# Improving the Egomotion Estimation by Correcting the Calibration Bias

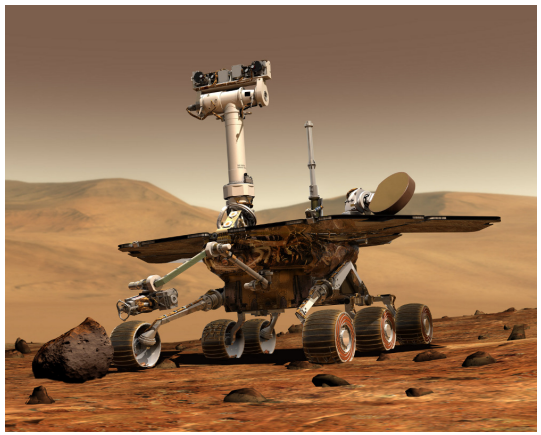
Ivan Krešo

Advisor: Siniša Šegvić

16/01/2015

# Motivation

- Egomotion - estimating the 3D rigid motion (rotation and translation) of the camera from an image sequence
- Basis for more involved structure and motion (SaM) estimation approaches
- Alternatives:
  - Inertial and differential GPS: accuracy depends on signal quality and coverage, expensive
  - Wheel odometry: wheel slippage
- Interesting applications:
  - Autonomous vehicle and robot navigation
  - Automotive systems: driver assistance, road safety inspection
  - Mars Rovers



# Stereo Camera Calibration

- Goal: map every image pixel to a 3D ray emerging from the focal point of the camera and spreading out to the physical world
- Linear intrinsic parameters:

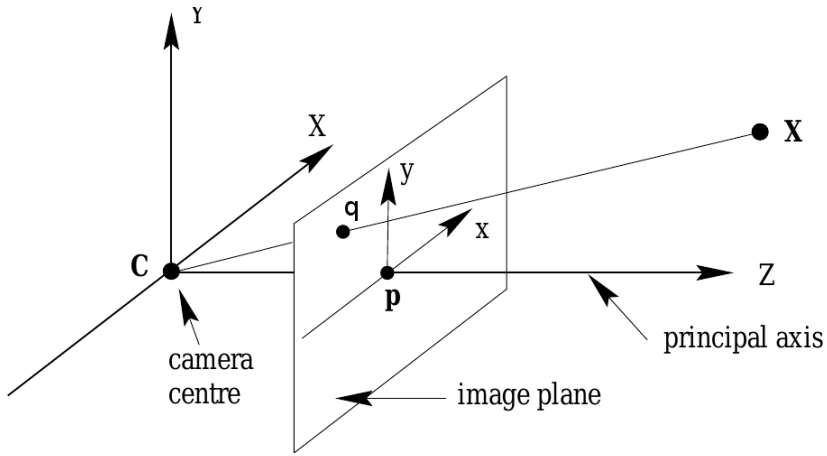
$$\mathbf{K} = \begin{pmatrix} f_x & 0 & c_u \\ 0 & f_y & c_v \\ 0 & 0 & 1 \end{pmatrix}$$

  - Camera matrix $\mathbf{K}$ transforms the normalized perspective projection coordinates to image coordinates
- Nonlinear intrinsic parameters: lens distortion
  - Radial: $k_1, k_2, k_3, k_4, k_5, k_6 \ldots$
  - Tangential: $p_1, p_2 \ldots$
- Extrinsic parameters: displacement between left and right cameras $(\mathbf{R}_c, \mathbf{t}_c)$

# Camera Model

$$\mathbf{X} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad \mathbf{q} = \begin{pmatrix} u \\ v \end{pmatrix}$$

$$x' = x/z \qquad y' = y/z$$




$$x'' = x' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + 2p_1 x' y' + p_2 (r^2 + 2(x')^2)$$

$$y'' = y' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + p_1 (r^2 + 2(y')^2) + 2p_2 x' y'$$

$$r^2 = (x')^2 + (y')^2$$

$$u = f_x \cdot x'' + c_u$$

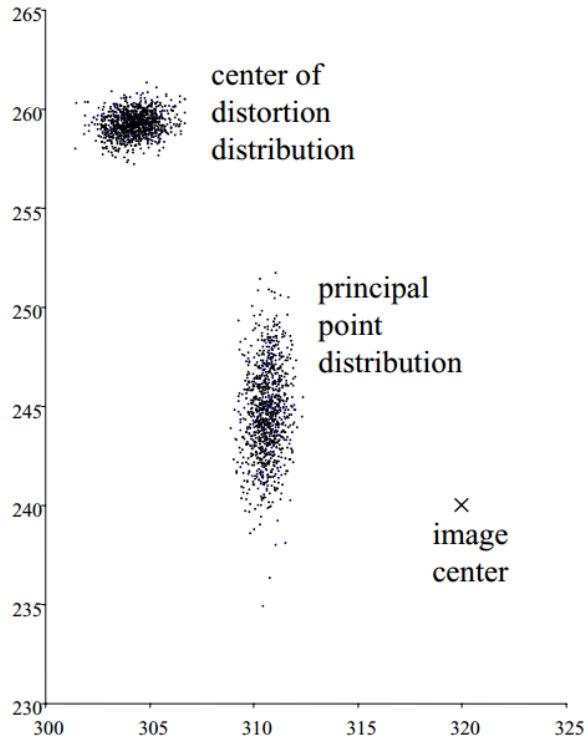
$$v = f_y \cdot y'' + c_v$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{K} \begin{pmatrix} x'' \\ y'' \\ 1 \end{pmatrix} ; \quad \mathbf{K} = \begin{pmatrix} f_x & 0 & c_u \\ 0 & f_y & c_v \\ 0 & 0 & 1 \end{pmatrix}$$

