Multi-objective Visual Odometry

Hsiang-Jen (Johnny) Chien and Reinhard Klette

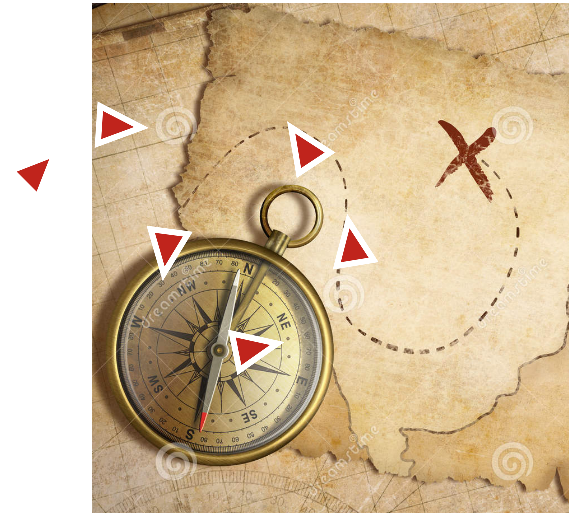
Centre for Robotics & Vision

Dept. of Electronic and Electric Engineering

School of Engineering, Computer, and Mathematical Sciences

Auckland University of Technology (AUT), Auckland, New Zealand

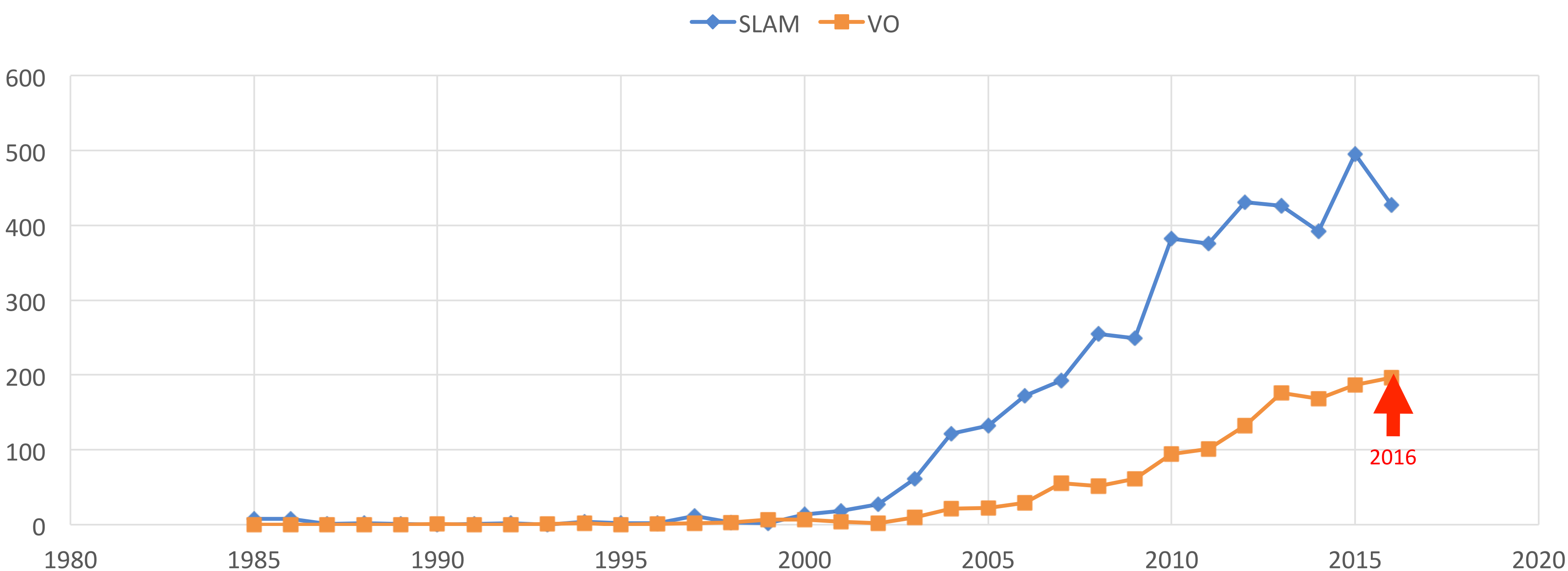
Visual Odometry



- Subsequently solve a system’s egomotion **ONLY** from two consequently taken image frames
- Current position of the system is determined by concatenating a series of previously solved poses
  - known as dead reckoning in terms of navigation
  - “dead” derived from deduced, or ded
- Related to *simultaneously locating and mapping* (SLAM) and *structure from motion* (SFM)

