# Robustness of 3D Point Positions to Camera Baselines in Markerless AR Systems

**Deepak Dwarakanath, Carsten Griwodz & Pål Halvorsen**

[ simula ]

UiO : University of Oslo

*ACM MMSYS 2016 – Austria*

*10-13 May 2016*

# AR Application

- **POPART project**
- **Quality of observer's position depends on accuracy of camera pose**
- **Markerless camera pose estimation is more challenging**

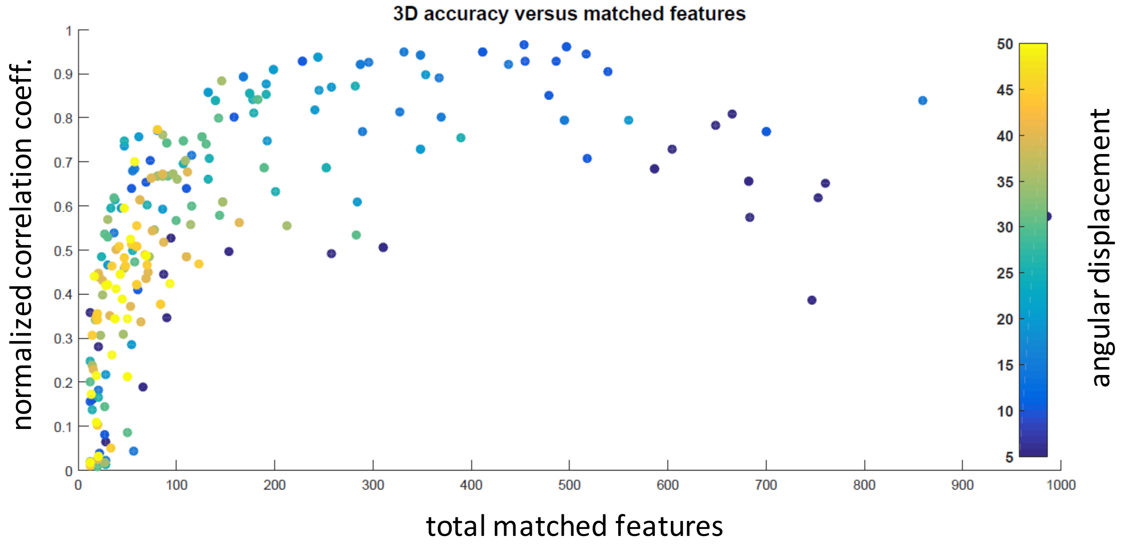

Augmented preview of the film set

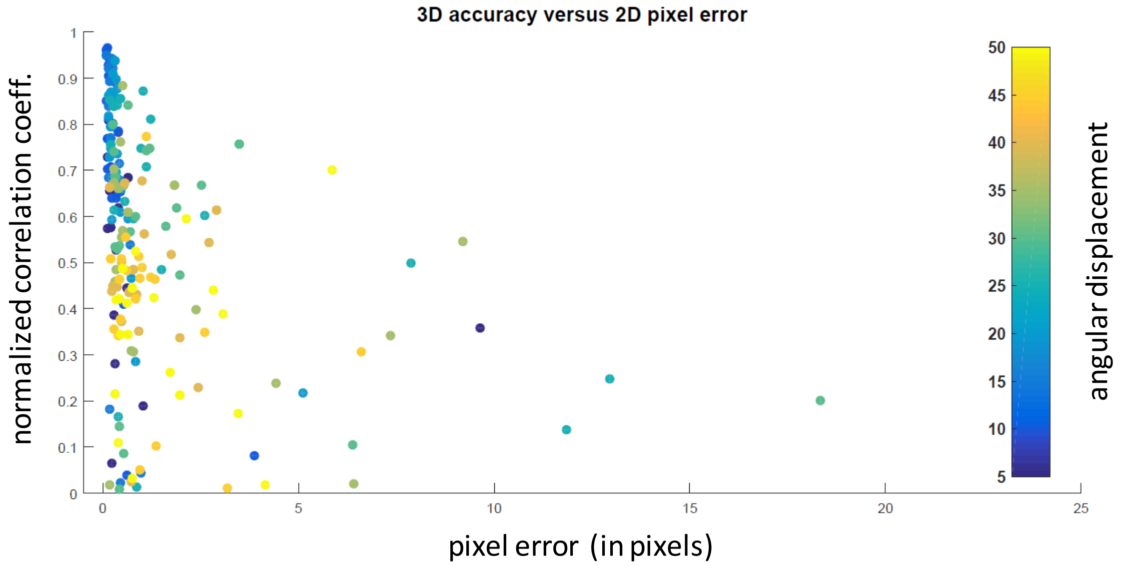
# Commonly Known

Feature based calibration – camera pose estimated using sparse feature points detected in the images

- **if the number of feature points is larger, the camera pose estimation is better**

- **minimizing the 2D error between the matched points yields better camera pose estimation**



