On Pairwise Costs for Network Flow Multi-Object Tracking

Visesh Chari, Simon Lacoste-Julien, Ivan Laptev, and Josef Sivic

CVPR 2015
Detect Then Track

Consider a two-stage approach to multi-object tracking

• Detect objects in each frame of a sequence
• Determine interframe correspondences between them to label objects and discover their trajectories
Data Association

Detections in frame 1

\[ c_{ij} = \text{cost to associate detection } i \text{ with detection } j \]

(lower cost is better)
Data Association

\[ c_{ij} = \text{cost to associate detection } i \text{ with detection } j \]

\[ c_i = \text{cost to include detection } i \text{ in the solution (e.g. negative detector confidence)} \]
Multi-frame Data Association
Pump $K$ units of flow from $S$ to $T$ while minimizing cost. Note: in practice, $S$ and $T$ are connected to all detections, since a trajectory can start or end in any frame.
Network Flow Approach

Involves solving for binary variables $x$ on each edge. Variable is 1 if edge is part of the solution, 0 if not.
Formulated as ILP

$$\min_x \sum_i c_i x_i + \sum_{ij \in E} c_{ij} x_{ij}$$

$$0 \leq x_i \leq 1, \ 0 \leq x_{ij} \leq 1$$

$$\sum_{i: ij \in E} x_{ij} = x_j = \sum_{i: ji \in E} x_{ji} \quad \left\{ \begin{array}{l} \mathbf{x} \in \text{FLOW}_K \\
\sum_i x_{it} = K = \sum_i x_{si} \\
x_i, x_{ij} \text{ are integer.} \end{array} \right.$$
Formulated as ILP

\[
\min_x \quad \sum_i c_i x_i + \sum_{ij \in E} c_{ij} x_{ij} \\
0 \leq x_i \leq 1, \quad 0 \leq x_{ij} \leq 1 \\
\sum_{i : i,j \in E} x_{ij} = x_j = \sum_{i : j,i \in E} x_{ji} \\
\sum_i x_{it} = K = \sum_i x_{si} \\
x_i, \ x_{ij} \text{ are integer.}
\]
Formulated as ILP

\[
\min_x \sum_i c_i x_i + \sum_{i,j \in E} c_{ij} x_{ij}
\]

\[
0 \leq x_i \leq 1, \ 0 \leq x_{ij} \leq 1
\]

\[
\sum_{i : ij \in E} x_{ij} = x_j = \sum_{i : ji \in E} x_{ji}
\]

\[
\sum_i x_{it} = K = \sum_i x_{si}
\]

\[
\{ x \in \text{FLOW}_K \}
\]

Implies each \( x \) must be 0 or 1

\( x_i, x_{ij} \) are integer.
Formulated as ILP

\[
\min_x \left\{ \sum_i c_i x_i + \sum_{i,j \in E} c_{ij} x_{ij} \right\}
\]

\[
0 \leq x_i \leq 1, \quad 0 \leq x_{ij} \leq 1
\]

\[
\sum_{i : ij \in E} x_{ij} = x_j = \sum_{i : ji \in E} x_{ji}
\]

\[
\sum_i x_{it} = K = \sum_i x_{si}
\]

\[x_i, x_{ij} \text{ are integer.}\]

Implies each x must be 0 or 1

Fun fact: we can drop (relax) this constraint to get an LP rather than ILP, and still be guaranteed an integer 0,1 solution. (due to totally unimodular constraints)
Network Flow Approach

Pros:
- Efficient solution (guaranteed polynomial time algorithms)
- Uses all frames to achieve a global batch solution

Cons:
- Cost function is limited to unary and pairwise costs defined over nodes
- Cannot represent higher-order terms, such as pairwise costs that are defined over (pairs of) edges

This paper: add ability to use pairwise edge costs
Examples: Pairwise Costs on Edges

Cost($x_i, x_j$) = $q_{ij} x_i x_j$

Motivation: Discourage two overlapping detections from both being part of the solution. Aka: non-maximum suppression.
Examples: Pairwise Costs on Edges

\[
\text{Cost}(x_{ij}, x_{jk}) = q_{ijk} x_{ij} x_{jk}
\]

