1,3}^{p,r} + x_{3,1}^{r,q} &\leq 1 \quad \text{for } k = 3 \\
&\vdots \\
-x_{1,1}^{p,q} + x_{1,n}^{p,r} + x_{n,1}^{r,q} &\leq 1 \quad \text{for } k = n
\end{aligned}$$

The above set of of equations can also be written as,

$$\left[\begin{array}{cccc|ccccc|cccc}
-\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}}\cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}}\cdots & \underline{\mathbf{0}} \\
-\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}}\cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_2 & \underline{\mathbf{0}} & \underline{\mathbf{0}}\cdots & \underline{\mathbf{0}} \\
-\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}}\cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_3 & \underline{\mathbf{0}} & \underline{\mathbf{0}}\cdots & \underline{\mathbf{0}} \\
\vdots & \vdots & \vdots\ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots\ddots & \vdots \\
-\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}}\cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \cdots & \underline{\mathbf{1}}_1 & \underline{\mathbf{1}}_n & \underline{\mathbf{0}} & \underline{\mathbf{0}}\cdots & \underline{\mathbf{0}}
\end{array}\right]
\left[\begin{array}{c}
\mathbf{x}_1^{(p,q)} \\ \mathbf{x}_2^{(p,q)} \\ \mathbf{x}_3^{(p,q)} \\ \vdots \\ \mathbf{x}_n^{(p,q)} \\ \hline
\mathbf{x}_1^{(p,r)} \\ \mathbf{x}_2^{(p,r)} \\ \mathbf{x}_3^{(p,r)} \\ \vdots \\ \mathbf{x}_n^{(p,r)} \\ \hline
\mathbf{x}_1^{(r,q)} \\ \mathbf{x}_2^{(r,q)} \\ \mathbf{x}_3^{(r,q)} \\ \vdots \\ \mathbf{x}_n^{(r,q)}
\end{array}\right] \leq \underline{\mathbf{1}} \qquad (12)$$

For the same camera triplet (p, q, r), appending the rows corresponding to $i = 1, j = 2$ and $k = 1, 2, \cdots n$ we get,

$$\left[\begin{array}{cccc|cccc|cccc} -\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} \\ -\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_2 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} \\ -\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 \cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_3 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} \\ \vdots & \vdots & \vdots \ddots & \vdots & \vdots & \vdots & \vdots \ddots & \vdots & \vdots & \vdots & \vdots \ddots & \vdots \\ -\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{1}}_1 & \underline{\mathbf{1}}_n & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} \\ \hline \underline{\mathbf{0}} & -\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & -\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_2 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & -\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_1 \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{1}}_3 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} \\ \vdots & \vdots & \vdots \ddots & \vdots & \vdots & \vdots & \vdots \ddots & \vdots & \vdots & \vdots & \vdots \ddots & \vdots \\ \underline{\mathbf{0}} & -\underline{\mathbf{1}}_1 & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} \cdots & \underline{\mathbf{1}}_1 & \underline{\mathbf{0}} & \underline{\mathbf{1}}_n & \underline{\mathbf{0}} \cdots & \underline{\mathbf{0}} \end{array}\right] \left[\begin{array}{c} \mathbf{x}_1^{(p,q)} \\ \mathbf{x}_2^{(p,q)} \\ \mathbf{x}_3^{(p,q)} \\ \vdots \\ \mathbf{x}_n^{(p,q)} \\ \hline \mathbf{x}_1^{(p,r)} \\ \mathbf{x}_2^{(p,r)} \\ \mathbf{x}_3^{(p,r)} \\ \vdots \\ \mathbf{x}_n^{(p,r)} \\ \hline \mathbf{x}_1^{(r,q)} \\ \mathbf{x}_2^{(r,q)} \\ \mathbf{x}_3^{(r,q)} \\ \vdots \\ \mathbf{x}_n^{(r,q)} \end{array}\right] \leq \underline{\mathbf{1}} \tag{13}$$