# Scope

- **Accuracy of camera pose based on state-of-art feature detectors and descriptors cannot be guaranteed with variation in camera baselines**
- **This paper explores the magnitude of such inaccuracy**
- **Evaluation of several state-of-art feature extractors**
- **Helps system builders to understand the operational limits and make better choices to design multimedia system**
- **Helps also to determine camera density around a scene**

# Related evaluation work

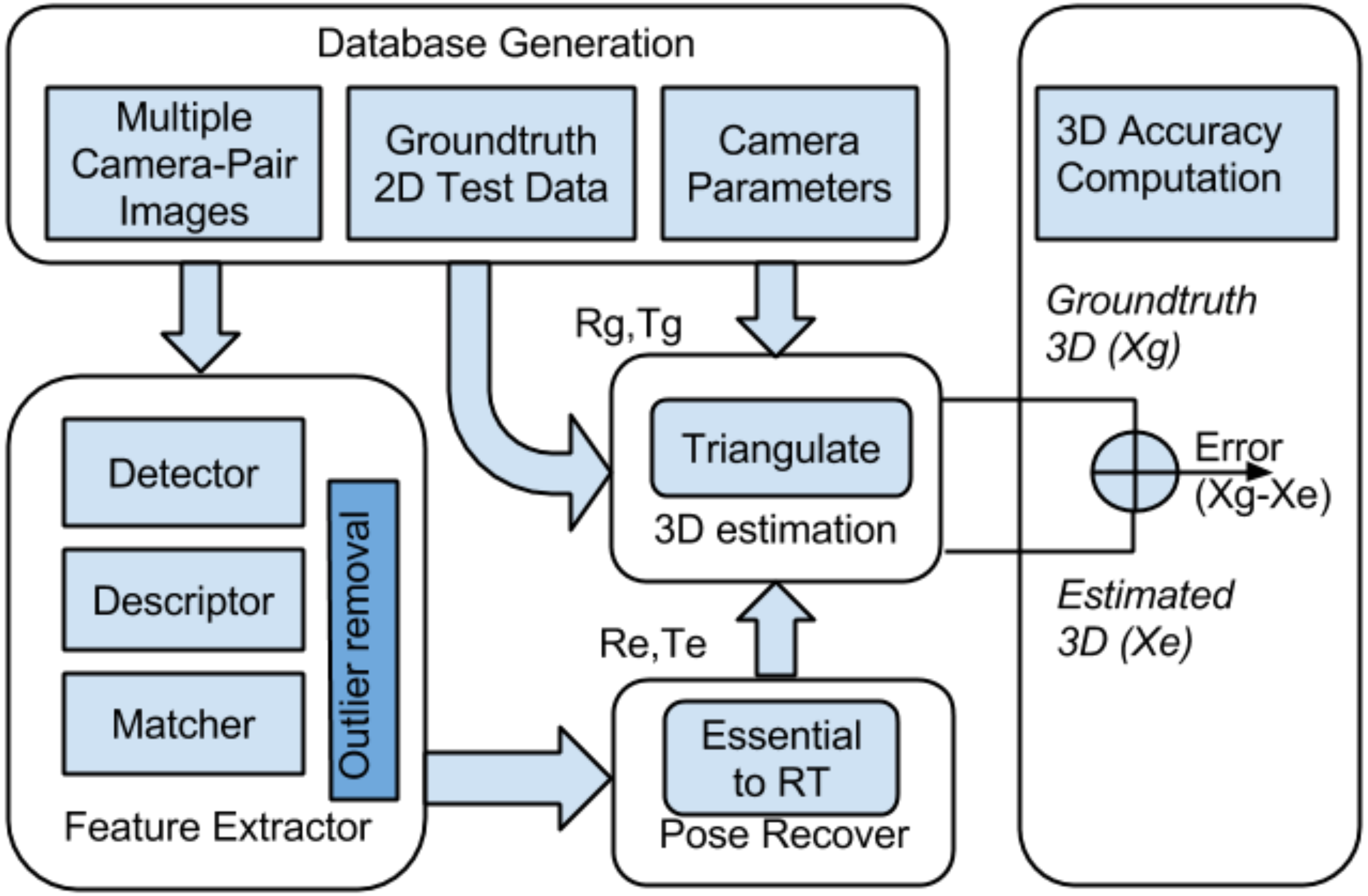
**Focus on:**

- Correctness of the feature matches
- Repeatability of features
- Reprojection error in 2D
- Limited candidates for evaluation