Trend

NUMBER OF PUBLICATIONS PER YEAR



Source: ELSEVIER SCOPUS

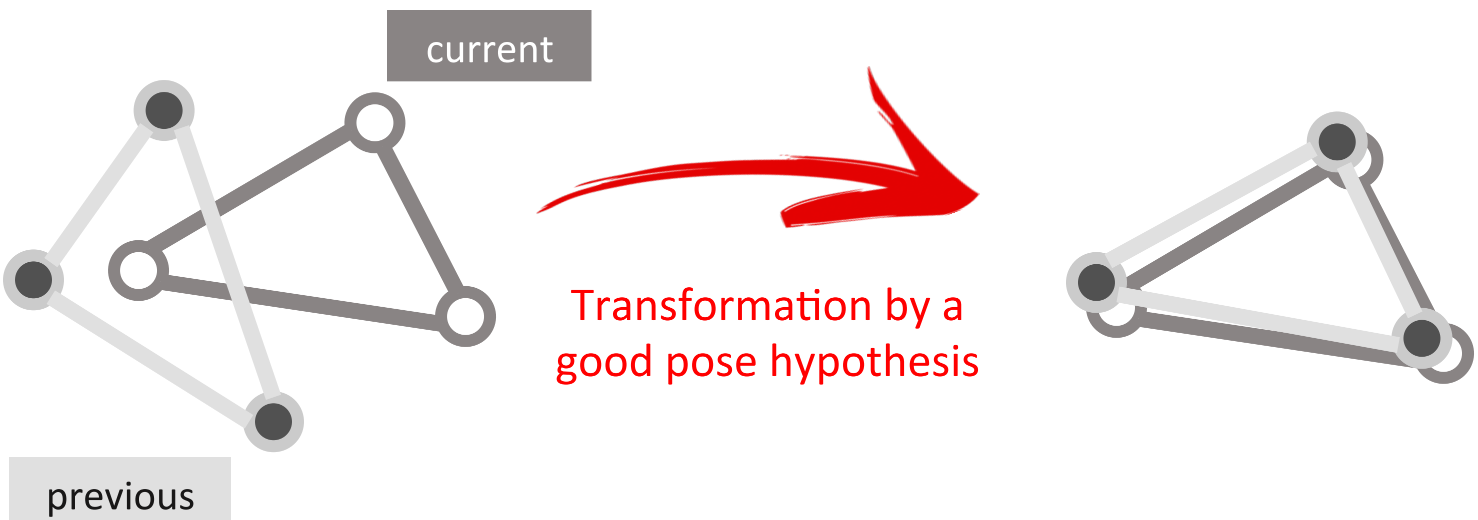
Two alternatives

- Indirect methods
  (feature-based)
  - Transform image pixels into a crafted **feature space**
  - Matching is performed before egomotion estimation
  - Use sparse key points
  - Faster and dominating VO/ SLAM for decades

- Direct methods
  (feature-free)
  - Use **pixel intensities** directly
  - Matching happens simultaneously during estimation
  - Use dense, semi-dense, or sparse pixels
  - Slow but becoming popular due to advances in parallel computing

Alignment problem

- Both alternatives treat pose estimation as an alignment problem
- Rational: the observed data in the current frame should be aligned well to the one transformed from the previous frame using a good pose estimate



Example for a test sequence on KITTI

Generated trajectory by proposed method

Comparison with given ground truth defines *drift* per frame

For test sequences, see

www.cvlibs.net/datasets/kitti/eval_odometry.php

Well-known alignment models

• Rigid alignment	RIGID
• Projective alignment	RPE
• Epipolar alignment	EPI
• Photometric alignment	PHOTO

