Multi-camera matching of spatio-temporal binary features

Alessio Xompero[1], Oswald Lanz[2], Andrea Cavallaro[1]

Centre for Intelligent Sensing

Queen Mary University of London

Queen Mary University of London[1]
Fondazione Bruno Kessler[2]

CIS centre for intelligent sensing

Queen Mary
University of London

Motivation
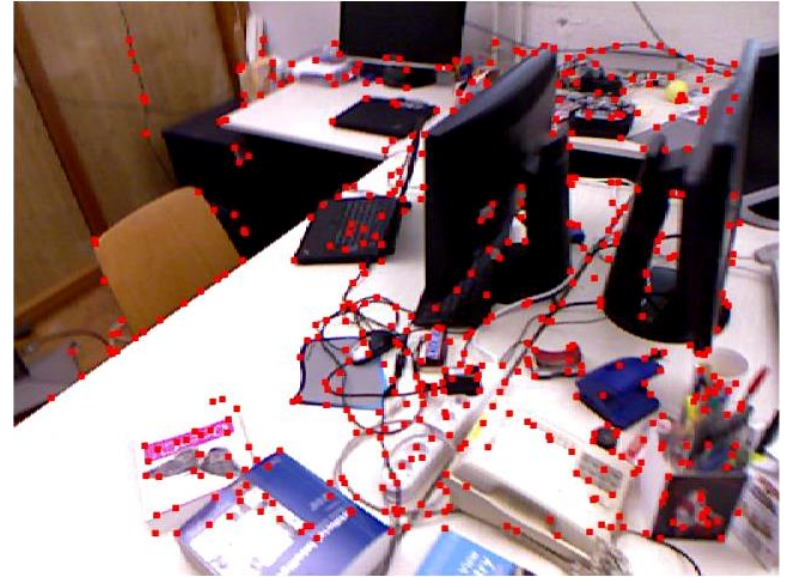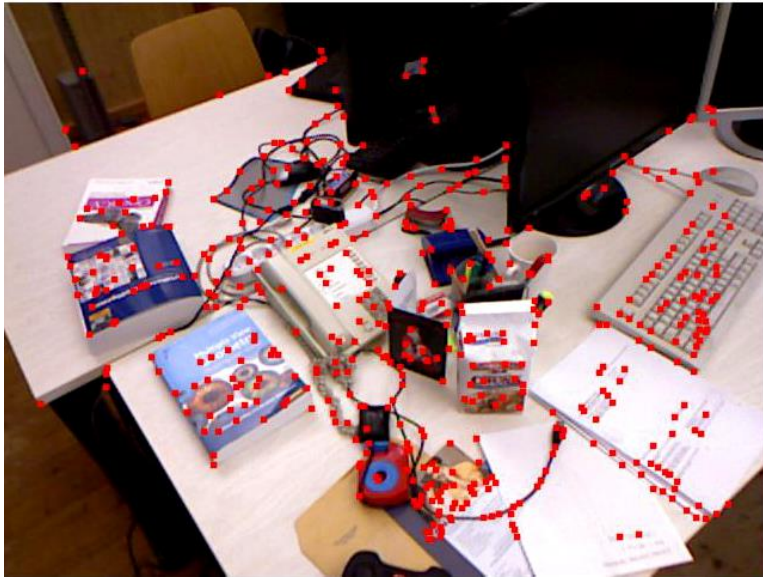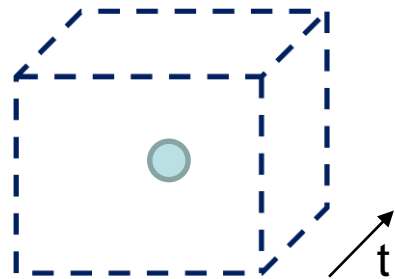

Objective: to match keypoints across moving cameras

severe view changes → performance degradation

CIS centre for intelligent sensing

Queen Mary
University of London

Spatio-temporal features: background

Detection in
space, scale and time
(batch)

+

description of
a fixed volume

t

Tracking of
local image features
(sequential)

t

Cuboid [Dollar2005]
HOG/HOF [Laptev2008]
HOG3D [Klaser2008]
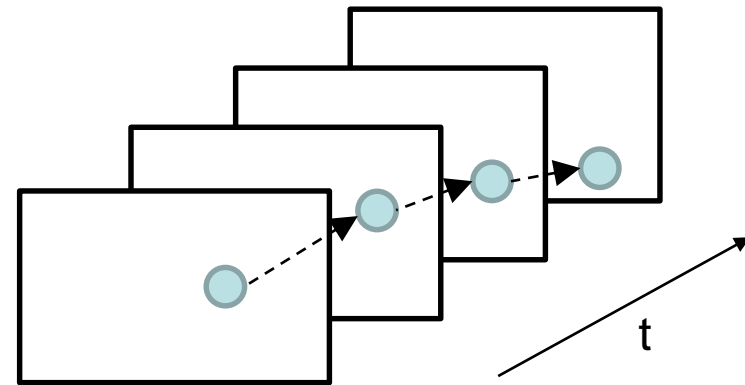
Daisy3D [Trulls2012]
ORB-SLAM [MurArtal2015]

