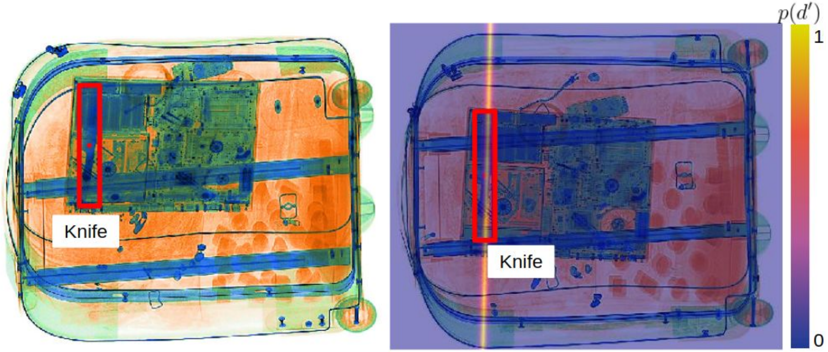
Multi-view Object Detection Using Epipolar Constraints within Cluttered X-ray Security Imagery

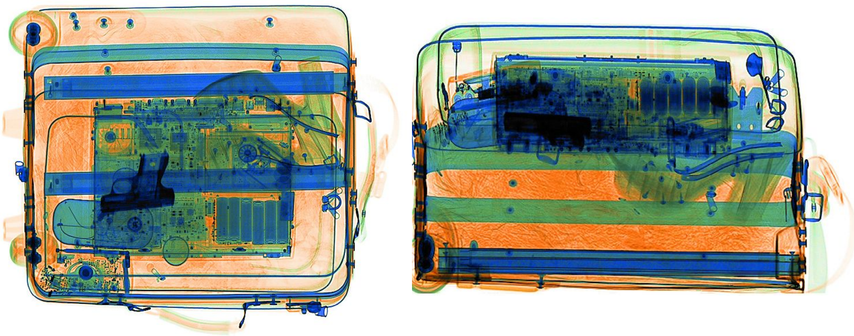
Brian K.S. Isaac-Medina*, Chris G. Willcocks*, Toby P. Breckon*[†]
Department of {*Computer Science, [†]Engineering}
Durham University, Durham, UK



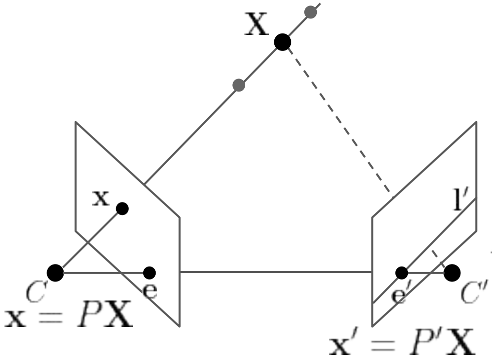
25th International Conference on Pattern Recognition
January 2021

