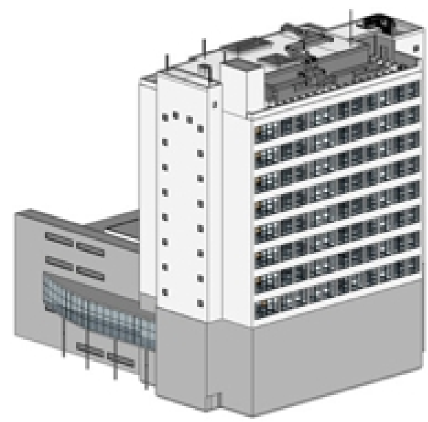
Labex Bézout meeting CEA LIST



Reconstructing digital models of buildings and cities in the IMAGINE group

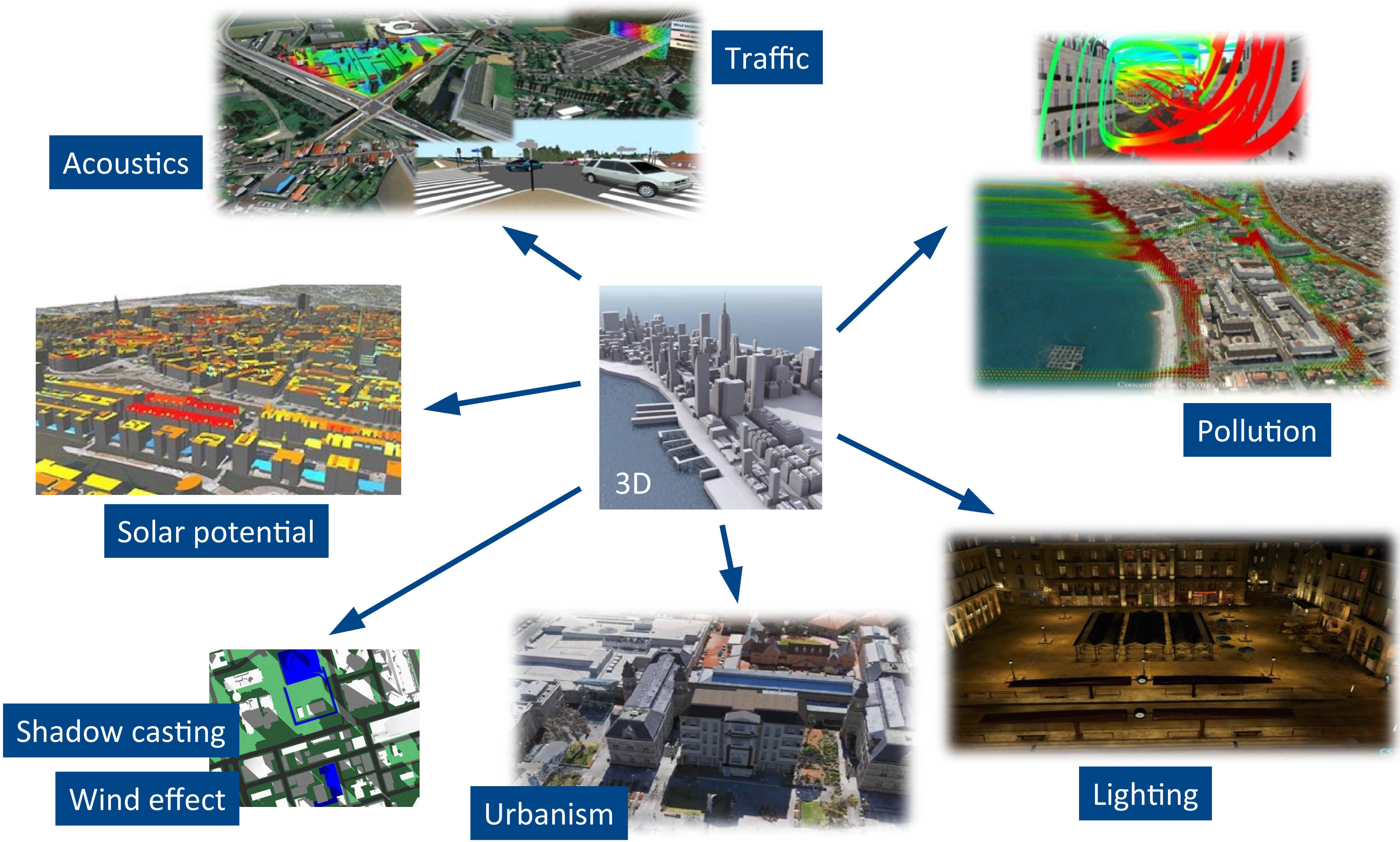
Renaud Marlet (LIGM/ENPC)

3/12/2015

Overview

- Motivation
- Background in the IMAGINE group
- Overview of recent results
- Perspectives

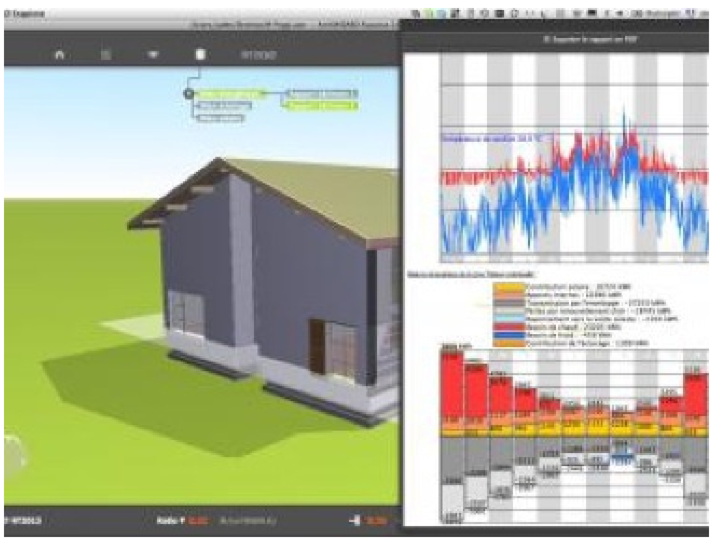
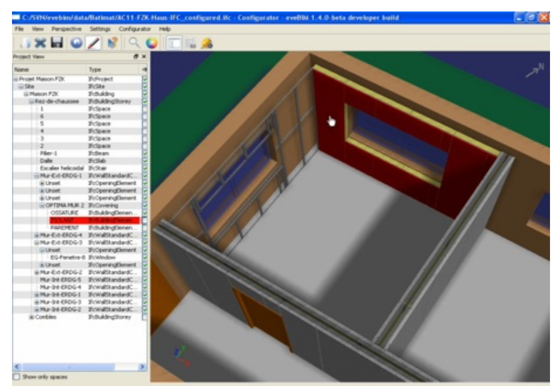
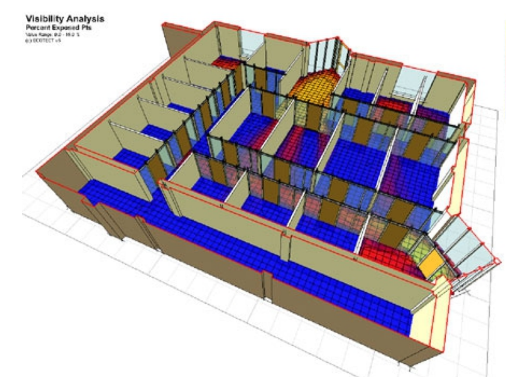
(Semantized) 3D models for cities



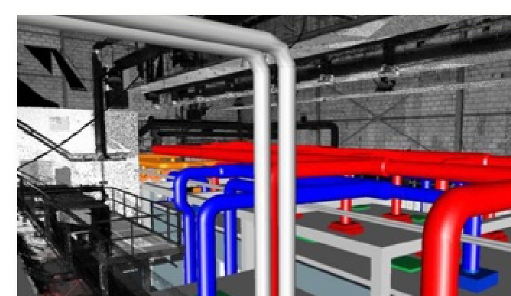
Semantized 3D models for buildings

Building Information Model (BIM)

- Planing, bidding
  - accurate, reliable evaluation of costs and performance
    - energy consumption, acoustics, lighting, security, regulations...
    - simulation, optimization

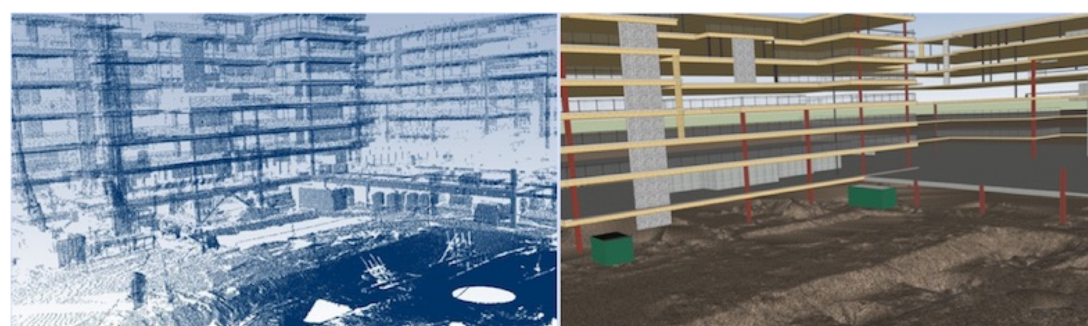
- Organizing
  - collision detection, coordination...



