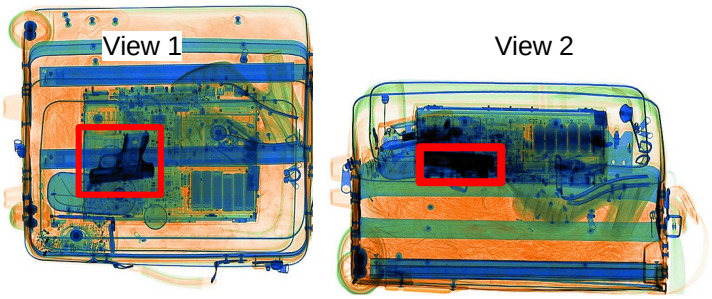
Multi-view Object Detection Using Epipolar Constraints within Cluttered X-Ray Security Imagery

Brian K. S. Isaac-Medina*, Chris G. Willcocks*, Toby P. Breckon*†
Department of {*Computer Science, †Engineering},
Durham University, UK

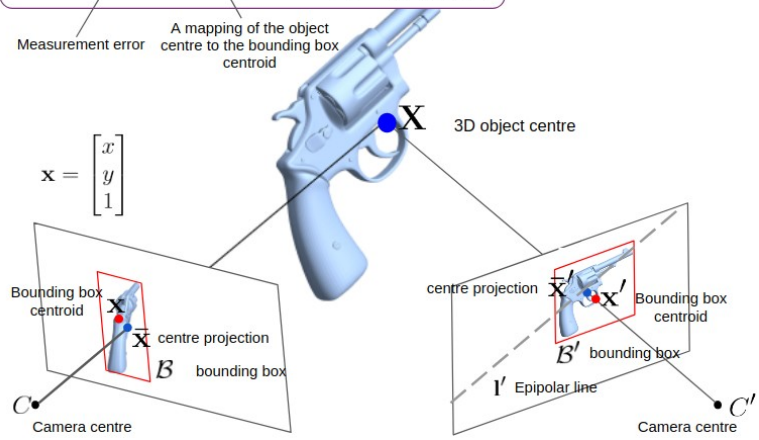
Issue: the use of information from multiple views in X-ray security object detection



Approach: epipolar constraints are imposed to improve object detection in uncalibrated multi-view X-ray baggage imagery (i - iv).

(i) Fundamental matrix is estimated using ground truth object-level bounding box centroids as point correspondences.

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

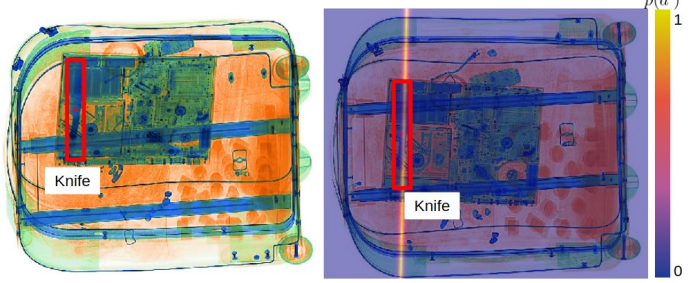


The fundamental matrix is estimated by taking the bounding box centroids as approximations to real correspondences

(ii) Detection distance to the epipolar line. Given a detection $\mathcal{B}$ in one view, the distance d' of the centroid $\mathbf{x}'$ of $\mathcal{B}'$ in the second view is modeled as $d'(\mathbf{x}', \mathbf{l}') \sim \mathcal{N}(\mu_{d'}, \sigma^2_{d'})$.

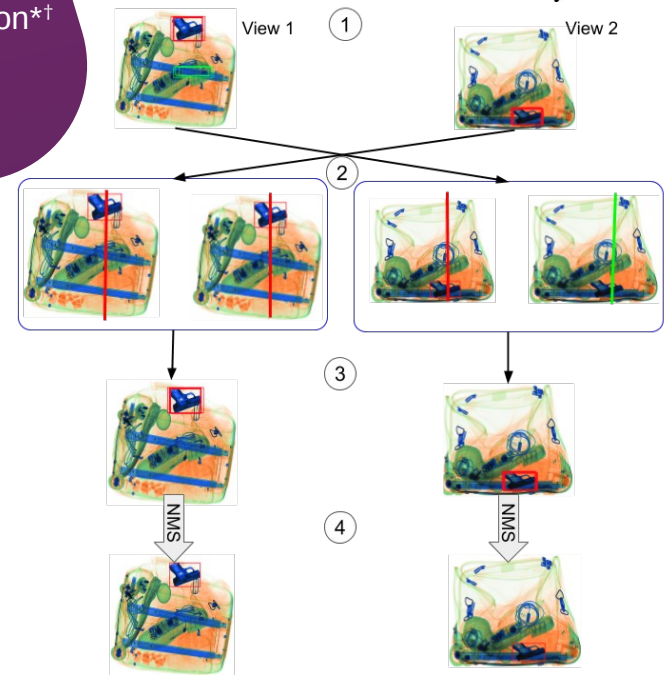
(iii) Epipolar detection confidence. The distance to the epipolar line is used to get an epipolar detection confidence.

