Overview of my research: image geometry and stereo

Pascal Monasse (pascal.monasse@enpc.fr)

LIGM/Imagine, École des Ponts ParisTech

May 31, 2023

My resume

- PhD from Université Paris Dauphine (2000)
- HDR from ENS Cachan (2013)
- 2001–2007: Researcher at Cognitech Inc., Pasadena, California
- 2007–2008: Researcher at ENS Cachan
- 2008–: Researcher (Professor since 2015) at LIGM, École des Ponts ParisTech
- Editor: IPOL, JMIV
- Publications: 1 book, 2 book chapters, 3 patents, 25 journal, 35 conference
- Regular reviewer: PAMI, TIP, CVPR, ICCV, ECCV, BMVC, 3DV...
- PhD students: 16 (14 graduates, 2 ongoing)
- Teaching: ENPC (C++ programming, algorithmics), MVA (3D computer vision)

Level lines of images

- Definition: if $u : \Omega \to \mathbb{R}$, we define $[u \geq \lambda] := \mathcal{X}^\lambda u$, $[u \leq \lambda] := \mathcal{X}_\lambda u$
- Monotone: $\lambda \to \mathcal{X}^\lambda u$ decreases and $\lambda \to \mathcal{X}_\lambda u$ increases from $(\mathbb{R}, \leq) \to (\mathcal{P}(\Omega), \subset)$
- Complete: reconstruction of u from its level sets

$$u(x) = \sup\{\lambda | x \in \mathcal{X}^\lambda\} = \inf\{\lambda | x \in \mathcal{X}_\lambda\}$$

Level lines of images

- Definition: if $u : \Omega \to \mathbb{R}$, we define $[u \geq \lambda] := \mathcal{X}^\lambda u$, $[u \leq \lambda] := \mathcal{X}_\lambda u$
- Monotone: $\lambda \to \mathcal{X}^\lambda u$ decreases and $\lambda \to \mathcal{X}_\lambda u$ increases from $(\mathbb{R}, \leq) \to (\mathcal{P}(\Omega), \subset)$
- Complete: reconstruction of u from its level sets

$$u(x) = \sup\{\lambda | x \in \mathcal{X}^\lambda\} = \inf\{\lambda | x \in \mathcal{X}_\lambda\}$$

Inversely:

Theorem (Reconstruction)

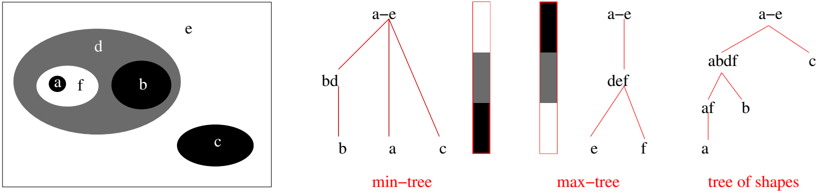
- *Let $X : \lambda \to X^\lambda$ decreasing (or $X : \lambda \to X_\lambda$ increasing), define u by: $u(x) = \sup\{\lambda | x \in X^\lambda\}$ (or $\inf\{\lambda | x \in X_\lambda\}$)*
- *Then $X^\lambda = \mathcal{X}^\lambda u$ or $X_\lambda = \mathcal{X}_\lambda u$*
- *Supplementary assumption of semi-continuity when $X(\mathbb{R})$ is infinite: $X^\lambda = \bigcap_{\mu > \lambda} X^\mu$ or $X_\lambda = \bigcap_{\mu < \lambda} X_\mu$*

Theorem (Ballester, Caselles, PM, COCV2003)

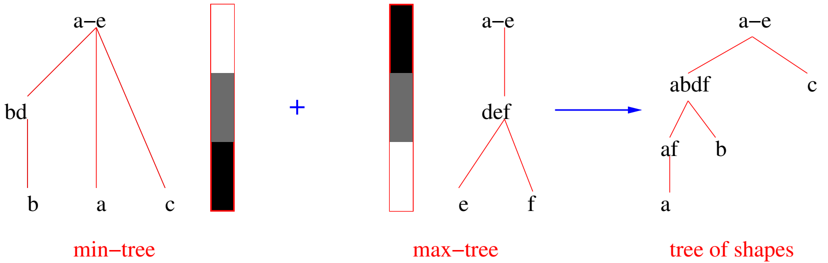
- *Call shapes the saturations $(Sat(X),\ X \in \mathcal{CC}([u \geq \lambda]))$ and $(Sat(X),\ X \in \mathcal{CC}([u < \lambda]))$*
- *The set of shapes has an inclusion tree structure*

Remarks:

- We need a topological property of Ω, the unicoherency
- u is upper semi-continuous, ensuring each shape is open or closed
- In discrete setting, $<$ is replaced by $\leq$

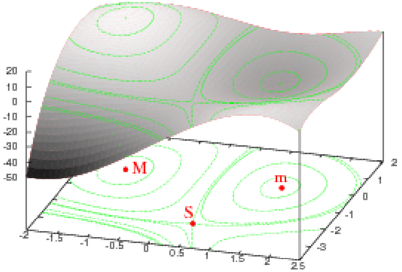


Fusion of component trees [Caselles, Meinhardt, PM, *Positivity* 2007]



- The most natural method to compute the tree of shapes.
- Works in any dimension.
- But may not be optimal...