Semantized 3D models for buildings (BIM)

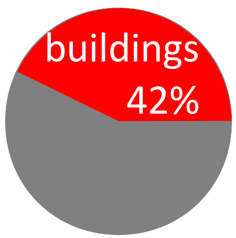
- Virtual presentation
  - decision-makers, customers, population

- Monitoring
  - augmented reality at construction site
  - progress/conformity analysis

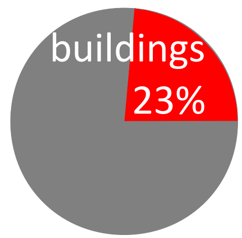


Renovation market: High stakes for energy savings

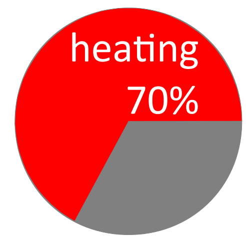
- Housing = largest energy consumer in France



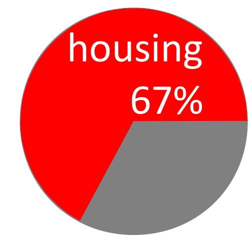
Energy consumption



CO_2 emissions

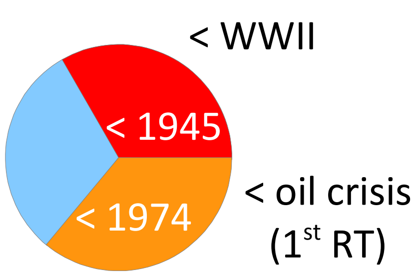


Housing consumption

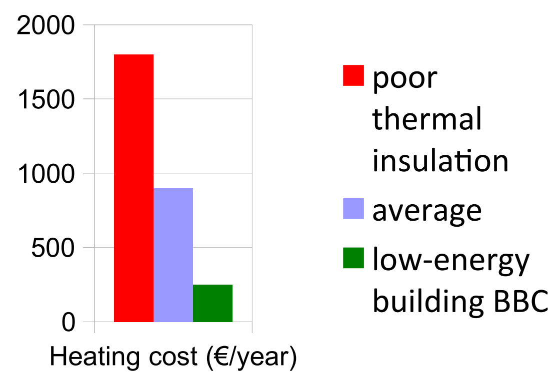


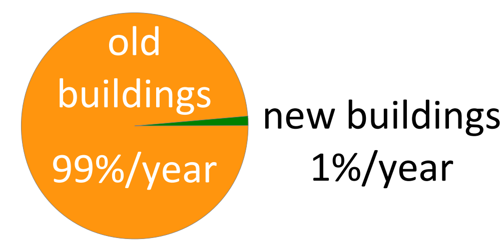
Buildings

- Housing stock = mostly old buildings



Housing stock in France

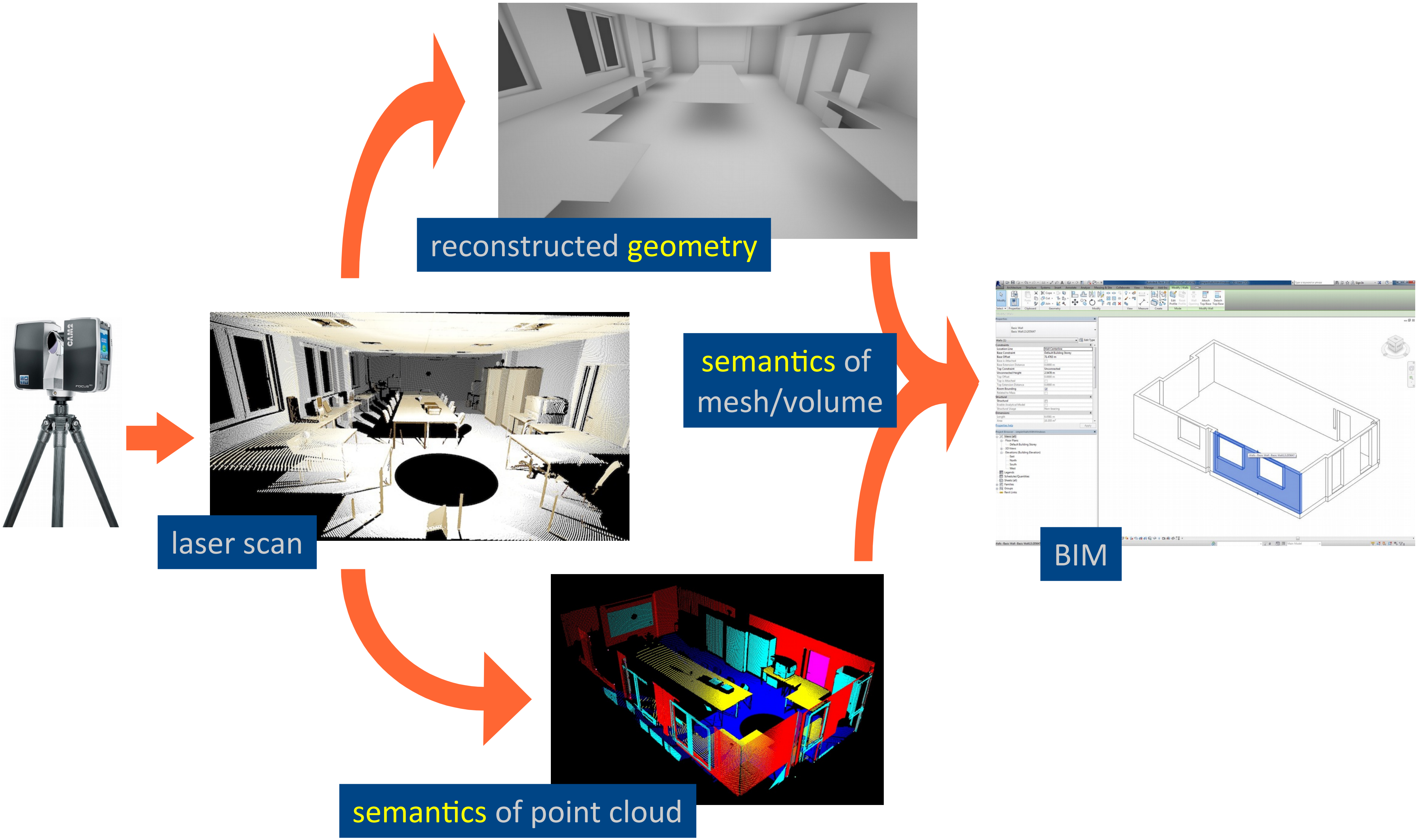




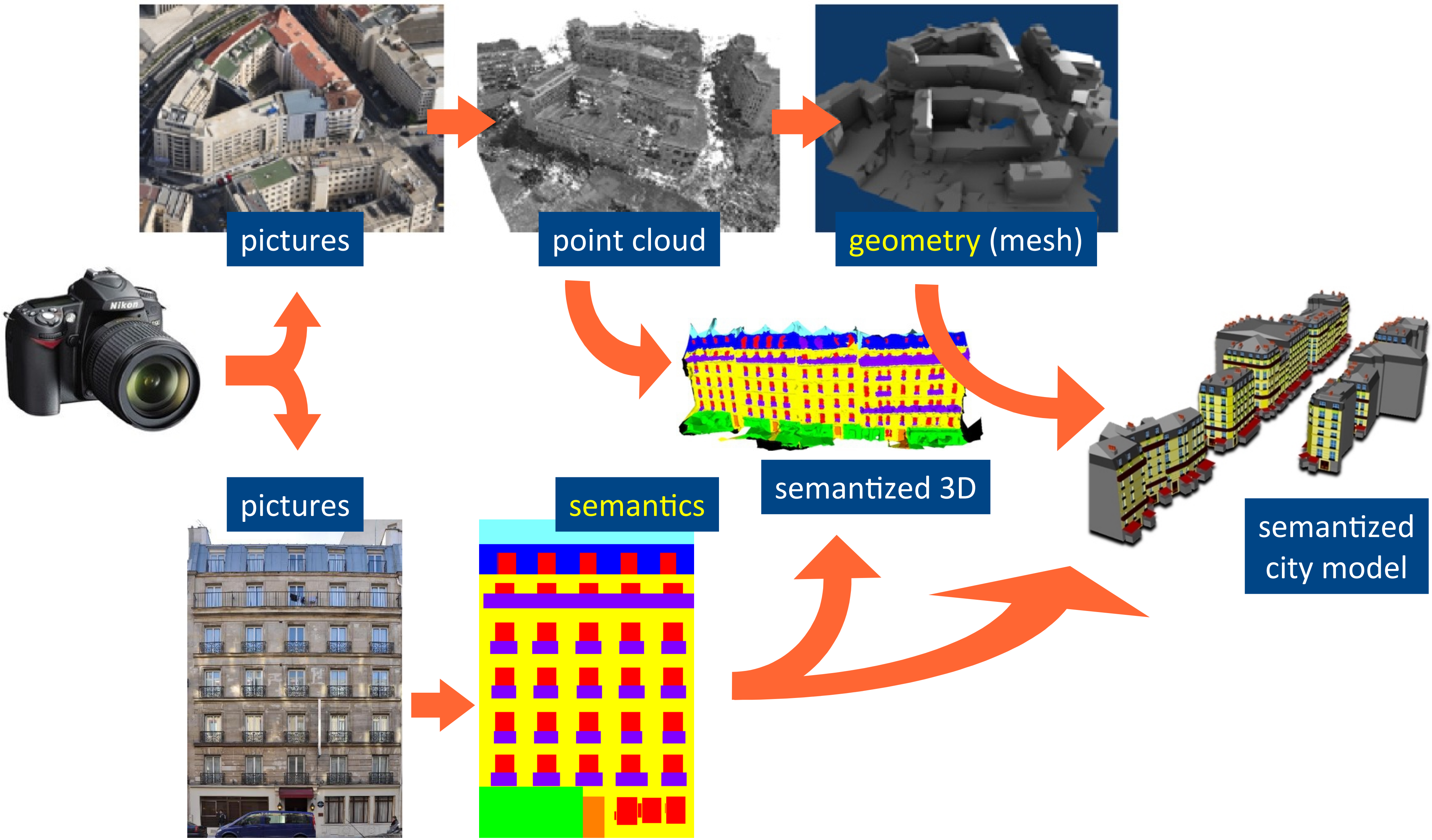
New building construction rate

- RT 2012 (5th thermal regulations) : new buildings only
  300,000/year = 1% ì impact in 50 years unless renovations

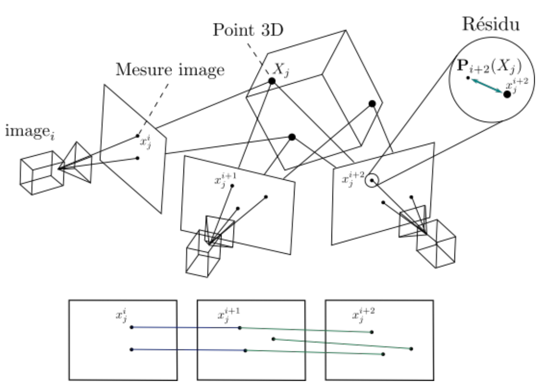
From laser scans to semantized 3D models



