Multiple Target Localisation at over 100 FPS

Simon Taylor

Tom Drummond

Department of Engineering

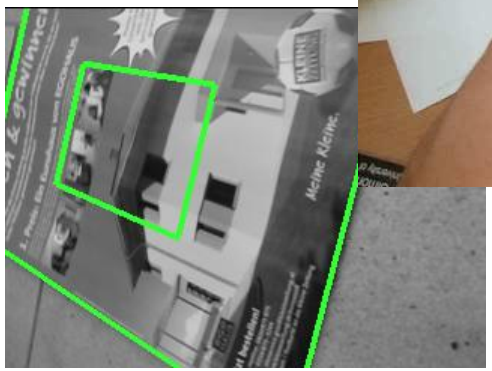
University of Cambridge, UK

This work was supported by The Boeing Company

Problem: Fast Localisation

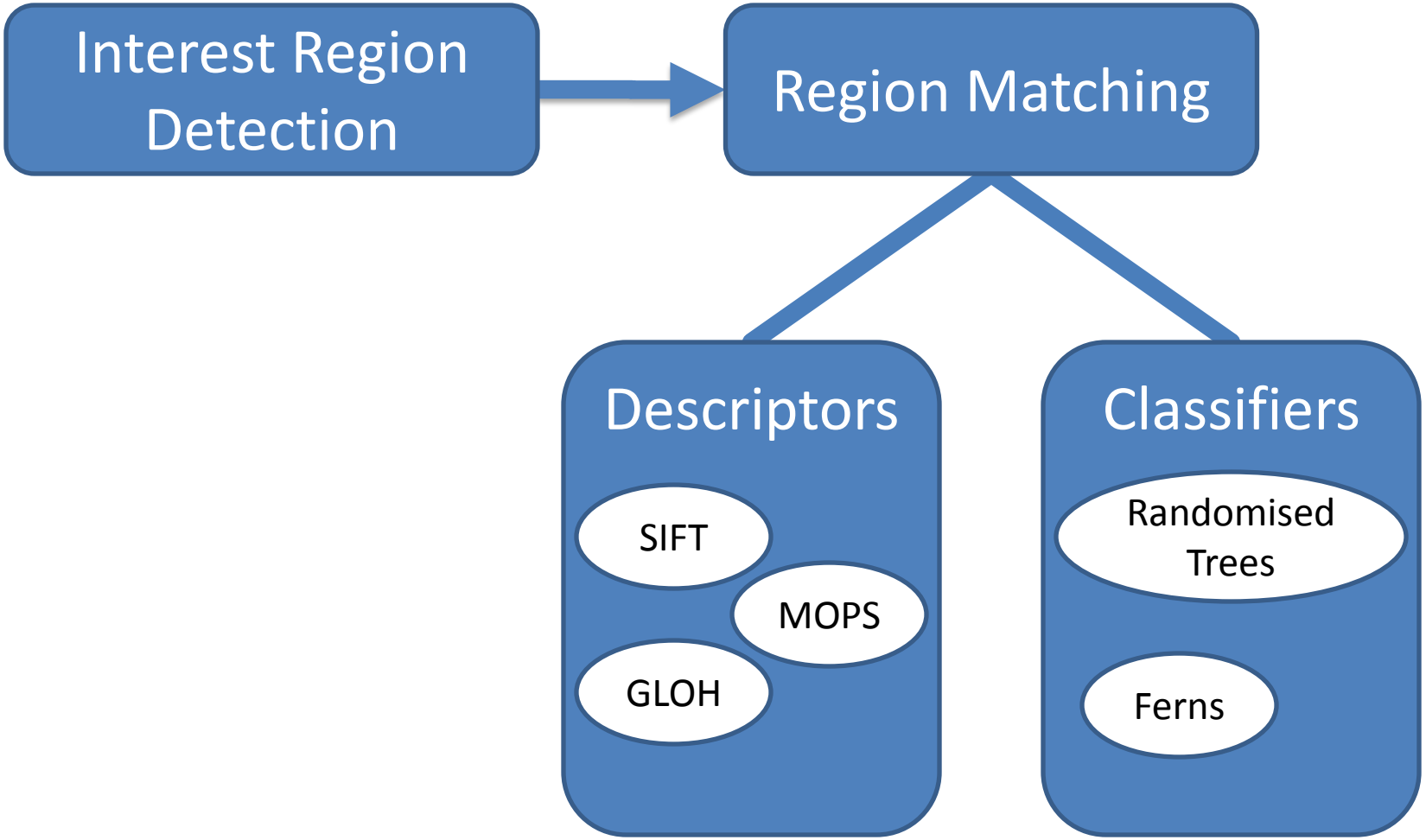
Known Target

Ca



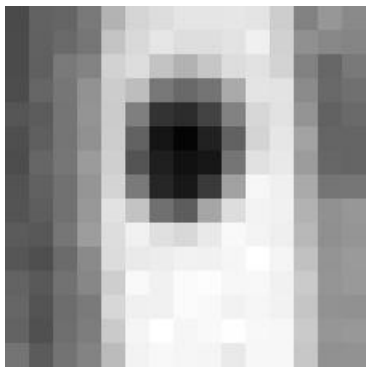
Localised Target

Local Feature Methods

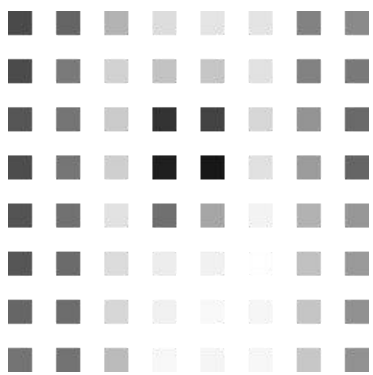


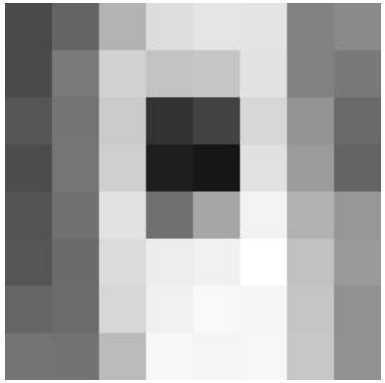
Quantised Patches

- Sparsely sampled 8 x 8 rectangular patch
- Quantised relative to μ and σ of samples