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



$\mathbf{l}' = F\mathbf{x} \quad F = [P'C]_{\times} P' P^{+}$

P, P' are the projection matrices

- **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

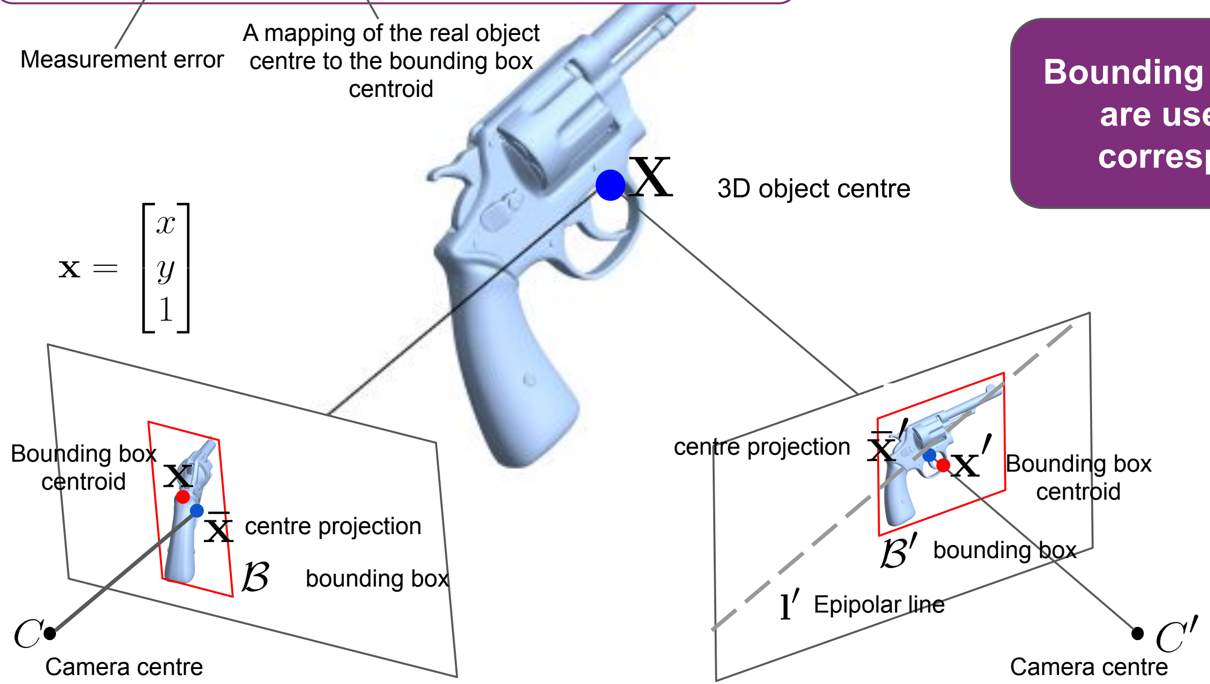
$$x = \bar{x} + \Delta x + \Psi \quad \Rightarrow \quad \begin{aligned} x &= \bar{x} + \Delta \hat{x} \\ \Delta \hat{x} &\sim \mathcal{N}(\mu_\Psi, \hat{\sigma}^2) \end{aligned}$$

Measurement error

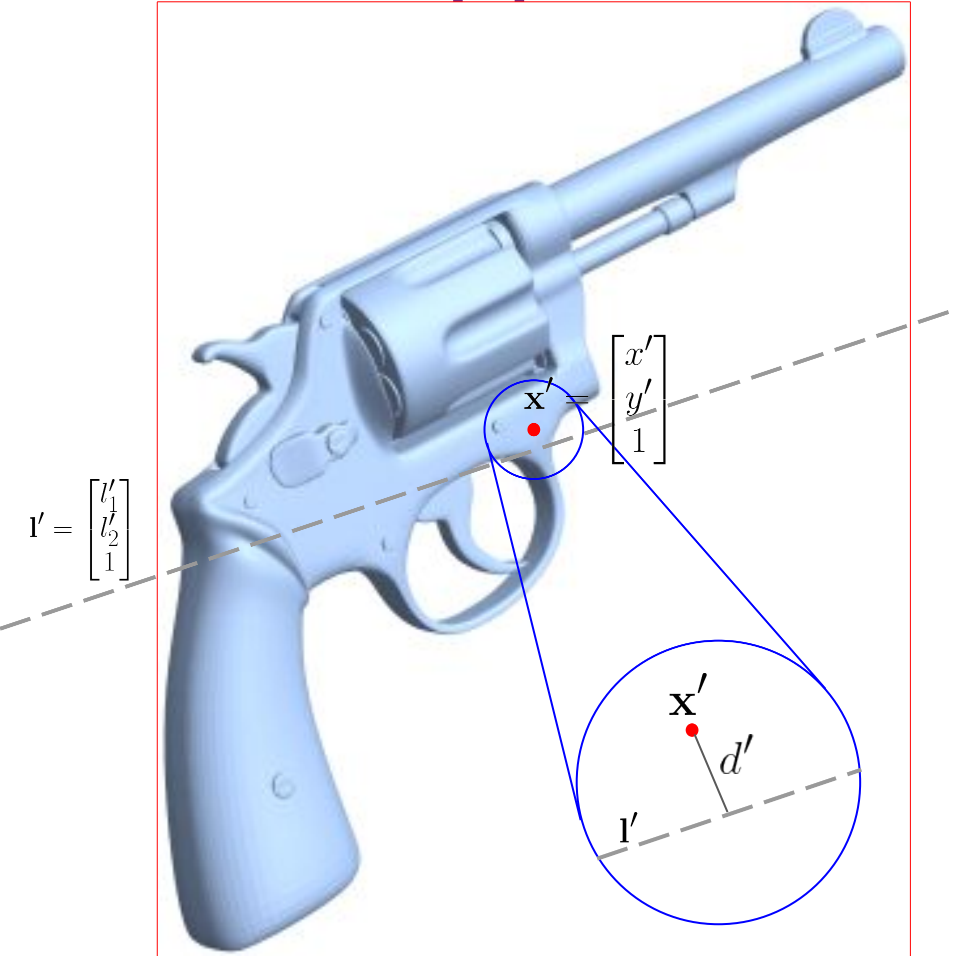
A mapping of the real object centre to the bounding box centroid

