3D Computer Vision

Radim Šára   Martin Matoušek

Center for Machine Perception
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague

https://cw.fel.cvut.cz/wiki/courses/tdv/start

http://cmp.felk.cvut.cz
mailto:sara@cmp.felk.cvut.cz
phone ext. 7203

rev. December 19, 2023

Open Informatics Master's Course

▶Basic Stereoscopic Matching Models

- notice many small isolated errors in the ML matching
- Q: how to reduce the noisiness? A: a stronger model

Potential models for M (from weaker to stronger)

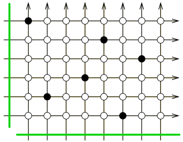
1. Uniqueness: Every image point matches at most once
   - excludes semi-transparent objects
   - used in the ML matching algorithm (but not in the WTA algorithm)

2. Monotonicity: Matched pixel ordering is preserved →189
   - for all $(i,j) \in M, (k,l) \in M, \quad k > i \Rightarrow l > j$

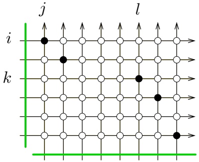
     Notation: $(i,j) \in M$ or $j = M(i)$ – left-image pixel i matches right-image pixel j
   - excludes thin objects close to the cameras
   - used in 3-Label Dynamic Programming (3LDP) [SP]

3. Coherence: Objects occupy well-defined 3D volumes
   - concept by [Prazdny 85]
   - algorithms are based on image/disparity map segmentation
   - a popular model (segment-based, bilateral filtering and their successors)
   - used in Stable Segmented 3LDP [Aksoy et al. PRRS 2008]

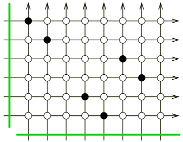
4. (Piecewise) binocular continuity: The scene images continuously w/o self-occlusions
   - disparities do not differ much in neighboring pixels (except at object boundaries)
   - full binocular continuity too strong, except in some applications
   - piecewise binocular continuity is combined with monotonicity in 3LDP



incoherent



monotonic coherent

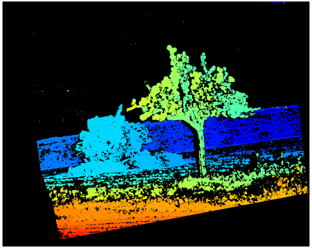


non-monotonic coherent

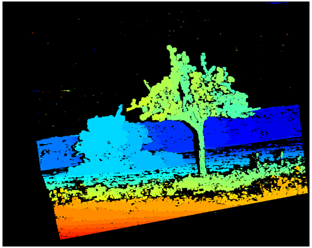
Some Results: AppleTree

left image

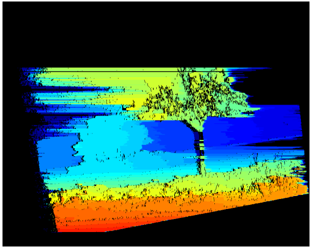
right image



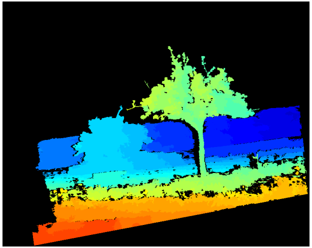
ML →184



3LDP with ordering
[SP]



naïve DP
[Cox et al. 1992]



Stable Segmented 3LDP
[Aksoy et al. PRRS 2008]

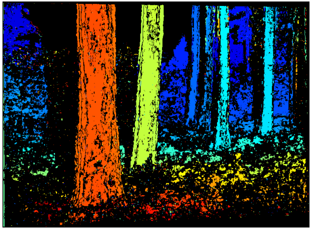
- 3LDP parameters α_i, V_e learned on Middlebury stereo data

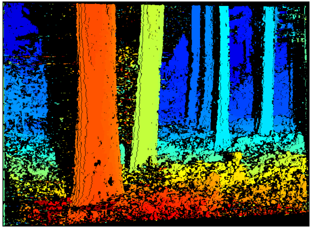
http://vision.middlebury.edu/stereo/

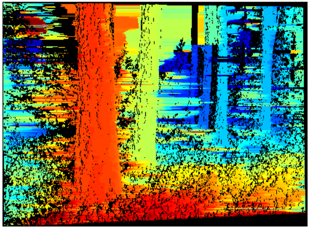
Some Results: Larch

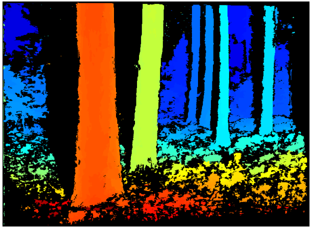

left image


right image


ML →184

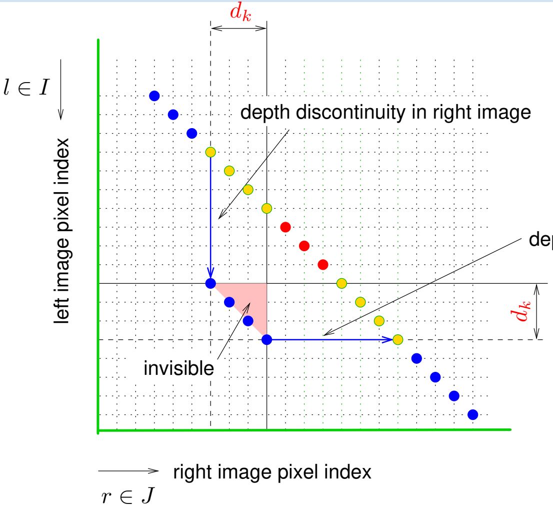

3LDP w/ordering [SP]


naïve DP

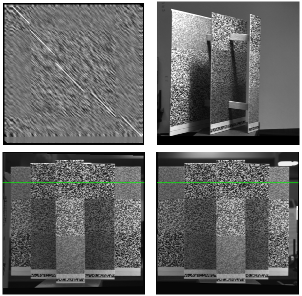

Stable Segmented 3LDP

- naïve DP: no mutual occlusion model, ignores symmetry, has no similarity distribution model, ignores $T \setminus M$
- but even 3LDP has errors in mutually occluded region
- Stable Segmented 3LDP: few errors in mutually occluded region since it uses a coherence model

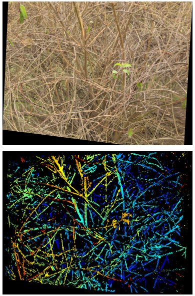
Binocular Discontinuities in Matching Table



- binocularly visible foreground points
- binocularly visible background pts violating ordering
- monocularly visible points (half-occluded in the other cam)
- d_k critical disparity



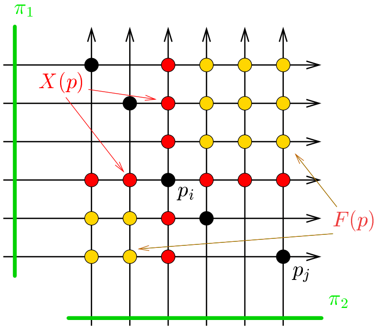
ordering holds



no ordering; Alg: GCS

- this leads to the concept of 'forbidden zone'

Formally: Uniqueness and Ordering in Matching Table T



$p_j \notin X(p_i), \quad p_j \notin F(p_i)$

- **Uniqueness Constraint:**

  A set of pairs $M = \{p_i\}_{i=1}^n$, $p_i \in T$ is a matching iff
  $\forall p_i, p_j \in M: \; p_j \notin X(p_i)$.