Not designed for matching between cameras

CIS centre for intelligent sensing

Queen Mary
University of London

Localisation and descriptor extraction

ORB sampling pattern
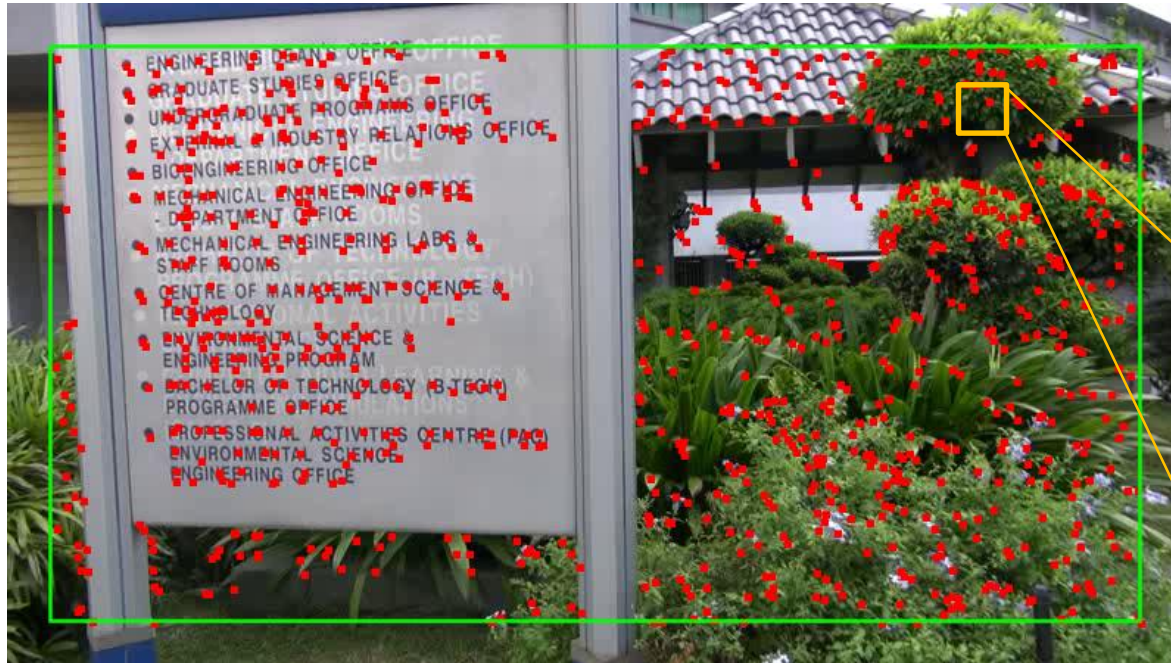
Localisation of FAST keypoints
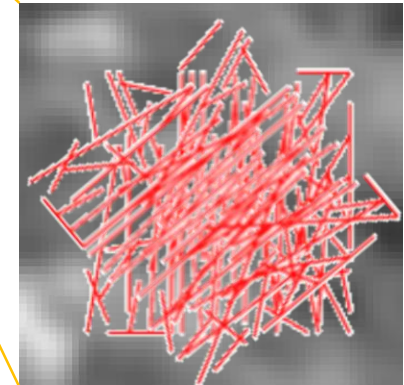
Orientation assignment

Rublee et al., "ORB: an efficient alternative to SIFT and SURF", IEEE Intern. Conf. on Computer Vision, 2011

CIS centre for intelligent sensing

Queen Mary
University of London

Spatio-temporal binary features
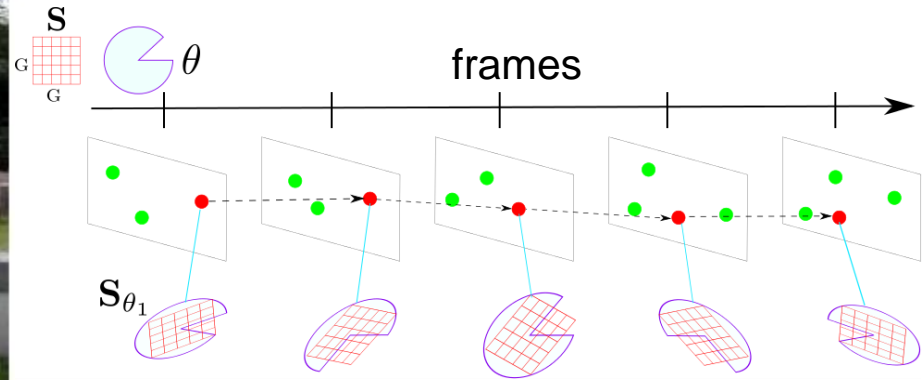
Feature tracking: frame-to-frame matching with nearest neighbour