Bounding box centroids are used as point correspondences

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$



Multi-view Epipolar Detection Confidence



The distance to the epipolar line is

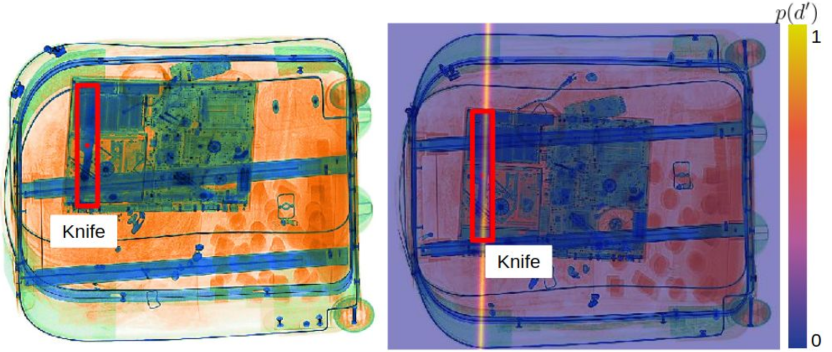
$$d'(\mathbf{x}', \mathbf{l}') = \frac{\mathbf{x}'^{\intercal}\mathbf{l}'}{\sqrt{l_1'^2 + l_2'^2}} = \frac{1}{\sqrt{l_1'^2 + l_2'^2}}(\bar{\mathbf{x}}'^{\intercal}\mathbf{l}' + l'_1\Delta\hat{x} + l'_2\Delta\hat{y})$$

If $\bar{\mathbf{x}}'$ is the true correspondence $\bar{\mathbf{x}}'^{\intercal}\mathbf{l}' = 0$. Then