**In this paper:**

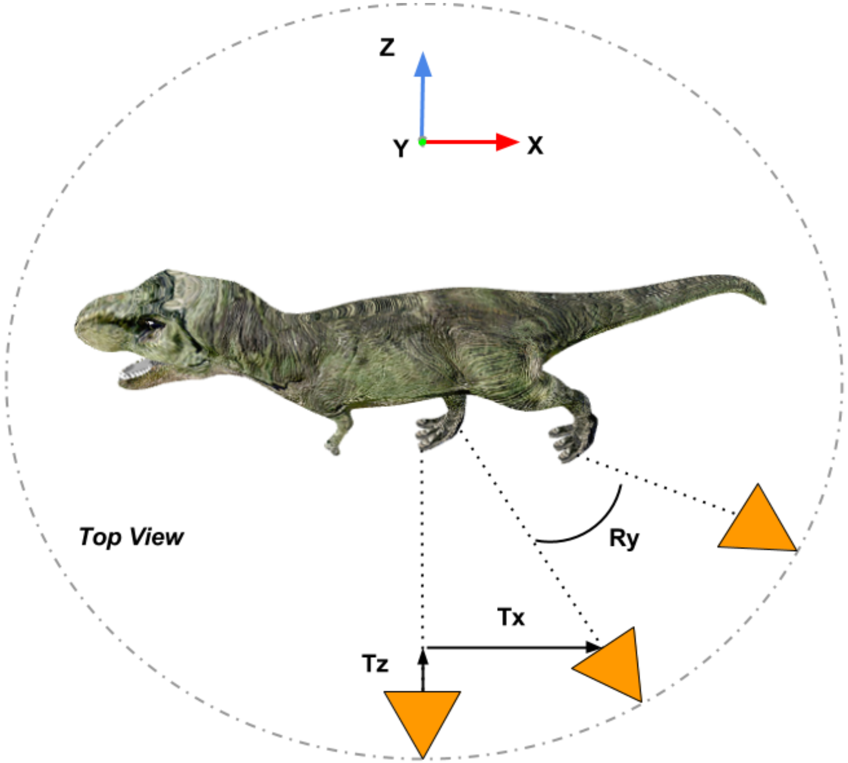
- Accuracy measured in 3D space metrics – relates to the problem directly
- Several well-known feature extractors
- Obtain operational limits for all tested feature extractors (under specific conditions)

# Experimental - Overview



# Experimental - Datasets

- **Turn-table configuration to keep the object size / distance constant**
  - Camera centers 500 units from model's geometric center in model coordinate system
- **450 stereo pairs from 9 known models are captured at 60x600 resolution**
- **Known values**
  - 3D mesh vertices
  - Corresponding 2D pixel positions on stereo images
  - Camera focal length and principal axes
  - Cameras' relative rotation and translation



# Experimental - Feature Extractors

- **26 feature extractor combinations using several detectors and descriptors**
  - Detectors - MSER, STAR, FAST
  - Descriptors - BRIEF, FREAK
  - Detectors and Descriptors - SIFT, SURF, BRISK, KAZE, AKAZE and ORB

- **Brute force matching**
- **RANSAC – outlier removal**

# Experimental - Pose Estimation

**Based on feature matching points in a stereo pair**

- Essential matrix (E) is estimated
- Using SVD, E=[T]R
- Cheirality constraint to select optimal solution
- Hence,
  - Relative Rotation (R)
  - Relative Translation (T)

  are estimated
- All measurements are in model coordinates and in model units

# Experiments - 3D Estimation and Accuracy Computation

- **Using feature-matched points + camera pose, *triangulation* is performed**
- **Resulting sparse 3D points are compared with ground truth points**
- **Computation in 3D space**
  - Normalized Correlation Co-efficient error
    (used for comparative study)
  - Mean Squared Error
    (used for design recommendation along with some penalties)

# Results - overview

- **Evaluation pipeline**
  - 2D pixel error

    Expressed as Sampson Error – second order approximation of geometric error

$$P_{error} = \sum_{i=1}^{N_p} \frac{(x_i' F x_i)^2}{(F x_i)_1^2 + (F x_i)_2^2 + (F^T x\prime_i)_1^2 + (F^T x\prime_i)_2^2}$$

  - Camera pose error

    Comparing estimated rotation and translation with known values (in 3 axes)