Level lines analysis

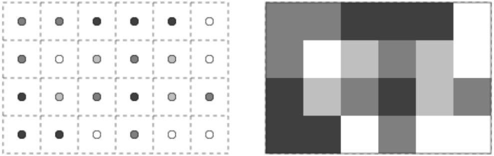


Singular points:

- M: maximum
- m: minimum
- S: saddle

- We can study the topological variation of level lines according to level (Morse theory)
- Morse function: C^2 with full rank Hessian at singular points
- This model is too constraining for images (no plateaux for example)
- Extension to images: [Caselles, PM, LNCS 2010]

Semi-continuous and continuous interpolation



Semi-continuous (nearest neighbor with SCS condition)

Continuous (order 1 spline, bilinear)

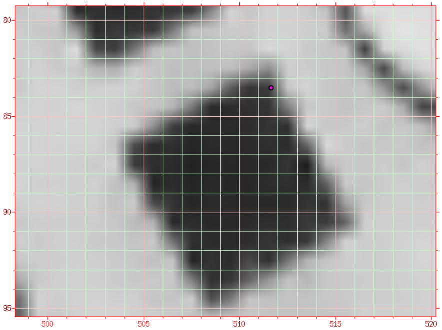
- Nearest neighbor is contrast invariant, but pixelized.
- Bilinear is not contrast invariant, but less pixelized.

Tree of bilinear level lines

- Between two adjacent pixels, linear variation, thus a single point at level λ
- We follow the level line from dual pixel to dual pixel.

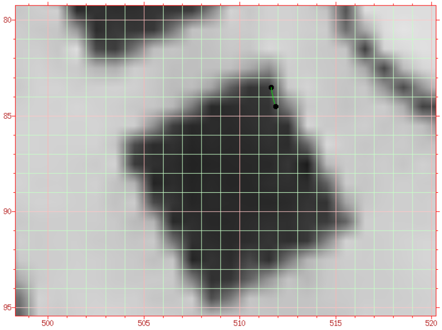
Tree of bilinear level lines

- Between two adjacent pixels, linear variation, thus a single point at level λ
- We follow the level line from dual pixel to dual pixel.



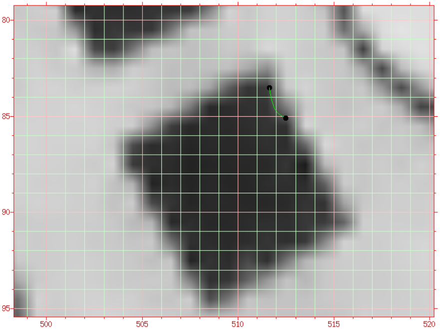
Tree of bilinear level lines

- Between two adjacent pixels, linear variation, thus a single point at level λ
- We follow the level line from dual pixel to dual pixel.



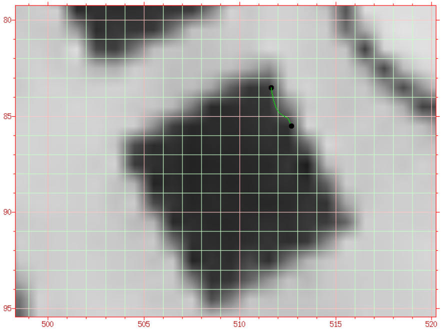
Tree of bilinear level lines

- Between two adjacent pixels, linear variation, thus a single point at level λ
- We follow the level line from dual pixel to dual pixel.



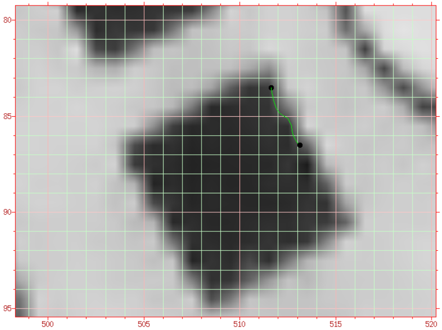
Tree of bilinear level lines

- Between two adjacent pixels, linear variation, thus a single point at level λ
- We follow the level line from dual pixel to dual pixel.