# Calibration Bias

- The most popular methods today assume the center of distortion is at the image principal point [Zhang, 2000]
- It has been shown that this does not hold in real cameras [Hartley, 2005]

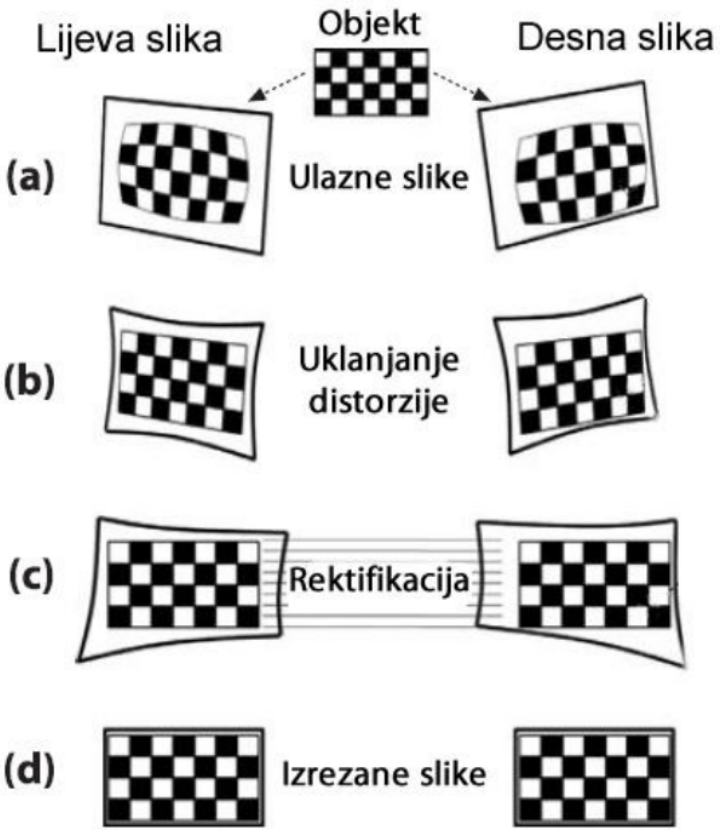
- Offset of the lens centre from the sensor centre
- Slight tilt of the sensor plane with respect to the lens
- Misalignment of individual components of a compound lens

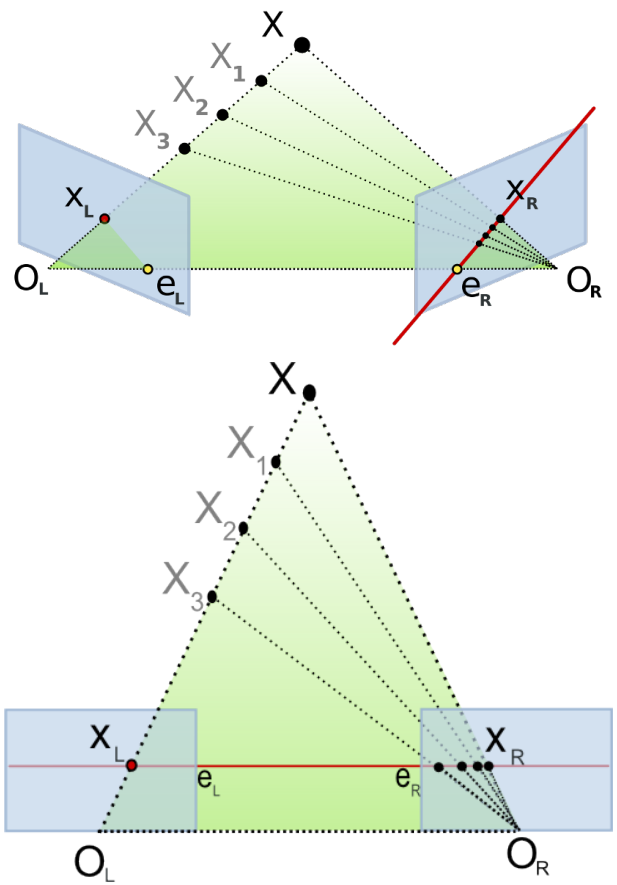



Taken from [Hartley, 2005]