$$d'(\mathbf{x}', \mathbf{l}') \sim \mathcal{N}(\mu_{d'}, \sigma^2_{d'})$$

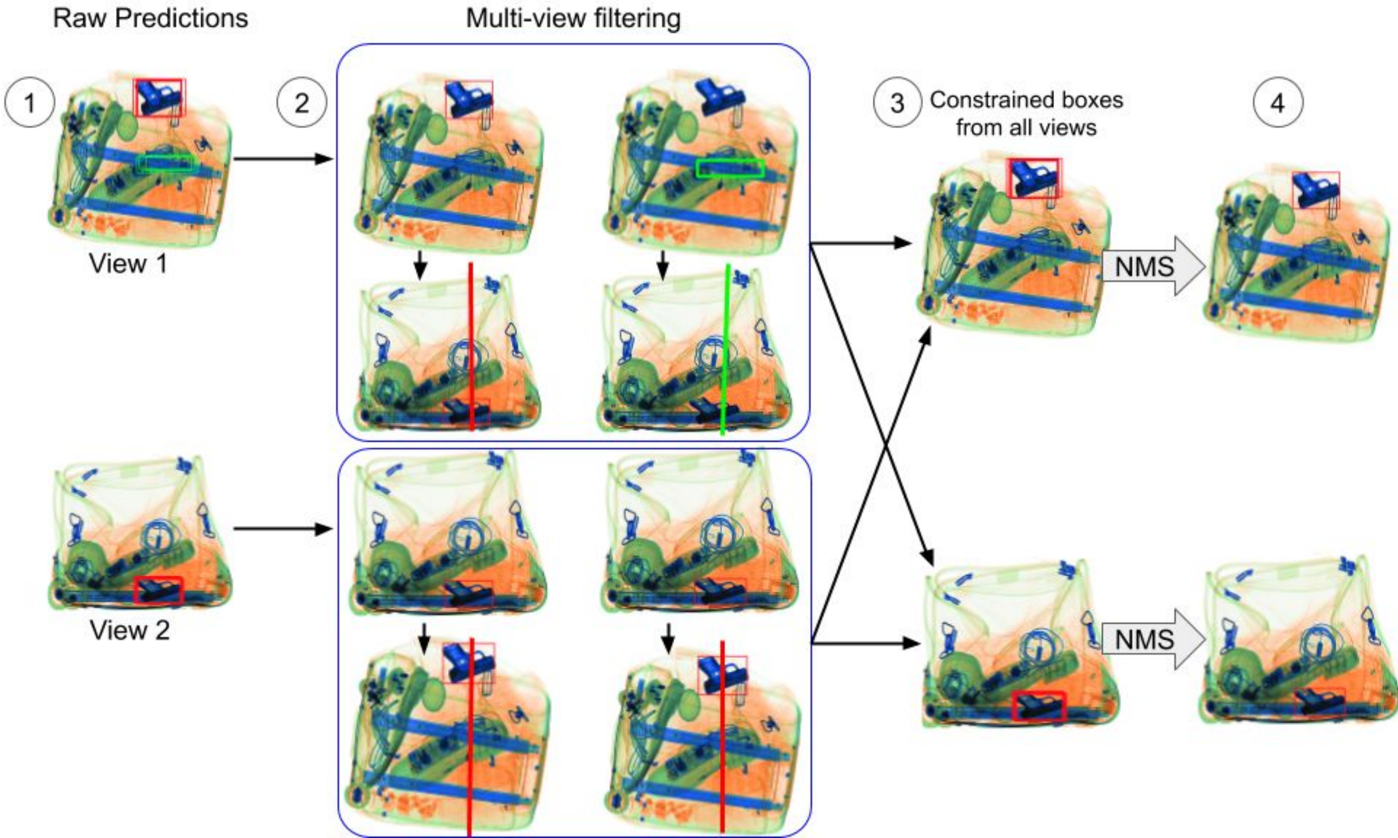
The sum of the tails for a given *d* is taken as a multi-view epipolar confidence