  X-zone, $p_i \notin X(p_i)$

- **Ordering Constraint:**

  Matching M is monotonic iff
  $\forall p_i, p_j \in M: \; p_j \notin F(p_i)$.

  F-zone, $p_i \notin F(p_i)$

- ordering constraint: matched points form a monotonic set in both images
- ordering is a powerful constraint: in $n \times n$ table we have: monotonic matchings $O(4^n) \ll O(n!)$ all matchings

  ⊛ 2: how many are there maximal monotonic matchings? (e.g. 27 for $n = 4$; hard!)

- uniqueness constraint is a basic occlusion model
- ordering constraint is a weak continuity model — and partly also an occlusion model
- monotonic matchings can be found by **dynamic programming**

Algorithm Comparison

Marroquin's Winner-Take-All (WTA →178)

- the ur-algorithm — very weak model
- dense disparity map
- $O(N^3)$ algorithm, simple but it rarely works

Maximum Likelihood Matching (ML →184)

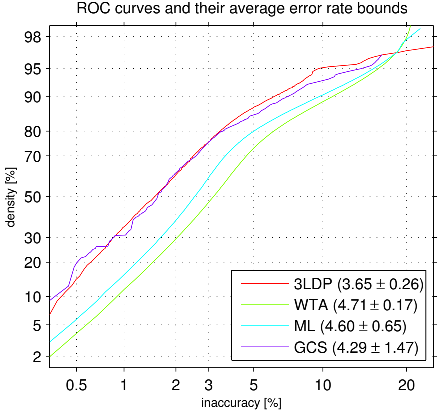
- semi-dense disparity map
- many small isolated errors
- models basic occlusion
- $O(N^3 \log(NV))$ algorithm — max-flow by cost scaling

MAP with Min-Cost Labeled Path (3LDP)

- semi-dense disparity map
- models occlusion in flat, piecewise binocularly continuous scenes
- has 'illusions' if ordering does not hold
- $O(N^3)$ algorithm

Stable Segmented 3LDP

- better than 3LDP — fewer errors at any given density
- $O(N^3 \log N)$ algorithm
- requires image segmentation — itself a difficult task

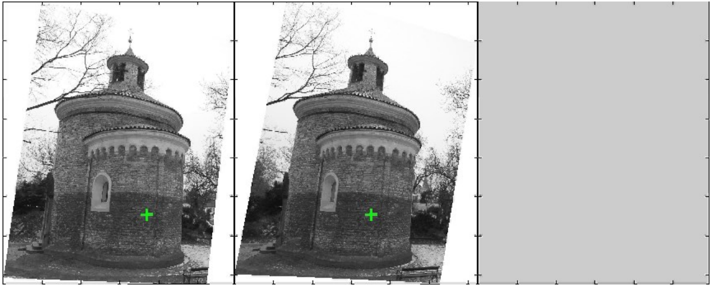


- ROC-like curve captures the density/accuracy tradeoff
- numbers: AUC (smaller is better)
- GCS is the one used in the exercises
- more algorithms at http://vision.middlebury.edu/stereo/ (good luck!)

GCS: Growing Correspondence Seeds

Alg: [Cech & Sara, BenCOS 2007]

1. Grow seed correspondences until they violate uniqueness severely — by a X-zone test
2. Select final unique matches by match competition in the X/FX-zone — by the stable matching algorithm



click for video

- explores only the "promising" regions in disparity space
- does not need "good" seeds because the competition revises them
- requires good-accuracy epipolar rectification — as all the algorithms mentioned do

Module IX

Additional Topics

9.1 Real Camera with Radial Distortion

covered by

[H&Z] Sec 7.4

Real Camera with Radial Distortion

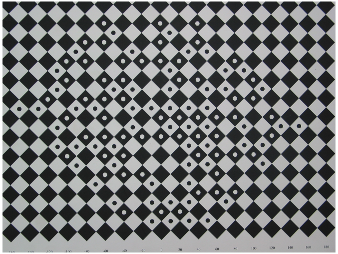
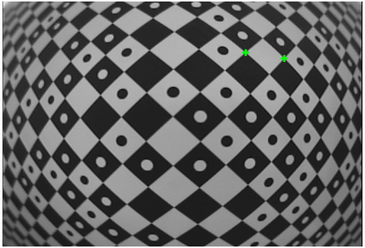


image with no radial distortion



an extreme case of barrel radial distortion

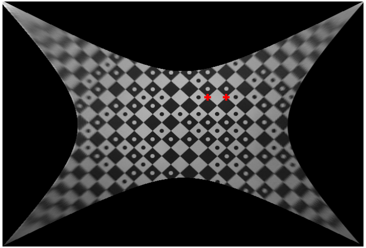
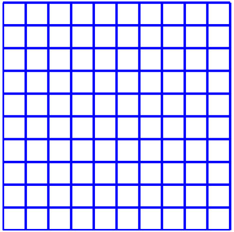
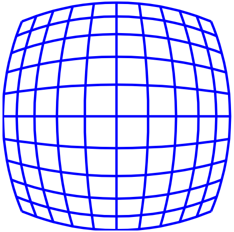


image undistorted by division model

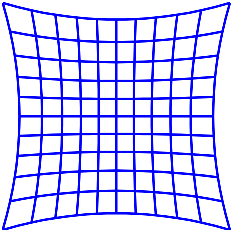
distortion types



none ($\lambda = 0$)

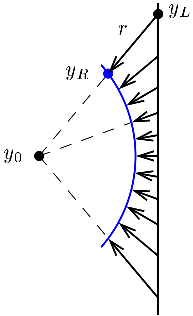


barrel ($\lambda = 0.3$)



pincushion ($\lambda = -0.3$)

The Radial Distortion Mapping



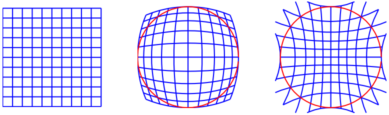
- everything is happening in the image plane

  y_0 – center of radial distortion (usually the principal point)

  y_L – linearly projected point (unknown)

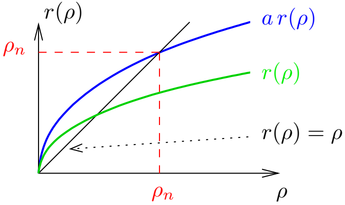
  y_R – radially distorted point (known)