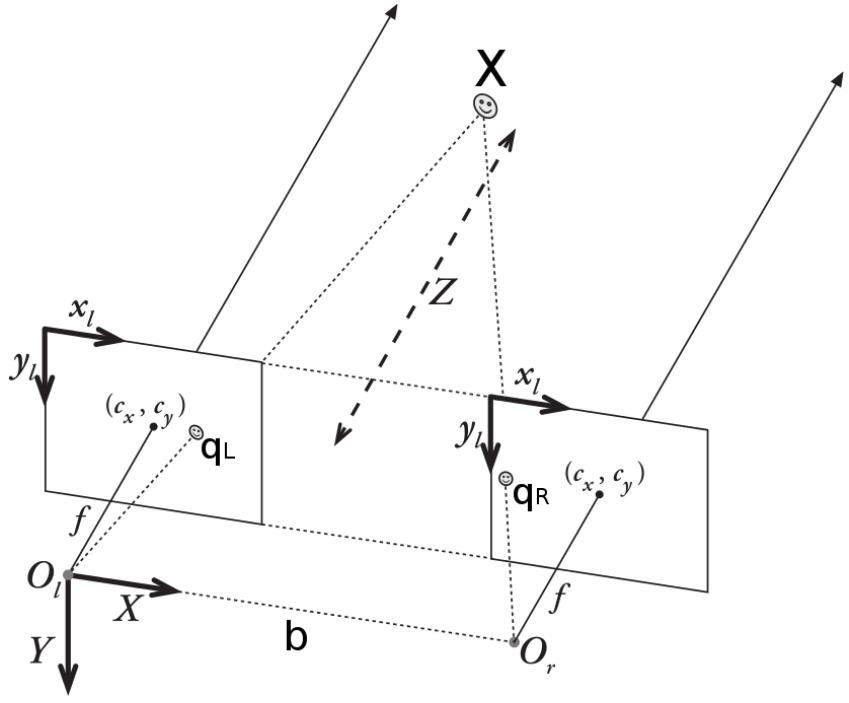
# Image Rectification

After the rectification all epipolar lines are horizontal and because the image planes now share the same x-axis all corresponding points will be located in the same image row as shown



# Rectified Stereo Camera Model

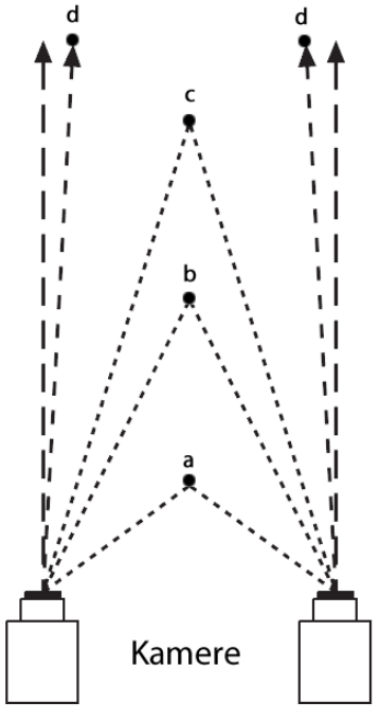


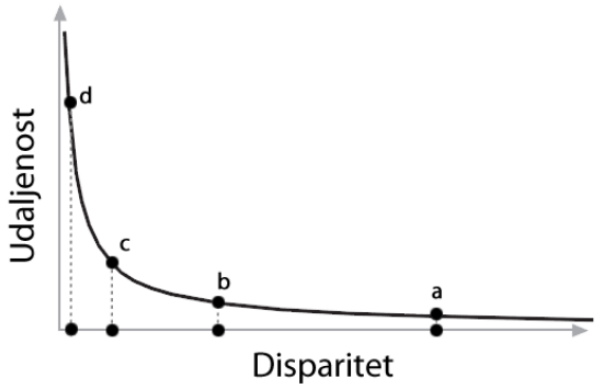

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \pi(\mathbf{X}, \mathbf{R}, \mathbf{t}) = \begin{pmatrix} f & 0 & c_u \\ 0 & f & c_v \\ 0 & 0 & 1 \end{pmatrix} [\mathbf{R}|\mathbf{t}] \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

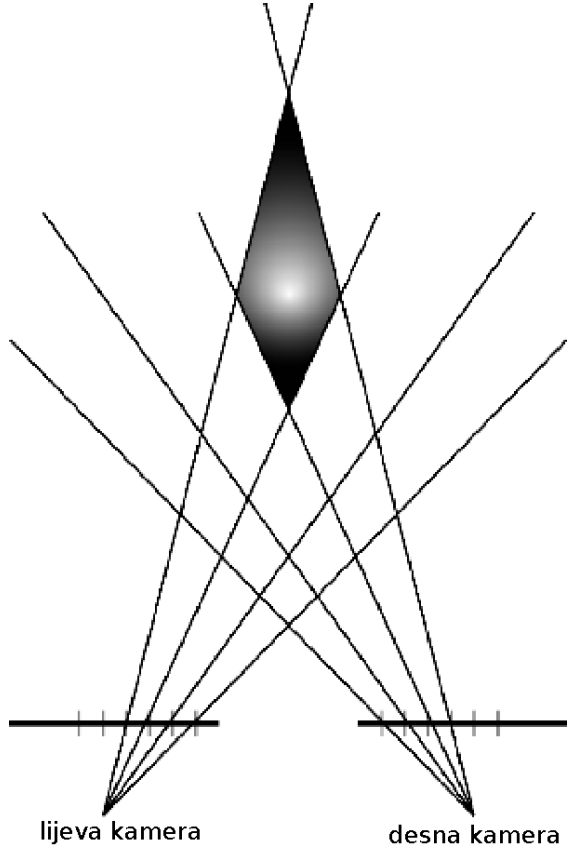
$$\mathbf{q}_l = \pi(\mathbf{X}, \mathbf{R}, \mathbf{t})$$

$$\mathbf{q}_r = \pi(\mathbf{X}, \mathbf{R}, \mathbf{t}_r) \qquad \mathbf{t}_r = \mathbf{t} - (b, 0, 0)^\top$$