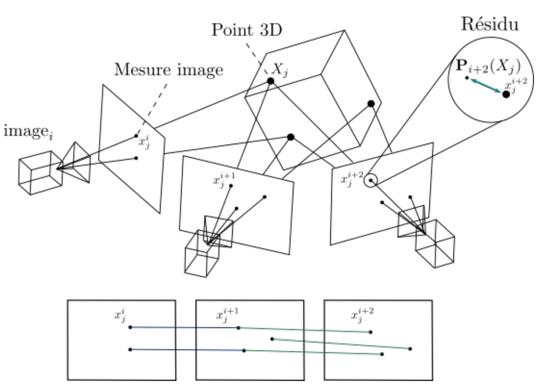
From pictures to semantized 3D models



3D reconstruction from images



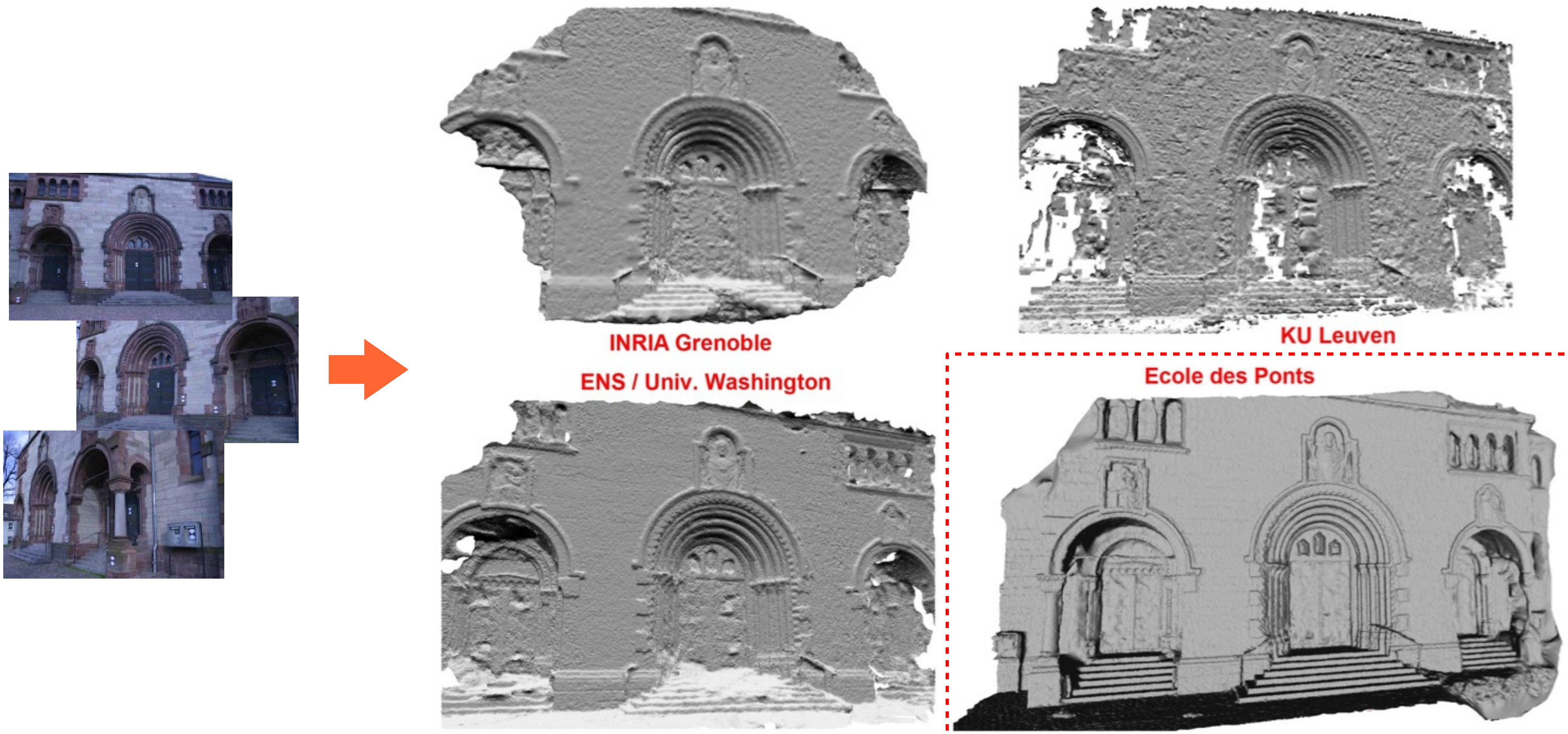
3D reconstruction from images



Performance on international benchmarks: 3D reconstruction

Vu et al., CVPR 2009

- Best results on Strecha et al.'s benchmarks
  - most complete & most accurate models (CVPR 2009)

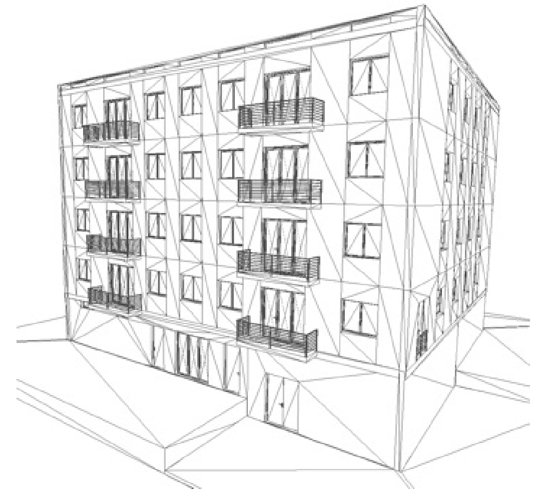


Transfer to startup company Acute3D

- 2011: creation
  - 2 researchers from IMAGINE (R. Keriven, J.-Ph. Pons)
  - 25 man.months of research
  - major contract with Autodesk
- 2015: buyout
  - 10 employees
  - >2M€ sales/year, 90% abroad (China, Japan...)
  - bought by Bentley Systems

Remaining issues

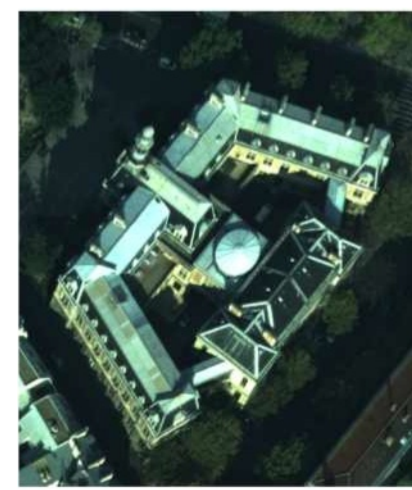
incomplete data
(access difficult, inefficient, expensive...)



geometry only, semantics missing

camera: flexible,
but not accurate, not robust
(low texture, ambiguity...)

laser: accurate,
but little flexibility



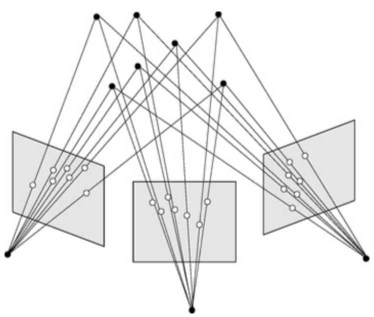
heterogeneous data
(street/air, in/outside, laser+photo)

Adaptive Structure from Motion with a contrario model estimation

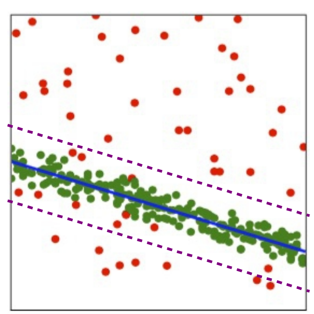
Moulon et al., ACCV 2012

openMVG

- Parameterless, adaptive SfM
  - automatic in/outlier thresholds (whole pipeline: + E, pose)
  - Heltmoltz principle: structure = unlikely coincidence

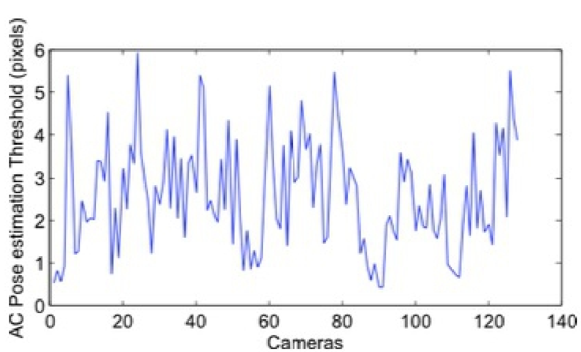


outliers among matches

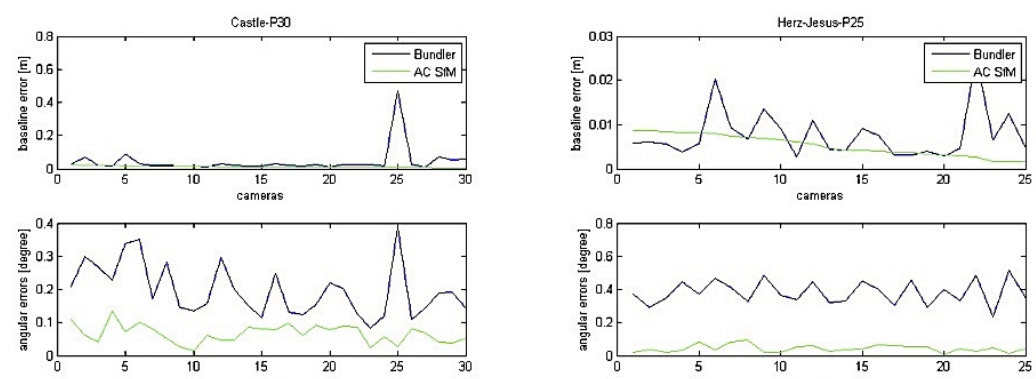


outlier thresholding

Scene		Bundler T_F threshold					AC-SfM T_F threshold			
		1	3	6	9	12	auto	min	med	max
FountainP11	error	0.002	0.003	0.003	0.004	0.005	**0.001**			
	ranking	**1**	3	2	4	5		0.57	1.00	10.5
HerzJesusP8	error	0.004	0.003	0.003	0.007	0.003	**0.001**			
	ranking	4	**1**	3	5	2		0.63	1.88	5.26
HerzJesusP25	error	0.004	0.010	0.005	**0.004**	0.004	0.005			
	ranking	3	5	4	**1**	2		0.23	1.53	82.8
CastleP19	error	8.22	0.029	0.032	0.039	X	**0.015**			
	ranking	4	**1**	2	3	X		0.69	0.91	15.7
CastleP30	error	0.055	0.057	0.043	0.042	0.045	**0.011**			
	ranking	4	5	2	**1**	3		0.55	0.92	284