Original
Image Patch





Sparse Sampling

Quantised
Image Patch

Training Set

- Artificially warped views of entire target from different viewpoints
- Grouped into "**viewpoint bins**"

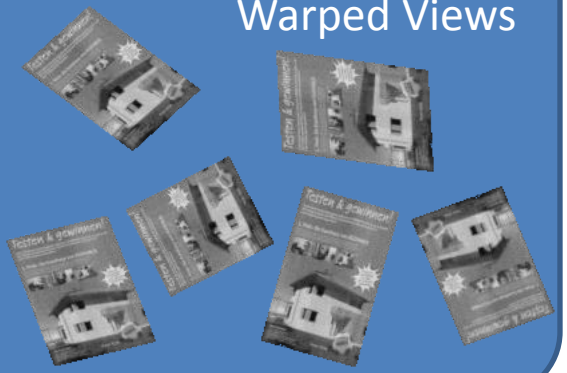
Scale Bin 4

Viewpoint Centre
"Reference Frame"

Warped Views

Scale Bin 8

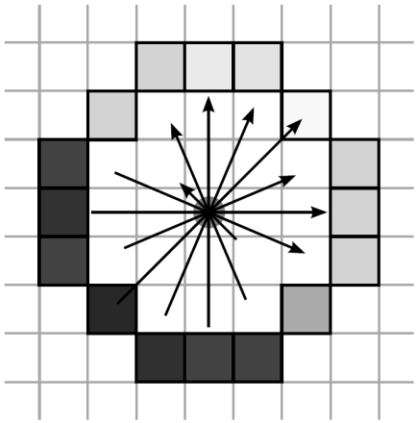
Viewpoint Centre

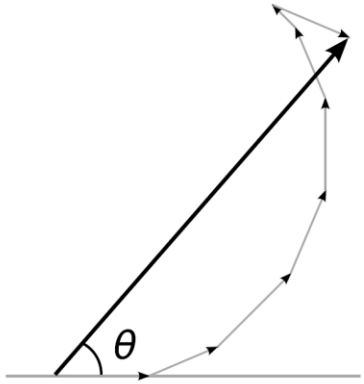


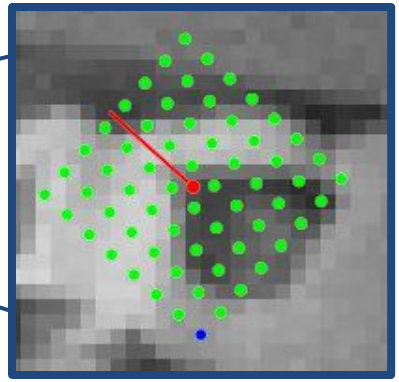
Warped Views

Extracting Subfeatures

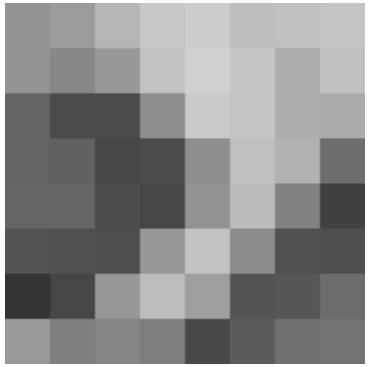
- FAST-9 Interest Points
- Orientation Estimation→