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



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



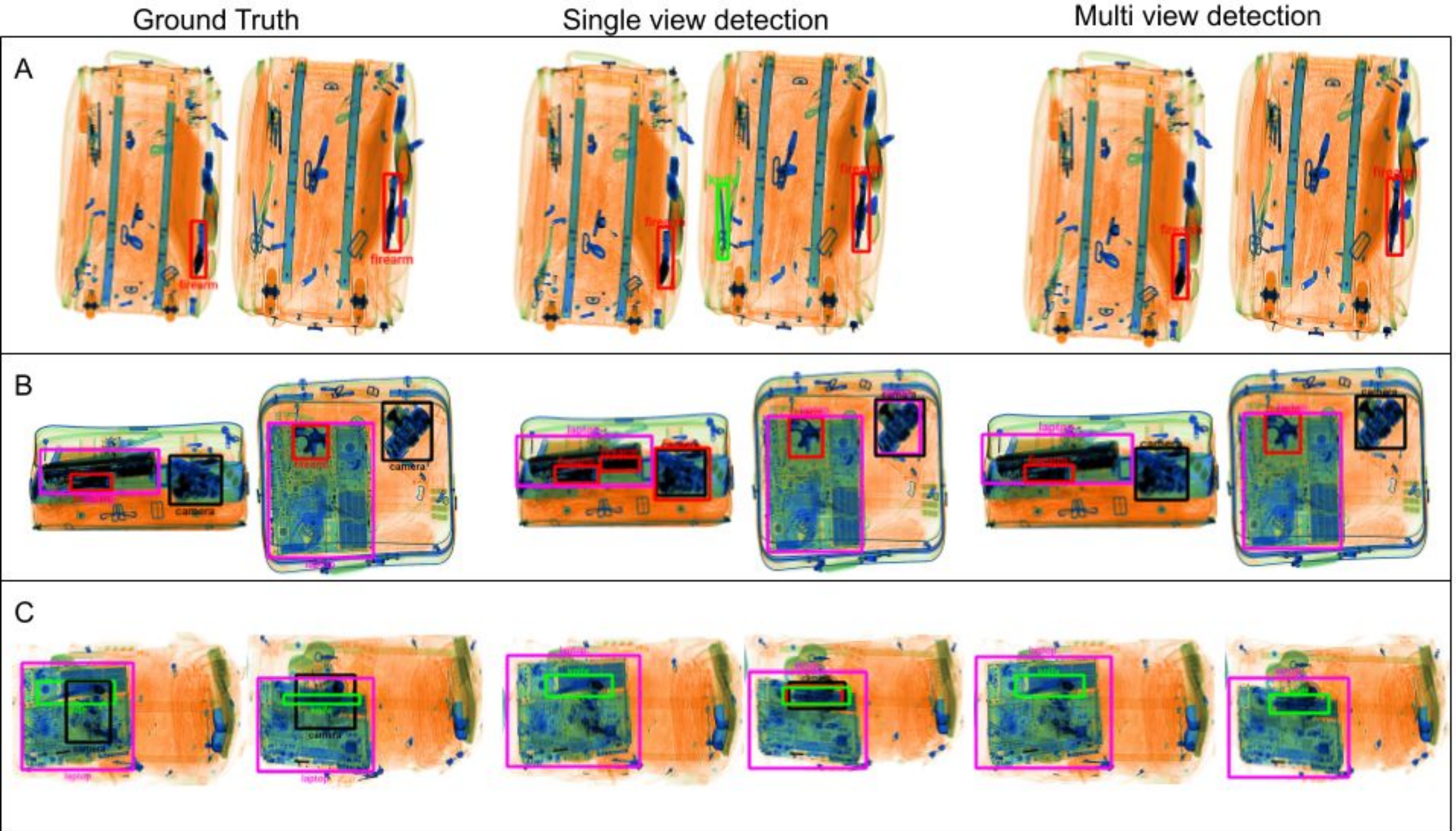
Results

Object Detector: YOLOv3[1]

Cageory	Method	AP	$AP_{0.5}$	$AP_{0.75}$	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_S	AR_M	AR_L
Firearm	SV	0.670	0.983	0.816	-	0.681	0.630	0.743	0.747	-	0.744	**0.776**
	MV	**0.691**	**0.988**	**0.848**	-	**0.702**	**0.679**	**0.746**	**0.749**	-	**0.747**	0.775
Laptop	SV	**0.705**	0.972	**0.886**	-	-	**0.705**	**0.770**	**0.772**	-	-	**0.772**
	MV	0.697	**0.973**	0.872	-	-	0.697	0.764	0.766	-	-	0.766
Knife	SV	0.320	0.756	0.236	0.083	0.349	**0.175**	0.440	0.447	0.112	0.464	0.263
	MV	**0.382**	**0.800**	**0.322**	**0.125**	**0.412**	0.138	**0.455**	**0.463**	**0.154**	**0.478**	**0.287**
Camera	SV	0.530	0.848	0.621	-	**0.700**	0.530	**0.605**	**0.605**	-	**0.700**	**0.605**
	MV	**0.546**	**0.881**	**0.633**	-	**0.700**	**0.546**	0.603	0.603	-	**0.700**	0.602
All	SV	0.557	0.882	0.640	0.083	0.577	0.510	0.640	0.643	0.112	0.636	0.604
	MV	**0.579**	**0.910**	**0.669**	**0.125**	**0.605**	**0.515**	**0.642**	**0.645**	**0.154**	**0.641**	**0.608**

[1] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.

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 $AP_{0.5}$ increased 2.8%
  - Recall was unaffected