276°

Queen Mary
University of London

Spatio-temporal binary features

Queen Mary
University of London
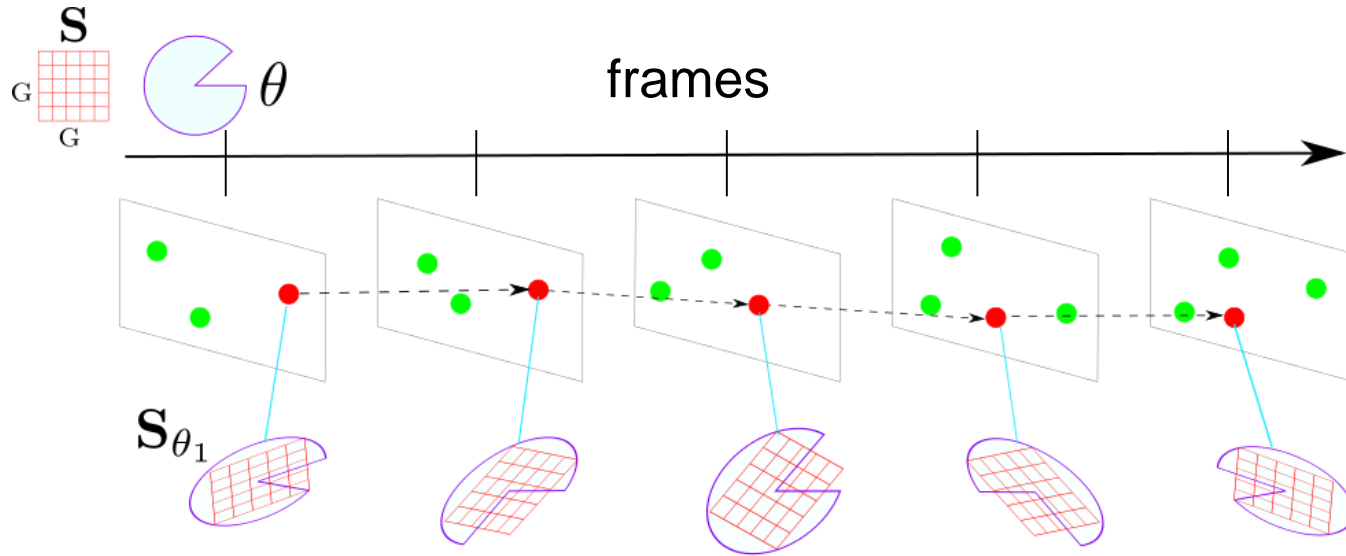
Descriptor reduction (dominant bits)

$$[1 \ 0 \ \ldots \ 1 \ 1] + [0 \ 0 \ \ldots \ 1 \ 1] + [1 \ 0 \ \ldots \ 0 \ 1] + [0 \ 1 \ \ldots \ 0 \ 1] + [0 \ 1 \ \ldots \ 0 \ 0] \ \left(<\right) \rightarrow \mathbf{z}_i$$