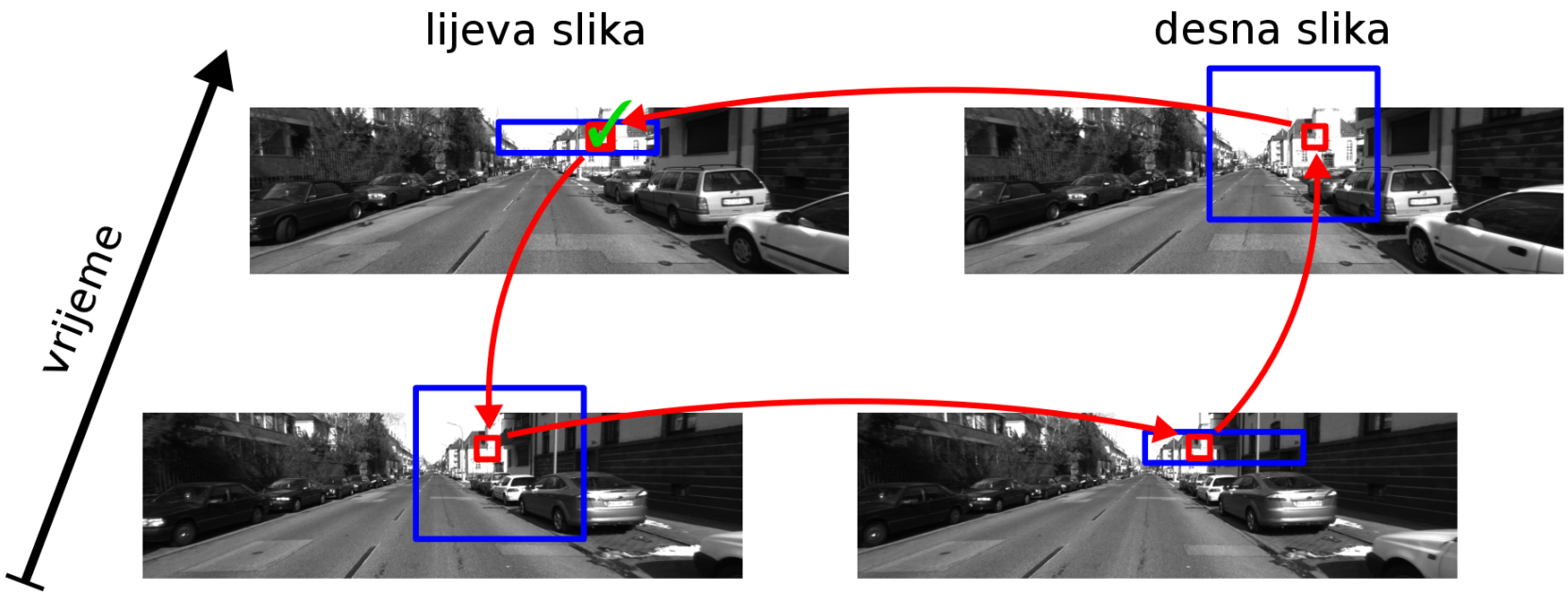
# Triangulation Error



# Finding Point Correspondences




- Matching between four images:
  - Previous left and right and current left and right

# Egomotion Optimization

$$\mathbf{X}_{i,t-1} = \boldsymbol{t}(\mathbf{q}^{l}_{i,t-1}, \mathbf{q}^{r}_{i,t-1}) \qquad \mathbf{X}_{i,t} = \boldsymbol{t}(\mathbf{q}^{l}_{i,t}, \mathbf{q}^{r}_{i,t})$$

Error in world space - 3D point alignment

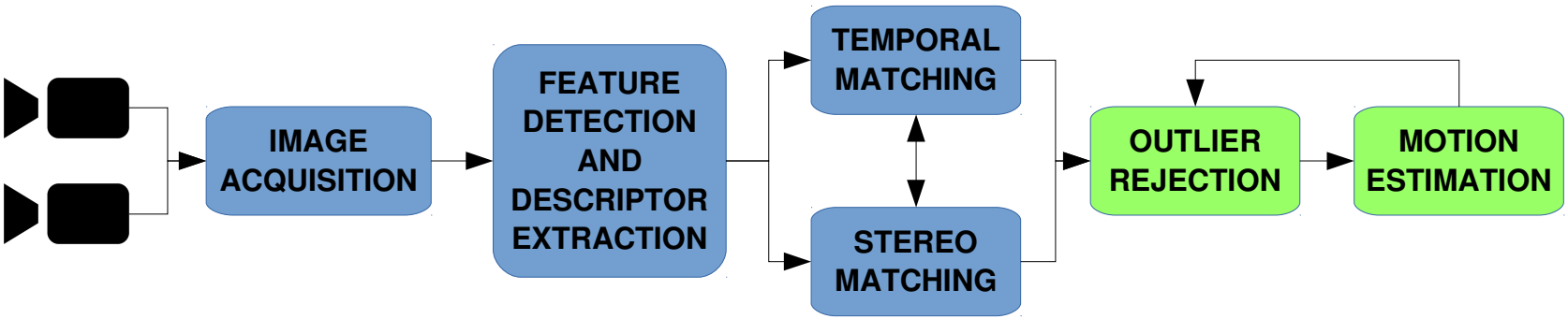
$$\underset{\mathbf{R},\mathbf{t}}{\operatorname{argmin}} \sum_{i=1}^{N} \|\mathbf{X}_{i,t} - (\mathbf{R}\mathbf{X}_{i,t-1} + \mathbf{t})\|^2$$

Error in image space - 2D point alignment [Nistér, 2006]

$$\underset{\mathbf{R},\mathbf{t}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{k}^{\{l,r\}} \|\mathbf{q}^{k}_{i,t} - \pi(\mathbf{X}_{i,t-1}, \mathbf{R}, \mathbf{t}_k)\|^2$$

$$\mathbf{t}_l = \mathbf{t} \qquad \mathbf{t}_r = \mathbf{t}_l - (b, 0, 0)^\top$$