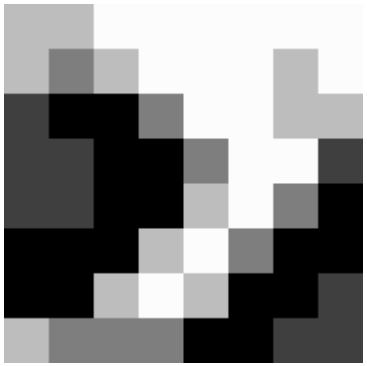




Rotate Sampling Grid



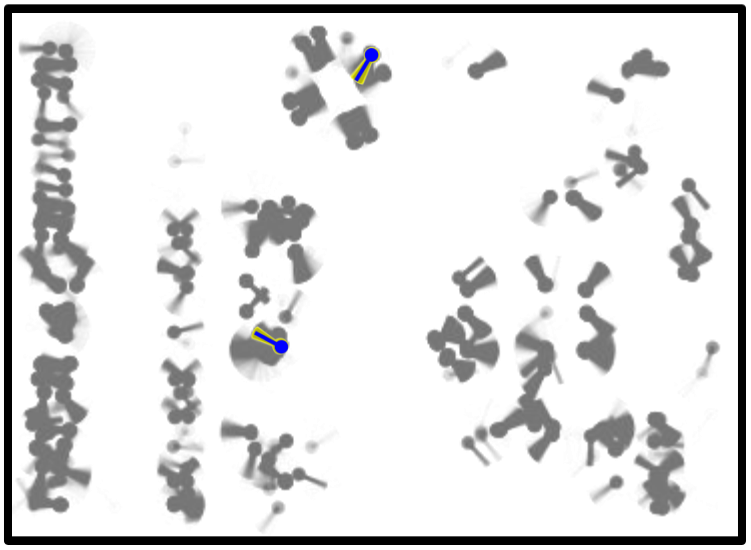
Bilinear Sampling



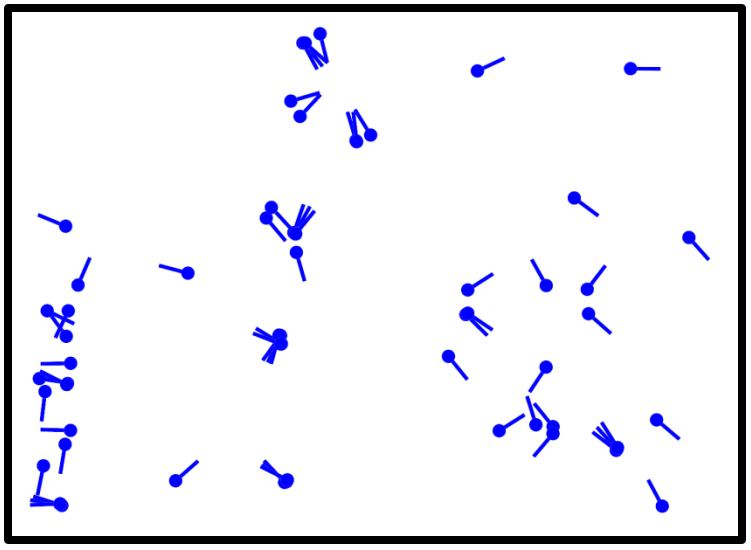
Quantisation

Clustering Subfeatures

Known warp

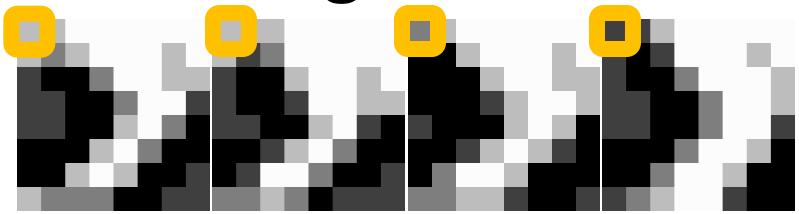


Select most repeatable