- radial distortion r maps y_L to y_R along the radial direction
- magnitude of the transfer depends on the radius $\rho = \|y_R - y_0\|$ only

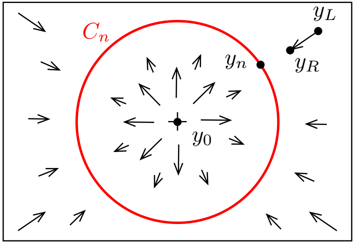


- circles centered at y_0 map to centered circles, lines incident on y_0 map on themselves
- the mapping $r()$ can be scaled to $a\,r()$ so that a particular circle C_n of radius ρ_n does not scale

distortion	inside C_n	outside C_n
barrel	expanding	contracting
pincushion	contracting	expanding

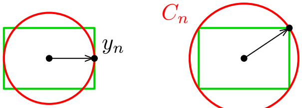


Radial Distortion Models



barrel distortion
arrows represent $\mathbf{z}_R - r(\mathbf{z}_L)$

- let $\mathbf{z} = \mathbf{y} - \mathbf{y}_0$ — non-homogeneous
- we have $\mathbf{z}_R = r(\mathbf{z}_L)$ — $\mathbf{z}_L$ – linear, $\mathbf{z}_R$ – distorted
- but we are often interested in $\mathbf{y}_L = r^{-1}(\mathbf{y}_R)$
- $\mathbf{y}_n$ – a no-distortion point on C_n: $r(\mathbf{y}_n) = \mathbf{y}_n$
- $\mathbf{z}_n = \mathbf{y}_n - \mathbf{y}_0$
- $\mathbf{y}_n$: a boundary point that preserves image content within the image size



in pincushion — in barrel

Division Model

$$\mathbf{z}_L = \frac{1-\lambda}{1-\lambda\frac{\|\mathbf{z}_R\|^2}{\|\mathbf{z}_n\|^2}}\,\mathbf{z}_R \quad \text{and} \quad \mathbf{z}_R = \frac{\hat{\mathbf{z}}}{1+\sqrt{1+\lambda\frac{\|\hat{\mathbf{z}}\|^2}{\|\mathbf{z}_n\|^2}}}, \quad \text{where } \hat{\mathbf{z}} = \frac{2\,\mathbf{z}_L}{1-\lambda}$$

- single parameter $-1 \le \lambda < 1$: $\lambda > 0$ – barrel distortion, $\lambda < 0$ – pincushion distortion
- has a closed-form inverse
- models even some fish-eye lenses

Polynomial Model

$$\mathbf{z}_L = \frac{D(\mathbf{z}_R; \mathbf{z}_n, \mathbf{k})}{1 + \sum_{i=1}^{n} k_i} \, \mathbf{z}_R \,,$$

$$D(\mathbf{z}_R; \mathbf{z}_n, \mathbf{k}) = 1 + k_1 \rho^2 + k_2 \rho^4 + \cdots + k_n \rho^{2n}, \quad \rho = \frac{\|\mathbf{z}_R\|}{\|\mathbf{z}_n\|}, \quad \mathbf{k} = (k_{1:n})$$

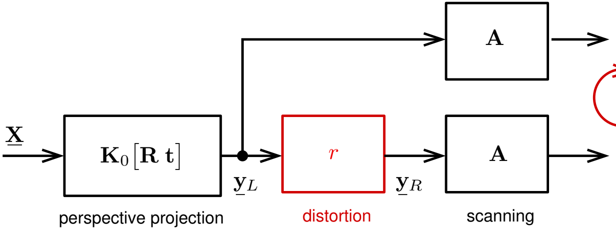
- e.g. $k_i \geq 0$ – barrel distortion, $k_i \leq 0$ – pincusion distortion, $i = 1, \dots, n$
- typically $n = 3$
- no closed-form inverse
- may loose monotonicity without requiring equal signs in all k_i — the undistorted image may then fold over itself
- hard to calibrate — higher coefficients tend to dominate

- Zernike orthogonal polynomials R_i^0 are a better choice

$$R_2^0(\rho) = 2\rho^2 - 1, \quad R_4^0(\rho) = 6\rho^4 - 6\rho^2 + 1, \quad R_6^0(\rho) = 20\rho^6 - 30\rho^4 + 12\rho^2 - 1, \cdots$$

- then $D(\mathbf{z}_R; \mathbf{z}_n, \mathbf{k}) = 1 + k_1 R_2^0(\rho) + k_2 R_4^0(\rho) + \cdots + k_n R_{2n}^0(\rho)$
- must know the field of view of the lens in the image plane; since ρ must satisfy $0 \leq \rho \leq 1$
- coefficients k_i will typically decrease in magnitude with increasing i — unlike in the plain polynomial model

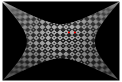
Real and Linear Camera Models

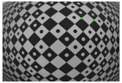


$\underline{\mathbf{x}}_L \simeq \mathbf{A}\mathbf{K}_0\left[\mathbf{R}\ \mathbf{t}\right]\underline{\mathbf{X}}$ (linear camera)

undistortion: $\underline{\mathbf{x}}_L = \mathbf{A}\, r^{-1}(\mathbf{A}^{-1}\underline{\mathbf{x}}_R)$

$\underline{\mathbf{x}}_R \simeq \mathbf{A}\, r\Big(\mathbf{K}_0\left[\mathbf{R}\ \mathbf{t}\right]\underline{\mathbf{X}}\Big)$ (real camera)





$$\mathbf{K}_0 = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ 'ideal' calibration matrix $$\mathbf{A}\mathbf{K}_0 = \begin{bmatrix} f & s\,f & u_0 \\ 0 & a\,f & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & s & u_0 \\ 0 & a & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$ everything affecting radial distortion — center, skew, aspect ratio

r radial distortion function — including the conversion from/to homogeneous representation!