Motivation: Encourage smooth trajectories, e.g. constant velocity motion.
This is not considered in this paper.
Examples: Pairwise Costs on Edges

Frame 1

Frame 2

Frame 3

Head detection flow network

Body detection flow network

Cost\((x^h_i, x^b_j) = -q_{ij} x^h_i x^b_j\)

Motivation: Encourage two compatible detections of different types to both be part of the solution.
E.g. co-occurring head and body detections
Approach: Form an IQP

Let \( z = [z_1, \ldots, z_M] \)
\[
= [x_1, \ldots, x_i, \ldots, x_n, x_{11}, \ldots, x_{ij}, \ldots, x_{mn}]^
\]

n detection vars   M-n connection vars

Form the integer quadratic program:

\[
\min_{z} c^T z + z^T Q z
\]

\[
s.t. \quad z \in FLOW_K
\]
\[z \quad \text{integer}\]

\( c \) is linear costs (on edges)
\( Q \) is sparse matrix of quadratic costs (on edge pairs)
Integer Quadratic Programs are NP-hard in general

This paper therefore discusses approximate solution methods.
Solution Approach

Relax the integer constraints
Unfortunately, sol’n vars can take values between 0 and 1.
“Fix up” the solution so that all variables are 0 or 1, while maintaining feasibility (satisfy the Flow constraints).
Solution Method 1

Modify diagonal terms of $Q$ so that it is positive semidefinite

$$Q_{ii}^{\text{new}} = \sum_{j \neq i} |Q_{ij}|$$

and adjust linear costs $c_i$ to keep same objective function value

$$c_i^{\text{new}} = c_i - Q_{ii}^{\text{new}} + Q_{ii}^{\text{old}}$$

Note: $z_i^2 = z_i$

Problem is now convex, and a global solution can be found efficiently, e.g. by gradient descent or by the Frank-Wolfe algorithm which iteratively minimizes a linearization of the convex quadratic problem.
Frank-Wolfe Algorithm

Convex function
(quadratic in our case)

Tangent plane approx at current solution estimate

Convex domain of feasible solutions.

\[ \argmin_{s \in D} s^T \nabla f(x_k) \]

\[ x_{k+1} \leftarrow x_k + \gamma (s_k - x_k) \]
Frank-Wolfe Algorithm

\[
\min_z \quad c^T z + z^T Qz \\
\text{s.t.} \quad z \in FLOW_K
\]

Let current solution estimate be \( z^* \)

Tangent is \( (c + (Q + Q^T)z^*)^T z \)

To find \( s = \arg\min z \) we solve the network flow problem

\[
\min_z \quad (c + (Q + Q^T)z^*)^T z \\
\text{s.t.} \quad z \in FLOW_K.
\]
Solution Method 2

Introduce additional variables $u_{ij} = z_i z_j$ and constraints in order to form an equivalent integer linear program.

$$\min_{z, u} \quad c^T z + q^T u$$

$$z \in \text{FLOW}_K$$

$$0 \leq u_{ij} \leq 1, \forall ij \in Q \quad \{ (z, u) \in \text{LOCAL}(Q) \}$$

$$u_{ij} \leq z_i, u_{ij} \leq z_j$$

$$z_i + z_j \leq 1 + u_{ij}$$

$$z, u \quad \text{integer}.$$

- $u_{ij}$ is 0 if either $z_i$ or $z_j$ are 0
- $u_{ij}$ is 1 if both $z_i$ and $z_j$ are 1
Solution Method 2

Introduce additional variables $u_{ij} = z_i z_j$ and constraints in order to form an equivalent integer linear program.

\[
\begin{align*}
\min_{z,u} & \quad c^T z + q^T u \\
\text{subject to} & \quad z \in \text{FLOW}_K \\
& \quad 0 \leq u_{ij} \leq 1, \forall i,j \in Q \\
& \quad u_{ij} \leq z_i, u_{ij} \leq z_j \\
& \quad z_i + z_j \leq 1 + u_{ij} \\
& \quad z, u \quad \text{integer}.
\end{align*}
\]

This is then relaxed to a linear program (non-integer solns)
Rounding the Solution

To get back an integer 0,1 solution:

1) round the values  $\rightarrow$ bad idea, may not satisfy FLOW

2) Hamming rounding – look for closest solution in FLOW
   $\rightarrow$ also not great, since that solution may not have a good objective function value.