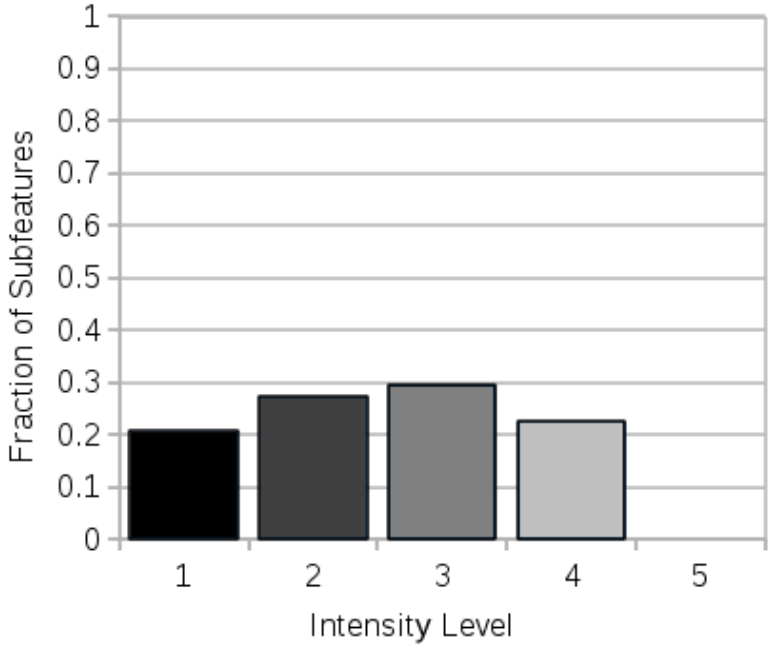


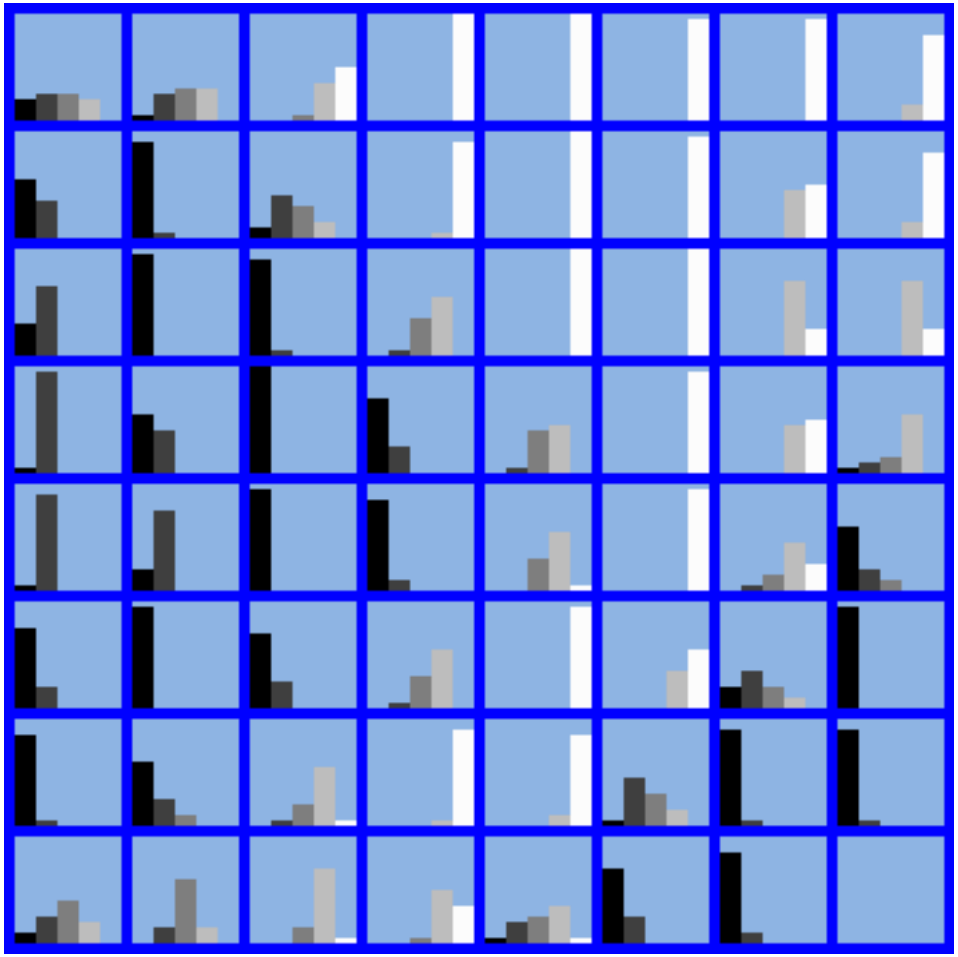
Histogrammed Intensity Patches

- Histogram of intensity for each sample



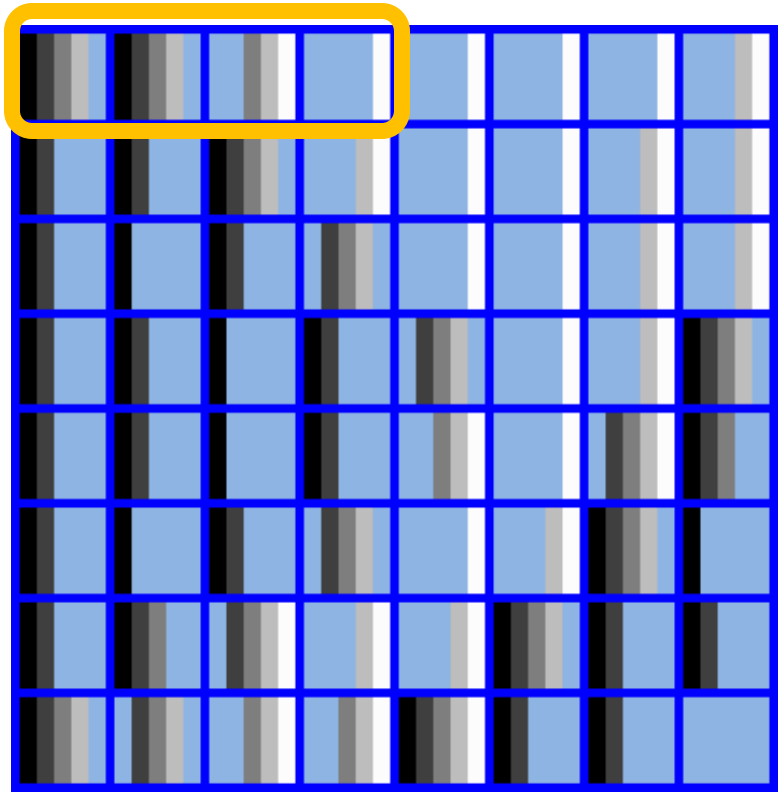
... 1081 subfeatures in total

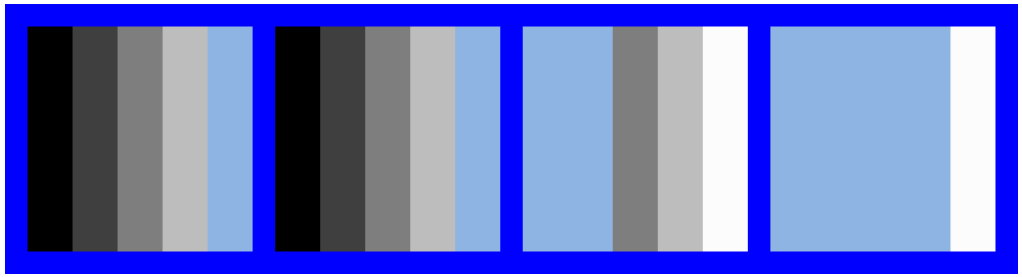




Binary Representation