- assumption: the principal point and the center of radial distortion coincide
- f included in $\mathbf{K}_0$ to make radial distortion independent of focal length
- $\mathbf{A}$ makes radial lens distortion an elliptic image distortion
- it suffices in practice that r^{-1} is an analytic function (r need not be)

Calibrating Radial Distortion

- radial distortion calibration includes at least 5 parameters: $\theta = (\lambda,\, u_0,\, v_0,\, s,\, a)$
- we may asume ORUA: $s = 0,\ a = 1$

Alg:

1. detect a set of straight line segment images $\{s_i\}_{i=1}^{n}$ from a calibration target — checkerboard patterns have many
2. select a suitable set of k measurement points per segment — checkerboard: given, in general: how to select k?
3. define (rotation/translation-) invariant radial transfer error per measurement point $e_{i,j}$ in segment i:

$$e_i^2(\theta) = \sum_{j=1}^{k-2} e_{i,j}^2(\theta) \qquad \text{eg. line fit residual (closed form)}$$

4. minimize total radial transfer error <u>while preserving y_n</u> — to avoid collapse to a point

$$\arg \min_{\theta=(\lambda,\, u_0,\, v_0,\, s,\, a)} \sum_{i=1}^{n} e_i^2(\theta) \qquad \text{s.t.} \quad y_n \text{ preserved}$$

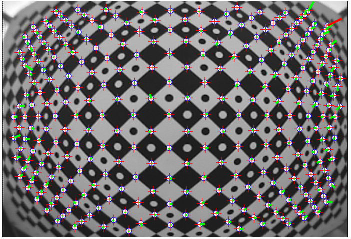
- line segments from real-world images requires segmentation to inliers/outliers — inliers = lines that are straight in reality
- marginalisation over the hidden inlier/outlier label gives a 'robust' error, e.g. → Slide 121

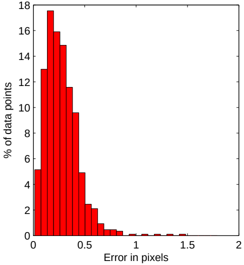
$$\varepsilon_i^2 = -\log\left(e^{-\frac{e_i^2}{2\sigma^2}} + t\right), \qquad t > 0$$

- direct optimization usually suffices but in general such optimization is unstable

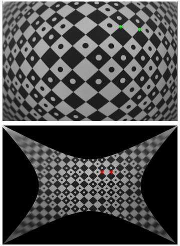
Low-resolution (VGA) wide field-of-view (130°) camera

Camera 0, im. 6: Reprojection errors (16x)

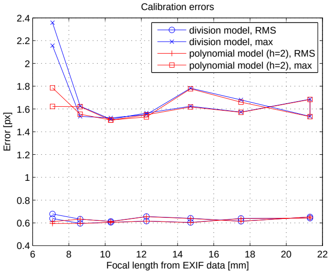


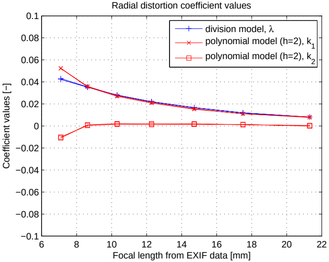


Cam	0
RMS [px]	0.33
max [px]	1.97
f [px]	94.59
a [-]	1.0951
u_0 [px]	243.26
v_0 [px]	353.37
(poly) k_1	$+0.8256$
k_2	-0.2261
k_3	$+1.2524$



4 Mpix consumer camera with a zoom





- polynomial model suffices for greater focal lengths
- above: alternating signs and similar-magnitude coefficients k_i are a sign of a low efficiency of the plain polynomial model
- below: radial distortion is slightly dependend on focal length

A Summary of This Course Highlights

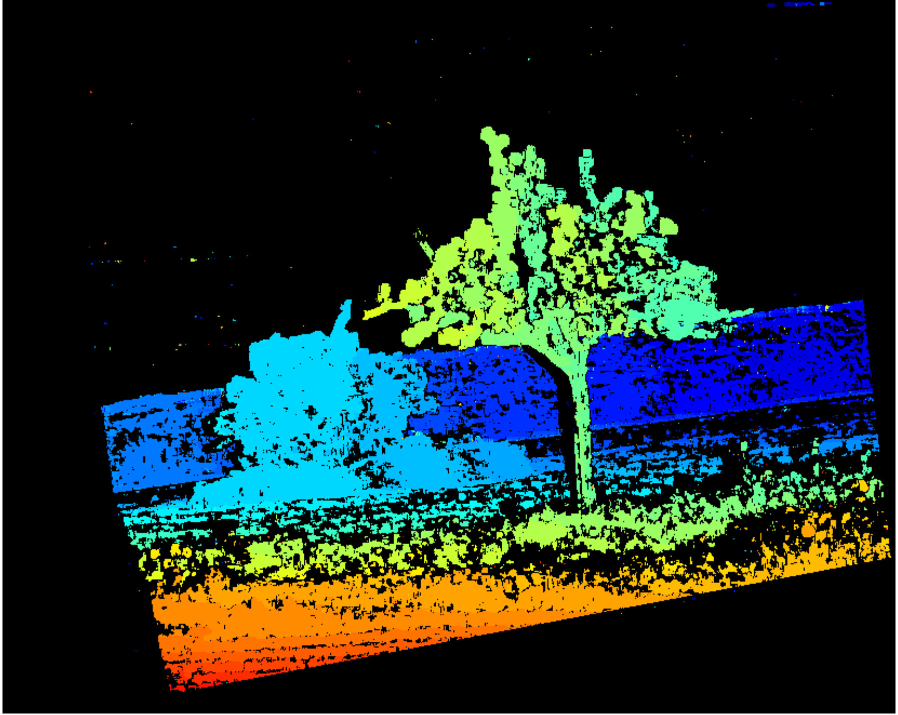
- homography as a two-image model
- epipolar geometry as a two-image model
- core algorithms for 3D vision:
  - simple intrinsic calibration methods
  - 6-pt alg for camera resection and 3-pt alg for exterior orientation (calibrated resection)
  - 7-pt alg for fundamental matrix, 5-pt alg for essential matrix
  - essential matrix decomposition to rotation and translation
  - efficient accurate triangulation
  - robust matching by RANSAC sampling
  - camera system reconstruction
  - efficient bundle adjustment
  - stereoscopic matching basics
- statistical robustness as a way to work with partially unknown information

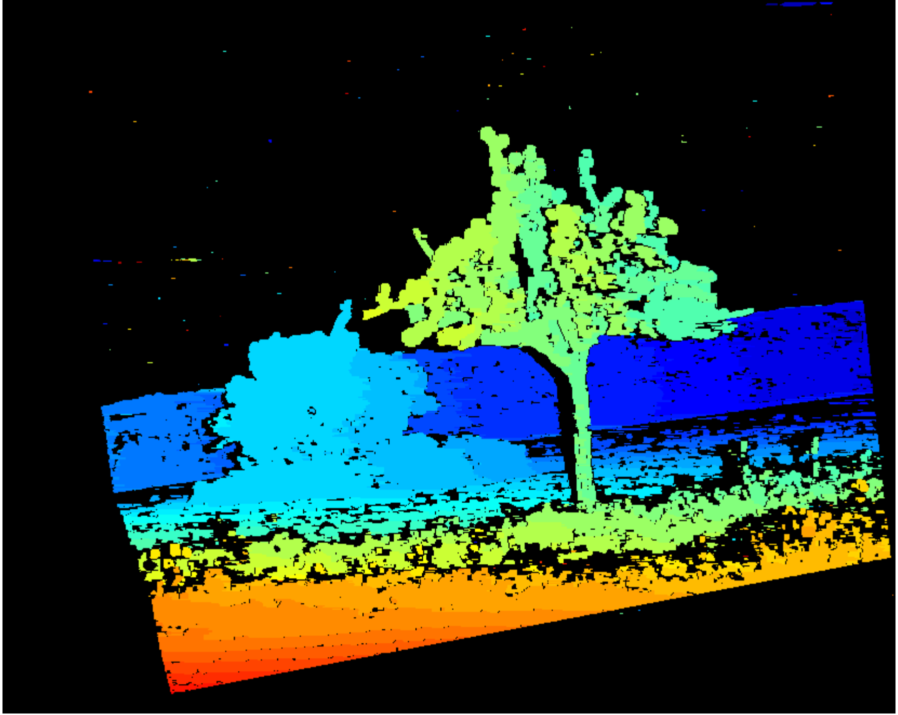
What can we do with these tools?