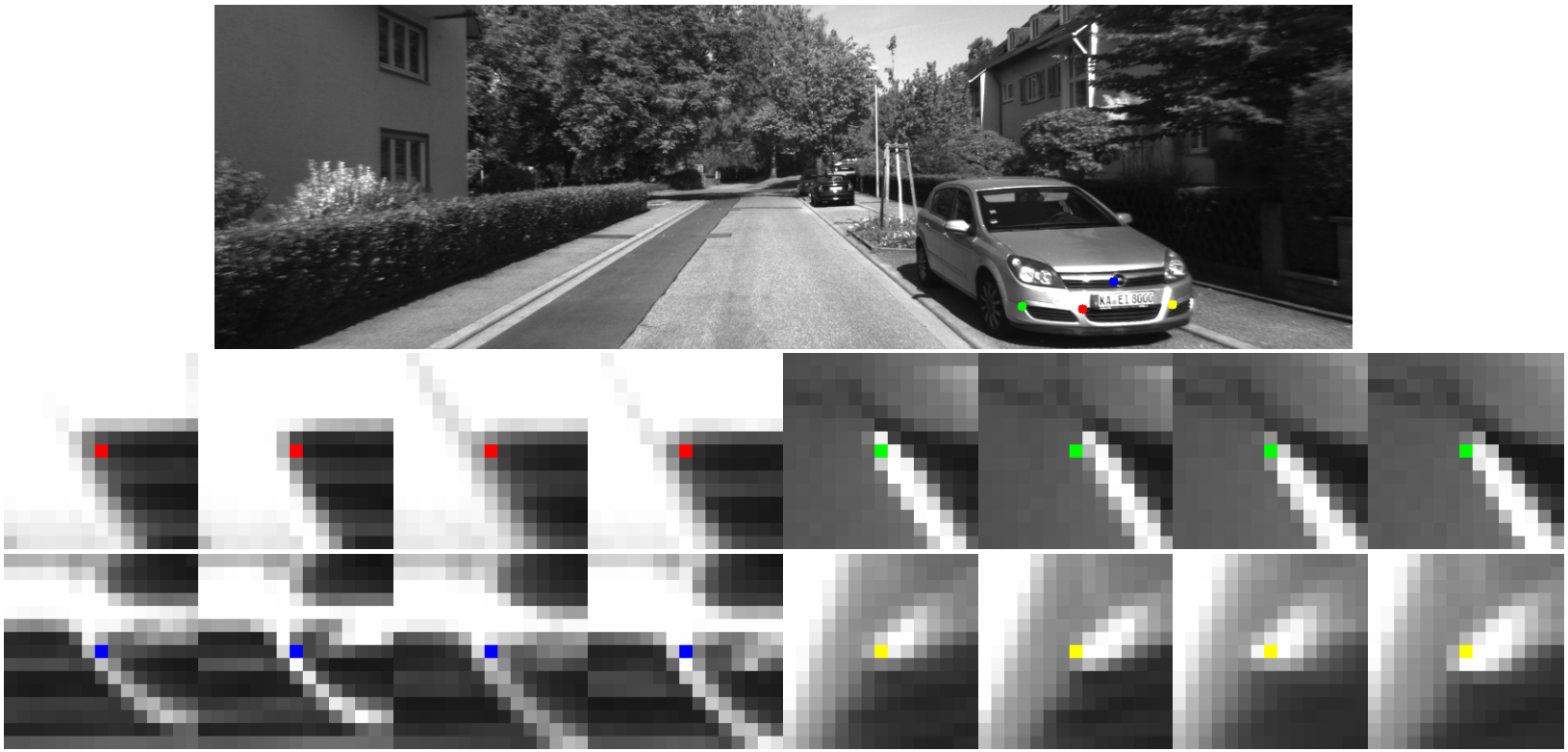
# Egomotion Pipeline




- We match Harris corners in left camera with NCC metric on 15x15 patches
- Disparity is measured by matching along epipolar line in right image
- RANSAC based outlier rejection:
  - Gauss-Newton method is used for hypothesis generation and final optimization

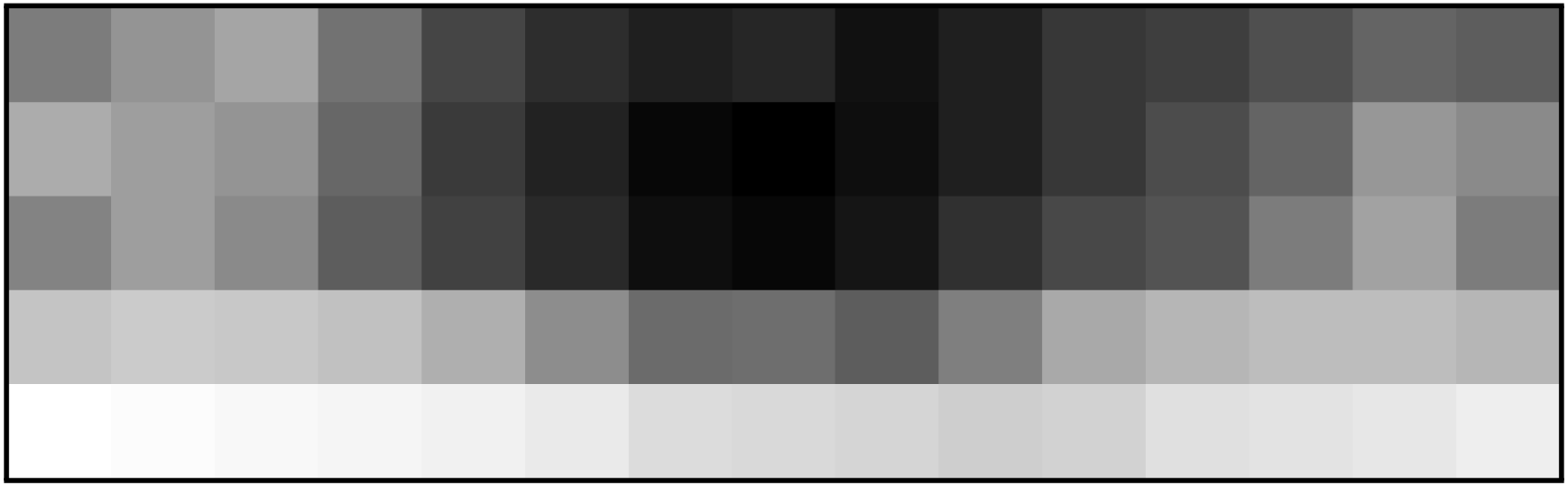
# Reprojection Error on the KITTI Dataset