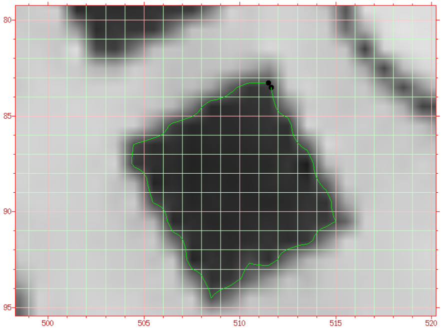
Tree of bilinear level lines

- Between two adjacent pixels, linear variation, thus a single point at level λ
- We follow the level line from dual pixel to dual pixel.



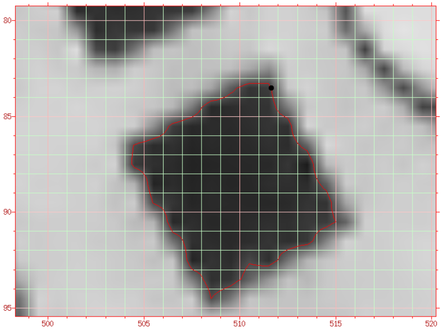
Tree of bilinear level lines

- Between two adjacent pixels, linear variation, thus a single point at level λ
- We follow the level line from dual pixel to dual pixel.



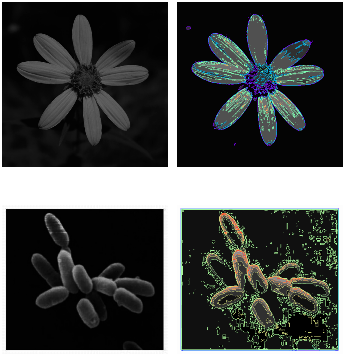
Tree of bilinear level lines

- Between two adjacent pixels, linear variation, thus a single point at level λ
- We follow the level line from dual pixel to dual pixel.



- Tree structure: We sort intersections of level lines with horizontal lines of pixels ($y = k + 1/2$)
- Reconstruction: We fill the interior of each level line by walking the tree in preorder

[PM, Extraction of the level lines of a bilinear image, IPOL 2019]



Curvature computation [Ciomaga, PM, Morel, SIAM MMS 2011]

At a regular point of u, we have two formulas for mean curvature:

1. $\text{curv}(u)(x_0) = \frac{u_{xx}u_y^2 - 2u_{xy}u_x u_y + u_{yy}u_x^2}{(u_x^2+u_y^2)^{3/2}}(x_0)$
2. $|\text{curv}(u)(x_0)| = |x''(s)|$ with $x(s)$ the level line through x_0 and s the curvilinear abscissa

Curvature computation [Ciomaga, PM, Morel, SIAM MMS 2011]

At a regular point of u, we have two formulas for mean curvature:

1. $\mathrm{curv}(u)(x_0) = \frac{u_{xx}u_y^2 - 2u_{xy}u_x u_y + u_{yy}u_x^2}{(u_x^2+u_y^2)^{3/2}}(x_0)$
2. $|\mathrm{curv}(u)(x_0)| = |x''(s)|$ with $x(s)$ the level line through x_0 and s the curvilinear abscissa

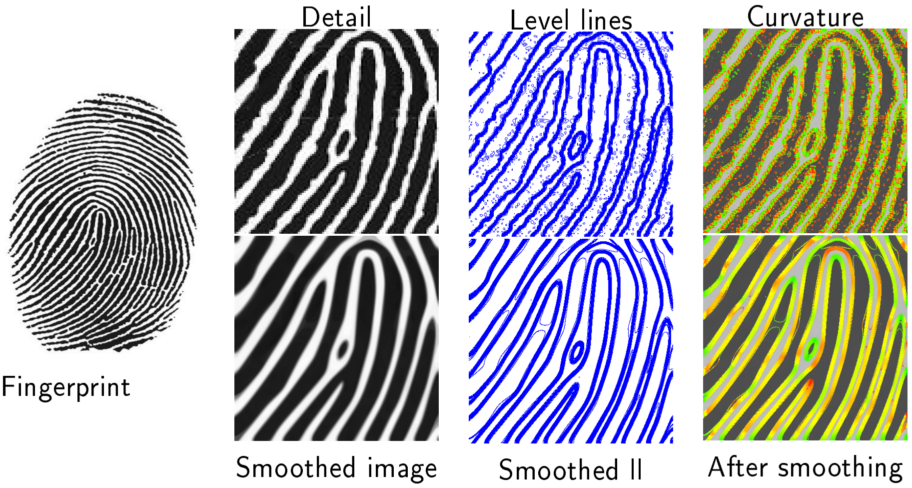
- (1) suggests a finite difference scheme (with Euclidean invariance!)
- (2) is indirect.