$$p(d') = \operatorname{erfc}\left(\frac{d' - \mu_{d'}}{\sqrt{2}\sigma_{d'}}\right)$$



An epipolar confidence is given to a detection by the distance of its centroid to the epipolar line

(iv) Multi-view Filtering. The detection confidence is multiplied by $p(d')$ and filtered by a threshold before Non Maximum Suppression (NMS).



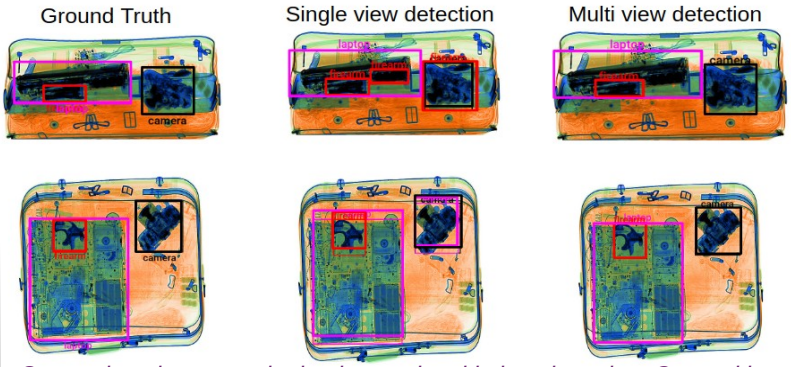
Multi-view object detection: 1. Single-view detections 2. Epipolar lines are obtained for each detection 3. Multi-view filtering 4. NMS.

Results

- Detector: YOLOv3[1]. Metric: MS COCO[2].
- AP increased +2.2% and $AP_{0.5}$ increased +2.8%, without affecting the recall.
- Precision improvement is due to elimination of false positives.

Category	Method	AP	$AP_{0.5}$	AP_S	AP_M	AP_L	AR_{10}	AR_S	AR_M	AR_L
Firearm	SV	0.670	0.983	-	0.681	0.630	0.747	-	0.744	**0.776**
	MV	**0.691**	**0.988**	-	**0.702**	**0.679**	**0.749**	-	**0.747**	0.775
Laptop	SV	**0.705**	0.972	-	-	**0.705**	**0.772**	-	-	**0.772**
	MV	0.697	**0.973**	-	-	0.697	0.766	-	-	0.766
Knife	SV	0.320	0.726	0.083	0.349	**0.175**	0.447	0.112	0.464	0.263
	MV	**0.382**	**0.800**	**0.125**	**0.412**	0.138	**0.463**	**0.154**	**0.478**	0.287
Camera	SV	0.530	0.848	-	**0.700**	0.530	**0.605**	-	**0.700**	**0.605**
	MV	**0.546**	**0.881**	-	**0.700**	**0.546**	0.603	-	**0.700**	0.602
All	SV	0.557	0.882	0.083	0.577	0.510	0.643	0.112	0.636	0.604
	MV	**0.579**	**0.910**	**0.125**	**0.605**	**0.515**	**0.645**	**0.154**	**0.641**	**0.608**

Multi-view object detection MS COCO[2] metrics



Comparison between single-view and multi-view detection. Our multi-view filtering approach removes false positives in complex scenes.

Conclusions

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

[1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *CVPR* 2016, pp. 779-788.
[2] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D.Ramanan, P. Dollar, and C.L. Zitnick, "Equationft COCO: Common objects in context," in *ECCV*. Springer, 2014, pp. 740-755.

25th International Conference on Pattern Recognition (ICPR 2020) – Jan 2021