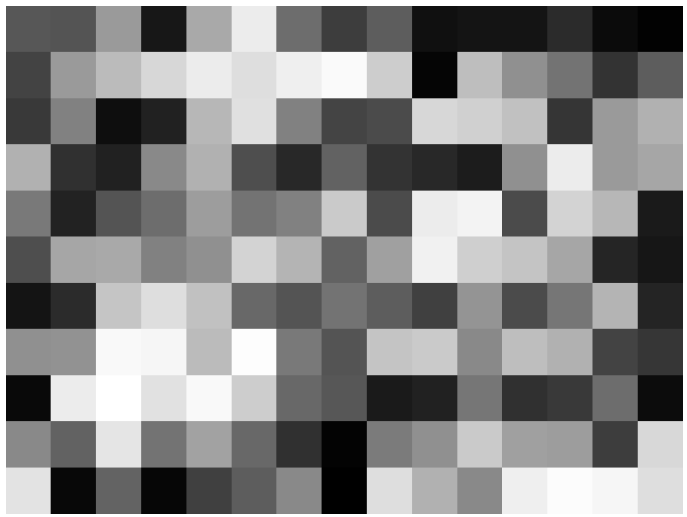
- We found many perfect correspondences with large reprojection errors under the groundtruth camera motion



| | | |
|---|---|---|
| Red | 7.10 | 7.01 |
| Green | 6.68 | 6.56 |
| Blue | 7.40 | 7.32 |
| Yellow | 6.75 | 6.35 |

# Error Heat Map on KITTI and Tsukuba Datasets





# Error Statistics on KITTI

- Mean values of the vector errors deviate from zero and contain bias following some specific patterns

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.35 | 0.367 | 0.374 | 0.344 | 0.301 | 0.277 | 0.267 | 0.272 | 0.262 | 0.267 | 0.292 | 0.295 | 0.314 | 0.334 | 0.332 |
| 0.383 | 0.369 | 0.367 | 0.337 | 0.293 | 0.27 | 0.249 | 0.245 | 0.254 | 0.267 | 0.292 | 0.308 | 0.334 | 0.367 | 0.36 |
| 0.359 | 0.369 | 0.36 | 0.333 | 0.297 | 0.274 | 0.254 | 0.249 | 0.263 | 0.283 | 0.305 | 0.326 | 0.349 | 0.373 | 0.35 |
| 0.417 | 0.439 | 0.436 | 0.411 | 0.386 | 0.363 | 0.342 | 0.343 | 0.333 | 0.352 | 0.381 | 0.395 | 0.398 | 0.398 | 0.394 |
| 0.606 | 0.587 | 0.581 | 0.573 | 0.538 | 0.508 | 0.477 | 0.474 | 0.465 | 0.455 | 0.462 | 0.483 | 0.492 | 0.494 | 0.518 |

a) Reprojection error L2-norm means

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.129 | -0.144 | -0.171 | -0.159 | -0.118 | -0.0729 | -0.0487 | -0.0328 | -0.0125 | -0.0196 | -0.0245 | -0.0176 | -0.0328 | -0.0658 | -0.0713 |
| -0.141 | -0.124 | -0.153 | -0.132 | -0.0902 | -0.0489 | -0.0238 | -0.022 | -0.0277 | -0.0394 | -0.053 | -0.0626 | -0.0824 | -0.112 | -0.125 |
| -0.112 | -0.116 | -0.138 | -0.115 | -0.0814 | -0.0509 | -0.031 | -0.0291 | -0.0455 | -0.0638 | -0.0922 | -0.0887 | -0.109 | -0.152 | -0.144 |
| -0.182 | -0.188 | -0.188 | -0.168 | -0.152 | -0.0982 | -0.0741 | -0.0574 | -0.0696 | -0.0849 | -0.115 | -0.121 | -0.116 | -0.146 | -0.134 |
| -0.304 | -0.255 | -0.243 | -0.234 | -0.191 | -0.131 | -0.0874 | -0.0892 | -0.0945 | -0.105 | -0.108 | -0.122 | -0.116 | -0.148 | -0.18 |