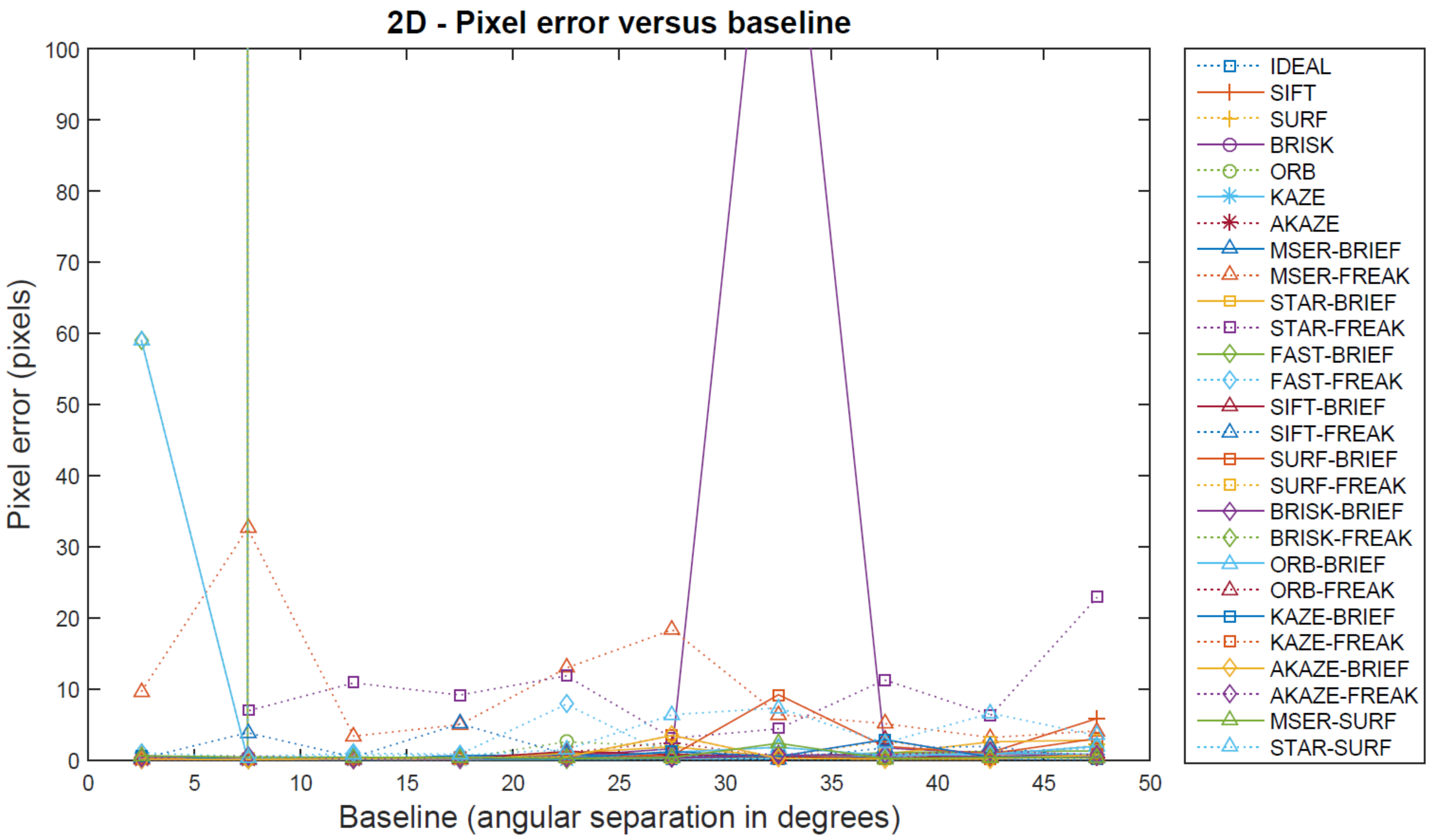
  - 3D estimation error

    Determines performance evaluation and helps in design recommendation

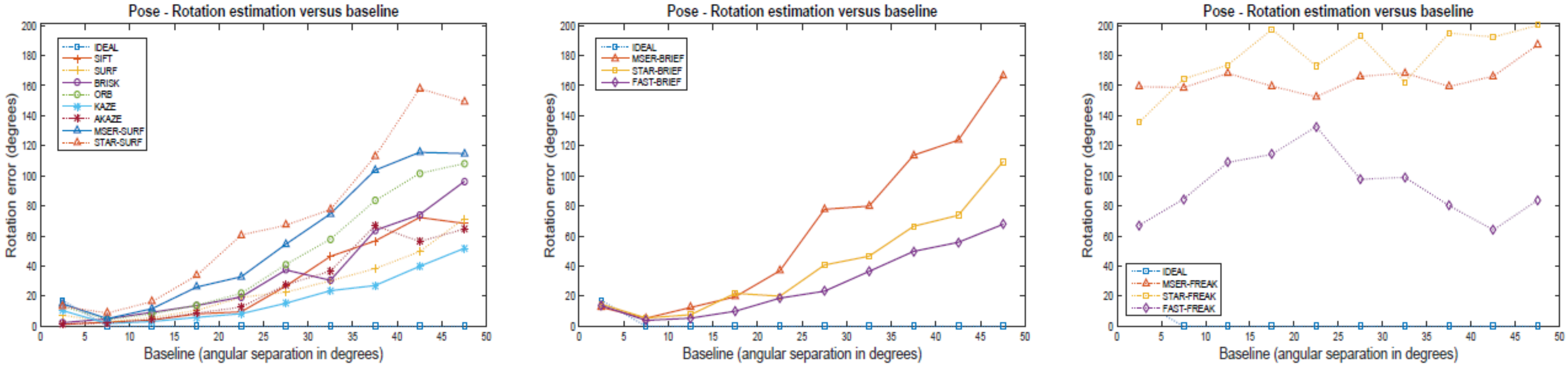
# Results – 2D pixel error

- **Pixel errors in 2D for matched features points <u>are fairly low</u> for varied baselines**