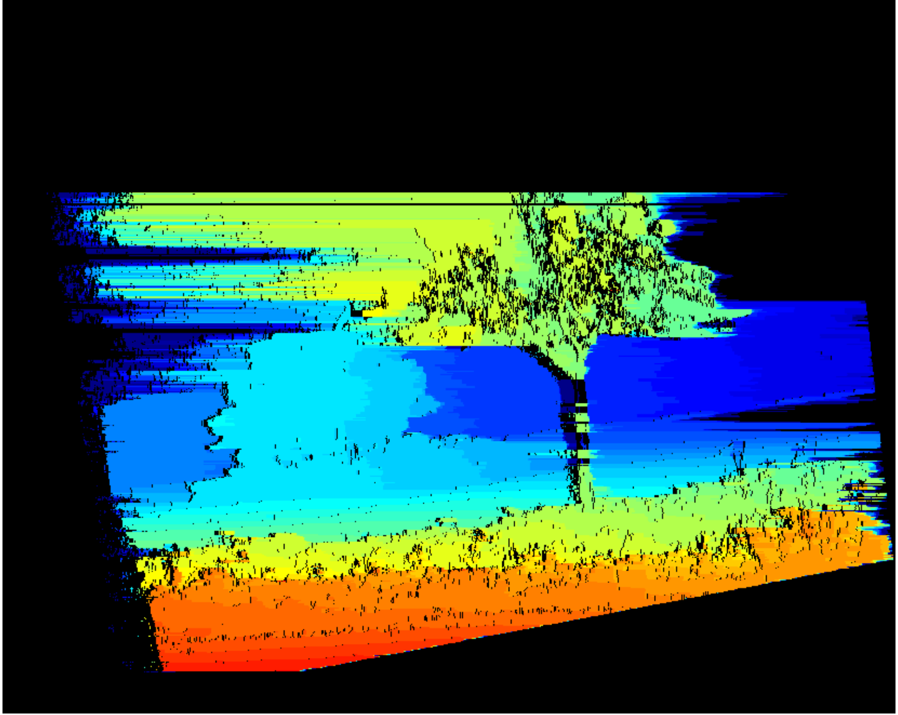
- perspective image rectification
- 3D scene reconstruction
- motion capture
- visual odometry
- robotic self-localization and mapping (SLAM) for navigation and motion planning

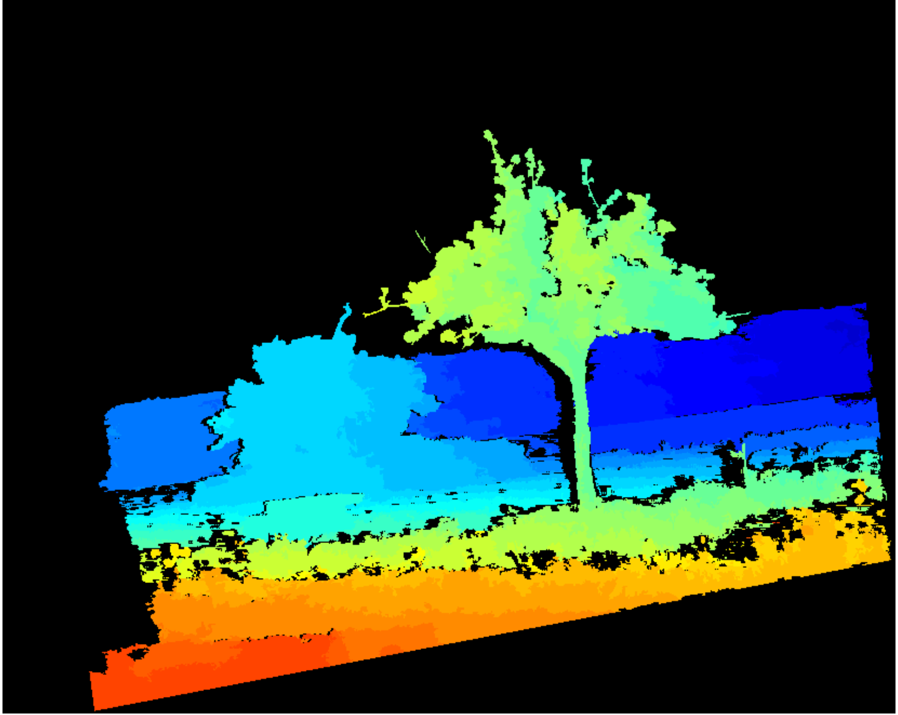
we did not cover 3D aggregation in scene maps