Progressing in this way for all triplets of cameras and all persons, and denoting the resulting matrix (similar to the one in the left hand side of Eqn. (13)) as $\underline{\mathbf{B}}$, we get the loop constraints as,

$$\underline{\mathbf{B}}\underline{\mathbf{X}} \leq \underline{\mathbf{1}} \tag{14}$$

where, the number of rows of $\underline{\mathbf{B}}$ is the total number of loop constraint equations which is $\binom{m}{3}\binom{n}{2}(n-2) = \frac{m(m-1)(m-2)n(n-1)(n-2)}{12}$

Thus, using the objective function from Eqn. (4) and the constraints from Eqns. (10) and (14) we can write the binary integer program in standard form as,

$$\begin{aligned} &\underset{\underline{\mathbf{X}}}{\operatorname{argmax}}\, \underline{\mathbf{C}}^T \underline{\mathbf{X}} \\ &\text{subject to } \underline{\mathbf{A}}\underline{\mathbf{X}} = \underline{\mathbf{1}} \\ &\underline{\mathbf{B}}\underline{\mathbf{X}} \leq \underline{\mathbf{1}} \\ &\underline{\mathbf{X}} \text{ is composed of binary variables.} \end{aligned} \tag{15}$$

where, the dimensions of the matrices and vectors are as follows,

$\underline{\mathbf{C}}$ is $\frac{m(m-1)n^2}{2} \times 1$, $\underline{\mathbf{A}}$ is $m(m-1)n \times \frac{m(m-1)n^2}{2}$ and $\underline{\mathbf{B}}$ is $\frac{m(m-1)(m-2)n(n-1)(n-2)}{12} \times \frac{m(m-1)n^2}{2}$.

In a similar way, the standard binary integer program formulation of the optimization problem given in Eqn. (12) of the main paper for a variable number of targets across cameras can be expressed as,

$$\begin{aligned} \underset{\underline{\mathbf{X}}}{\operatorname{argmax}} \, & (\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}} \\ \text{subject to } & \underline{\mathbf{A}}\underline{\mathbf{X}} \leq \underline{\mathbf{1}} \\ & \underline{\mathbf{B}}\underline{\mathbf{X}} \leq \underline{\mathbf{1}} \\ \underline{\mathbf{X}} \text{ is composed of binary variables.} \end{aligned} \tag{16}$$

where $\underline{\mathbf{C}}, \underline{\mathbf{X}}, \underline{\mathbf{A}}$ and $\underline{\mathbf{B}}$ are as explained above with the dimensions of them changed to incorporate variable number of persons across cameras.

3 Equivalence Between One-to-One NCDA (Eqn.(9) in the main manuscript) and The Generalized NCDA (Eqn.(12) in the main manuscript)

As shown in the previous section, if the similarity score matrix and the assignment matrix are vectorized, one can rewrite the problems in Eqn.(9) and Eqn.(12) (in the main manuscript) in standard binary integer program form. The one-to-one NCDA problem in Eqn.(9) (in the main manuscript) can be rewritten as

$$\begin{aligned} &\underset{\underline{\mathbf{X}}}{\operatorname{argmax}}\, \underline{\mathbf{C}}^T \underline{\mathbf{X}} \\ &\text{subject to } \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}},\ \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}} \\ &\underline{\mathbf{X}} \text{ is composed of binary variables.} \end{aligned} \tag{17}$$

where $\mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}$ is the pairwise association constraint (same as Eqn.(4) in the main manuscript) and $\mathbf{B}\underline{\mathbf{X}} \leq \underline{\mathbf{1}}$ is the rewritten loop constraint (same as Eqn.(7) in the main manuscript). 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) in the main manuscript).

It was also shown that the generalized form of NCDA (Eqn.(12) in the main manuscript) can, similarly be rewritten as,

$$\begin{aligned} &\underset{\underline{\mathbf{X}}}{\operatorname{argmax}}\, (\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}} \\ &\text{subject to } \mathbf{A}\underline{\mathbf{X}} \ \leq \underline{\mathbf{1}},\ \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}} \\ &\underline{\mathbf{X}} \text{ is composed of binary variables.} \end{aligned} \tag{18}$$

Now let us prove that the problem expressed by Eqn. (18) is equivalent to the problem expressed by Eqn. (17) under the condition that the number of data-points/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. (18). To prove the equivalence, we have to show that $\underline{\mathbf{X}}^*$ also maximizes the problem expressed by Eqn. (17).