frames

thresholding

$\mathbf{S}_{\theta_1}$

Temporally **dominant** bits
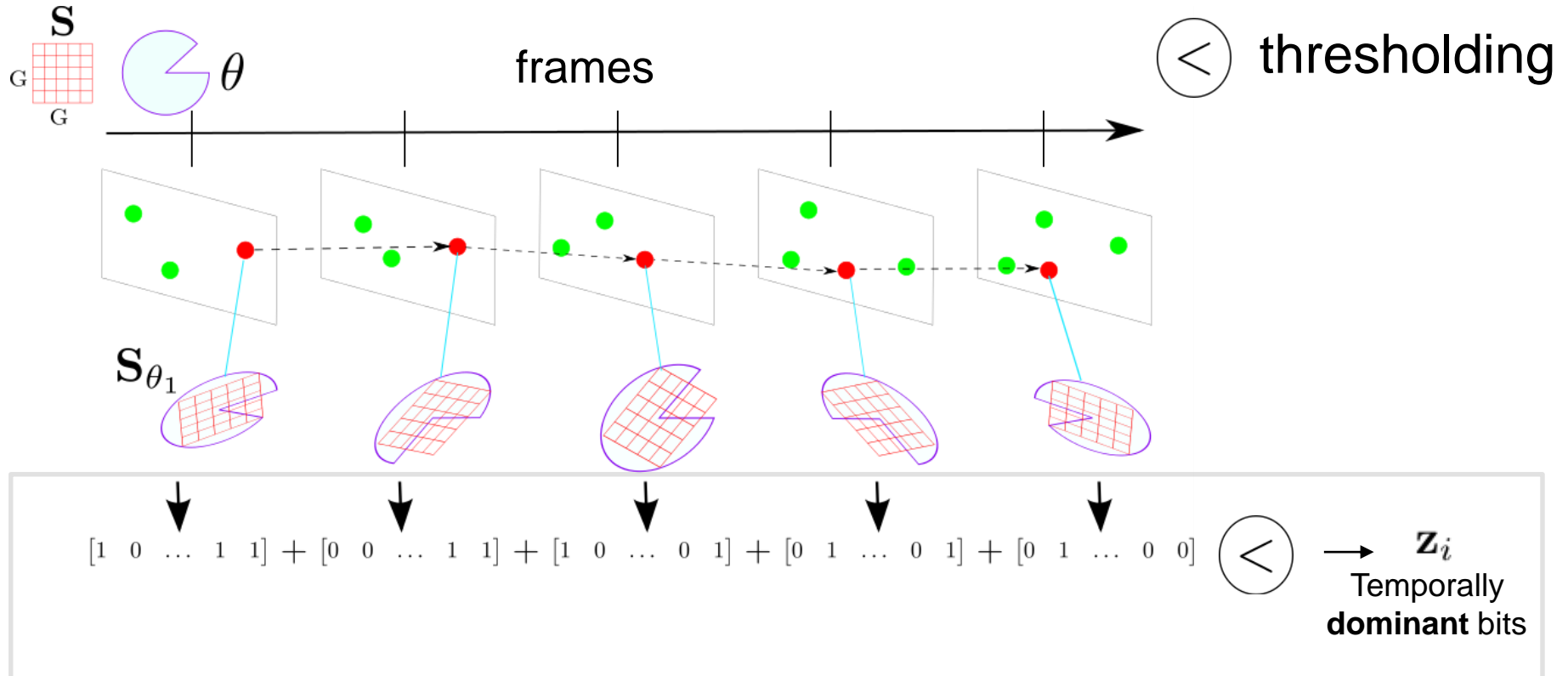
Descriptor reduction (stable bits)
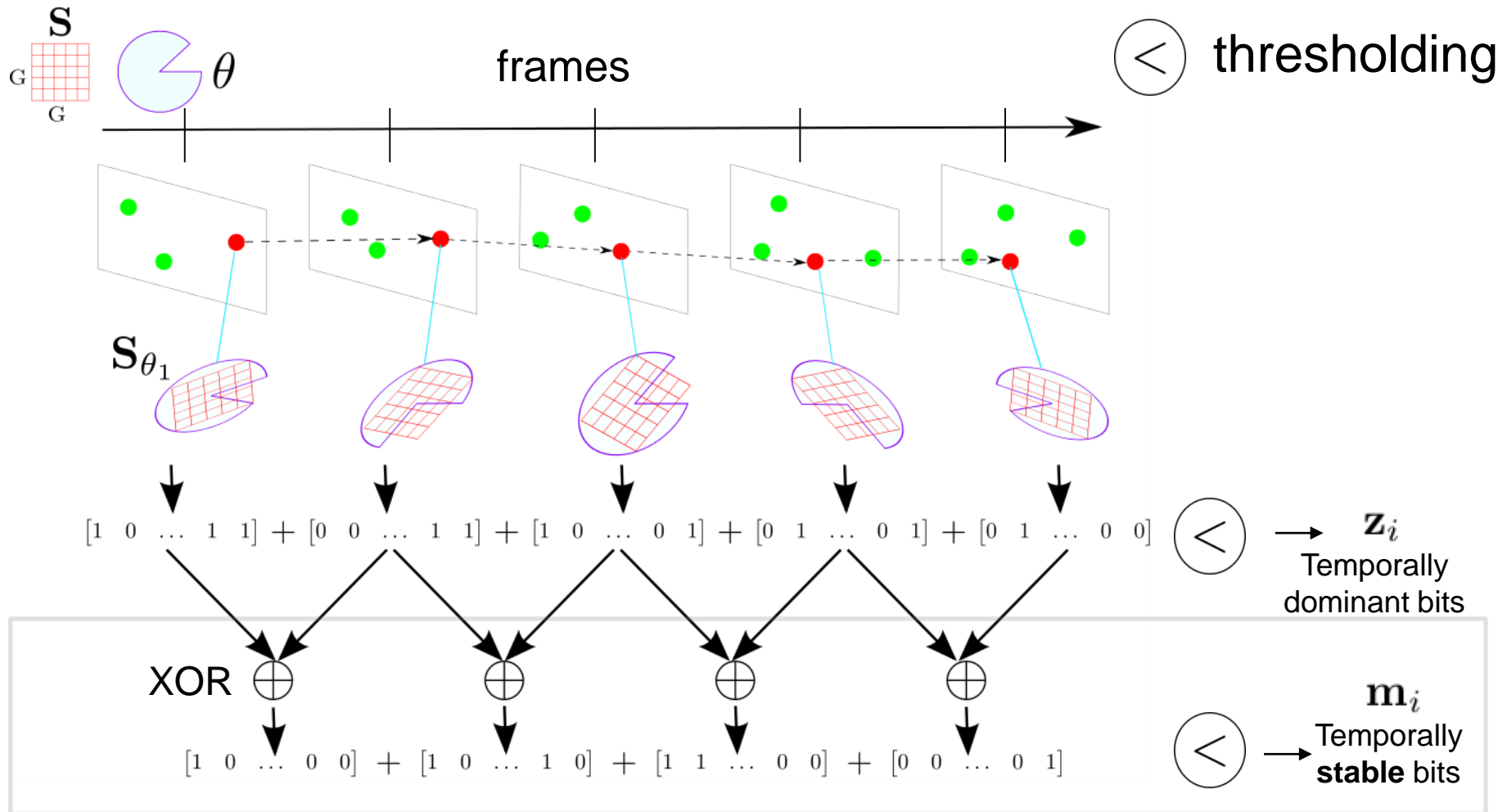
thresholding