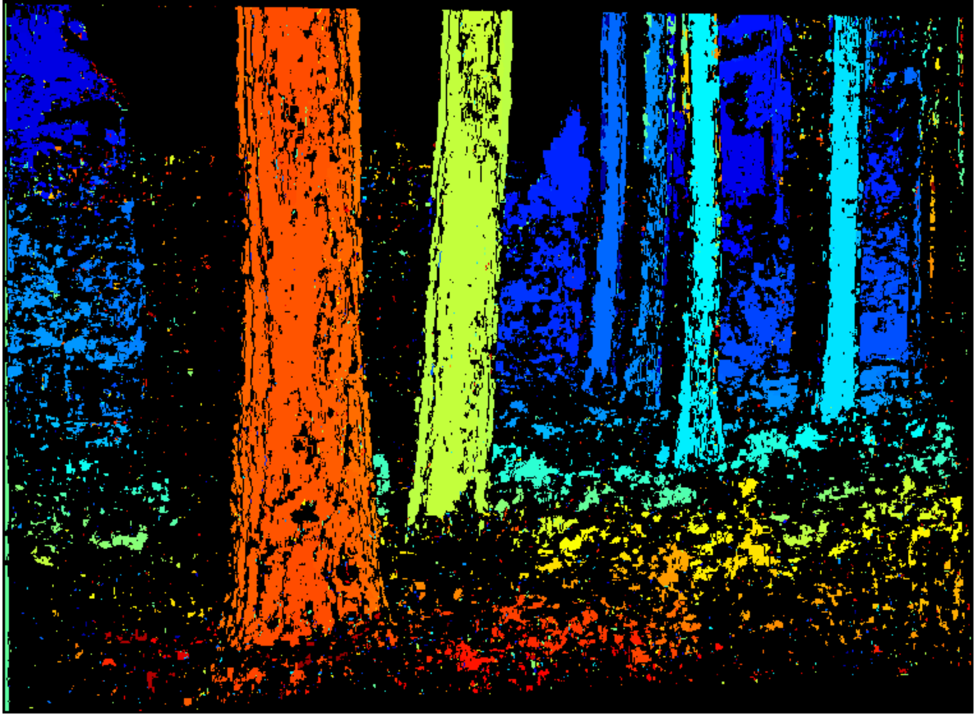
Thank You

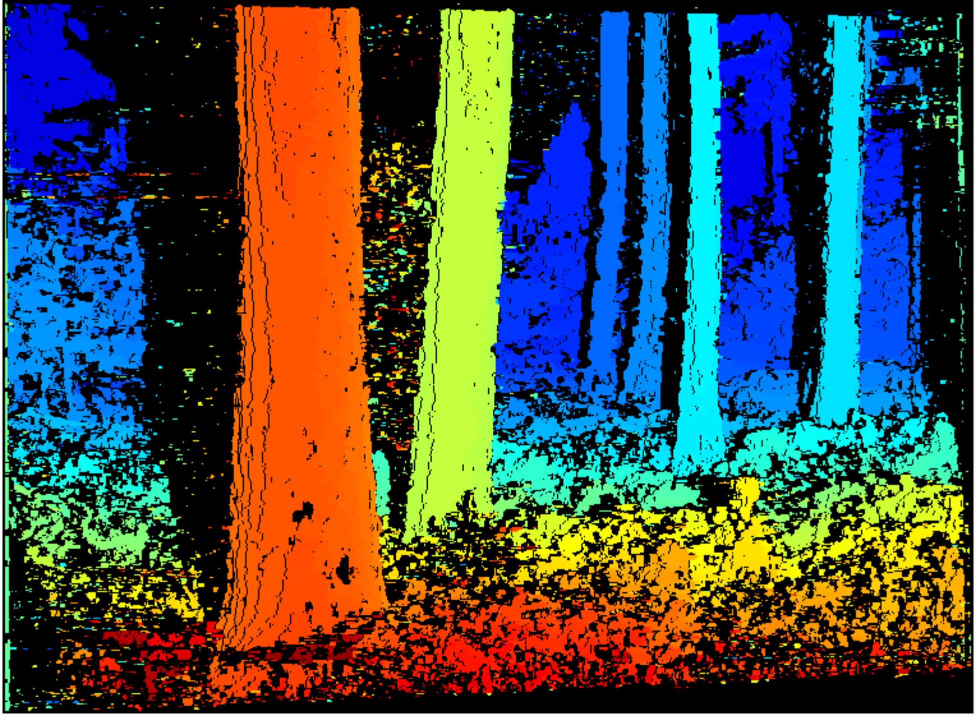


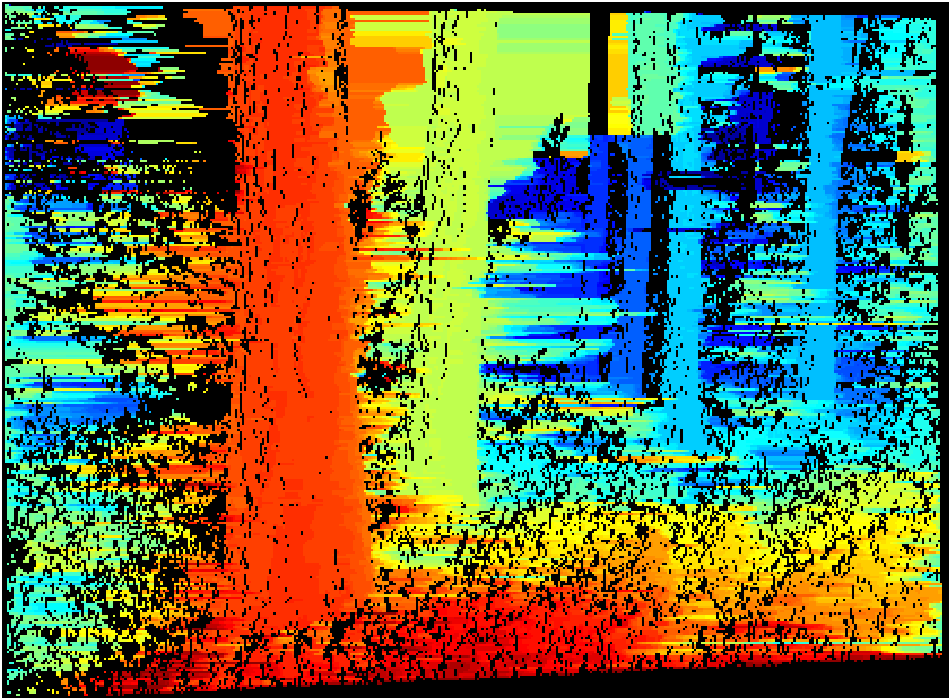


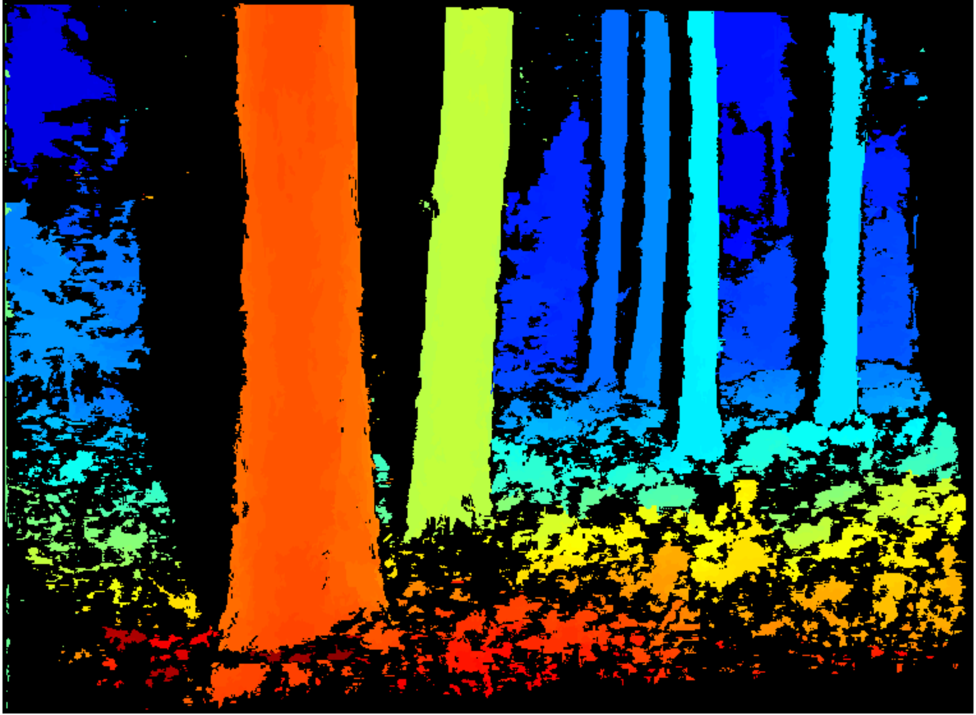


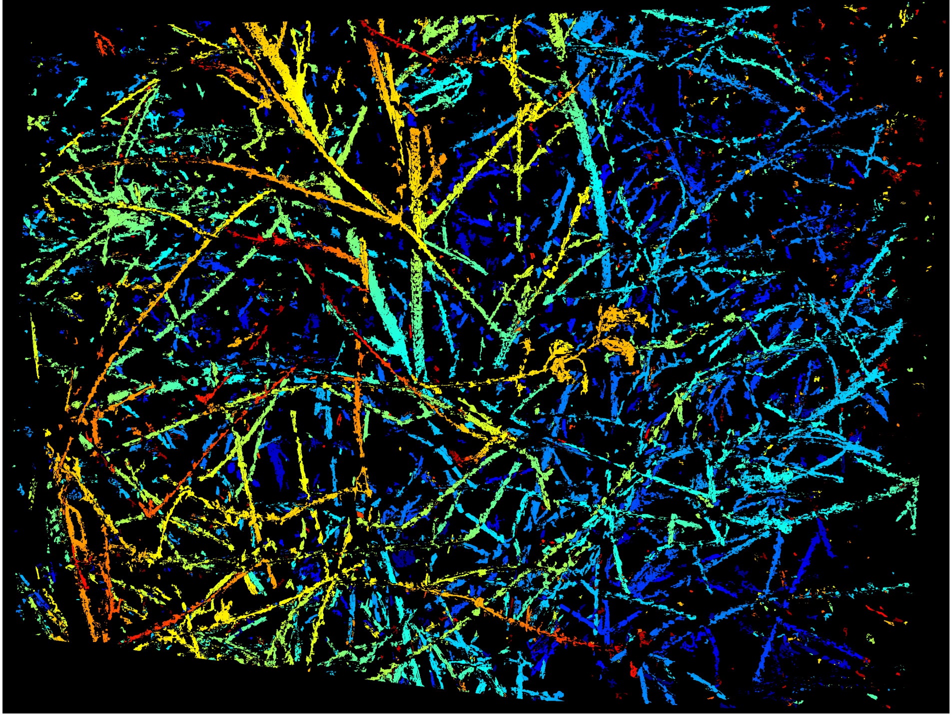


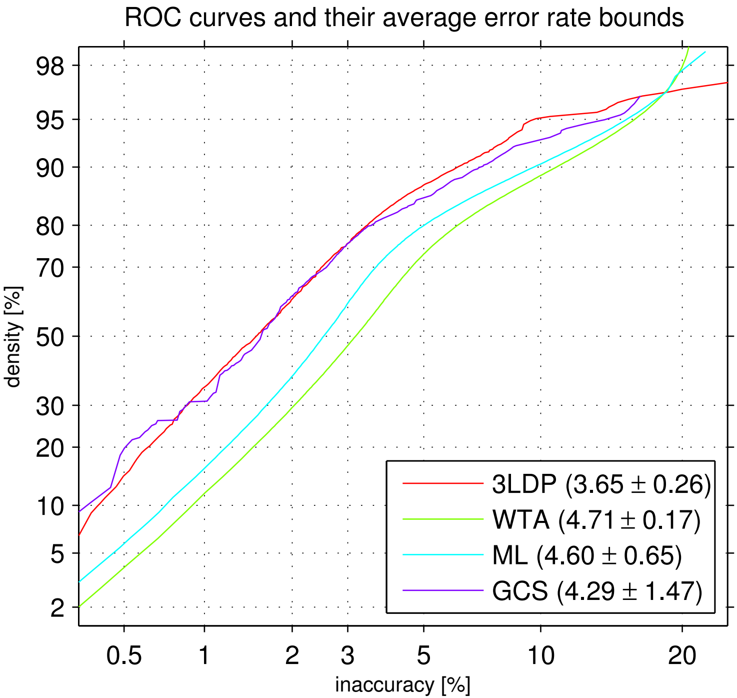




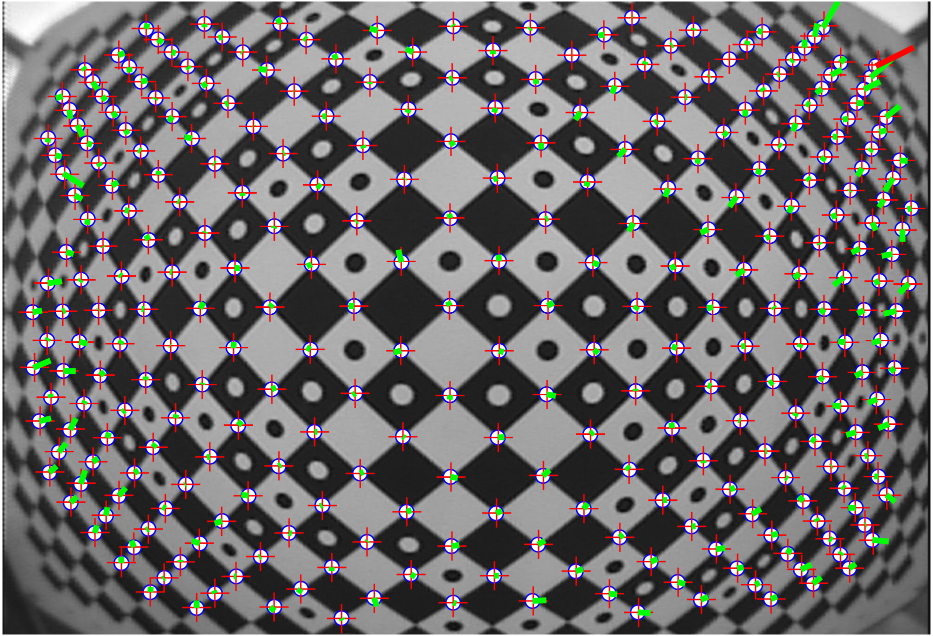


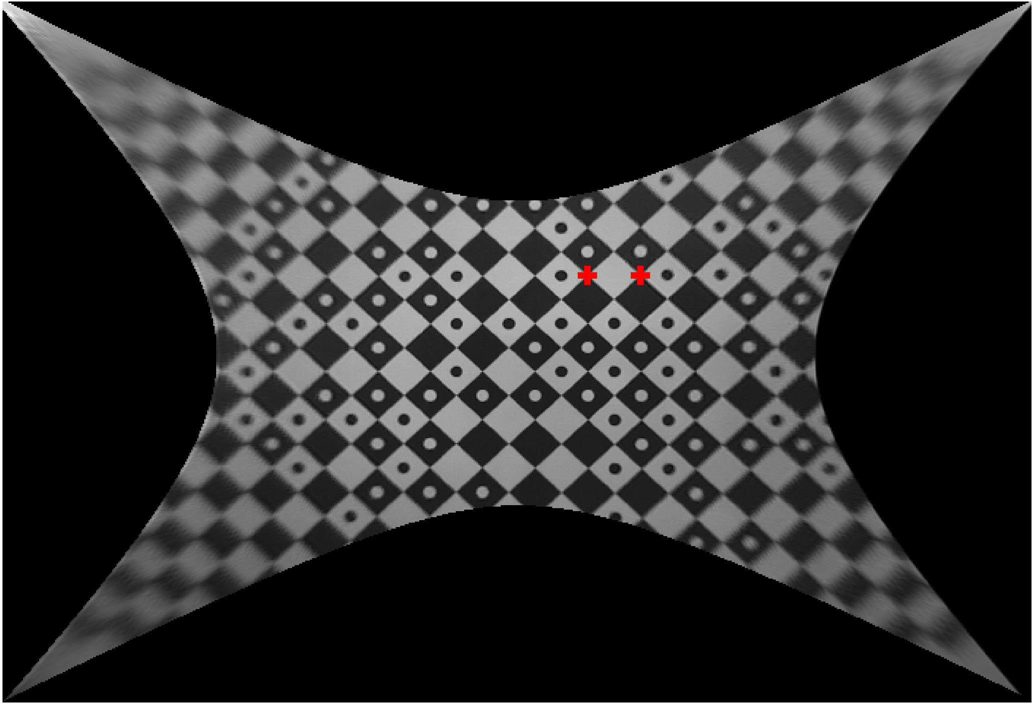


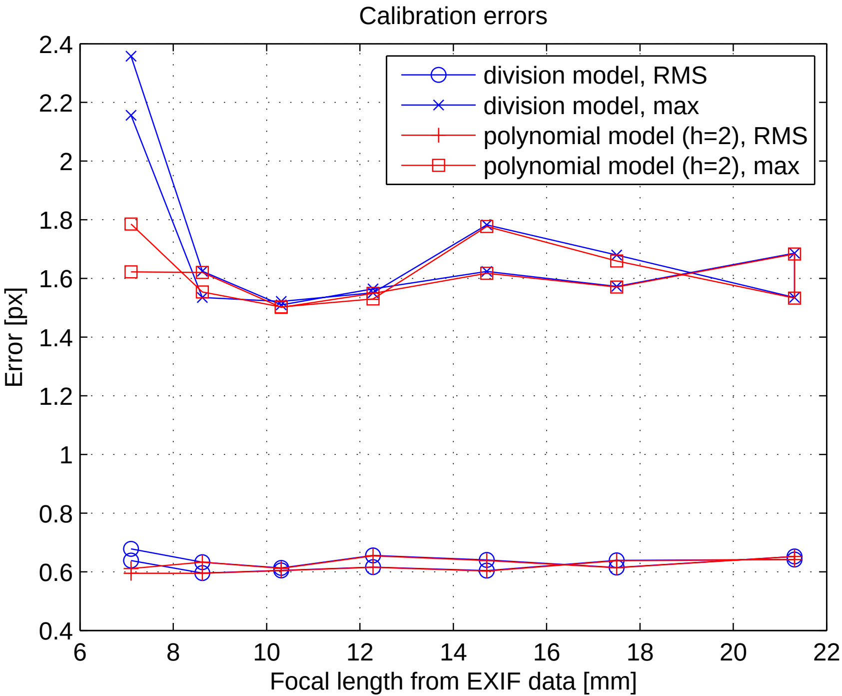



