Parameterizing Object Detectors in the Continuous Pose Space

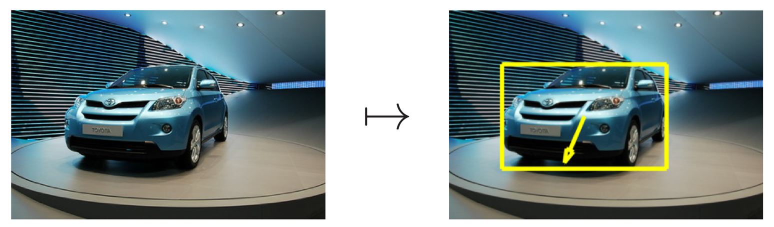
Kun He [1], Leonid Sigal [2], Stan Sclaroff [1]

[1] {hekun,sclaroff}@cs.bu.edu [2] lsigal@disneyresearch.com

Boston University · Disney Research Pittsburgh

Motivation

Simultaneous object detection and continuous pose estimation: $x \mapsto y = (B, \theta)$



Most existing approaches:

- Regression: need localization as input
- View-specific detectors: arbitrary discretization, expensive when fine-grained

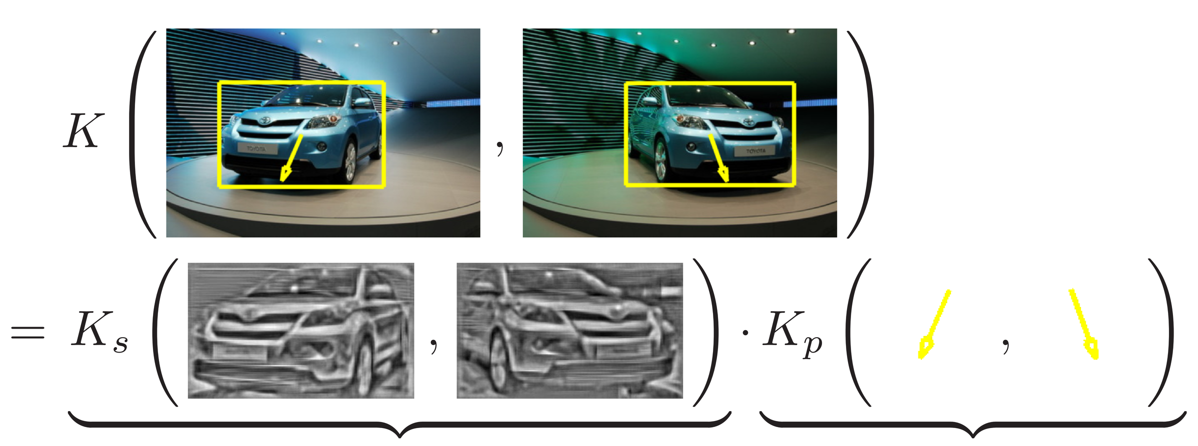
Our proposal: build a *unified* model to perform both tasks in a mutually beneficial way

Model

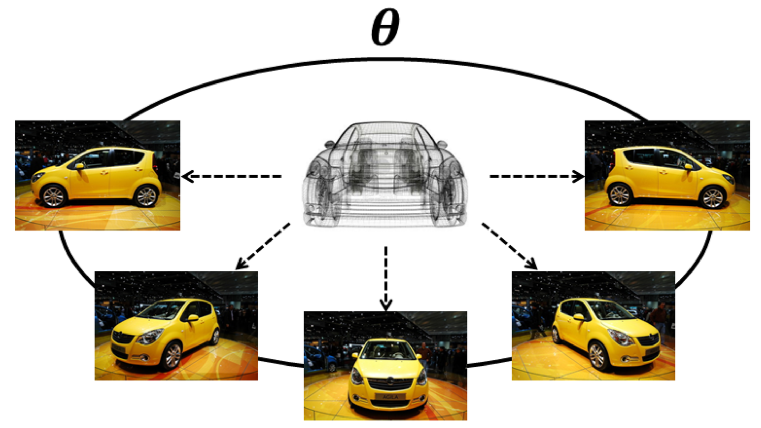
Modeling approach: structured kernel machine, learned using structural SVM

$$f(x,y) = \langle \mathbf{w}, \Psi(x,y) \rangle = \sum_{j \in \mathcal{SV}} \alpha_j K(x, y, x_j, y_j)$$

Joint kernel function (multiplicative kernel [1]):



Parametric detectors in the continuous pose space.
No discretization!