frames

$\mathbf{S}$

θ

$\mathbf{S}_{\theta_1}$

$[1 \quad 0 \quad \dots \quad 1 \quad 1] + [0 \quad 0 \quad \dots \quad 1 \quad 1] + [1 \quad 0 \quad \dots \quad 0 \quad 1] + [0 \quad 1 \quad \dots \quad 0 \quad 1] + [0 \quad 1 \quad \dots \quad 0 \quad 0]$

$\mathbf{z}_i$
Temporally dominant bits

XOR $\oplus$

$[1 \quad 0 \quad \dots \quad 0 \quad 0] + [1 \quad 0 \quad \dots \quad 1 \quad 0] + [1 \quad 1 \quad \dots \quad 0 \quad 0] + [0 \quad 0 \quad \dots \quad 0 \quad 1]$

$\mathbf{m}_i$
Temporally **stable** bits

CIS centre for intelligent sensing

Queen Mary
University of London

Matching descriptors between cameras

Spatio-temporal patches
from camera i

Spatio-temporal patches
from camera j
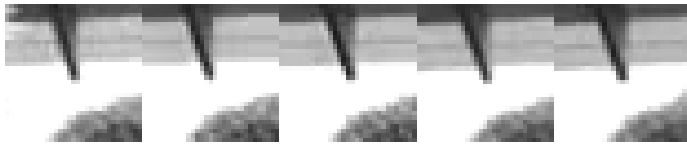
$\mathbf{z}_i$ | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1

$\mathbf{m}_i$ | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0

$\mathbf{z}_j$ | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1

$\mathbf{m}_j$ | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0

CIS centre for intelligent sensing

Queen Mary
University of London

Matching descriptors between cameras

Spatio-temporal patches
from camera i

Spatio-temporal patches
from camera j

XOR

$\mathbf{z}_i$ | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |

$\mathbf{z}_j$ | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |

$\mathbf{m}_i$ | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
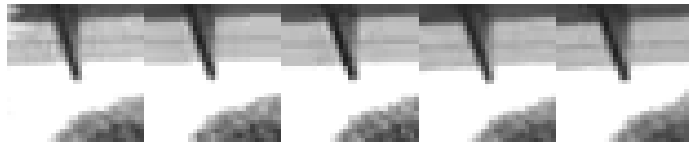
$\mathbf{m}_j$ | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |

| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |

Matching descriptors between cameras

Spatio-temporal patches
from camera i

Spatio-temporal patches
from camera j

XOR
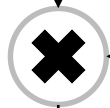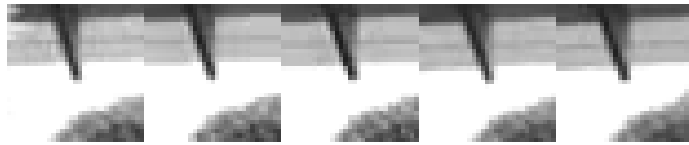
$\mathbf{z}_i$ | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | ⊕ | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | $\mathbf{z}_j$

$\mathbf{m}_i$ | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | $\mathbf{m}_j$

AND ✖ ← | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |

| 0 | | | 0 | 1 | | 1 | 0 |

CIS centre for intelligent sensing

Queen Mary
University of London

Matching descriptors between cameras

Spatio-temporal patches
from camera i

Spatio-temporal patches
from camera j

XOR
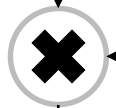
$\mathbf{z}_i$ | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | ⊕ | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | $\mathbf{z}_j$