- Histogram values quantised to 1-bit
- 40 bytes per feature model

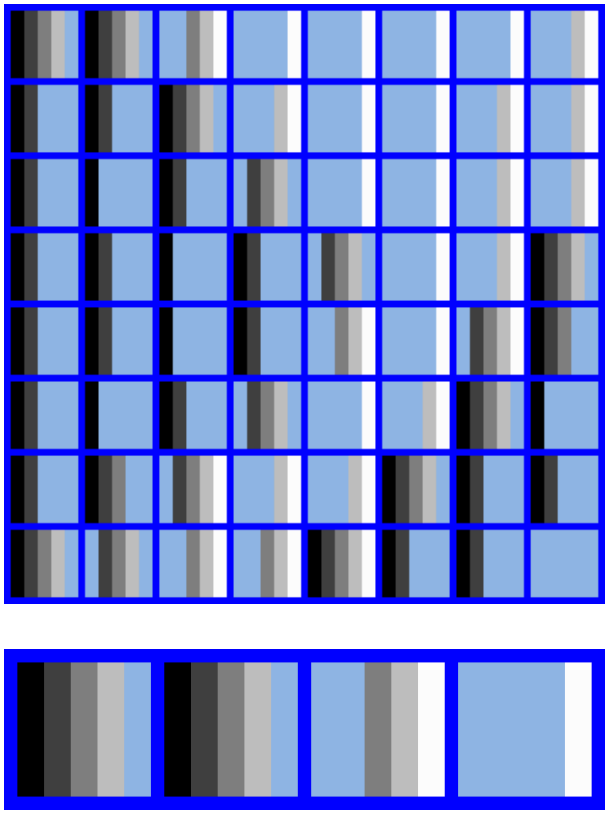




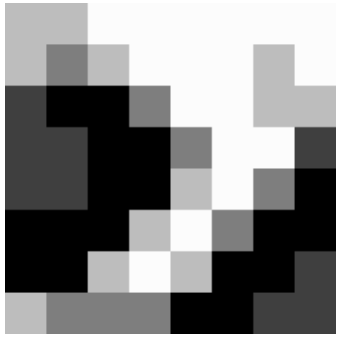
00001 00001 11000 11110

Dissimilarity Score

- Count of pixels in “rare” bins for a feature



00001 00001 11000 11110



Matching Patch