b) Reprojection error means on the $u$ axis

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0173 | 0.0252 | 0.0219 | 0.0049 | -0.00545 | -0.00162 | 0.0048 | 0.00391 | 0.00424 | 0.0178 | 0.0288 | 0.0401 | 0.0601 | 0.0779 | 0.0624 |
| 0.0169 | 0.0372 | 0.0431 | 0.0345 | 0.0197 | 0.0146 | 0.00939 | 0.0117 | 0.0173 | 0.0265 | 0.0467 | 0.0613 | 0.0726 | 0.0801 | 0.0512 |
| 0.0255 | 0.05 | 0.0545 | 0.0499 | 0.0425 | 0.0345 | 0.0234 | 0.0204 | 0.0325 | 0.0407 | 0.0517 | 0.0581 | 0.07 | 0.069 | 0.0369 |
| 0.0641 | 0.0826 | 0.0884 | 0.0924 | 0.0891 | 0.0773 | 0.0769 | 0.065 | 0.0779 | 0.0842 | 0.0821 | 0.0921 | 0.0918 | 0.0811 | 0.053 |
| 0.144 | 0.131 | 0.131 | 0.148 | 0.126 | 0.141 | 0.128 | 0.128 | 0.121 | 0.133 | 0.149 | 0.156 | 0.166 | 0.175 | 0.15 |

c) Reprojection error means on the $v$ axis

# Learning the Stereoscopic Deformation Field

- Idea: learn the camera calibration correction by exploiting groundtruth motion

$$\tilde{\mathbf{X}}_{i,t-1} = \boldsymbol{t}(\tilde{\mathbf{q}}^{l}_{i,t-1}, \tilde{\mathbf{q}}^{r}_{i,t-1})$$

$$\underset{\mathbf{D}^{l}_{u},\mathbf{D}^{l}_{v},\mathbf{D}^{r}_{u},\mathbf{D}^{r}_{v}}{\operatorname{argmin}} \sum_{t=1}^{M}\sum_{i=1}^{N}\sum_{k}^{\{l,r\}} \left\|\tilde{\mathbf{q}}^{k}_{i,t} - \pi(\tilde{\mathbf{X}}_{i,t-1}, \mathbf{R}_t, \mathbf{t}_{t,k})\right\|^2 \qquad (1)$$

$$\tilde{\mathbf{q}} = \boldsymbol{d}(\mathbf{q}, \mathbf{D}_u, \mathbf{D}_v)$$

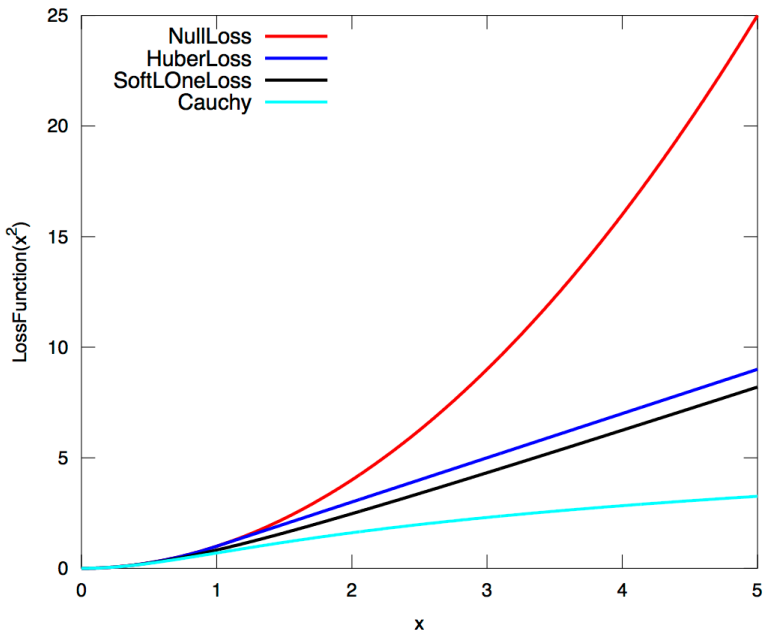
- $\mathbf{D}^{l}_{u}, \mathbf{D}^{l}_{v}, \mathbf{D}^{r}_{u}, \mathbf{D}^{r}_{v}$ - the deformation field matrices for $u$ and $v$ axes and left and right cameras respectively
- $\mathbf{R}_t$ and $\mathbf{t}_t$ in each frame $t$ are taken from groundtruth data
- $\boldsymbol{d}$ - function which applies the deformation to the point
- Levenberg-Marquardt method from Ceres Solver is used to optimize (1)

# Learning the Stereoscopic Deformation Field

- We collect the tracks in all sequences by filtering them with groundtruth motion using a fixed error threshold
- Threshold has to be large enough to capture the impact of larger reprojection errors further from the image center
- Filtering with a permissive threshold produces outliers which can exert a significant impact to the least-squares optimization
- To reduce their impact we wrap the square loss into the robust Cauchy loss function:

$$\rho(s) = a^2 log(1 + \frac{s}{a^2})$$

- $s$ - square loss output
- $a$ - parameter which determines the scale at which robustification takes place



# Integrating the Deformation Field in Motion Estimation

$$\underset{\mathbf{R},\mathbf{t}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{k}^{\{l,r\}} \|\widehat{\mathbf{q}}_{i,t}^{k} - \pi(\mathbf{X}_{i,t-1}, \mathbf{R}, \mathbf{t}_k)\|^2$$

$$\widehat{\mathbf{q}} = \boldsymbol{i}(\mathbf{q}, \mathbf{D}_u, \mathbf{D}_v)$$

- $\boldsymbol{i}$ - computes the bilinear interpolation between the four cells of the deformation field which surround the point $\mathbf{q}$