The problem is that we always need to smooth slightly the image. For that, we use the mean curvature motion or the affine erosion:

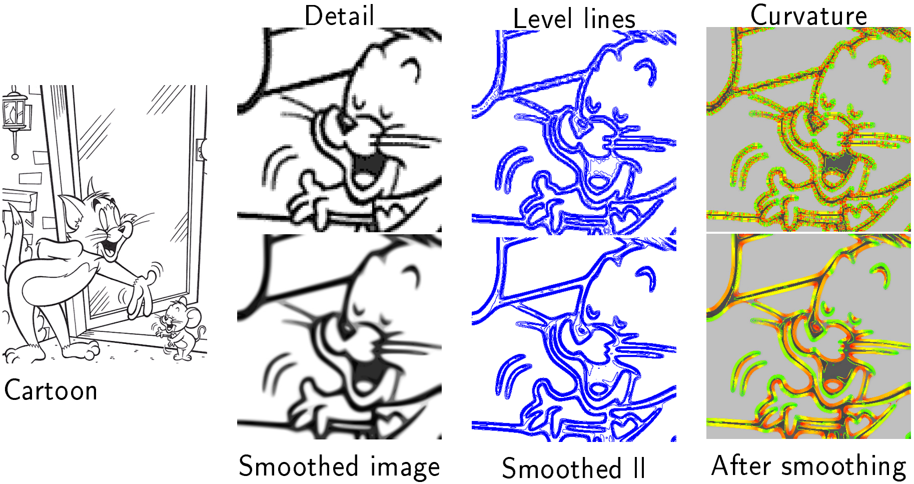
$$\frac{\partial u}{\partial t} = |Du|\,\mathrm{curv}(u)^{\alpha} \quad (\alpha = 1, 1/3)$$

[Alvarez, Guichard, Lions, Morel, Arch. for Rational Mech. 1993]
[Moisan, TIP 1998]

[Ciomaga, PM, Morel, The image curvature microscope, IPOL 2017]

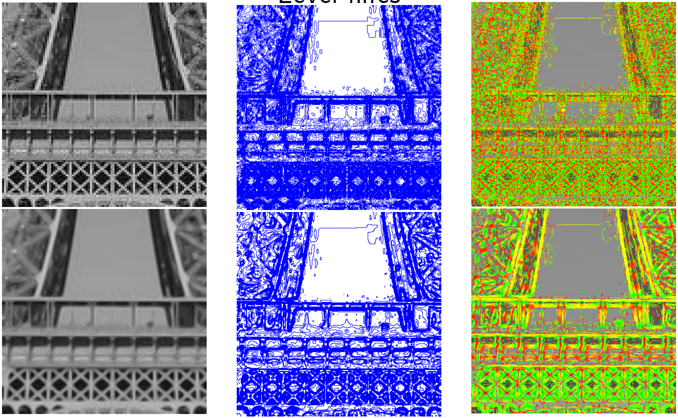


[Ciomaga, PM, Morel, The image curvature microscope, IPOL 2017]



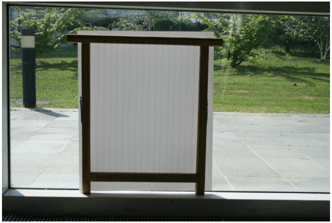
[Ciomaga, PM, Morel, The image curvature microscope, IPOL 2017]

Eiffel tower

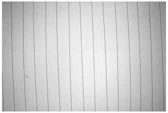
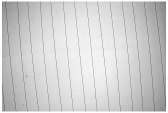
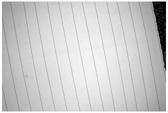
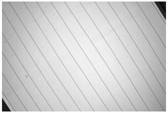


Geometric distortion correction [Grompone, PM, Morel, Tang, ICIP 2010]

- The ideal pinhole camera model is not faithful to reality
- The most detrimental source of error in the 3D pipeline is the geometric distortion of the lens



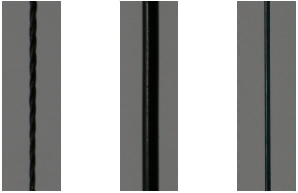
Harp of calibration [Tang, Grompone, PM, Morel, JOSA 2012]

Results measured with harp

Opaque background

Translucent background



(a) (b) (c)

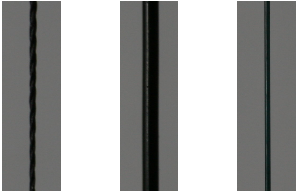
String:

- (a) Sewing
- (b) Tennis
- (c) Fishing

Results measured with harp

Opaque background

Translucent background

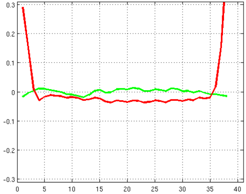


(a) (b) (c)

String:

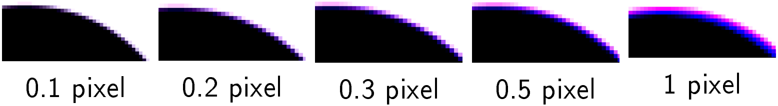
- (a) Sewing
- (b) Tennis
- (c) Fishing

method	d (pixels)	$\bar{d}_{\max}$
original distortion	2.21	6.70
Textured pattern	**0.04**	**0.16**
BA (4 parameters)	0.07	0.30
BA (6 parameters)	0.60	3.00
DxO Optics Pro	0.32	0.99
PTLens	0.46	1.51



Chromatic aberration correction [Rudakova, PM, PSIVT 2013]

There is a shift between the color channels, even with high quality lenses, it must be corrected to subpixel accuracy

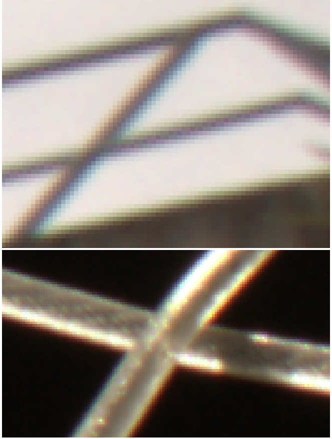


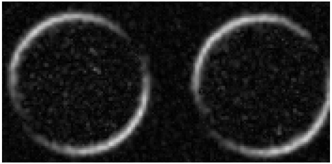
- Decompose image into red, green, blue channels
- Take photograph of a pattern of black disks
- Detect centers of disks precisely
- Register these anchor points between channels
- Interpolate with bivariate polynomial vector field

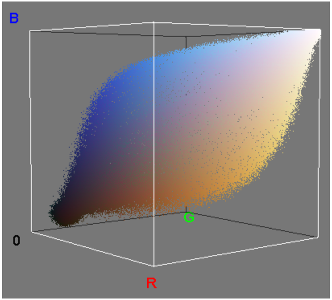
Chromatic aberration correction

Keys of success:

- Precise disk center detection (< 0.01 pixel error) robust to noise and aliasing
- Rich correction model (not simply radial polynomial)





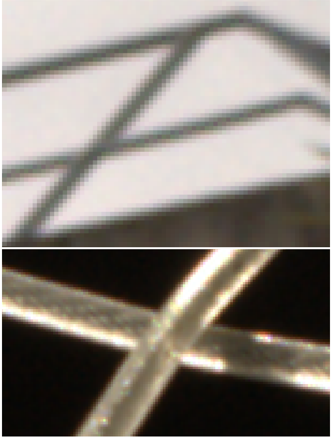


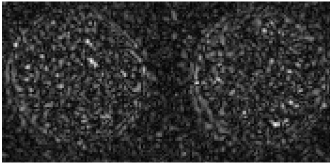
Original

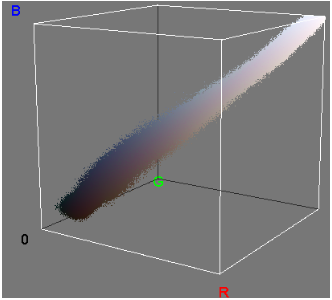
Chromatic aberration correction

Keys of success:

- Precise disk center detection (< 0.01 pixel error) robust to noise and aliasing
- Rich correction model (not simply radial polynomial)







Corrected

a contrario SfM [Moulon, PM, Marlet, ACCV 2012]

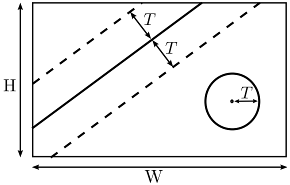
- Helmoltz principle: "Could the observation have occured by chance?"
- Provide a threshold T and a confidence score for the model
- *Background model* $\mathcal{H}_0$: uniform distribution
- Strong deviation from $\mathcal{H}_0$ deemed meaningful
- Given model M, assuming k inliers among n, T_k being the k^{th} smallest residual and $N = N_{\text{sample}}$

$$NFA(M,k) = N_{\text{tests}}(n,k,N)\,\mathbb{P}(\text{residual} \leq T_k | M, \mathcal{H}_0)^{k-N}$$

- Expectation: $NFA(M) = \min_{k=N_{\text{sample}}+1\ldots n} NFA(M,k) \leq 1$

Application to SfM:

- Homography
- Pose/resection
- Essential matrix



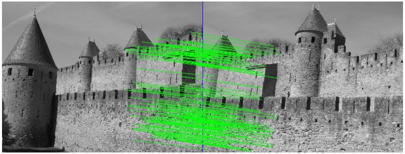
[Moisan, Moulon, PM, Automatic Homographic Registration of a pair of images, with a contrario elimination of outliers, IPOL 2012]

SIFT ratio = 1.1 (all feature points in im1 have their preferred match in im2 and matches that are at most 10% worse)