00010 00010 00001 00001 … - Patch
00001 00001 11000 11110 … - HIP
00000 00000 00000 00000 … - ANDed

Bitcount(ANDed) = 0+

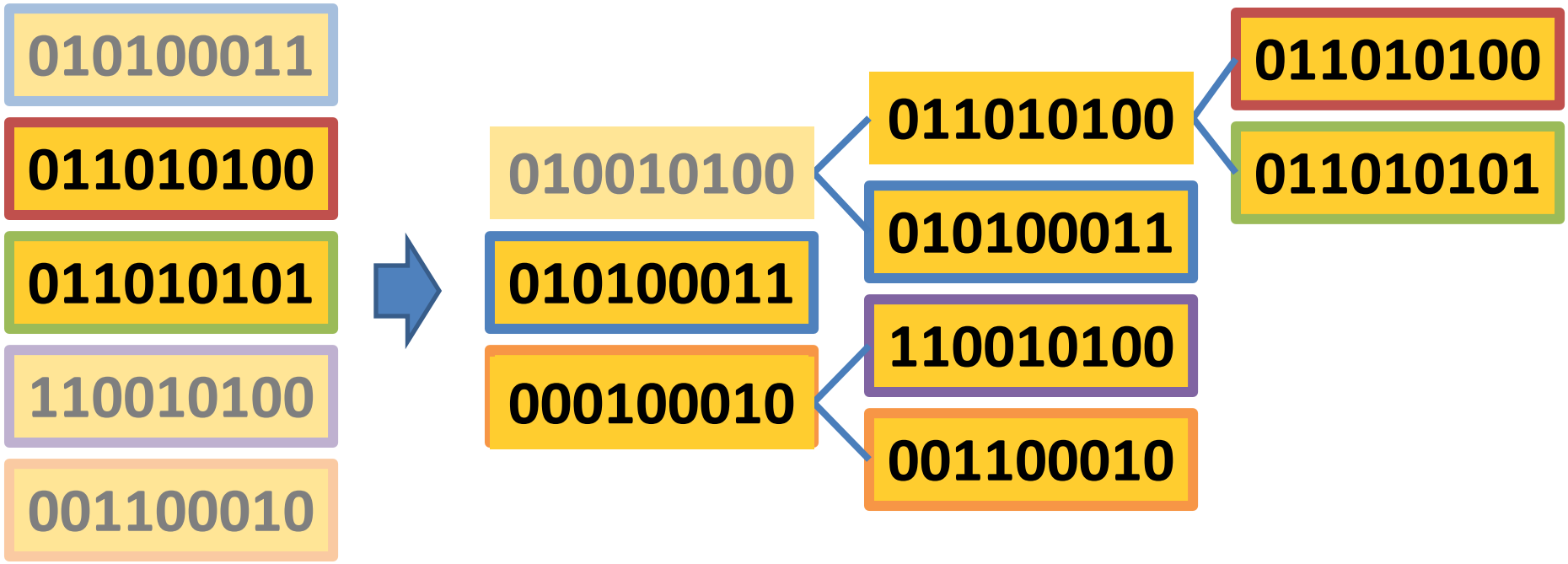
Non-matching Patch

10000 10000 10000 10000 … - Patch
00001 00001 11000 11110 … - HIP
00000 00000 10000 10000 … - ANDed