- **This does not guarantee a high 3D accuracy**



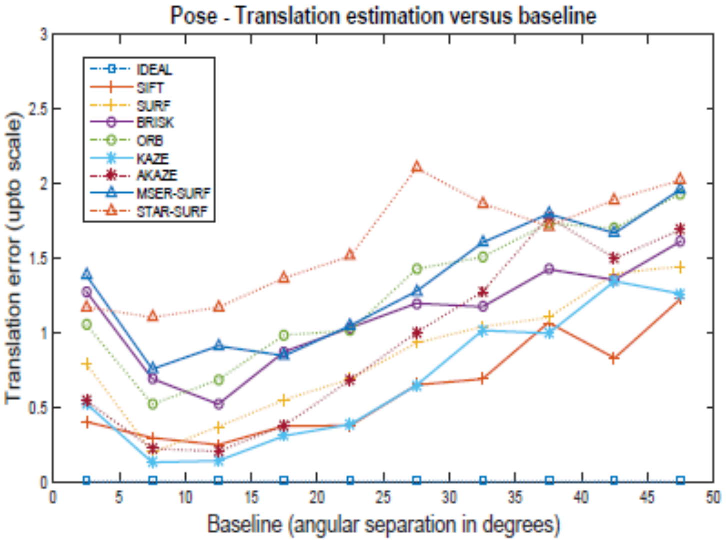
# Results – Rotational Error



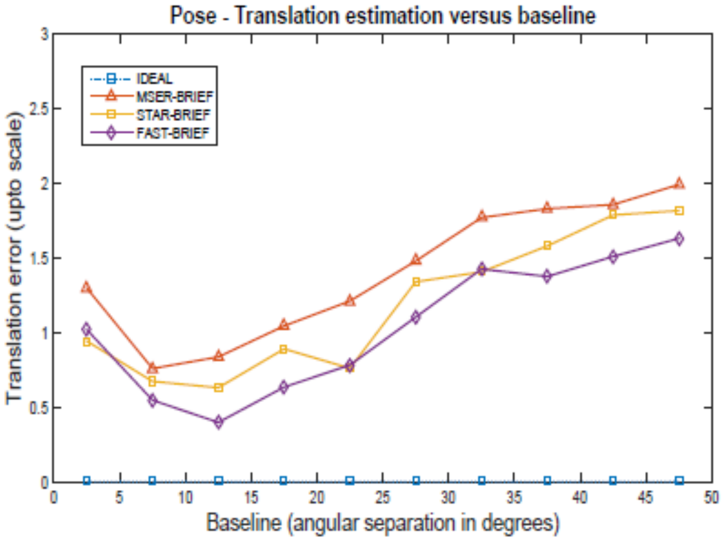

(a) Own Descriptors (b) Brief Descriptor (c) Freak Descriptor

- **Rotational Error increases with the increase in camera baseline (a) & (b)**
- **Although baseline refers to Ry, estimation of Rx,Rz results in non-zeros**
- **FREAK descriptor performs poorly**