ROC curves and their average error rate bounds
98
95
90
80
70
50
30
20
10
5
2
density [%]
0.5
1
2
3
5
10
20
inaccuracy [%]
3LDP (3.65 ± 0.26)
WTA (4.71 ± 0.17)
ML (4.60 ± 0.65)
GCS (4.29 ± 1.47)

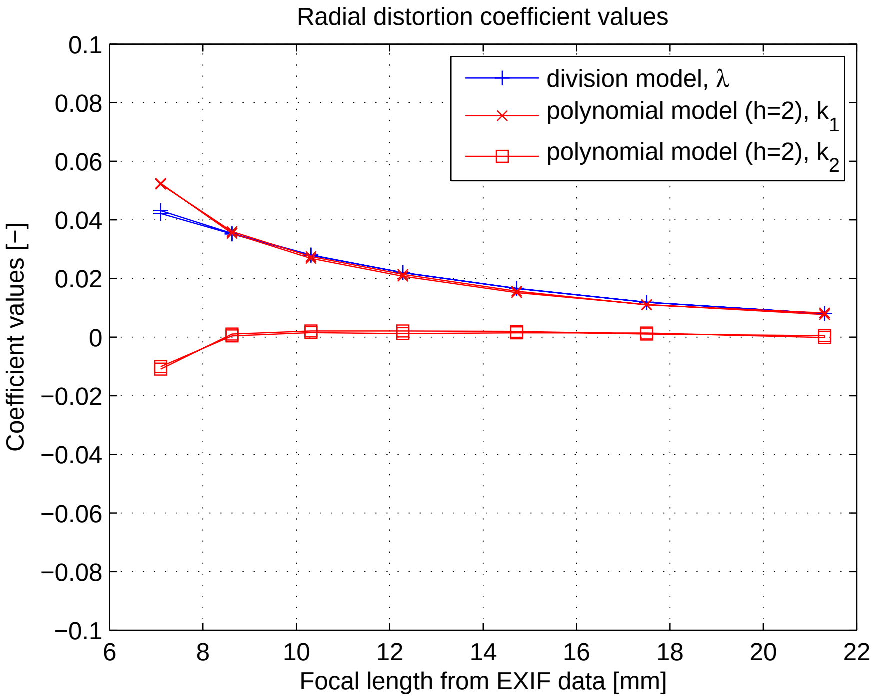
Camera 0, im. 6: Reprojection errors (16x)







Calibration errors
division model, RMS
division model, max
polynomial model (h=2), RMS
polynomial model (h=2), max
Error [px]
Focal length from EXIF data [mm]
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
6
8
10
12
14
16
18
20
22



Radial distortion coefficient values
division model, λ
polynomial model (h=2), k1
polynomial model (h=2), k2
0.1
0.08
0.06
0.04
0.02
0
−0.02
−0.04
−0.06
−0.08
−0.1
Coefficient values [−]
6
8
10
12
14
16
18
20
22
Focal length from EXIF data [mm]