$\mathbf{m}_i$ | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | $\mathbf{m}_j$

AND ✖ ← | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | → ✖ AND

| 0 | | 0 | 1 | | 1 | 0 |

| 1 | | 0 | 0 | | 0 | 1 |

Queen Mary
University of London

Matching descriptors between cameras

Spatio-temporal patches
from camera i

| 0 | | 0 | 1 | | 1 | 0 |

Spatio-temporal patches
from camera j

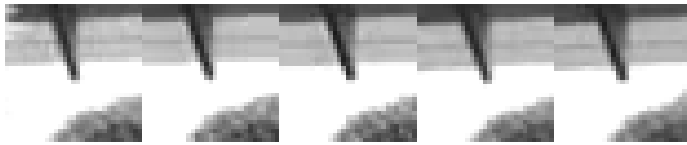| 1 | | 0 | 0 | | 0 | 1 |

CIS centre for intelligent sensing

Queen Mary
University of London
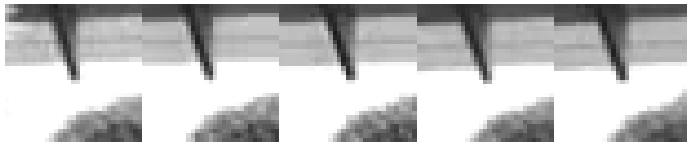
Matching descriptors between cameras

Spatio-temporal patches
from camera *i*

Spatio-temporal patches
from camera *j*

$$h(i,j) = \frac{M_i \langle \mathbf{m}_i, \mathbf{z}_i \oplus \mathbf{z}_j \rangle + M_j \langle \mathbf{m}_j, \mathbf{z}_i \oplus \mathbf{z}_j \rangle}{M_i + M_j}$$

$\langle \cdot, \cdot \rangle$: logical dot product

Weighted linear combination of two masked
Hamming distances [Balntas2017]

CIS centre for intelligent sensing

Queen Mary
University of London

Experimental setup: methods and dataset

- T-DS: temporally dominant + stable bits descriptor
- T-D: temporally dominant bits descriptor
- S-ORB: set of temporally ORB descriptors
- LMED: single ORB with least median dist [MurArtal2015]

R. Mur-Artal et al., "ORB-SLAM: a versatile and accurate monocular SLAM system", IEEE Trans. on Robotics, 2015

CIS centre for intelligent sensing

Queen Mary
University of London

Performance evaluation

- ## Similarity matching
  - nearest neighbor with ratio test (**1-to-1**)
  - threshold-based (**M-to-M**)

- ## Dissimilarity measures
  - T-DS $\rightarrow$ weighted Hamming distance
  - T-D, LMED $\rightarrow$ Hamming distance
  - S-ORB $\rightarrow$ set2set min dist

- ## Performance measures
  - P: precision $\quad\quad\quad$ *# correct matches / # matches*
  - R: recall $\quad$ *# correct matches / # true correspondences*
  - F-score $\quad\quad\quad\quad$ 2 (P x R) / (P + R)

CIS centre for intelligent sensing

Queen Mary
University of London

Feature tracking results

Short

Long

Trajectory length

Queen Mary
University of London
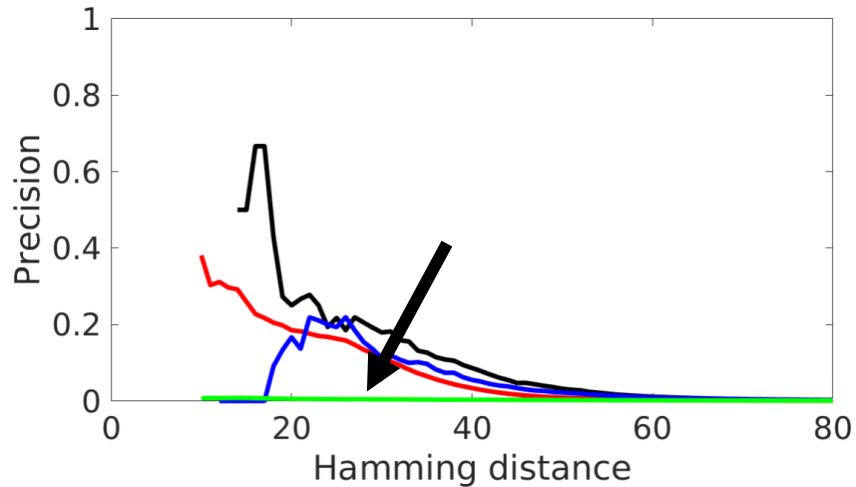
Matching results