Bitcount(ANDed) = 2+

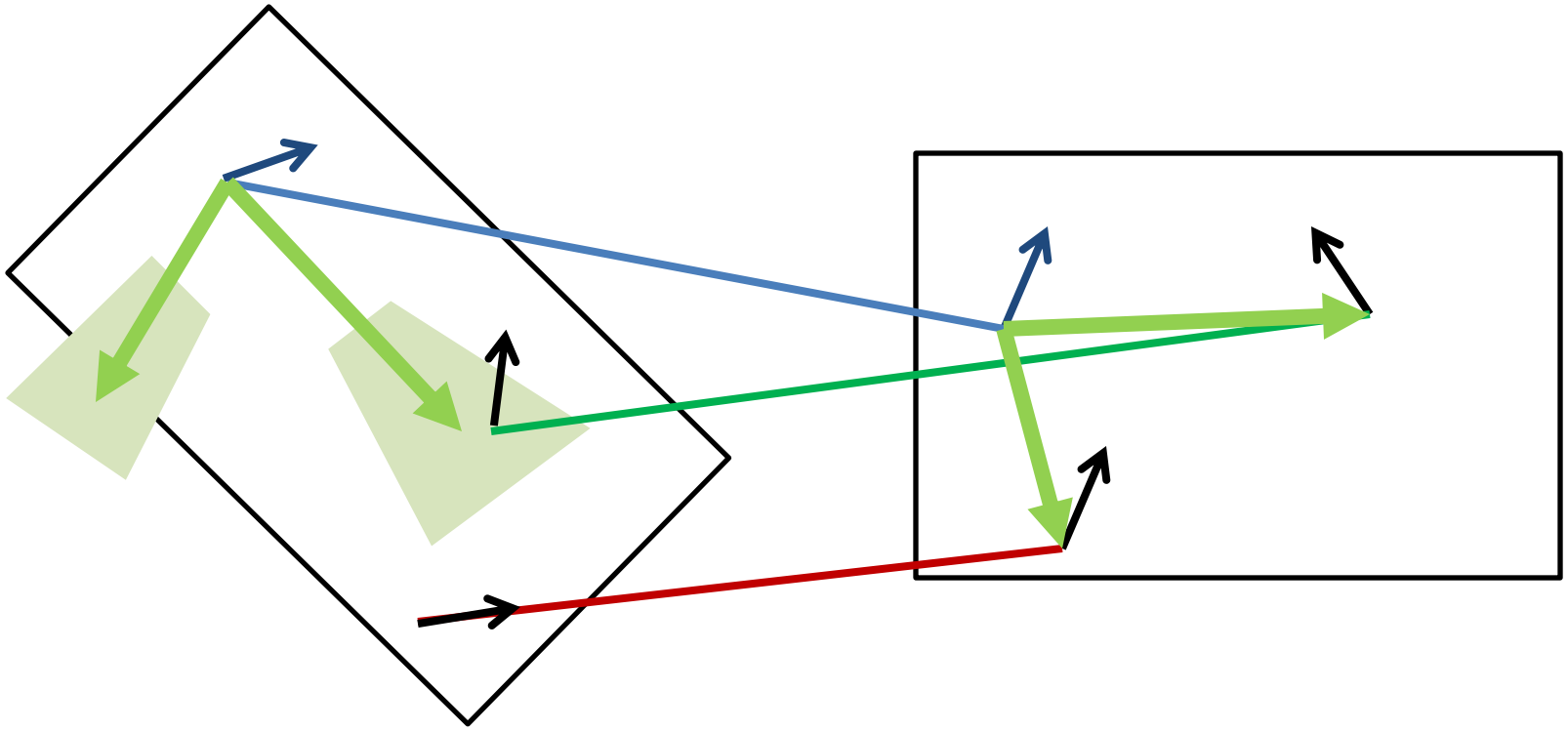
Binary Tree Lookup

- Exploits similarity between HIPs to avoid exhaustive search



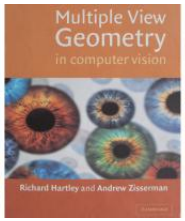
Viewpoint Consistency

- Hough transform on target ID, scale, rotation
- Check vector between matches is reasonable