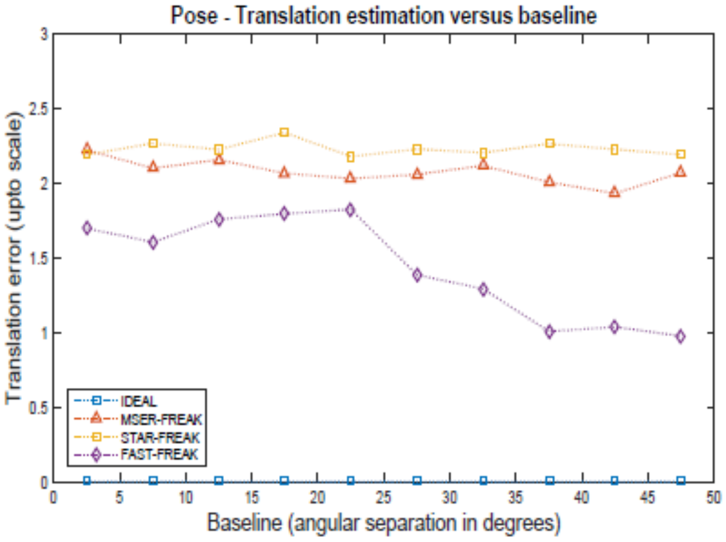
# Results – Translational Error




(a) Own Descriptors




(b) Brief Descriptor

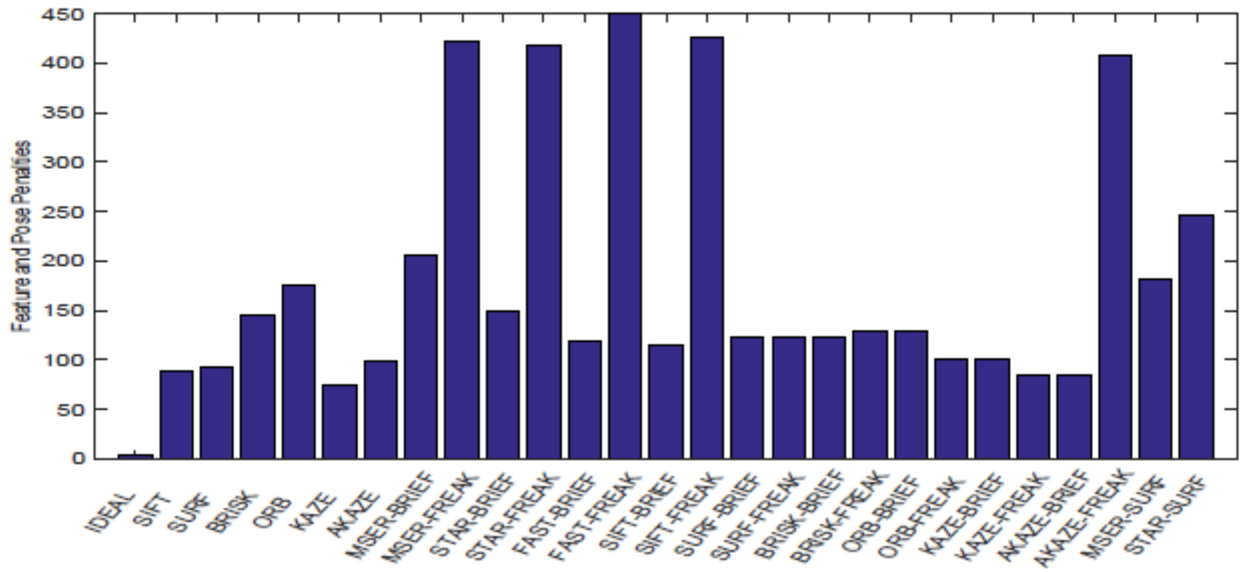



(c) Freak Descriptor

- **Translational Error increases with the increase in camera baseline (a) & (b)**
- **FREAK descriptor performs poorly**

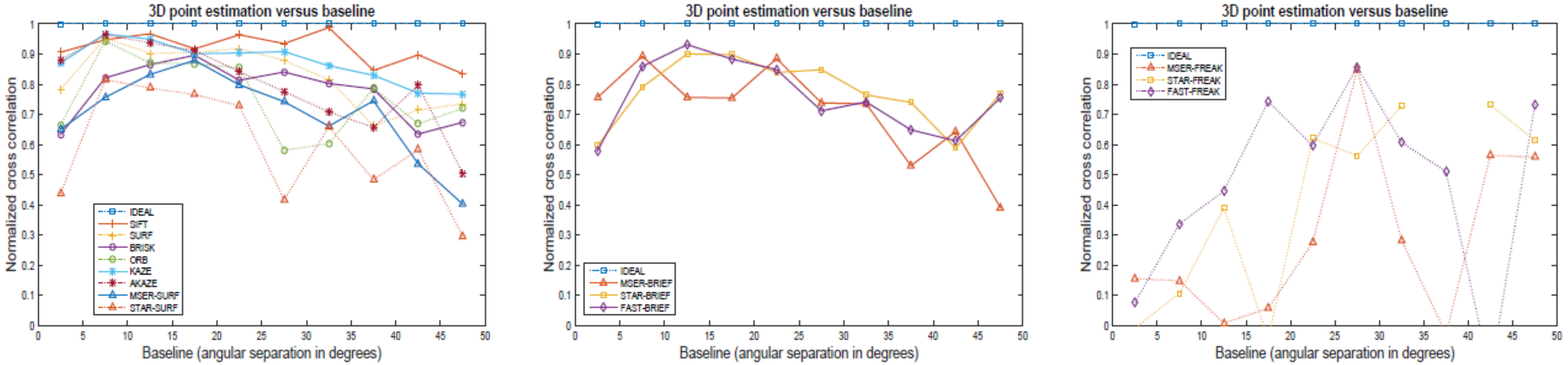
# Results – camera pose error

- **Possible reasons for camera pose error**
  - Wrong matches even after outlier removal – wrong essential matrix
  - Feature point matches confined to an area – gives a wrong rotational estimation in terms of perspective
- **Penalities occur when:**
  - Translation error is more than unity
  - Rotation is more than 90 degrees
  - No matches were found