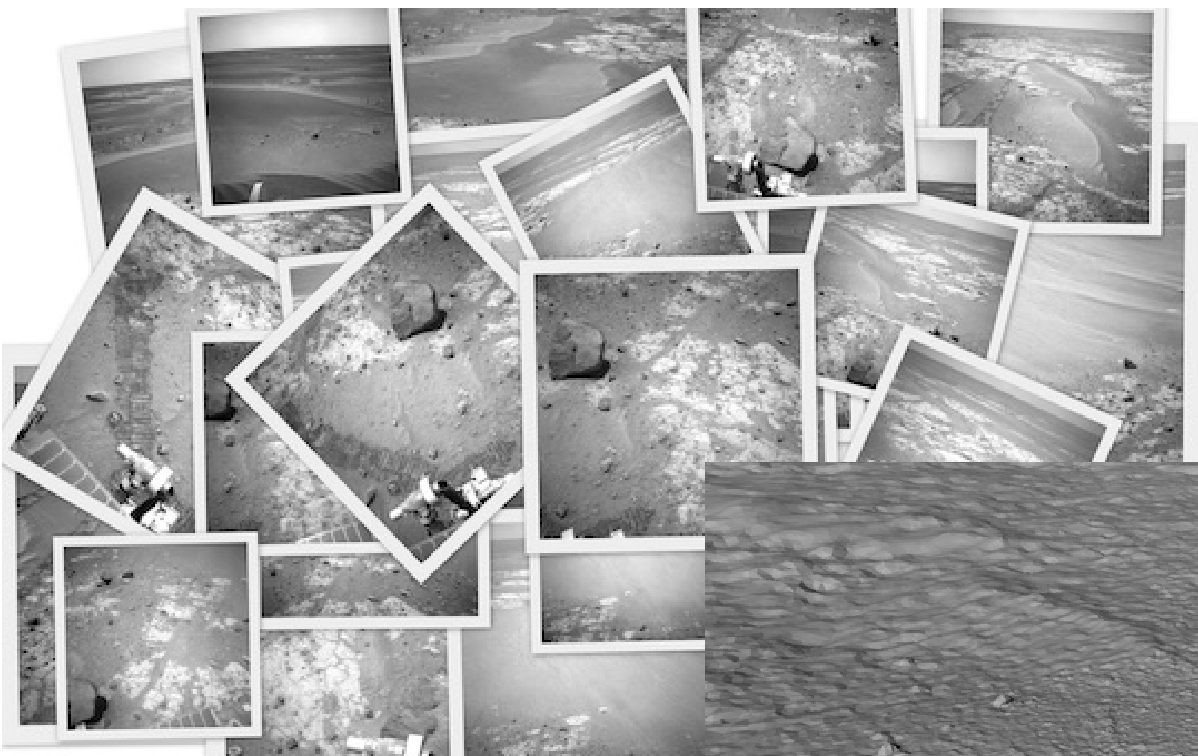


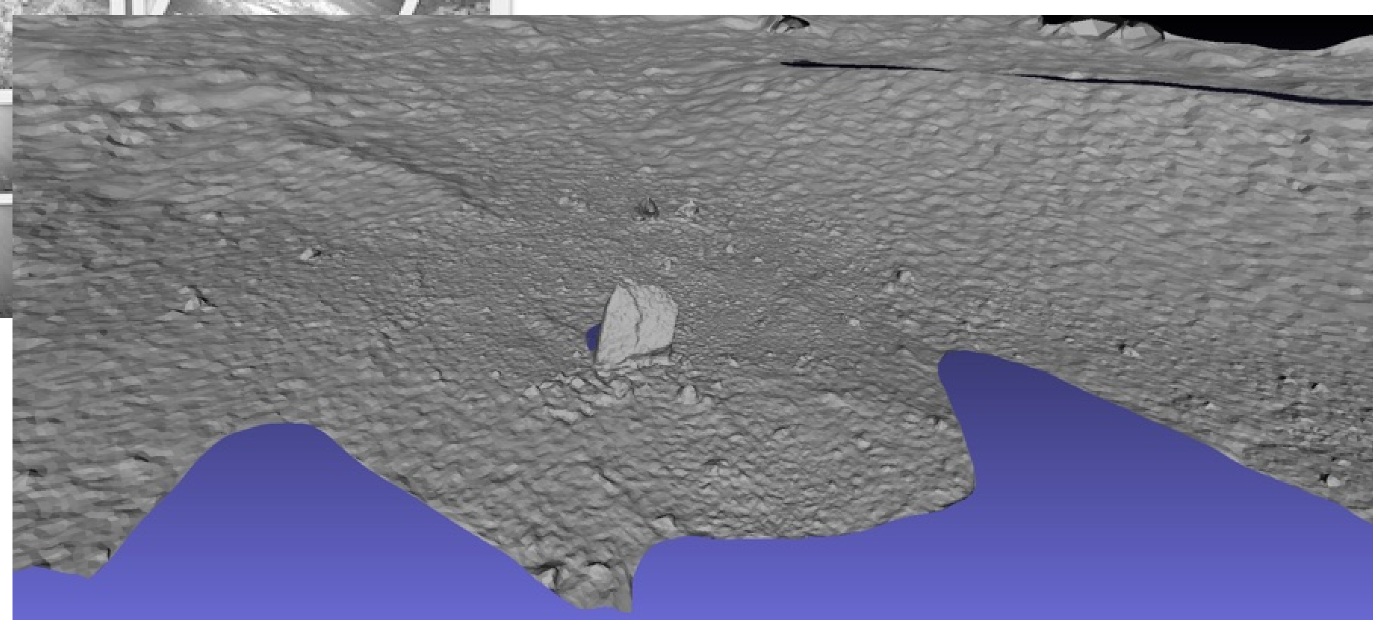
automatic thresholding



Performance on international benchmarks: Calibration & 3D reconstruction

- Winner PRoVisG MARS 3D Challenge (2011)

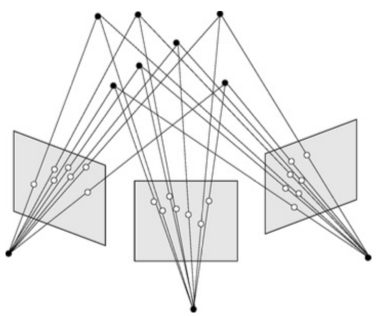




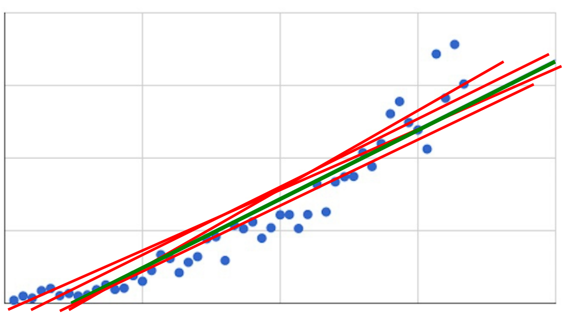
Match selection and refinement for highly accurate 2-view Structure from Motion

Liu et al., ECCV 2014 (oral)

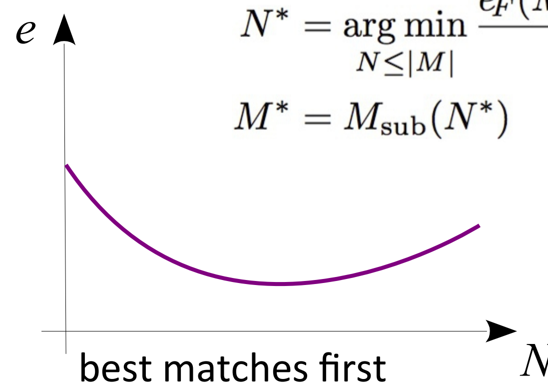
- Fewer data with higher accuracy ?
  Or more data with less accuracy?



matches: varied accuracy



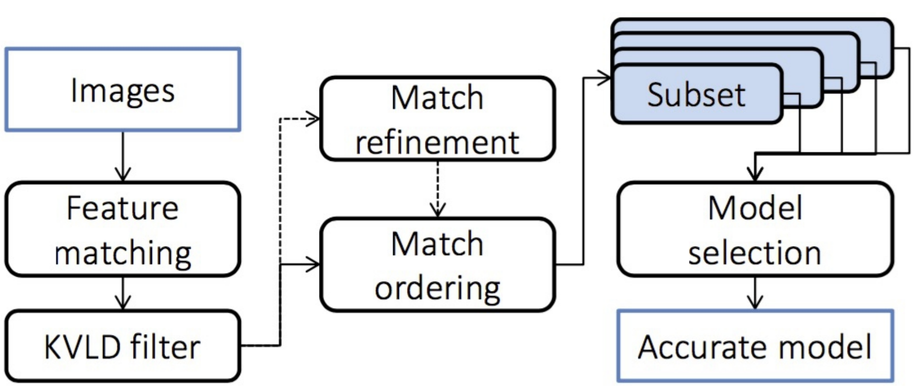
different choices of inliers



$$N^* = \arg\min_{N \le |M|} \frac{e_F(M_{\mathrm{sub}}(N))^2}{N}$$

$$M^* = M_{\mathrm{sub}}(N^*)$$

best matches first



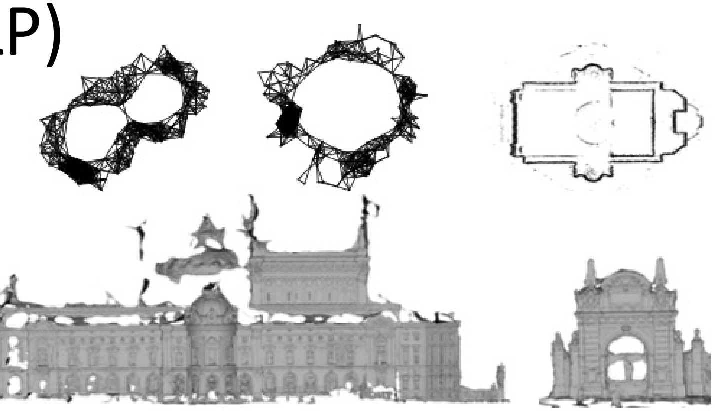
Dataset	Strecha et al. [23]					DTU robot [1]				
e_R (deg $\times 10^{-2}$)	raw	MS	MR	MR+MS	gain	raw	MS	MR	MR+MS	gain
RANSAC	16.4	9.52	10.3	8.87	1.9	26.5	22.3	21.5	21.3	1.2
MSAC	14.1	9.53	8.86	8.43	1.7	21.3	21.7	20.4	20.1	1.1
LO-RANSAC	16.4	9.54	10.3	8.97	1.8	26.8	22.2	21.5	21.3	1.3
MLESAC	15.8	7.81	9.50	7.76	2.0	21.8	22.6	20.8	20.2	1.1
ORSA	12.2	7.24	6.48	6.60	1.9	21.9	21.7	20.8	20.3	1.1
e_t (deg)	raw	MS	MR	MR+MS	gain	raw	MS	MR	MR+MS	gain
RANSAC	1.85	1.09	1.23	1.04	1.8	3.83	2.12	1.81	1.02	3.7
MSAC	1.59	1.08	1.03	0.96	1.6	1.27	1.03	0.93	0.70	1.8
LO-RANSAC	1.83	1.10	1.21	1.05	1.7	3.89	2.14	1.76	1.02	3.8
MLESAC	2.16	0.95	1.09	0.87	2.5	2.02	1.34	1.23	0.77	2.6
ORSA	1.38	0.81	0.68	0.74	1.9	1.22	0.88	0.66	0.66	1.8

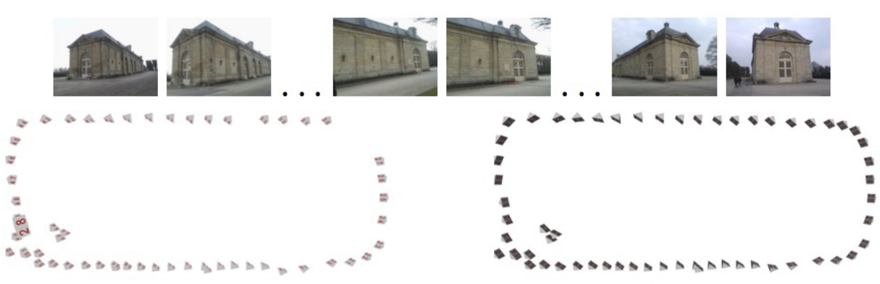
Global fusion of relative motions for robust, accurate and scalable Structure from Motion