# Experimental Results

- We compare our approach to two variants based on the Libviso library [Geiger, 2011]:
  - Baseline - Libviso without feature weighting
  - Libviso with feature weighting

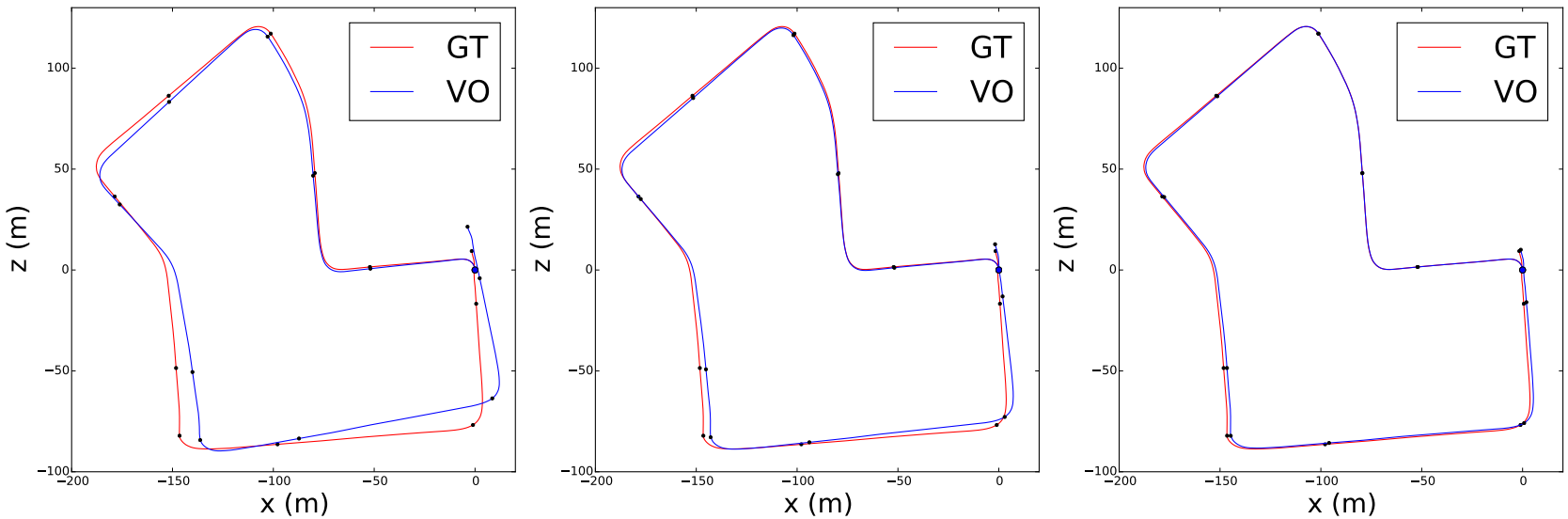
$$w_i = \left( \frac{|u_i - c_u|}{|c_u|} + 0.05 \right)^{-1}$$

# Experimental Results

- Leave-one-out evaluation on the training sequences 04-10
- Relative translation error in percents of the traveled distance
- Relative rotation error in degrees per meter

| **KITTI** | | **Baseline** | | **With FW** | | **With DF** | |
|---|---|---|---|---|---|---|---|
| Seq. | length | trans. | rot. | trans. | rot. | trans. | rot. |
| 04 | 394 m | 1.14 | 0.0094 | 0.63 | 0.0027 | 0.78 | 0.0068 |
| 05 | 2206 m | 1.29 | 0.0095 | 0.70 | 0.0047 | 0.43 | 0.0030 |
| 06 | 1233 m | 1.30 | 0.0069 | 0.75 | 0.0041 | 0.53 | 0.0047 |
| 07 | 695 m | 2.02 | 0.0221 | 0.86 | 0.0083 | 0.40 | 0.0034 |
| 08 | 3223 m | 1.45 | 0.0087 | 1.10 | 0.0056 | 1.02 | 0.0048 |
| 09 | 1705 m | 1.51 | 0.0067 | 1.16 | 0.0041 | 0.97 | 0.0050 |
| 10 | 920 m | 0.80 | 0.0067 | 0.65 | 0.0042 | 0.70 | 0.0040 |
| All | 10376 m | 1.386 | 0.0089 | 0.933 | 0.0051 | 0.771 | 0.0043 |

# Experimental Results - Case study: the sequence 07




(a) Baseline (b) With feature weighting (c) With deformation field

# Future Work

- Add robust loss to the egomotion cost function
- Evaluate different regularization approaches in the loss function used to calibrate the stereoscopic deformation field
- Evaluate the impact of calibration bias correction to the multi-frame bundle adjustment optimization

# Reference



Zhang, Zhengyou (2000)
A Flexible New Technique for Camera Calibration


Hartley, R. I. and Kang, S. B. (2005)
Parameter-free radial distortion correction with centre of distortion estimation


David Nistér and Oleg Naroditsky and James R. Bergen (2006)
Visual odometry for ground vehicle applications


Bernd Kitt and Andreas Geiger and Henning Lategahn (2011)
Visual Odometry based on Stereo Image Sequences with RANSAC-based Outlier Rejection Scheme


Andreas Geiger and Philip Lenz and Raquel Urtasun (2012)
Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite


Ivan Kreso and Marko Sevrovic and Sinisa Segvic (2013)
A Novel Georeferenced Dataset for Stereo Visual Odometry


Hernán Badino and Akihiro Yamamoto and Takeo Kanade (2013)
Visual Odometry by Multi-frame Feature Integration