Multi-view Object Detection Using Epipolar Constraints within Cluttered X-ray Security Imagery

Department of {*Computer Science, †Engineering}
Durham University, Durham, UK

25th International Conference on Pattern Recognition
January 2021
Introduction

Contemporary X-ray scanners used for aviation security screening provide two or more views.

The geometry of two views of the same scene is related by epipolar geometry.

\[ y' = Fx \quad F = \left[ P'C \right] \cdot P' \cdot P^+ \]

- **Uncalibrated** cameras.
- Feature detection and matching is not suitable for transmission imagery.
- Multi-view information has not been integrated before.
- We use **object-level annotations** to estimate the fundamental matrix.
Fundamental Matrix Estimation

\[ x = \bar{x} + \Delta x + \Psi, \quad \Rightarrow \quad x = \bar{x} + \Delta \hat{x} \]
\[ \Delta \hat{x} \sim \mathcal{N}(\mu_\Psi, \sigma^2) \]

Measurement error
A mapping of the real object centre to the bounding box centroid

\[ x = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \]

Bounding box centroids are used as point correspondences
Multi-view Epipolar Detection Confidence

The distance to the epipolar line is

\[
d'(x', l') = \frac{x'^\top l'}{\sqrt{l_1^2 + l_2^2}} = \frac{1}{\sqrt{l_1^2 + l_2^2}}(\ddot{x}'\top l' + l_1' \Delta \dot{x} + l_2' \Delta \dot{y})
\]

If \( \ddot{x}'l' \) is the true correspondence \( \ddot{x}'l' = 0 \). Then

\[
d'(x', l') \sim \mathcal{N}(\mu_d, \sigma_d^2)
\]

The sum of the tails for a given \( d \) is taken as a multi-view epipolar confidence

\[
p(d') = \text{erfc} \left( \frac{d' - \mu_d}{\sqrt{2} \sigma_d} \right)
\]
Multi-view filtering

1. Object detector predictions
2. Find epipolar lines for each detected object
3. Validate bounding boxes by their epipolar confidence
4. Perform NMS
### Results

**Object Detector: YOLOv3**

<table>
<thead>
<tr>
<th>Category</th>
<th>Method</th>
<th>AP</th>
<th>AP&lt;sub&gt;0.5&lt;/sub&gt;</th>
<th>AP&lt;sub&gt;0.75&lt;/sub&gt;</th>
<th>AP&lt;sub&gt;S&lt;/sub&gt;</th>
<th>AP&lt;sub&gt;M&lt;/sub&gt;</th>
<th>AP&lt;sub&gt;L&lt;/sub&gt;</th>
<th>AR&lt;sub&gt;1&lt;/sub&gt;</th>
<th>AR&lt;sub&gt;10&lt;/sub&gt;</th>
<th>AR&lt;sub&gt;S&lt;/sub&gt;</th>
<th>AR&lt;sub&gt;M&lt;/sub&gt;</th>
<th>AR&lt;sub&gt;L&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firearm</td>
<td>SV</td>
<td>0.670</td>
<td>0.983</td>
<td>0.816</td>
<td>-</td>
<td>0.681</td>
<td>0.630</td>
<td>0.743</td>
<td>0.747</td>
<td>-</td>
<td>0.744</td>
<td>0.776</td>
</tr>
<tr>
<td></td>
<td>MV</td>
<td>0.691</td>
<td>0.988</td>
<td>0.848</td>
<td>-</td>
<td>0.702</td>
<td>0.679</td>
<td>0.746</td>
<td>0.749</td>
<td>-</td>
<td>0.747</td>
<td>0.775</td>
</tr>
<tr>
<td>Laptop</td>
<td>SV</td>
<td>0.705</td>
<td>0.972</td>
<td>0.886</td>
<td>-</td>
<td>-</td>
<td>0.705</td>
<td>0.770</td>
<td>0.772</td>
<td>-</td>
<td>-</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td>MV</td>
<td>0.697</td>
<td>0.973</td>
<td>0.872</td>
<td>-</td>
<td>-</td>
<td>0.697</td>
<td>0.764</td>
<td>0.766</td>
<td>-</td>
<td>-</td>
<td>0.766</td>
</tr>
<tr>
<td>Knife</td>
<td>SV</td>
<td>0.320</td>
<td>0.756</td>
<td>0.236</td>
<td>0.083</td>
<td>0.349</td>
<td>0.175</td>
<td>0.440</td>
<td>0.447</td>
<td>0.112</td>
<td>0.464</td>
<td>0.263</td>
</tr>
<tr>
<td></td>
<td>MV</td>
<td>0.382</td>
<td>0.800</td>
<td>0.322</td>
<td>0.125</td>
<td>0.412</td>
<td>0.138</td>
<td>0.455</td>
<td>0.463</td>
<td>0.154</td>
<td>0.478</td>
<td>0.287</td>
</tr>
<tr>
<td>Camera</td>
<td>SV</td>
<td>0.530</td>
<td>0.848</td>
<td>0.621</td>
<td>-</td>
<td>0.700</td>
<td>0.530</td>
<td>0.605</td>
<td>0.605</td>
<td>-</td>
<td>0.700</td>
<td>0.605</td>
</tr>
<tr>
<td></td>
<td>MV</td>
<td>0.546</td>
<td>0.881</td>
<td>0.633</td>
<td>-</td>
<td>0.700</td>
<td>0.546</td>
<td>0.603</td>
<td>0.603</td>
<td>-</td>
<td>0.700</td>
<td>0.602</td>
</tr>
<tr>
<td>All</td>
<td>SV</td>
<td>0.557</td>
<td>0.882</td>
<td>0.640</td>
<td>0.083</td>
<td>0.577</td>
<td>0.510</td>
<td>0.640</td>
<td>0.643</td>
<td>0.112</td>
<td>0.636</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>MV</td>
<td>0.579</td>
<td>0.910</td>
<td>0.669</td>
<td>0.125</td>
<td>0.605</td>
<td>0.515</td>
<td>0.642</td>
<td>0.645</td>
<td>0.154</td>
<td>0.641</td>
<td>0.608</td>
</tr>
</tbody>
</table>

Results

A) Elimination of false positives
B) More challenging examples
C) Incorrectly eliminating previously detected objects
Conclusions

- Fundamental matrix estimation using bounding box centroids
- Epipolar confidence reduces false positives
- Improved benchmark against single-view
  - AP increased 2.2% and AP\textsubscript{0.5} increased 2.8%
  - Recall was unaffected