Since $\underline{\mathbf{X}}^*$ maximizes the objective function under the constraints as expressed by Eqn. (18), we can write,

$$\begin{aligned} (\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}}^* \geq (\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}} \\ \text{for } \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} \leq \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \end{aligned} \tag{19}$$

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 (19) holds true for the feasible set of Eqn. (17), *i.e.*,

$$\begin{gathered}(\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}}^* \geq (\underline{\mathbf{C}}^T - k\underline{\mathbf{1}}^T)\underline{\mathbf{X}} \\ \text{for } \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \\ \implies \underline{\mathbf{C}}^T\underline{\mathbf{X}}^* - k\underline{\mathbf{1}}^T\underline{\mathbf{X}}^* \geq \underline{\mathbf{C}}^T\underline{\mathbf{X}} - k\underline{\mathbf{1}}^T\underline{\mathbf{X}} \\ \text{for } \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \\ \text{with both } \underline{\mathbf{X}}^* \text{ and } \underline{\mathbf{X}} \text{ composed of binary variables.}\end{gathered} \tag{20}$$

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. (20) we get,

$$\begin{gathered}\underline{\mathbf{C}}^T\underline{\mathbf{X}}^* \geq \underline{\mathbf{C}}^T\underline{\mathbf{X}} \\ \text{for } \{\underline{\mathbf{X}} : \mathbf{A}\underline{\mathbf{X}} = \underline{\mathbf{1}}, \mathbf{B}\underline{\mathbf{X}} \leq d\underline{\mathbf{1}}\} \\ \text{with both } \underline{\mathbf{X}}^* \text{ and } \underline{\mathbf{X}} \text{ composed of binary variables.}\end{gathered} \tag{21}$$

Therefore, $\underline{\mathbf{X}}^*$ also maximizes the problem (17), thus proving the equivalence.

4 Settings for Person Re-Identification Experiments

In our implementation we used the following settings:

- To be consistent with the evaluations carried out by state-of-the-art methods, images were normalized to 128×64. 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.

5 Methodology for Pairwise Similarity Score Generation for Spatio-temporal Cell Tracking

Each 2D image slice in the 4D confocal image stack is segmented into individual cell slices using an adaptive Watershed segmentation method [1] 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 [2].

The similarity scores between 2D cell slices in spatio-temporally neighboring images are obtained using a probabilistic graphical model based method, 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.

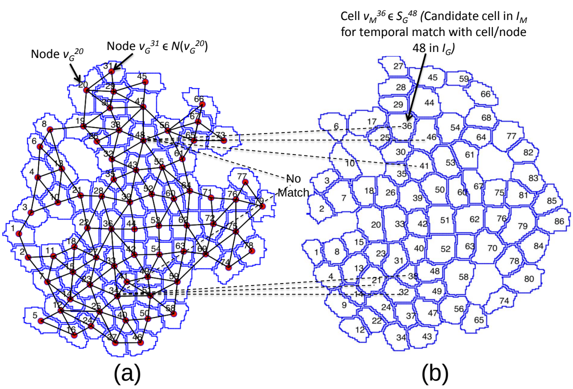

Fig. 1: Graph Structure. (a) For tracking cells between two spatially and temporally consecutive image slices, a graph is built on one of the images, where the nodes of the graph are the segmented cells and two neighboring cells share an edge between them. For temporal tracking, the cells undergoing 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.

5.1 Cell division detection:

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

5.2 Formation of CRF between pairs of image slices:

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. 1.

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 [3].

5.3 Similarity score generation:

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.

References

1. K. Mkrtchyan, D. Singh, M. Liu, G. V. Reddy, A. K. Roy-Chowdhury, and M. Gopi, "Efficient cell segmentation and tracking of developing plant meristem," in *International Conference on Image Processing*, 2011, pp. 2165–2168.
2. K. Mkrtchyan, A. Chakraborty, and A. K. Roy-Chowdhury, "Automated registration of live imaging stacks of arabidopsis," in *International Symposium on Biomedical Imaging*, 2013, pp. 672–675.
3. A. Chakraborty and A. K. Roy-Chowdhury, "A conditional random field model for tracking in densely packed cell structures," in *International Conference on Image Processing*, 2014, pp. 451–455.