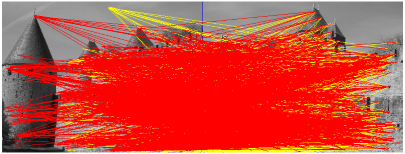
image 1

image 2

panorama



inliers (129=7.5%)



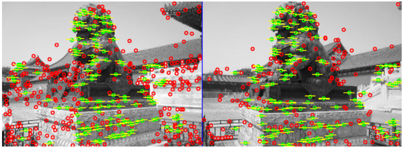
outliers (1553=92.5%)

[Moisan, Moulon, PM, Fundamental matrix of a stereo pair, with a contrario elimination of outliers, IPOL 2016]

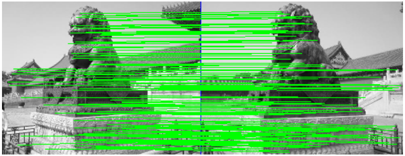
SIFT ratio $= 1.1$ (all feature points in im1 have their preferred match in im2 and matches that are at most 10% worse)

image 1

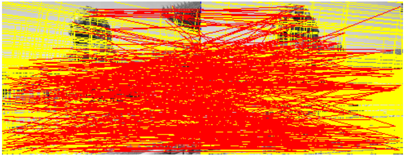
image 2



epipolar lines



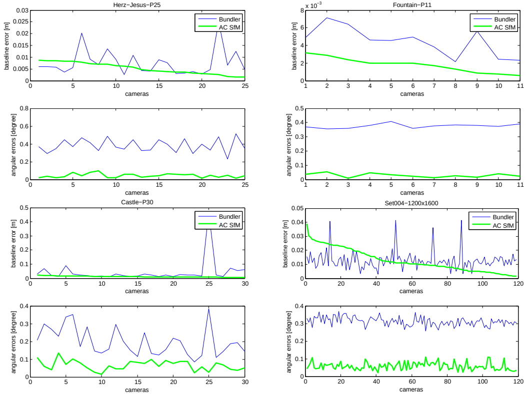
inliers (135=27%)



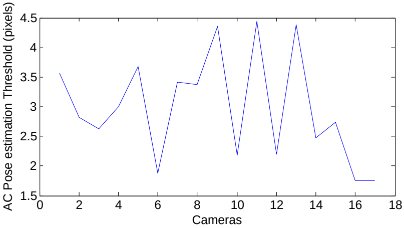
outliers (372=73%)

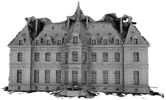
Comparison with Bundler

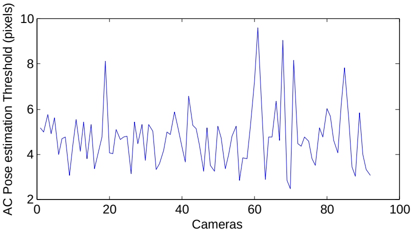
We compare the estimated rotation and translation with respect to the ground-truth.

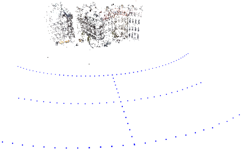


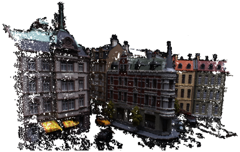
Comparison with Bundler

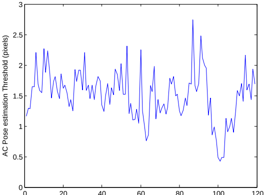












Global calibration [Moulon, PM, Marlet, ICCV 2013]

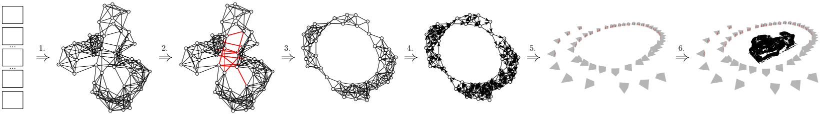
Problems of incremental calibration:

- depends on initial pair and order of image inclusion
- subject to drift
- involves mutliple BAs, therefore fairly slow

Global calibration [Moulon, PM, Marlet, ICCV 2013]

Problems of incremental calibration:

- depends on initial pair and order of image inclusion
- subject to drift
- involves mutliple BAs, therefore fairly slow



1. Estimate relative pairwise rotations R_{ij}
2. Global rotation consistency: eliminate outliers by checking loops
3. Global rotations R_i: $R_i \sim R_{i,j} R_j$
4. Relative translation directions t_{ij}
5. Global translations T_i: $T_j \sim R_{ij} T_i + \lambda t_{ij}$
6. Final structure with bundle adjustment

Comparison with other methods

- Accuracy $\geq$ best global methods (Olsson)
- Speed $\sim$ best incremental methods (Bundler, VisualSfM)

Inaccuracy (mm):