Rigid alignment

- **Known:** 3D-to-3D point correspondences
- **Given:** Pose hypothesis
- **Yield:** Geodesic error in world (3D space) units
- Commonly used in 3D registration

$$\varphi_{RIGID}\left(\underbrace{\mathbf{x},\mathbf{y}}_{\text{two corresponding 3D points}};\ \underbrace{\mathbf{R},\mathbf{t}}_{\text{pose hypothesis}}\right) = \underbrace{\left\|\mathbf{y}-\left(\mathbf{R}\mathbf{x}+\mathbf{t}\right)\right\|^2}_{\text{applying rigid transformation}}$$

Projective alignment

- **Known:** 3D-to-2D point correspondence
- **Given:** Pose hypothesis
- **Yield:** Geodesic error in image plane
- Known as *reprojection error* (RPE) in SFM and VO
- Minimisation of RPE in a least-square form is considered the "gold standard"

$$\varphi_{RPE}(\underbrace{\mathbf{x}}_{\text{3D point}},\underbrace{\mathbf{y}}_{\text{2D point}};\ \mathbf{R},\mathbf{t}) = \left\|\mathbf{y}-\underbrace{\pi}_{\text{perspective projection}}(\mathbf{R}\mathbf{x}+\mathbf{t})\right\|^2$$

Epipolar alignment

- **Known:** 2D-to-2D point correspondence
- **Given:** Pose hypothesis
- **Yield:** Epipolar error in (normalised) image plane
- Commonly used in uncalibrated two-view geometry
- Useful when lacking 3D information

$$\varphi_{EPI}\big(\underbrace{\mathbf{x},\mathbf{y}}_{\text{two corresponding 2D points (in canonical image coordinates)}};\ \mathbf{R},\mathbf{t}\big) = \left| \underbrace{\mathbf{y}^T [\mathbf{t}]_\times \mathbf{R}}_{\text{essential matrix}} \mathbf{x} \right|$$

Note: Here we show algebraic epipolar error. In practice a correction factor is applied to obtain geometric error.

Photometric alignment

- **Known:** 3D point and intensity images
- **Given:** Pose hypothesis
- **Yield:** Photometric error
- Used by all the direct methods
- No need to know point correspondences

$$\varphi_{PHOTO}\left(\underbrace{\mathbf{x}}_{\text{3D point}};\ \mathbf{R},\mathbf{t}\right) = \left|\underbrace{I}_{\substack{\text{intensity image}\\ \textbf{previous}\text{ frame}}}\left(\pi\left(\mathbf{x}\right)\right) - \underbrace{I'}_{\substack{\text{intensity image}\\ \textbf{current}\text{ frame}}}\left(\pi\left(\mathbf{R}\mathbf{x}+\mathbf{t}\right)\right)\right|$$

Multi-objective approach

- Use tracked image features and measured scene depth to instantiate four sub-objective functions
- Each sub-objective function φ_{SUB} computes the sum-of-squares of a corresponding residual function φ_{SUB}
- Can we simply sum them up?

$$\varphi_{RIGID}(\mathbf{R},\mathbf{t}) + \varphi_{RPE}(\mathbf{R},\mathbf{t}) + \varphi_{EPI}(\mathbf{R},\mathbf{t}) + \varphi_{PHOTO}(\mathbf{R},\mathbf{t})$$

They are even in different units!

Mahalanobis distance

- Generalised Euclidean distance measuring how likely an observation $\mathbf{x}$ belongs to a normal distribution with co-variance matrix Σ
- Can be used to represent each residual term in a covariance-normalised unit-free form
- Need to estimate error covariance now