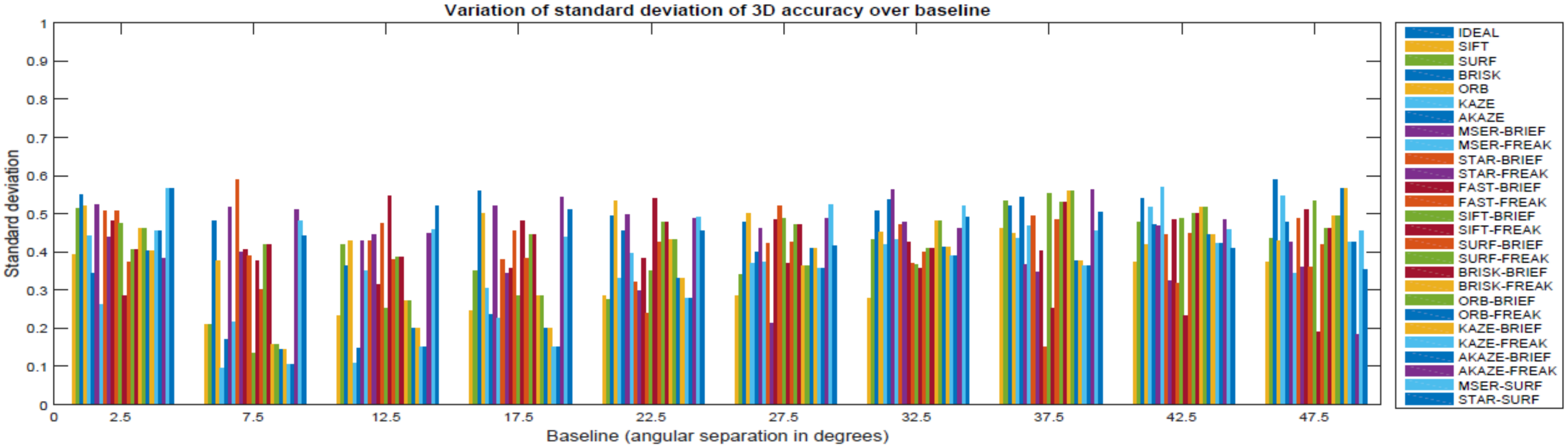
# Result - 3D error

- **Mean**

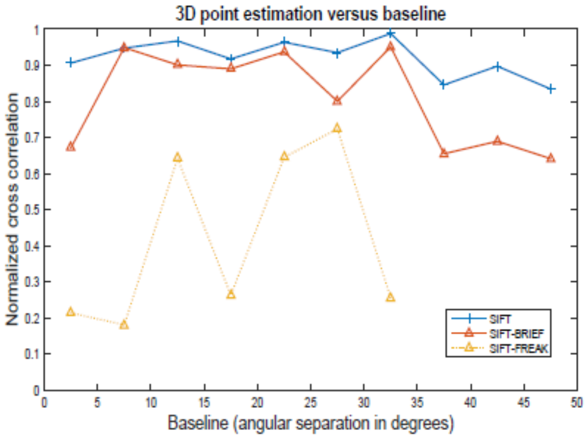



(a) Own Descriptors (b) Brief Descriptor (c) Freak Descriptor

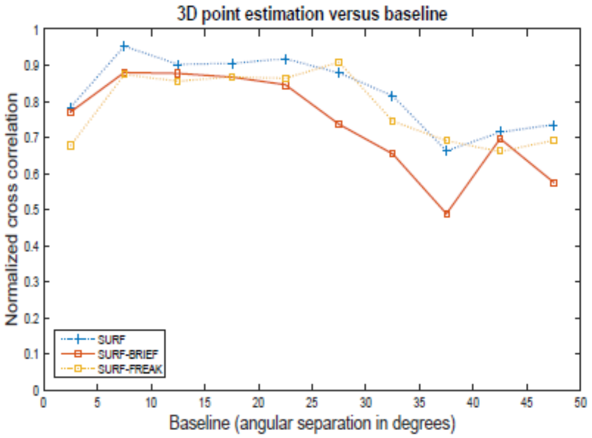
- **Standard Deviation**



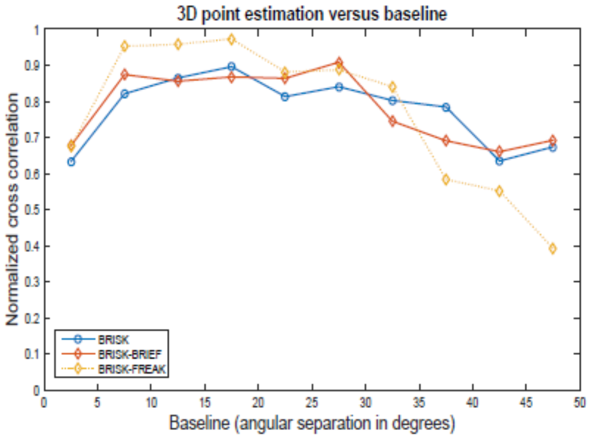
# Results – 3D error (More combinations)



