Confidence intervals for tracking performance scores

Ricardo Sánchez-Matilla and Andrea Cavallaro

Published in: IEEE International Conference on Image Processing (ICIP) 2018

Centre for Intelligent Sensing
Queen Mary University of London
Performance evaluation (ranking)
Annotations

Performance evaluation requires manual annotations

Manual annotations are [1]:

– Tedious
– Expensive (time and economically wise)
– (Potentially) inaccurate
– (Potentially) unfeasible to retrieve

Inaccurate annotations produce inaccurate evaluations

Tracking: Interpolated annotations

Ideal and interpolated GT

Manual (ideal) annotation
Interpolated annotation
Publicly available annotation (uses linear interpolation)

$\beta$: interpolation factor
Tracking: Datasets

- Most public datasets use linear interpolation for their annotations

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tool</th>
<th>Linear Interpolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAVIAR</td>
<td>CaviarGui</td>
<td>✓</td>
</tr>
<tr>
<td>TUD</td>
<td>NA</td>
<td>✓</td>
</tr>
<tr>
<td>ETH</td>
<td>NA</td>
<td>✓</td>
</tr>
<tr>
<td>PETS09</td>
<td>NA</td>
<td>✓</td>
</tr>
<tr>
<td>KITTI</td>
<td>MT</td>
<td>NA</td>
</tr>
<tr>
<td>i-LIDS</td>
<td>ViPER</td>
<td>✓</td>
</tr>
<tr>
<td>MOTB15</td>
<td>VATIC</td>
<td>✓</td>
</tr>
<tr>
<td>MOTB16</td>
<td>NA</td>
<td>✓</td>
</tr>
</tbody>
</table>

NA: not available
Confidence intervals

- Annotation inaccuracies should be taken into account
- Estimate uncertainty in annotations for a given dataset:
  - Unknown interpolation
  - Without doing further annotations
  - Estimate confidence interval
Separating manual/interpolated annotations

\[ Z = \{ z^\lambda_k : \lambda = 1 \ldots \Lambda; k = 0 \ldots K_\lambda - 1 \} \]

\[ z^\lambda_k = (u, v, w, h) \]

\[ \lambda: \text{target identity} \]

\[ k: \text{time} \]

\[ Z = \hat{Z} \cup \check{Z}, \quad \hat{Z} \cap \check{Z} = \emptyset \]

\( Z \): public dataset
\( \hat{Z} \): manually annotated subset
\( \check{Z} \): linearly interpolated subset

• Percentage of interpolation in the dataset:

\[ \frac{|\hat{Z}|}{|Z|} \]

Second derivatives equal to zero
Building interpolated versions

- Generate interpolated versions of the manual subset through a decimation-interpolation procedure

\[ Z \rightarrow \left[ \tilde{Z} \quad \hat{Z} \right] \rightarrow Z_\beta \]

\[ \beta \in \{3, 6, 9, 12\} : \text{interpolation factor} \]

- \( Z \): public dataset
- \( \tilde{Z} \): manually annotated subset
- \( \hat{Z} \): linearly interpolated subset
Confidence intervals

• Compare interpolated version against manual version

\[ \alpha_{s,\beta} = s(\tilde{Z}, Z_\beta) \]
\[ s(\cdot, \cdot) = 100 - \text{MOTA} \]
\[ s(\cdot, \cdot) = 100 - \text{MOTP} \]

\[ \text{MOTA} = 1 - \frac{1}{N} \sum_{\lambda=0}^{\Lambda} \sum_{k=0}^{\tilde{K}_\lambda - 1} (FN_k^\lambda + FP_k^\lambda + IDSW_k^\lambda) \]

\[ \text{MOTP} = \frac{1}{N} \sum_{\lambda=0}^{\Lambda} \sum_{k=0}^{\tilde{K}_\lambda' - 1} \frac{\tilde{Z}_k^\lambda \cap Z_{k,\beta}^\lambda}{\tilde{Z}_k^\lambda \cup Z_{k,\beta}^\lambda} \]

\( \tilde{Z} \): public dataset
\( \tilde{Z} \): manually annotated subset
\( \hat{Z} \): linearly interpolated subset
\( N \): \# annotations
\( \beta \): interpolation factor
Results on MOTB16

• 39.7% annotations are generated through linear interpolation
• MOTA confidence: 0.22
• MOTP confidence 3.14

We assume identity switches are negligible
Impact on MOTB16
Conclusion

• Interpolation:
  – Makes possible to annotate large scale datasets
  – Makes possible to annotate not visible objects
  – Introduces inaccuracies

• Confidence interval:
  – Allows to take into account annotation inaccuracies
  – From an already annotated dataset
  – With unknown annotation policy
  – Without the need of doing further annotations

• Ranking of methods should be redesign considering confidence intervals
Q&A Questions Answers