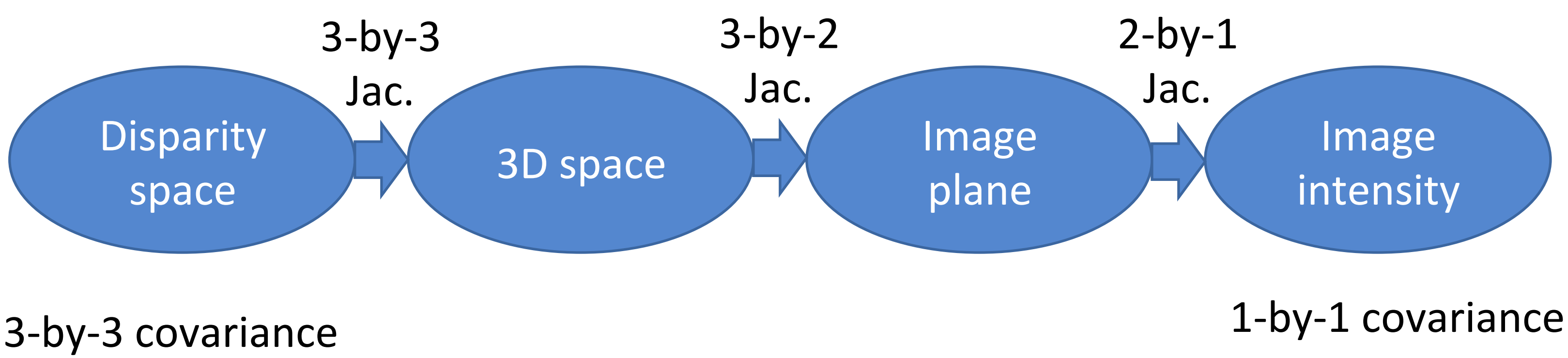
$$\delta\left(\mathbf{x};\mu,\Sigma\right) = \left\|\mathbf{x}-\mu\right\|_{\Sigma} = \sqrt{\left(\mathbf{x}-\mu\right)^{T}\Sigma^{-1}\left(\mathbf{x}-\mu\right)}$$

Propagation of uncertainty

- Error covariance Σ in the domain of a function f can be propagated to it's range by $\Sigma' = \mathbf{J} \Sigma \mathbf{J}^T$ where $\mathbf{J}$ is the Jacobian matrix of f at the point Σ is obtained
- The chaining of propagation is carried out for each point correspondence from the domain to the range of each residual function φ_{SUB}

Example: φ_{PHOTO}

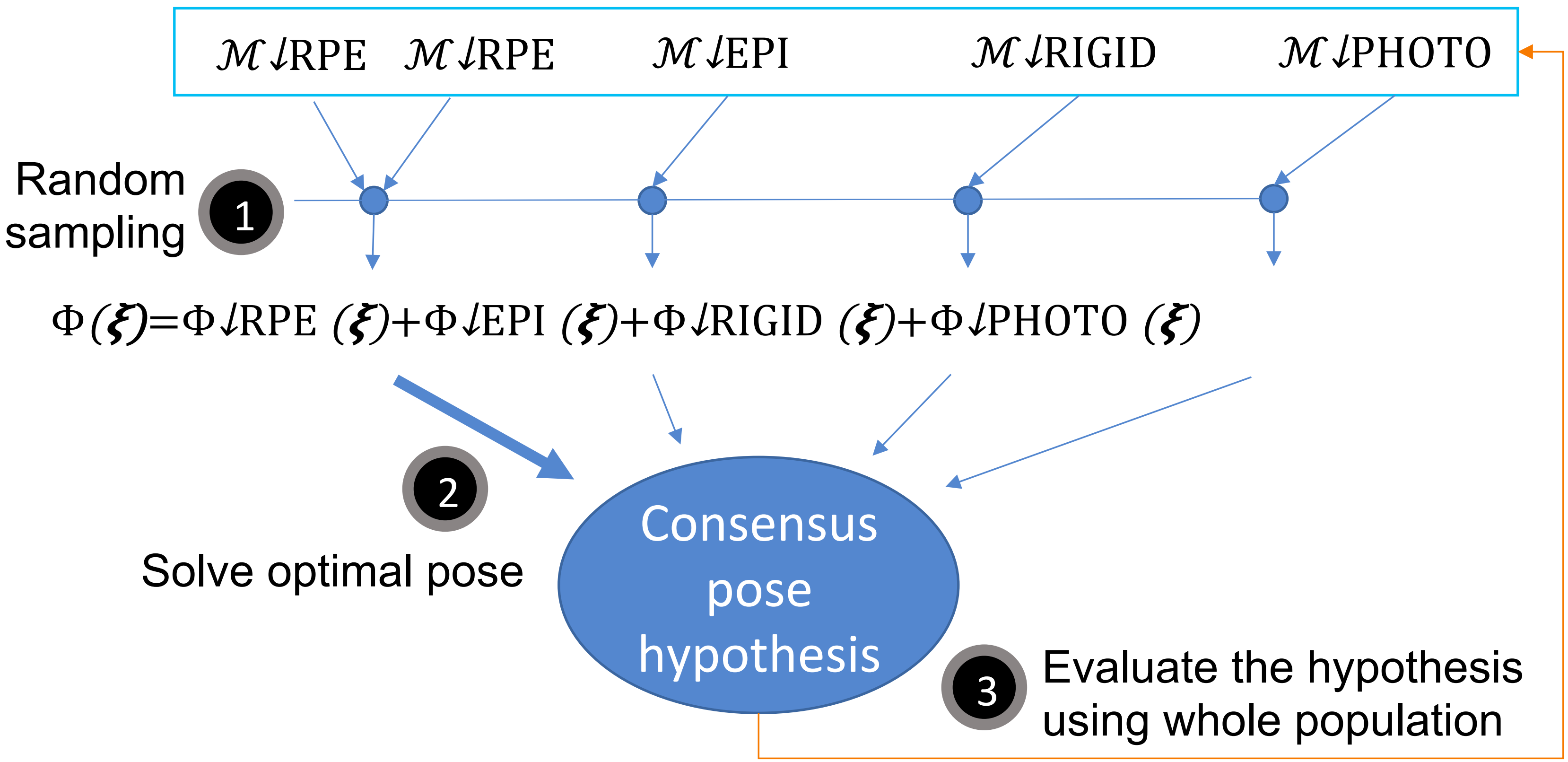
- Evaluation of the photometric error starts from a point in 3D space and ends up with an intensity difference
- In case the point is measured using stereo vision, the propagation has to back-trace to the disparity space