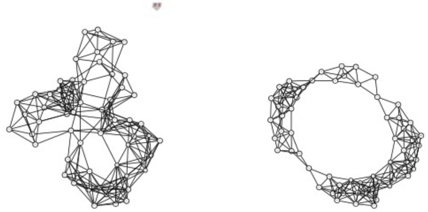
Moulon et al., ICCV 2013

openMVG

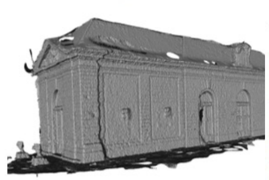
- Incremental SfM: drifts
- Global SfM: error smoothed over whole image graph
  - improved rejection of rotation outliers
  - a contrario trifocal estimation for translation (LP)
  - global registration of translations (LP)







Bundler vs GlobalSfM

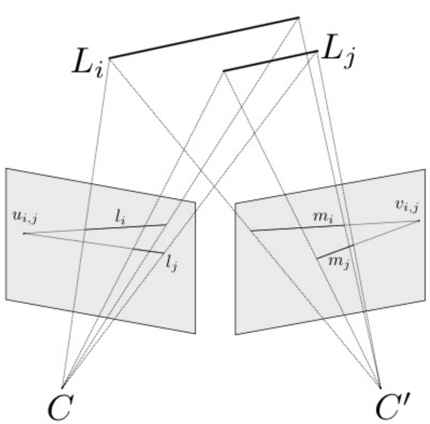


	Accuracy (mm)					Running times (s)						
Scene	Ours	Bundler [31]	VSfM [35]	Olsson [25]	Arie [3]	Ours	OursP	Bundler [31]	VSfM [35]	Olsson [25]	Ratio [25]/Ours	Ratio [25]/OursP
FountainP11	2.5	7.0	7.6	**2.2**	4.8	12	5	36	3	133	11.1	26
EntryP10	**5.9**	55.1	63.0	6.9	N.A.	16	5	16	3	88	5.5	17
HerzJesusP8	**3.5**	16.4	19.3	3.9	N.A.	6	2	10	2	34	5.6	17
HerzJesusP25	**5.3**	21.5	22.4	5.7	7.8	47	10	100	12	221	4.7	22
CastleP19	**25.6**	344	258	76.2	N.A.	20	6	78	9	99	4.9	16
CastleP30	**21.9**	300	522	66.8	N.A.	55	14	300	18	317	5.7	22

Robust and accurate camera pose estimation with lines and/or points

Salaün et al., submitted

- SfM failure: little texture, single plane
  → estimate pose from line segments
  - robust: 2 pairs of (matched) parallel lines
  - no Manhattan-world assumption
  - parameterless (a contrario thresholding)
  - ⊕ points if available

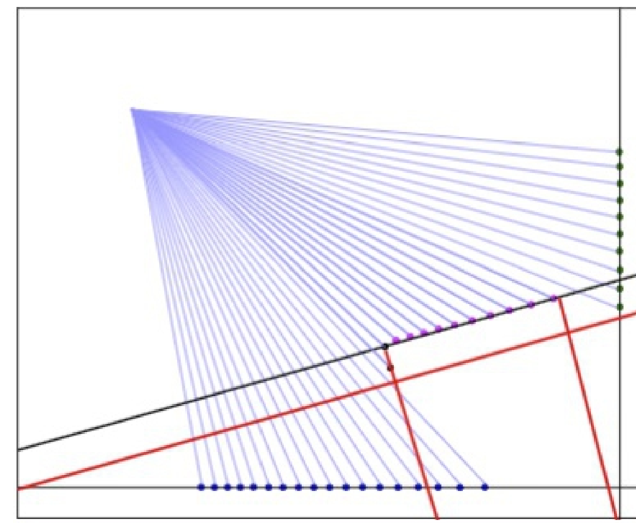


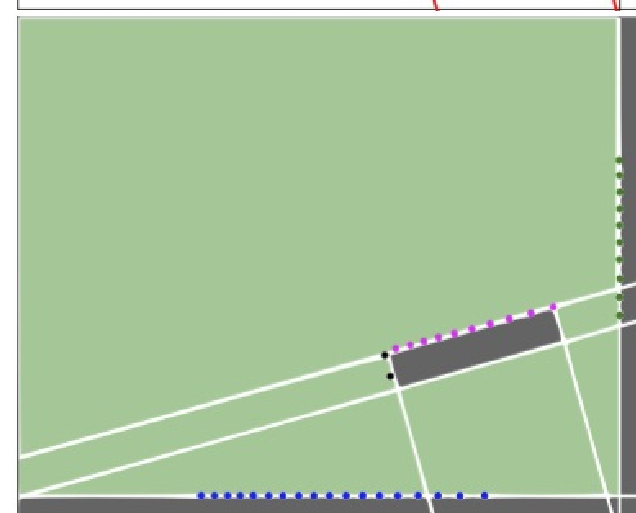
Method \ Dataset		Herz-Jesu P8	Castle P30	Herz-Jesu P25	Castle P18	Fountain P11	Office P8	Building P6	Car P4
Best VP	e_R	6.11	3.23	4.74	0.81	0.26	1.84	18.35	21.12
5-points [2]	e_R	**0.01**	**0.01**	**0.02**	**0.01**	**0**	5.23	**0.15**	**0.15**
	e_t	**0.22**	**0.12**	**0.2**	**0.1**	**0.12**	15.22	**0.39**	**0.3**
4-points	e_R	2.15	0.06	1.39	0.62	**0**	7.73	**0.19**	8.46
	e_t	11.78	0.31	9.98	**0.14**	**0.15**	22.68	**0.41**	28.41
3-lines [1] (no LM)	e_R	3.86	0.43	0.8	0.47	0.28	1.04	7.12	11.73
	e_t	20.26	22.86	23.2	18.19	2.74	15.3	57.18	32.3
3-lines + SIFT	e_R	4.16	**0.07**	0.36	0.19	**0.01**	2.68	3.49	7.93
	e_t	4.89	0.47	1.95	0.52	0.24	5.57	11.23	25.33
2x2-lines	e_R	4.04	0.17	0.72	0.21	**0.06**	0.64	0.51	1.16
	e_t	8.07	0.81	3.98	1.25	0.56	3.87	2.45	16.06
mixed	e_R	0.77	0.17	0.71	0.76	0.1	**0.55**	0.78	**0.23**
	e_t	8.37	0.8	4.1	5.07	0.62	3.82	0.56	0.67
AC-mixed	e_R	**0.01**	**0.01**	**0.02**	**0.01**	**0**	**0.57**	**0.12**	**0.19**
	e_t	**0.23**	**0.12**	**0.21**	**0.1**	**0.13**	**2.54**	**0.43**	**0.27**

Piecewise-planar 3D reconstruction with edge and corner regularization

Boulch et al., SGP/CGF 2014

- Watertight polygonal mesh, from laser scan
  - detected planes → 3D arrangement
    3D cell labeling: empty or full
  - completion of hidden area
    (ghost planes)
  - robustness to sampling anisotropy
  - insensitivity to plane insertion order

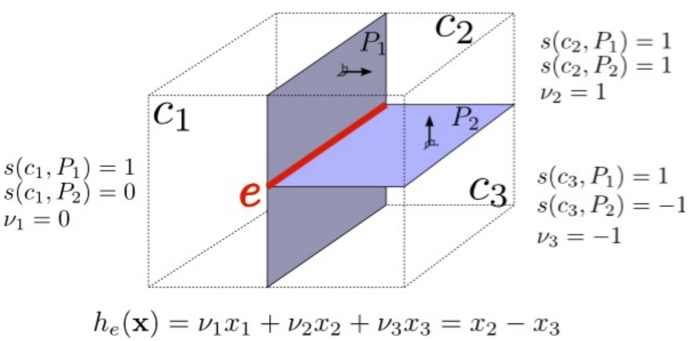


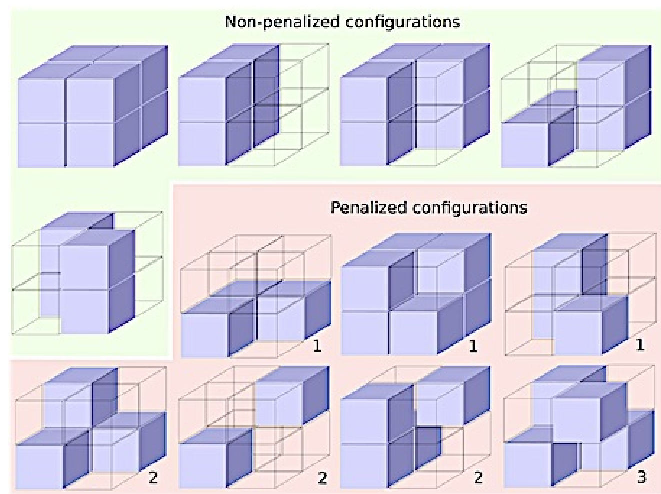


Piecewise-planar 3D reconstruction with edge and corner regularization

Boulch et al., SGP/CGF 2014

- Watertight piecewise-planar polygonal mesh
  - regularization: area, edge length, number of corners
  - 8th-order potentials → mixed integer programming problem