T-DS: a lot of wrong matches

Matching results
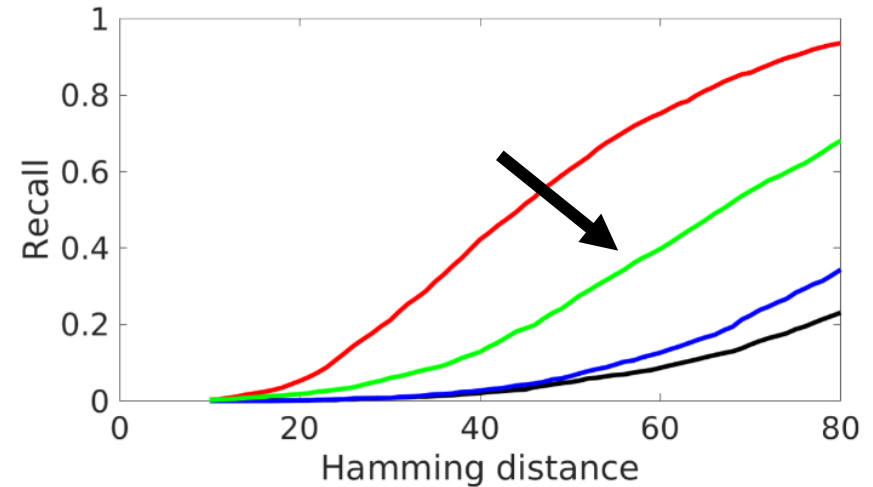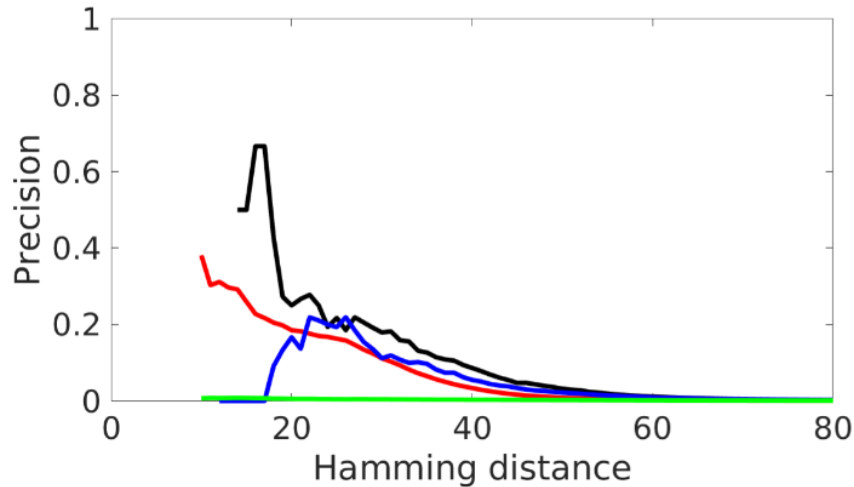
T-DS: a lot of wrong matches, but good recall

CIS centre for intelligent sensing

Queen Mary
University of London
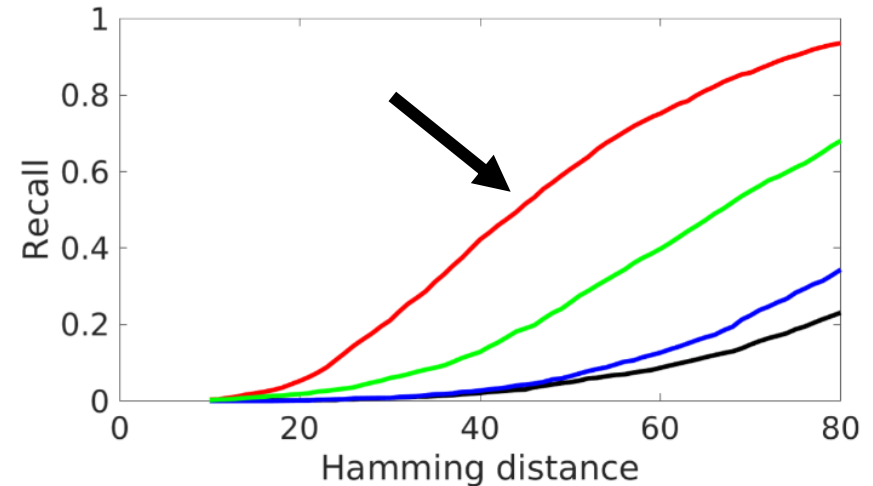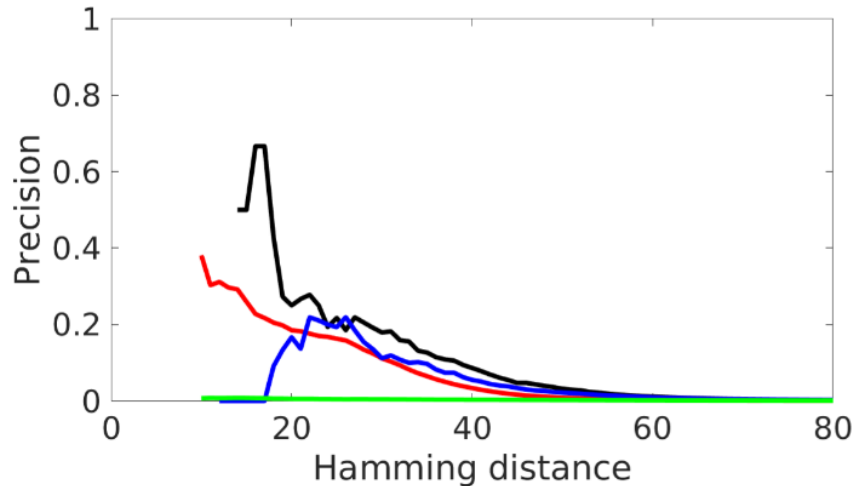
Matching results