Scene	Global		Inc.	
	Ours	Olsson	Bundler	VSfM
HerzJesusP8	**3.5**	3.9	16.4	19.3
HerzJesusP25	**5.3**	5.7	21.5	22.4
CastleP19	**25.6**	76.2	344	258
CaslteP30	**21.9**	66.8	300	522

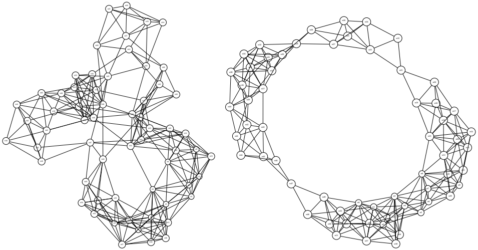
Running time (s):

Scene	Global		Inc.	
	Ours	Olsson	Bundler	VSfM
HerzJesusP8	**2**	34	10	2
HerzJesusP25	**10**	221	100	12
CastleP19	**6**	99	78	9
CaslteP30	**14**	317	300	18

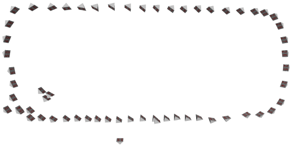
Examples of 3D reconstructions

Sceaux Orangerie

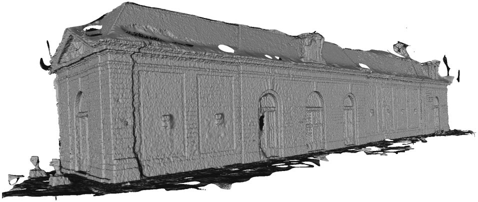
Bundler calibration



Initial and cleaned up visibility graphs



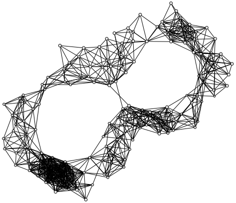
Global calibration

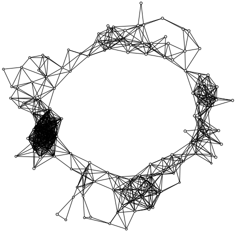


Reconstruction

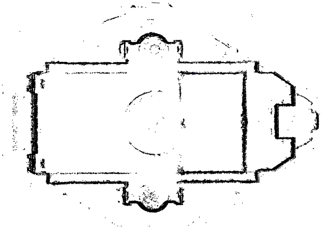
Examples of 3D reconstructions

Opera Garnier

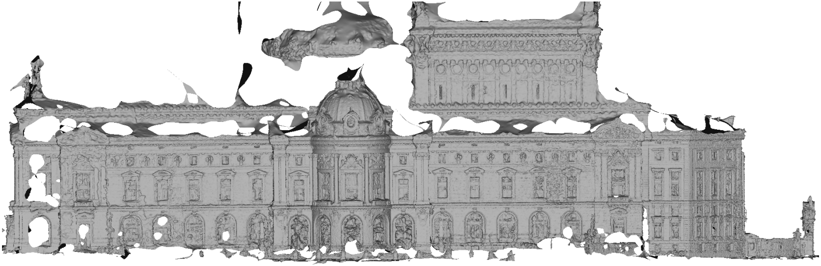




Initial and cleaned up visibility graphs



Cameras and 3D points

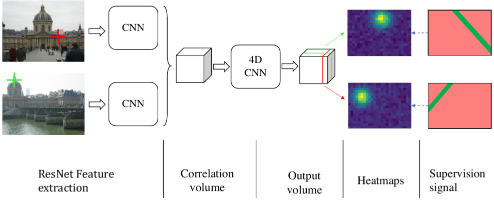


Reconstruction

Reproduction of stereo algorithms in IPOL

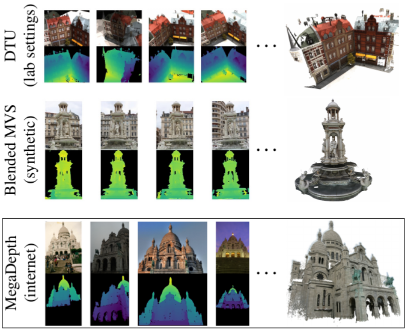
- Calibration: [Julià, PM, Pierrot-Deseilligny, The orthographic projection model for pose calibration of long focal images, 2019]
- Stereo pair rectification:
  - [PM, Quasi-Euclidean epipolar rectification, 2011]
  - [Darmon, PM, The polar epipolar rectification, 2021]
- Disparity map:
  - [Kolmogorov, PM, Tan, Kolmogorov and Zabih's graph cuts stereo matching algorithm, 2014]
  - [Tan, PM, Stereo disparity through cost aggregation with guided filter, 2014]
  - [Julià, PM, Bilaterally weighted patches for disparity map computation, 2015]
- RANSAC:
  - [Moisan, Moulon, PM, Automatic Homographic Registration of a pair of images, 2012]
  - [—, Fundamental matrix of a stereo pair, 2016]
  - [Riu, Nozick, PM, Automatic RANSAC by likelihood maximization, 2022]