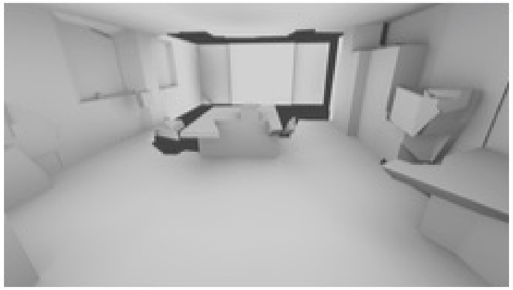




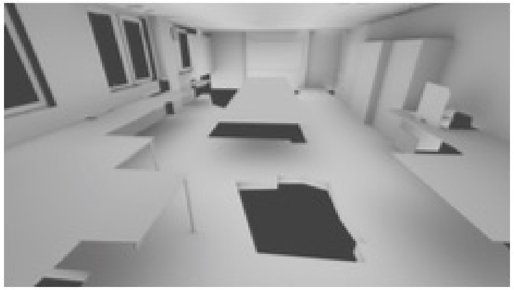
edge+corner regularization

laser scan

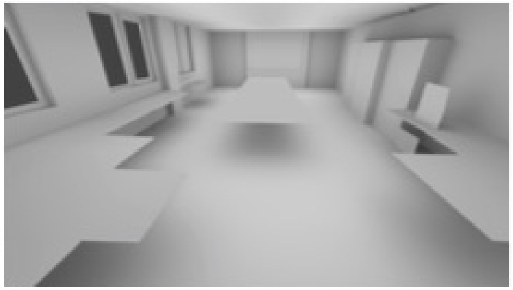
area regularization



edge regularization



corner regularization

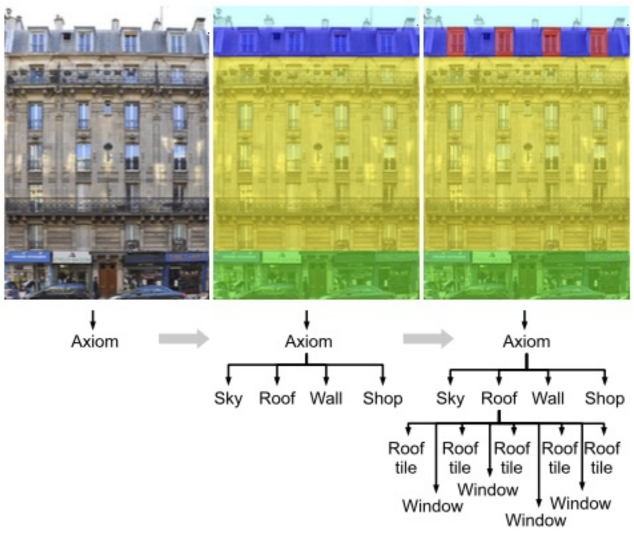


High-level bottom-up cues for top-down parsing of facade images

Ok et al., 3DIMPVT 2012

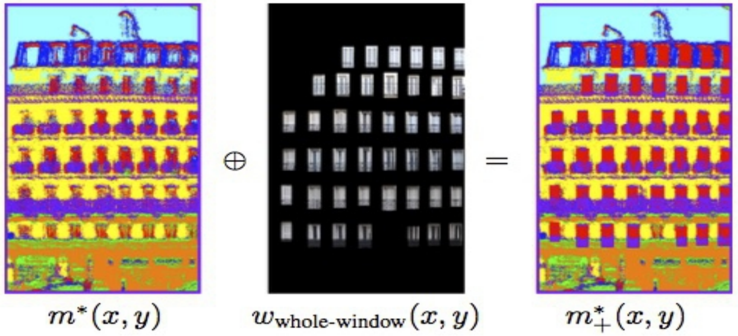
- Top-down parser with reinforcement learning: use of higher-level cues

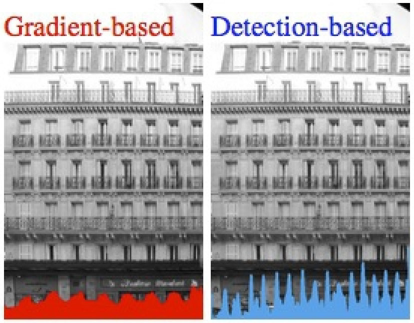
Parsing with a split grammar

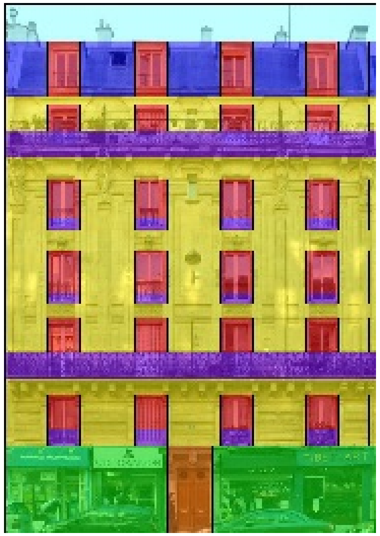


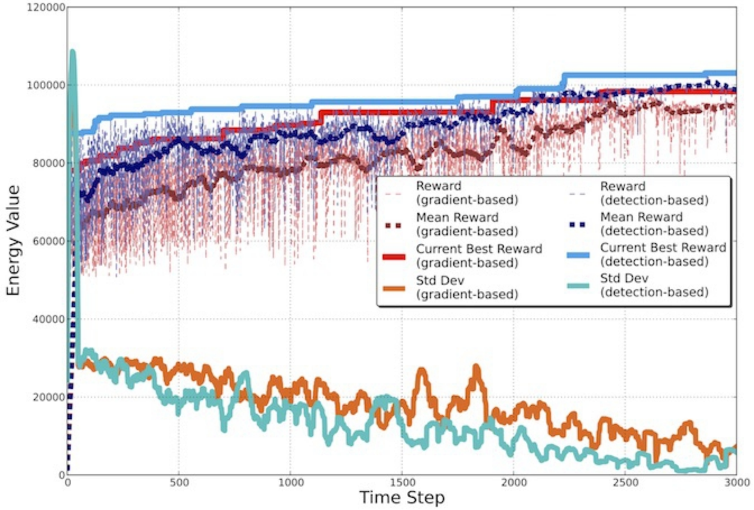
RL parser = Markov decision process:

- agent constructing a derivation tree
- learning of policy function $\pi(s,a)$: derivation choice a in state s
- $\pi(s,a) = (1-\varepsilon)\,\delta(a,a^*) + \varepsilon P(a|s)$





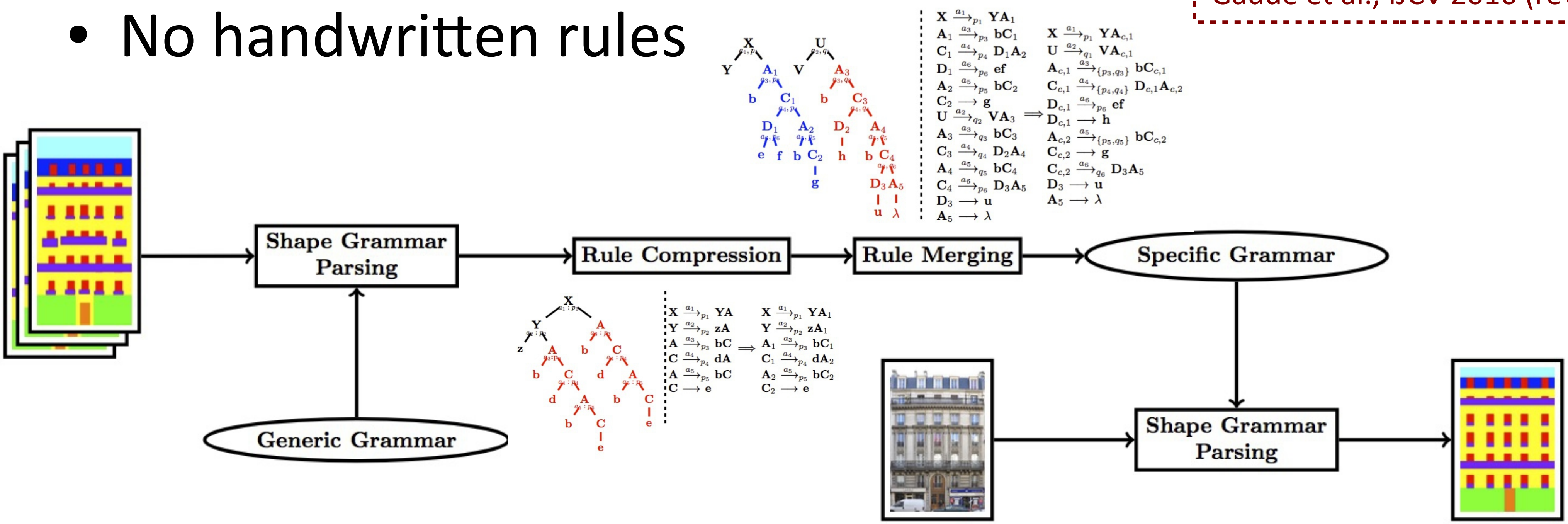


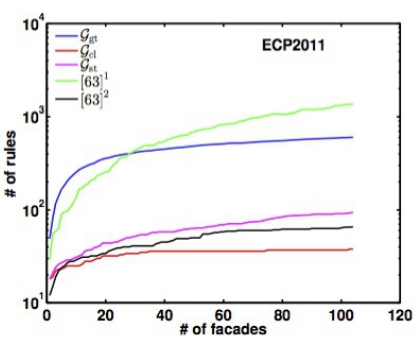


Learning grammars for architecture-specific facade parsing

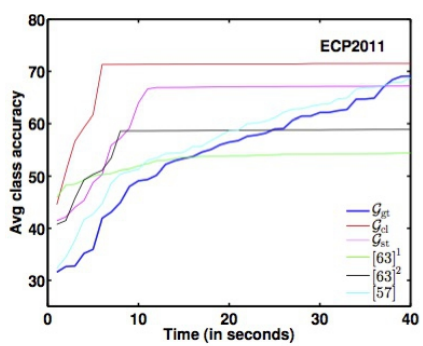
Gadde et al., IJCV 2016 (rev)

- No handwritten rules





scalability



convergence speed

			RF unaries — Grammar induced from $\mathcal{G}^1_{\text{gen}}$					DARWIN unaries — Grammar induced from $\mathcal{G}^2_{\text{gen}}$					State of art (no grammar)	
	[62]	[41]	$[69]^1$	$[69]^2$	$\mathcal{G}_{\text{gt}}$	$\mathcal{G}_{\text{st}}$	$\mathcal{G}_{\text{cl}}$	$[69]^1$	$[69]^2$	$\mathcal{G}_{\text{gt}}$	$\mathcal{G}_{\text{st}}$	$\mathcal{G}_{\text{cl}}$	[40]	[12]
Door	47	50	20	26	19	41	52	49	54	48	57	**62**	60	**79**
Shop	88	81	84	85	79	85	86	87	89	88	90	**94**	86	**94**
Balcony	58	49	30	42	24	51	55	58	69	66	78	**84**	71	**91**
Window	62	66	24	48	26	58	64	52	59	56	67	**72**	69	**85**
Wall	82	80	74	78	71	78	83	79	83	76	85	**89**	**93**	90
Sky	95	91	**99**	97	95	92	92	**99**	96	96	96	98	97	97
Roof	66	71	33	34	29	63	67	52	58	54	73	**79**	73	**90**
Average	71.1	69.7	51.9	58.6	49.1	66.9	71.3	67.9	72.6	66.5	78.1	**82.5**	78.4	**89.4**
Overall	74.7	74.8	62.9	69.3	59.9	73.1	76.2	74.2	78.6	71.8	82.6	**86.9**	85.1	**90.8**
IoU	-	-	36.5	42.1	34.3	55.4	57.6	54.8	57.3	52.3	67.7	**71.8**	-	-

Learning grammars for architecture-specific facade parsing

Gadde et al., IJCV 2016 (rev)

- Specificity of learned grammars

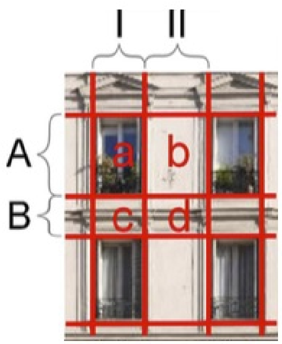
Haussmannian facade
parsed with
Art déco grammar

Art déco facade
parsed with
Haussmannian grammar

	$\mathcal{G}_{\mathrm{AA}}$	$\mathcal{G}_{\mathrm{AH}}$	$\mathcal{G}_{\mathrm{HA}}$	$\mathcal{G}_{\mathrm{HH}}$
Door	59	56	57	62
Shop	88	86	83	94
Balcony	63	51	54	84
Window	66	56	48	72
Wall	84	71	76	89
Sky	92	82	92	98
Roof	58	68	51	79
Average	72.9	67.1	65.9	82.5
Overall	78.8	71.9	70.8	87.0
IoU	59.4	55.8	57.6	71.8

A MRF shape prior for facade parsing with occlusions

Kozinski et al., CVPR 2015



$$\mathcal{R} = \{A, B\}, \quad \mathcal{C} = \{I, II\},$$
$$\Psi(A, I) = \text{window},$$
$$\Psi(A, II) = \text{wall},$$
$$\Psi(B, I) = \text{wall},$$
$$\Psi(B, II) = \text{wall},$$
$$\mathcal{V} = \{(A, B), (B, A)\},$$
$$\mathcal{H} = \{(I, II), (II, I)\}.$$

$(A, I) = a$, $(A, II) = b$, $(B, I) = c$, $(B, II) = d$.