Implementation

- For each two frames $k-1$ and k five data terms are built
    1. $\mathcal{M}_{\mathrm{RPE}}$: Mapping of 3D points in $k-1$ to 2D points in k
    2. $\mathcal{M}_{\mathrm{RPE}}$: Mapping of 3D points in k to 2D points in $k-1$
    3. $\mathcal{M}_{\mathrm{EPI}}$: Mapping of 2D points in $k-1$ to k
    4. $\mathcal{M}_{\mathrm{RIGID}}$: Mapping of 3D points in $k-1$ to k
    5. $\mathcal{M}_{\mathrm{PHOTO}}$: Mapping of 3D points to intensities in $k-1$
- A RANSAC-based outlier rejection is performed to kick out poor correspondences
- A nonlinear optimisation process then solves for the pose that minimises the total energy of four sub-objectives built from five (filtered) data terms

Multi-objective RANSAC



4 Repeat for several iterations; the hypothesis supported by the largest consensus set from population wins the election

Experiments

- A KITTI sequence is selected for evaluation
- No bundle adjustment, no loop closure
- Implemented using OpenCV in C++, with CPU-only parallelism
- Recovered egomotion is compared with GPU/IMU readings
- For each configuration, **five trials** are carried out and the average drift (in %) is calculated

Combinations

- We tried out all 16 combinations of 4 models
  - A four-letter label is assigned to each combination
  - **`B`**: backward RPE / **`P`**: photometric / **`R`**: rigid / **`E`**: epipolar
  - E.g. **`BxxE`** stands for backward RPE + epipolar objectives
  - Forward RPE, the classical objective, is always activated