3) Frank-Wolfe rounding – one iteration of Frank-Wolfe algorithm by solving the linear program

$$\min_{\mathbf{z}} \quad (\mathbf{c} + (\mathbf{Q} + \mathbf{Q}^\top)\mathbf{z}^*)^\top \mathbf{z}$$

s.t. $\mathbf{z} \in \text{FLOW}_K$.

Bob’s note: This Q is not convex, so this is only a heuristic.
Some Results

Figure 1: Results of network flow tracking using cost functions with/without pairwise terms. (a)-(b): a pairwise term that penalizes the overlap between different tracks helps resolving ambiguous tracks (shown in red) in crowded scenes. (c)-(d): a pairwise term that encourages the consistency between two signals (here head detections and body detections) helps eliminating failures (shown in red) of object detectors.
Re-detection measure. The proposed re-detection measure evaluates the ability of a tracker to find the correct location of a given object after $\Delta t$ frames. The measure is inspired by the common evaluation procedure for object detection in still images $[10]$ and extends it to tracking. For each pair of detections $A_t$ and $B_{t+\Delta t}$ associated to the same track by a tracker, we check if there exists a ground truth track that overlaps with $A_t$ and $B_{t+\Delta t}$ on frames $t$ and $t + \Delta t$ respectively. If the answer is negative, the subtrack $(A_t, B_{t+\Delta t})$ is labeled as false positive. Otherwise, it is labeled as true positive unless there exist multiple subtracks overlapping with the same ground truth. To avoid multiple responses, in the latter case only one subtrack is labeled as true positive while others are declared as false positives.
Some Results

"MarchingRally" tracks
- NF Basic
- Greedy + NMS
- NF + Overlap

"TownCenter" Head Tracks
- NF Basic
- Greedy + NMS
- NF + Overlap + Co-occurrence