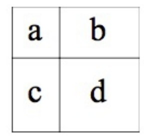
horiz. neighbors

h↱	a	b	c	d
a	+	+	−	−
b	+	+	−	−
c	−	−	+	+
d	−	−	+	+

vert. neighbors

v↱	a	b	c	d
a	+	−	+	−
b	−	+	−	+
c	+	−	+	−
d	−	+	−	+

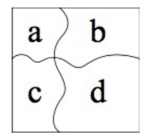
user-defined prior – enables 2D alignment



h↱	a	b	c	d
a	+	+	−	−
b	−	+	−	−
c	−	−	+	+
d	−	−	−	+

v↱	a	b	c	d
a	+	−	−	−
b	−	+	−	−
c	+	−	+	−
d	−	+	−	+

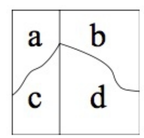
(a) A non-repeating pattern with straight, axis-aligned boundaries.



h↱	a	b	c	d
a	+	+	+	−
b	−	+	−	+
c	+	−	+	+
d	−	+	−	+

v↱	a	b	c	d
a	+	+	−	−
b	+	+	−	−
c	+	−	+	+
d	−	+	+	+

(b) A non-repeating pattern with winding, axis-driven boundaries.

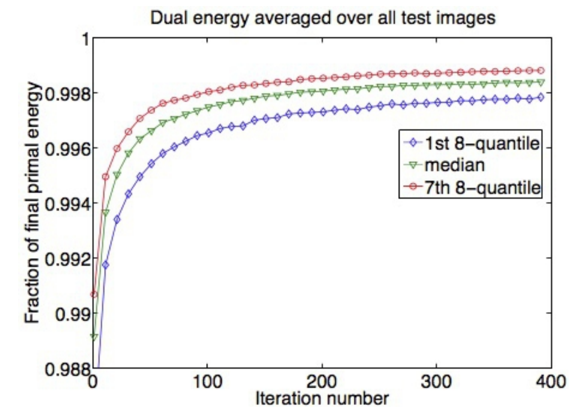


h↱	a	b	c	d
a	+	+	+	−
b	−	+	−	−
c	−	−	+	+
d	−	−	−	+

v↱	a	b	c	d
a	+	−	−	−
b	−	+	−	−
c	+	−	+	−
d	−	+	−	+

(c) A non-repeating pattern on grid with monotonic boundaries.

allows irregular shapes



efficient inference (dual decomposition)

state-of-the-art accuracy

	RNN unaries			TextonBoost unaries			
	raw	[7]	Ours	raw	[1]	[5]	Ours
roof	70	74	78	89	90	91	91
shop	79	93	90	95	94	95	97
balcony	74	70	76	90	91	90	91
sky	91	97	94	94	97	96	97
window	62	75	67	86	85	85	87
door	43	67	44	77	79	74	79
wall	92	88	93	90	90	91	90
pixel accur.	82.6	84.2	**86.2**	90.1	90.8	90.8	**91.3**

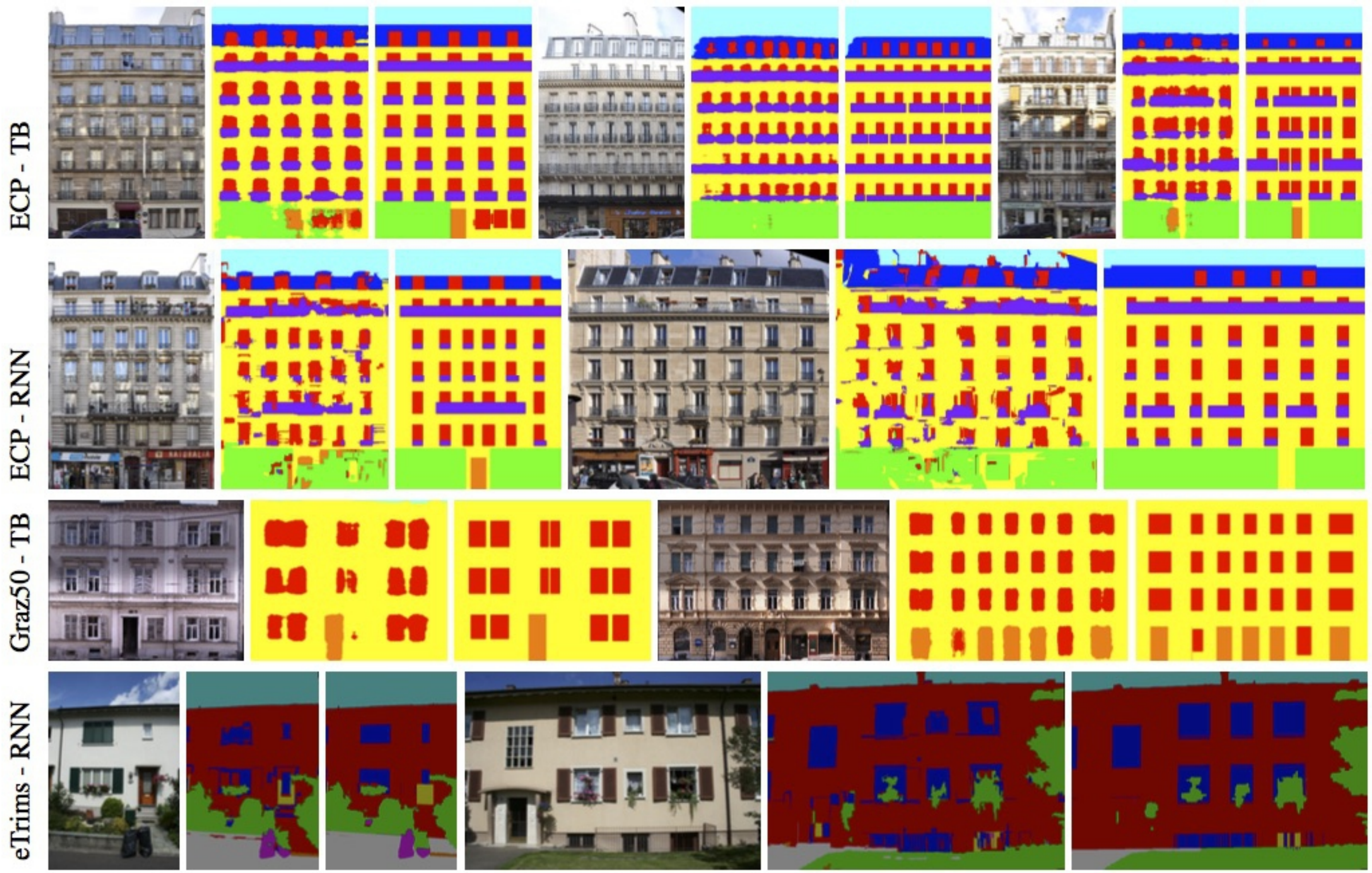
	Graz50		
	[9]	[5]	Ours
sky	91	93	93
window	60	82	84
door	41	50	60
wall	84	96	96
pix. acc.	78.0	91.8	**92.5**

	ArtDeco			
	raw[1]	raw[2]	ours[3]	ours[4]
roof	82	82	81	82
shop	96	95	97	97
balcony	88	87	82	87
sky	97	97	98	97
window	87	85	82	82
door	64	63	57	57
wall	77	87	89	88
vegetation	–	90	–	90
	83.5	88.4	**88.8**	**88.8**

	eTrims			
	raw	[7]-L3	[1]	Ours
building	88	87	91	92
car	69	69	70	70
door	25	19	18	20
pavement	34	34	33	33
road	56	56	57	56
sky	94	94	97	96
vegetation	89	88	90	91
window	71	79	71	70
pixel accur.	81.9	81.6	**83.8**	83.5

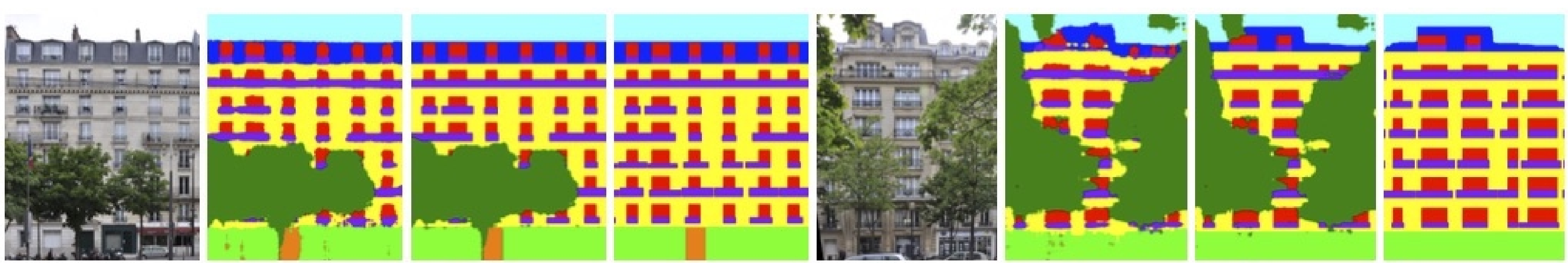
A MRF shape prior for facade parsing with occlusions