[Darmon, Aubry, PM, Learning to guide local feature matches, 3DV 2020]



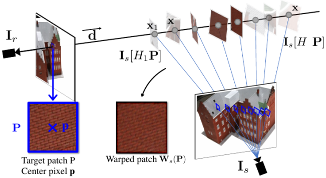
[Darmon, Bascle, Devaux, PM, Aubry, Deep multi-view stereo gone wild, 3DV 2021]

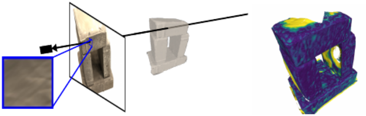
- Tricks for training unsupervised approaches on data in the wild
- Supervised depthmap-based MVS methods are SOTA with few internet images.
- Benchmark of MVSNet and variants



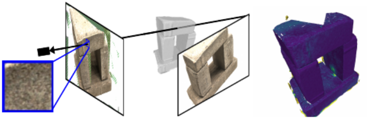
[Darmon, Bascle, Devaux, PM, Aubry, Improving neural implicit surfaces geometry with patch warping, CVPR 2022]

- NeRF methods unable to learn and render high frequency textures.
- Add patch warping and SSIM comparison





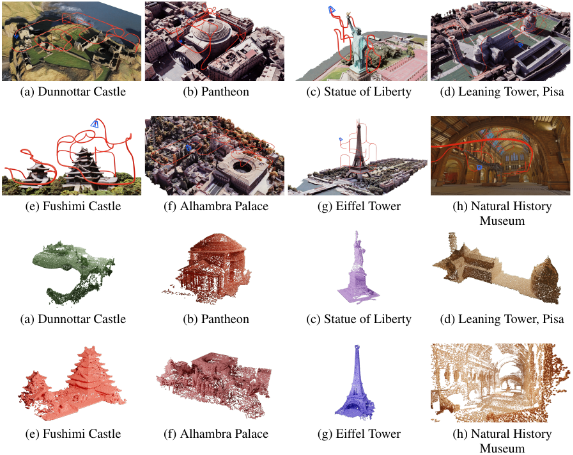
VolSDF and error map



ours

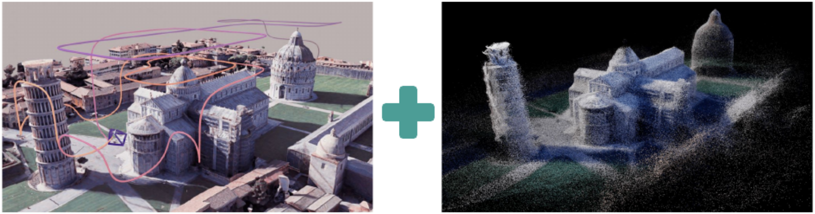
[Guédon, PM, Lepetit, SCONE: surface coverage optimization in unknown environments by volumetric integration, NeurIPS 2022]

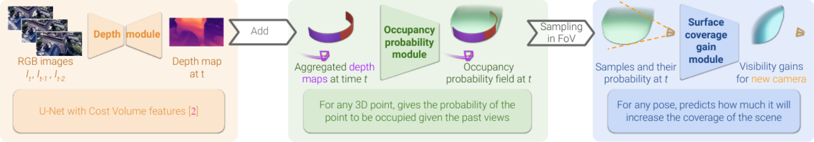
RGB-D camera → next best view



[Guédon, Monnier, PM, Lepetit, MACARONS: mapping and coverage anticipation with RGB online self-supervision, CVPR 2023]

RGB camera $\rightarrow$ Exploration+reconstruction

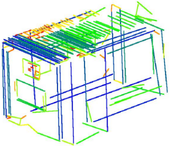
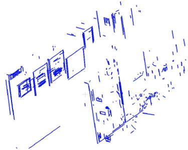




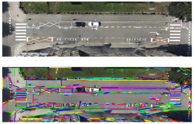
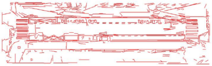
Multi-modal registration based on segments

[Djahel, PM, Vallet, A 3D segments based algorithm for heterogeneous data registration, ISPRS 2022]

- Indoor LiDAR/images


- Aerial/terrestrial images

We need long segments, hence our current work on multi-scale LSD

Other IPOL contributions

- [Briand, PM, Theory and practice of image B-spline interpolation, IPOL 2018]

image

order 1

order 3

order 5

order 11

- [Gay, Lecoutre, Menouret, Morillon, PM, Bilateral K-means for superpixel computation (the SLIC method), IPOL 2022]