Comparison with Wagner et al. 2008

Ad



Book


Cars


Map

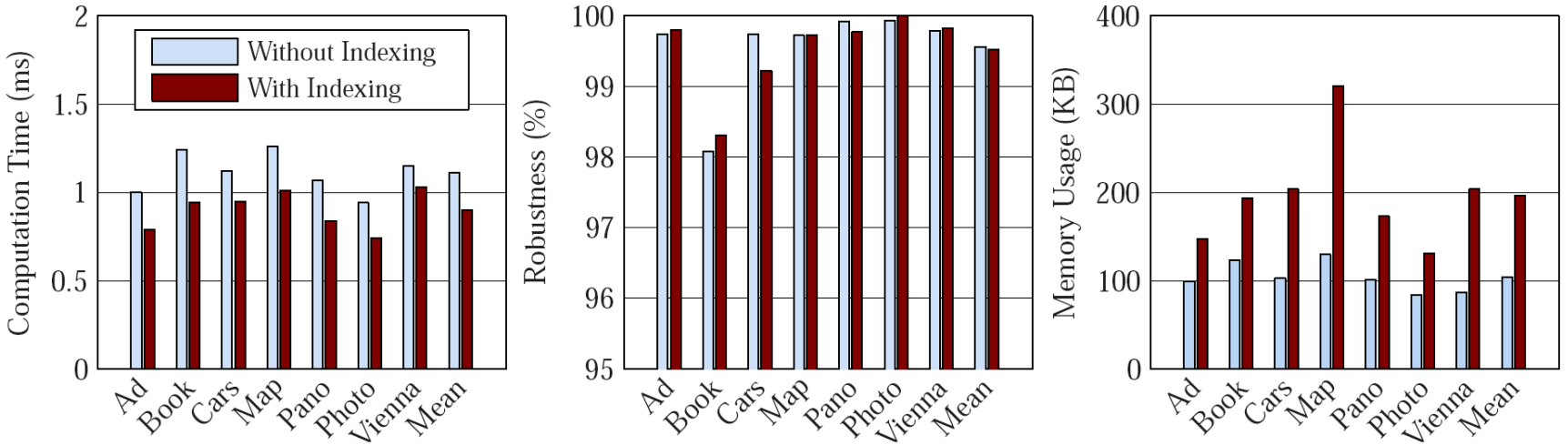

Panorama


Photo


Vienna

- Wagner et al. 2008 tested on 7 sequences
  - Total frame time around 5ms on average
  - Around 96% of frames localised successfully

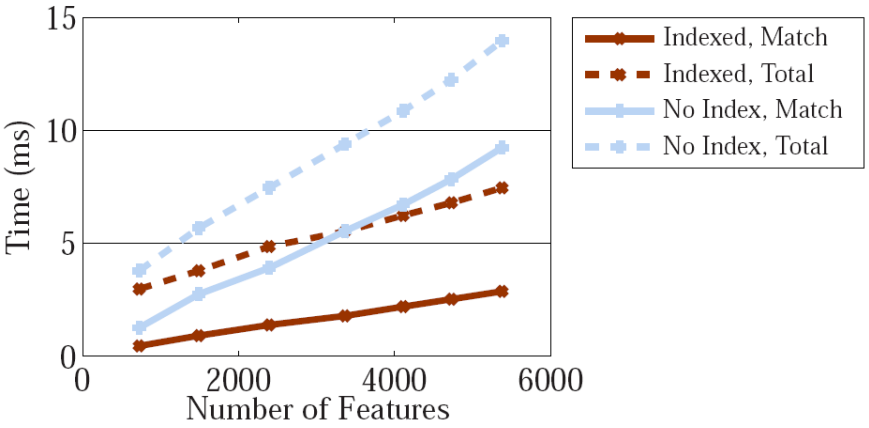
Comparison with Wagner et al. 2008

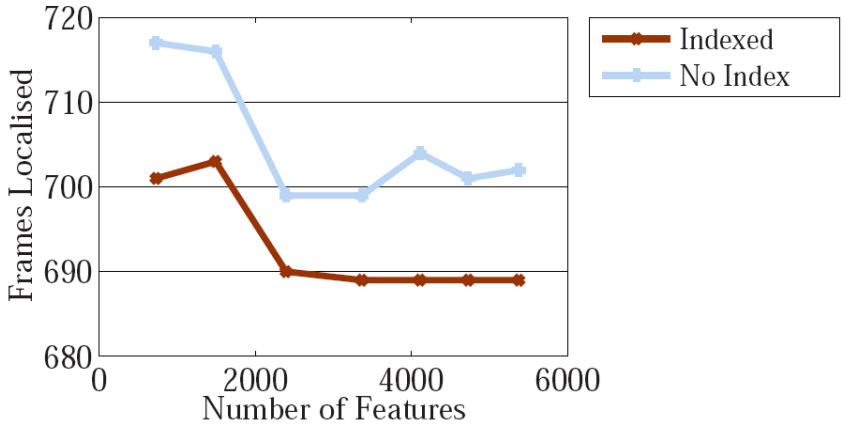


- Over 4x faster
- More robust
- Memory usage 5-10 times lower

Multiple Target Localisation

Multiple Target Localisation





- Under 7.5ms with all 7 targets in database
- Small robustness penalty from
  - Additional matching targets
  - Indexing

Conclusions

- Classification-based matching can provide robust and fast localisation
- HIPs provide a memory and computationally efficient model for features
- Training phase can limit the impact of fast but inaccurate region detection approaches
- Tree-based lookup allows exact classification results without exhaustive comparison