Kozinski et al., CVPR 2015



facade analysis

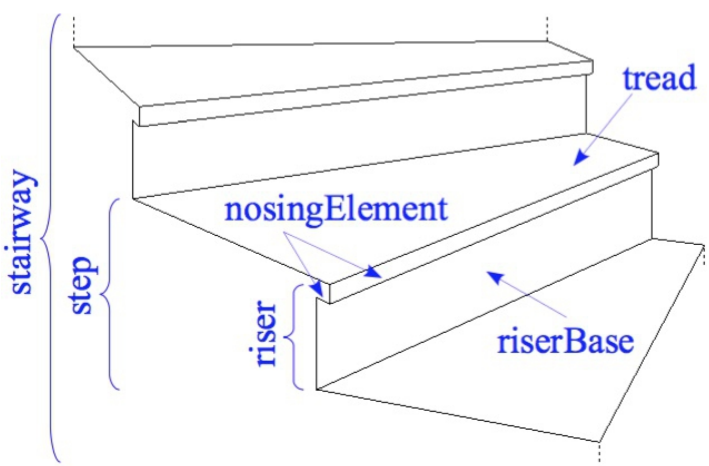
with occlusion

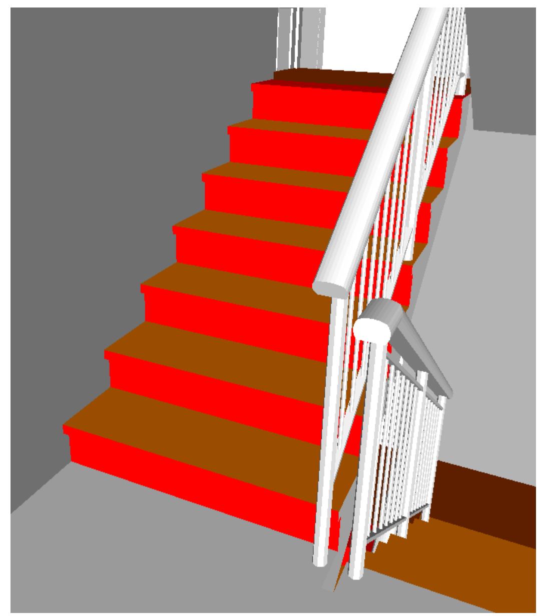


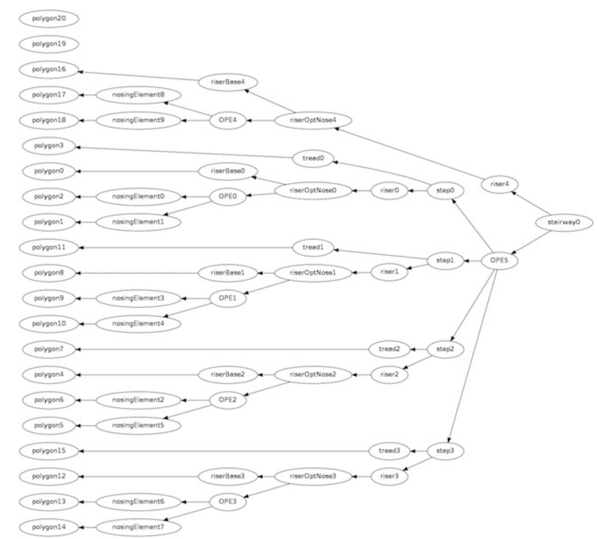
Semantizing complex 3D scenes using constrained attribute grammars

Boulch et al., SGP/CGF 2013

- Bottom-up generation of shared parsed forest





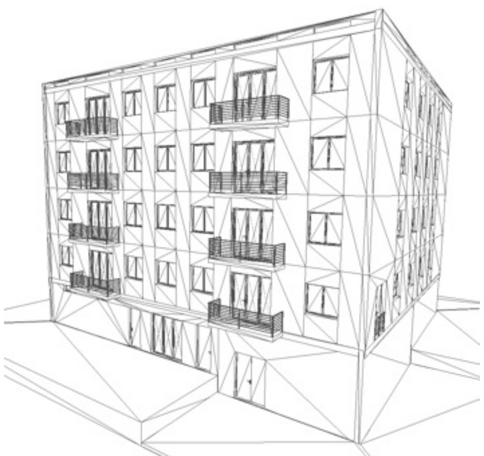


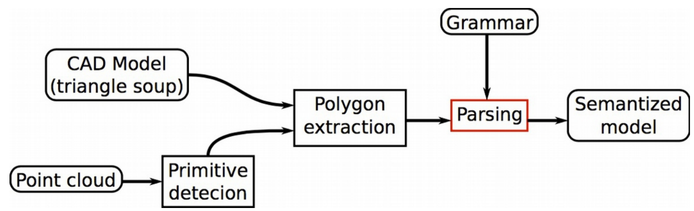
tread t	$\rightarrow$	polygon *p*	$\langle$horizontal(p), p.breadth <=2.0$\rangle$
riserBase b	$\rightarrow$	polygon *p*	$\langle$vertical(p.breadthVector), 0.05 <= p.breadth, p.breadth <=0.25$\rangle$
nosingElement e	$\rightarrow$	polygon *p*	$\langle$horizontal(p.lengthVector), p.breadth <=0.05$\rangle$
riser r	$\rightarrow$	*riserBase b*, maxseq(*nosingElement*, edgeAdj) *n*	$\langle$edgeAdj(b,n), above(b,n)$\rangle$
step s	$\rightarrow$	*riser r*, *tread t*	$\langle$edgeAdj(r,t), above(r,t)$\rangle$
stairway w	$\rightarrow$	maxseq(*step*, edgeAdj) *s*, optional *riser r*	$\langle$edgeAdj(s,r), above(s,r)$\rangle$

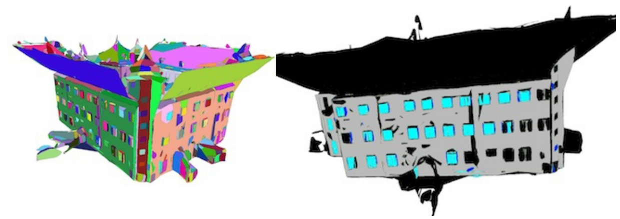
Semantizing complex 3D scenes using constrained attribute grammars

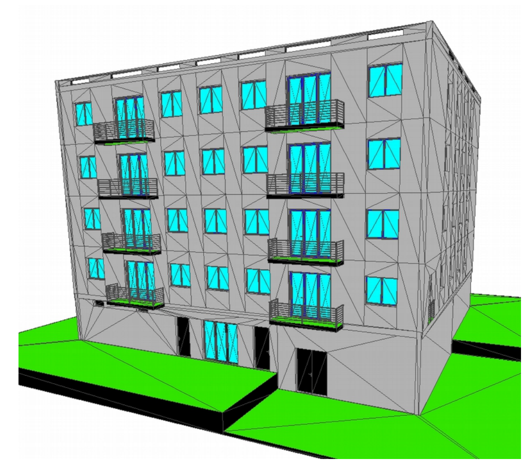
Boulch et al., SGP/CGF 2013

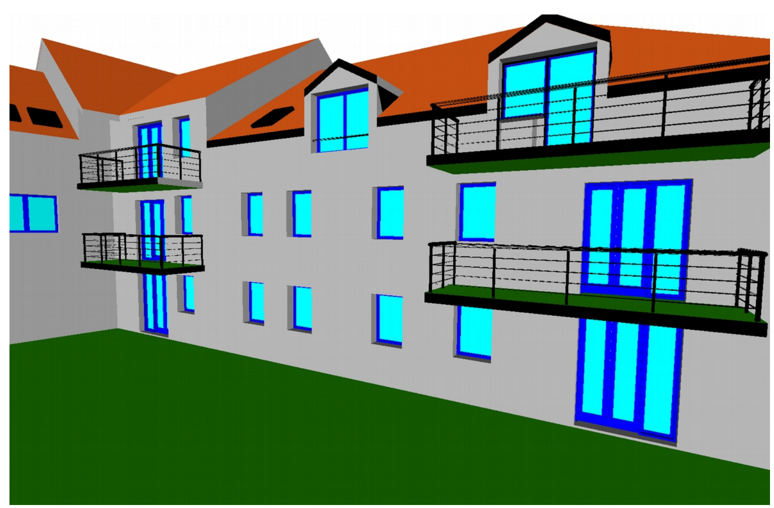
- Bottom-up generation of shared parsed forest

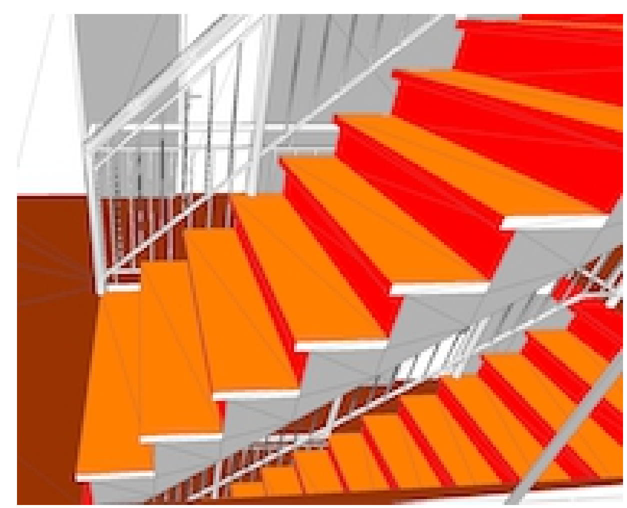










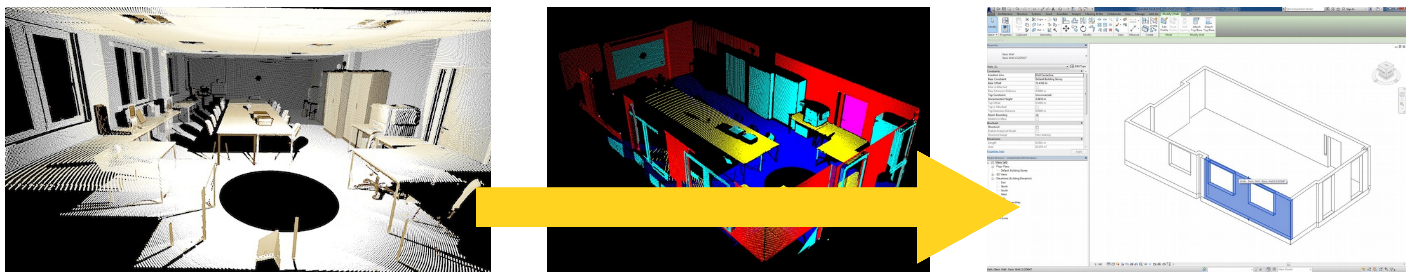


Projects

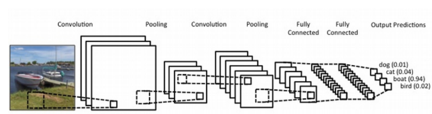
- IMAGINE: joint project ENPC-CSTB
  - IMage, Apprentissage et GéométrIe pour la Numérisation de l'Environnement
- Chaire Bouygues "Bâtir durable et innover"
  - maquettes numériques de bâtiments existants
- ANR Semapolis
  - semantic visual analysis and 3D reconstruction of urban environments

Perspectives

- Scan-to-BIM, picture-to-BIM, picture-to-CityGML...



- Deep learning



  - e.g., detection of furniture
  - 2015: rank 1 on PASCAL VOC2012 challenge
  - 2+ researchers, 7 PhD students (5 main topic + 2 auxiliary)

PASCAL2 Pattern Analysis, Statistical Modelling and Computational Learning

Gidaris et al., ICCV 2015

- Co-innovation lab
  - robotics for civil engineering, ...