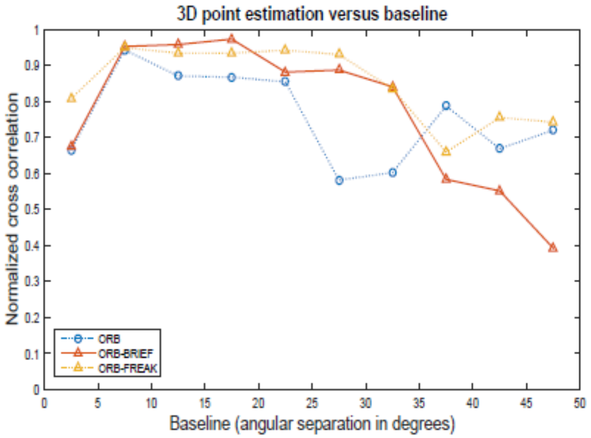
(a) SIFT



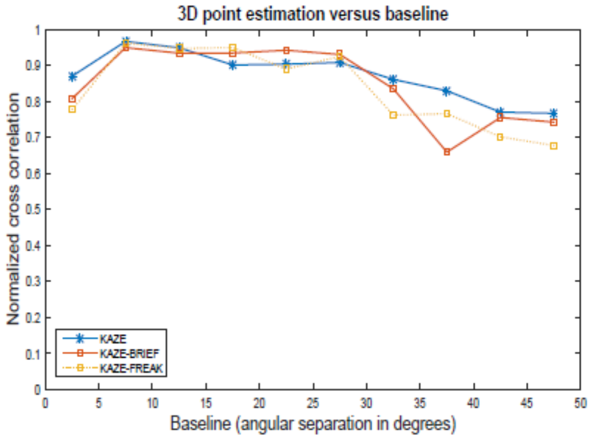
(b) SURF



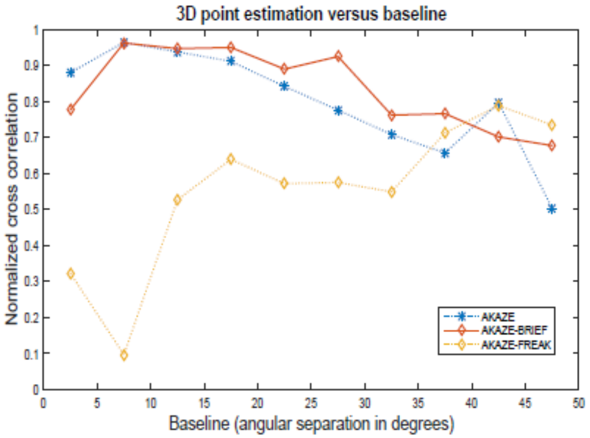
(c) BRISK



(d) ORB



(e) KAZE



(f) AKAZE

# Performance Evaluation

| Baseline (< 5) deg | Baseline (5 – 30) deg | Baseline (30 - 50) deg |
|---|---|---|
| SIFT, KAZE, AKAZE – good performers<br><br>Rotation – translation ambiguity exists | SIFT, SURF, KAZE with their own descriptors<br><br>BRIEF descriptor with all detectors except MSER, STAR, FAST<br><br>FREAK descriptor with SURF; BRISK ORB and KAZE | SIFT and KAZE perform better than any other |

- **NCC – Normalized Correlation error – only a relative measure for comparison**
- **However this is not sufficient to choose a feature extractor**

# Design recommendations

- **We consider MSE of the deviation is 3D reconstructed points**
- **We incorporate the penalties incurred by the feature extractors over all models in a range of baselines. This is presented as reliability of the feature**

| **Feature extractors** | **Baseline($5^o - 20^o$)** *Rotation* [degrees] | *Position* [model] | **Baseline($20^o - 35^o$)** *Rotation* [degrees] | *Position* [model] | **Baseline($35^o - 50^o$)** *Rotation* [degrees] | *Position* [model] | **Time** [seconds] | **Relia- -bility** [per-cent] |
|---|---|---|---|---|---|---|---|---|
| | mean(deviance) | | mean(deviance) | | mean(deviance) | | | |
| IDEAL | 0.00 (0.00) | 0.00 (0.00) | 0.00 (0.00) | 0.00 (0.00) | 0.00 (0.00) | 0.00 (0.00) | 0.00 | 99.11 |
| SIFT | 13.09 (7.17) | 8.23 (1.96) | 2.14 (1.06) | 2.83 (2.64) | 2.64 (0.56) | 4.22 (1.11) | 17.34 | 80.22 |
| SURF | 15.58 (5.94) | 12.04 (2.26) | 5.59 (0.64) | 6.27 (0.30) | 3.63 (0.89) | 5.33 (0.80) | 5.47 | 79.56 |
| BRISK | 20.21 (8.94) | 25.31 (13.48) | 6.43 (2.41) | 18.73 (13.03) | 3.82 (0.52) | 88.69 (141.00) | 1.75 | 67.56 |
| ORB | 21.04 (9.31) | 9.41 (0.47) | 8.29 (1.54) | 33.30 (41.34) | 3.93 (0.05) | 9.22 (6.91) | 0.85 | 61.11 |
| KAZE | 12.12 (3.84) | 7.76 (2.23) | 4.78 (1.25) | 6.34 (1.64) | 2.92 (0.26) | 7.27 (2.13) | 27.67 | 83.56 |
| AKAZE | 11.68 (2.76) | 6.91 (3.97) | 7.51 (0.74) | 12.36 (5.03) | 4.61 (1.22) | 14.48 (13.06) | 4.96 | 78.00 |
| MSER-SURF | 19.95 (11.36) | 261.94 (422.71) | 8.12 (1.33) | 20.55 (10.17) | 5.10 (0.19) | 16.09 (12.27) | 7.55 | 59.78 |
| STAR-SURF | 29.67 (12.08) | 53.51 (47.55) | 11.34 (3.75) | 16.08 (7.11) | 7.08 (0.98) | 22.15 (8.20) | 0.75 | 45.33 |

# Conclusion

- **SIFT and KAZE seem to be promising in terms robustness over large baselines**
- **Low pixel error in matched features <u>does not guarantee</u> a good 3D accuracy; especially with variation in the camera baseline**
- **26 feature combinations over 50 camera baselines were studied**
- **Design recommendation**
  - To select feature extractor based on acceptable accuracy, execution time and reliability
  - To design the camera density to capture a scene for a given quality of service

Thank you