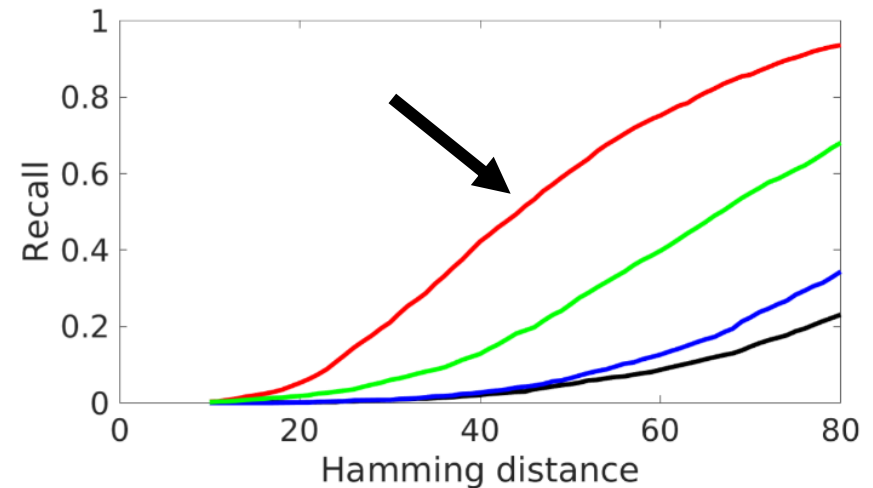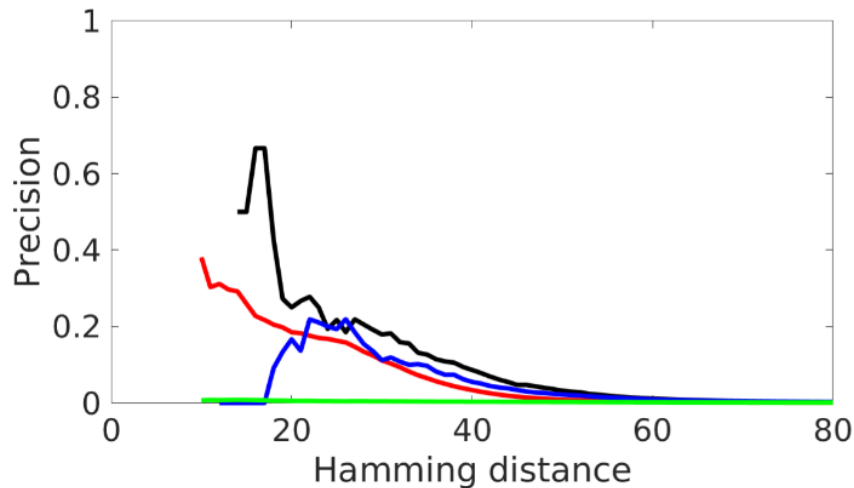
T-DS: a lot of wrong matches, but good recall
S-ORB: best, but computationally expensive

Matching results

T-DS: a lot of wrong matches, but good recall

S-ORB: best, but computationally expensive

$\downarrow$

extra cost of $O(L_i L_j)$ for each pair

L: trajectory length

CIS centre for intelligent sensing

Queen Mary
University of London

Conclusions

- New spatio-temporal binary descriptor (T-DS)
  - temporally dominant and stable bits from sets of ORB (S-ORB)
  - outperforms LMED [MurArtal2015]

- However, the set of ORB descriptors (S-ORB)
  - outperforms all other approaches
  - computationally expensive to match

- Future work
  - reduction that preserves matching efficiency
  - to consider scale

CIS centre for intelligent sensing

Queen Mary
University of London