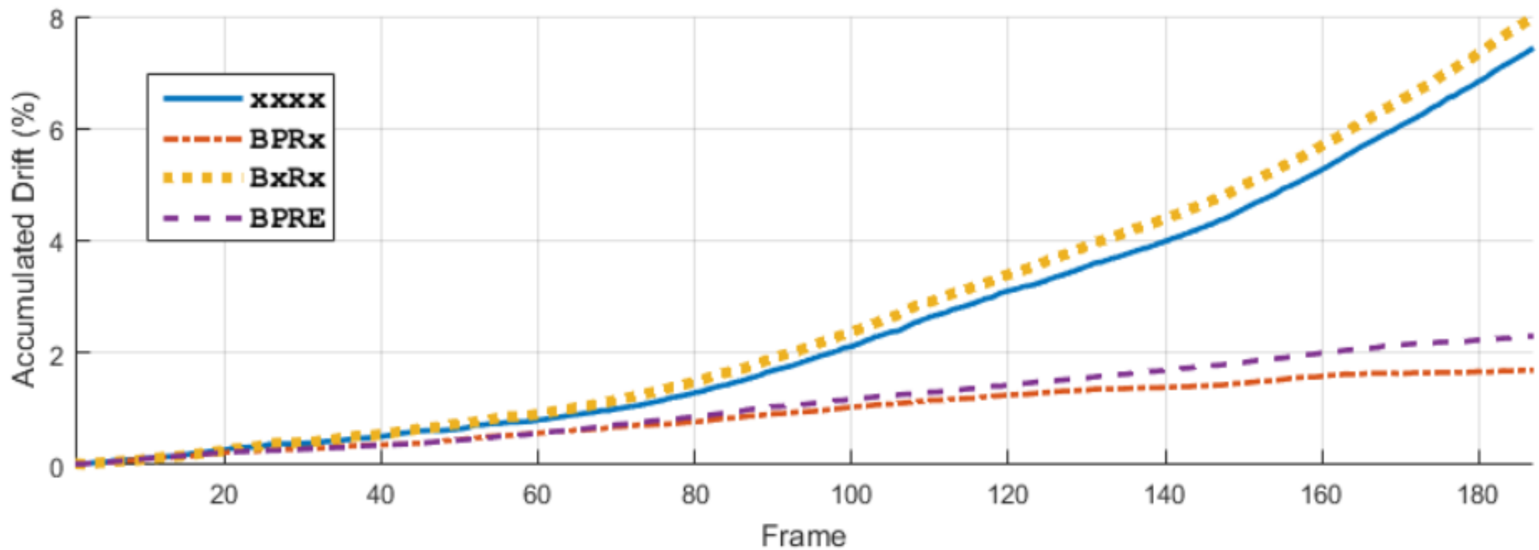
Results

- Using additional energy model(s) outperforms mono-objective VO in most cases
- The best record (63% improvement) is achieved by using photometric + rigid alignments (**`xPRx`**)
- When backward RPE is solely used (**`Bxxx`**), the result is slightly worse than the baseline by 0.17%

Model	Best	Worst	Mean	Std.	Model	Best	Worst	Mean	Std.
xxxx	4.97	5.54	5.21	0.27	Bxxx	**5.14**	5.99	5.41	0.34
xPxx	2.26	2.76	2.52	0.21	BPxx	1.99	2.50	2.23	0.21
xxRx	4.65	5.09	4.88	0.15	BxRx	5.10	**6.00**	**5.58**	**0.37**
xPRx	**1.84**	2.39	2.18	0.26	BPRx	1.96	2.56	**2.16**	0.26
xxxE	2.27	2.31	2.28	**0.01**	BxxE	2.21	**2.29**	2.24	0.03
xPxE	2.24	2.71	2.47	0.17	BPxE	2.17	2.48	2.31	0.11
xxRE	2.29	2.38	2.34	0.03	BxRE	2.18	2.31	2.24	0.05
xPRE	2.41	2.59	2.50	0.08	BPRE	2.21	2.40	2.33	0.08

Accumulated drift

- The all-enabled multi-objective VO is three times more accurate than the baseline model at the end of a sequence
- An interesting finding suggests the use of epipolar term is not necessary to achieve better estimation



Drift analysis of the best (BPRx), worst (BxRx), all-enabled (BPRE), and the baseline model (xxxx)

Conclusions

- We reviewed four alignment models used as objective functions in existing VO approaches
- A unifying framework (including error modelling) is proposed
- Experimental results indicate that at least **30%** improvement is attenable when multiple objectives are incorporated
- Time profiling shows that multi-objective VO incurs **13%** more computational cost compared to baseline

Sounds like a good deal!