Track detection AP vs Time interval (frames)
Some Results

<table>
<thead>
<tr>
<th></th>
<th>Rcll</th>
<th>Prcn</th>
<th>GT</th>
<th>MT</th>
<th>PT</th>
<th>ML</th>
<th>FP</th>
<th>FN</th>
<th>IDs</th>
<th>FM</th>
<th>MOTA</th>
<th>MOTP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TUD Stadtmitte</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>67.9</td>
<td>72.0</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>305</td>
<td>371</td>
<td>26</td>
<td>26</td>
<td>39.3</td>
<td>59.5</td>
</tr>
<tr>
<td>NF+pairwise</td>
<td>59.6</td>
<td><strong>89.9</strong></td>
<td>10</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>77</td>
<td>467</td>
<td>15</td>
<td>22</td>
<td>51.6</td>
<td><strong>61.6</strong></td>
</tr>
<tr>
<td>Milan [20]</td>
<td><strong>69.1</strong></td>
<td>85.6</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>134</td>
<td>457</td>
<td>15</td>
<td>13</td>
<td><strong>56.2</strong></td>
<td><strong>61.6</strong></td>
</tr>
<tr>
<td><strong>PETS S2L1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>93.7</td>
<td>83.4</td>
<td>19</td>
<td>17</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>870</td>
<td>293</td>
<td>64</td>
<td>66</td>
<td>73.6</td>
</tr>
<tr>
<td>NF+pairwise</td>
<td>92.4</td>
<td>94.3</td>
<td>19</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>262</td>
<td>354</td>
<td>56</td>
<td>74</td>
<td>85.5</td>
<td><strong>76.2</strong></td>
</tr>
<tr>
<td>Milan [20]</td>
<td><strong>96.8</strong></td>
<td>94.1</td>
<td>19</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>282</td>
<td>148</td>
<td>22</td>
<td>15</td>
<td><strong>90.3</strong></td>
<td>74.3</td>
</tr>
<tr>
<td><strong>PETS S2L2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>47.7</td>
<td>87.6</td>
<td>43</td>
<td>1</td>
<td>37</td>
<td>5</td>
<td>693</td>
<td>5383</td>
<td>291</td>
<td>531</td>
<td>38.1</td>
<td>60.7</td>
</tr>
<tr>
<td>NF+pairwise</td>
<td>60.6</td>
<td>88.6</td>
<td>43</td>
<td>6</td>
<td>34</td>
<td>3</td>
<td>807</td>
<td>4050</td>
<td>244</td>
<td>379</td>
<td>50.4</td>
<td><strong>60.6</strong></td>
</tr>
<tr>
<td>Milan [20]</td>
<td><strong>65.1</strong></td>
<td>92.4</td>
<td>43</td>
<td>11</td>
<td>31</td>
<td>1</td>
<td>549</td>
<td>3592</td>
<td>167</td>
<td>153</td>
<td><strong>58.1</strong></td>
<td>59.8</td>
</tr>
<tr>
<td><strong>PETS S2L3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>44.5</td>
<td>92.2</td>
<td>44</td>
<td>9</td>
<td>15</td>
<td>20</td>
<td>164</td>
<td>2428</td>
<td>121</td>
<td>189</td>
<td>38.0</td>
<td>69.3</td>
</tr>
<tr>
<td>NF+pairwise</td>
<td><strong>45.5</strong></td>
<td>91.2</td>
<td>44</td>
<td>12</td>
<td>15</td>
<td>17</td>
<td>155</td>
<td>2125</td>
<td>44</td>
<td>50</td>
<td><strong>40.3</strong></td>
<td>61.2</td>
</tr>
<tr>
<td>Milan [20]</td>
<td>43.0</td>
<td><strong>94.2</strong></td>
<td>44</td>
<td>8</td>
<td>17</td>
<td>19</td>
<td>115</td>
<td>2493</td>
<td>27</td>
<td>22</td>
<td>39.8</td>
<td><strong>65.0</strong></td>
</tr>
<tr>
<td><strong>PETS S1L1-2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>62.9</td>
<td>89.1</td>
<td>44</td>
<td>18</td>
<td>15</td>
<td>11</td>
<td>295</td>
<td>1425</td>
<td>289</td>
<td>140</td>
<td>47.8</td>
<td>65.2</td>
</tr>
<tr>
<td>NF+pairwise</td>
<td><strong>68.9</strong></td>
<td>92.0</td>
<td>44</td>
<td>20</td>
<td>16</td>
<td>8</td>
<td>230</td>
<td><strong>1198</strong></td>
<td>35</td>
<td>74</td>
<td><strong>62.0</strong></td>
<td>62.1</td>
</tr>
<tr>
<td>Milan [20]</td>
<td>64.9</td>
<td><strong>92.4</strong></td>
<td>44</td>
<td>21</td>
<td>12</td>
<td>11</td>
<td>169</td>
<td>1349</td>
<td>22</td>
<td>19</td>
<td>60.0</td>
<td>61.9</td>
</tr>
<tr>
<td><strong>PETS S1L2-1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF</td>
<td>31.3</td>
<td>87.4</td>
<td>42</td>
<td>4</td>
<td>15</td>
<td>23</td>
<td>208</td>
<td>3501</td>
<td>101</td>
<td>243</td>
<td>23.7</td>
<td>57.9</td>
</tr>
<tr>
<td>NF+pairwise</td>
<td><strong>37.9</strong></td>
<td>89.6</td>
<td>42</td>
<td>4</td>
<td><strong>20</strong></td>
<td>18</td>
<td>223</td>
<td><strong>3141</strong></td>
<td>67</td>
<td>122</td>
<td>32.2</td>
<td>55.0</td>
</tr>
<tr>
<td>Milan [20]</td>
<td>30.9</td>
<td><strong>98.3</strong></td>
<td>42</td>
<td>2</td>
<td>19</td>
<td>21</td>
<td>27</td>
<td>3494</td>
<td>42</td>
<td>34</td>
<td>29.6</td>
<td><strong>58.8</strong></td>
</tr>
</tbody>
</table>

Table 1: Table summarizing results over PETS and TUD sequences. Bold indicates best value for each column for each dataset. Abbreviations are as follows GT - ground truth tracks. MT - Mostly tracked. PT - partially tracked. ML - mostly lost. FP - false positives. FN - false negatives. IDs - ID swaps. FM - fragmentation.