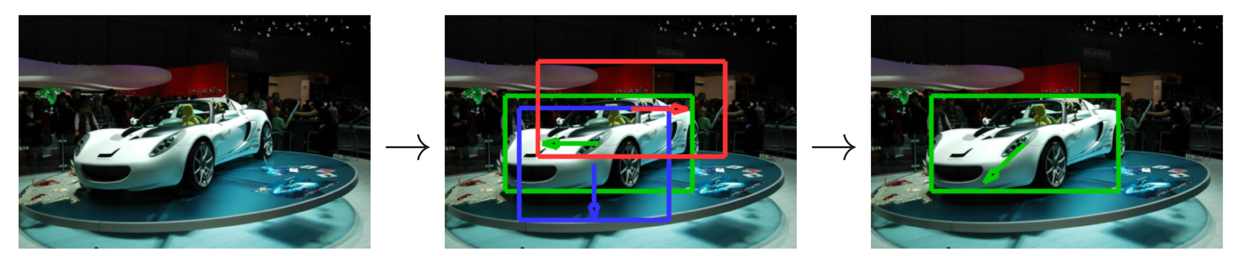
Cascaded Inference

Joint inference problem:

$$\max_{B,\theta} \sum_{j \in \mathcal{SV}} \alpha_j \underbrace{\phi(x,B)^T \phi(x_i, B_j)}_{K_s} \underbrace{\exp\left(-\gamma d(\theta, \theta_j)^2\right)}_{K_p}$$

- Large number of B, continuous θ
- Non-convex problem

→ use a two-step cascade!



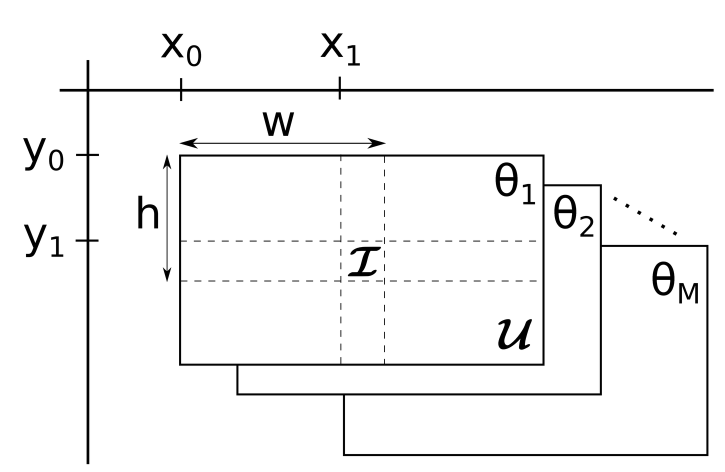
Initialization/pruning: $\mathcal{Y} \to \{(B_k, \theta_k)\}_{k=1}^{K}$

1. sample "seed poses" $\{\theta_1, \ldots, \theta_M\}$,
2. construct corresponding detectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_M\}$,
3. evaluate $\{\mathbf{w}_m\}$ to give detection proposals.

Proposal generation: branch-and-bound algorithm that generalizes [2].

state representation:
$$s = (w, h, x_0, x_1, y_0, y_1, \theta), \quad \theta \in \{\theta_1, \ldots, \theta_M\}$$

bounding detector scores: $\forall B \in s$,
$$\mathbf{w}_\theta^\top \Phi_{bow}(\cap_{B \in s} B) \le f_s(B, \theta) \le \mathbf{w}_\theta^\top \Phi_{bow}(\cup_{B \in s} B)$$



Refinement: solve

$$\max_k \max_{\theta \in \Theta_k} \sum_{j \in \mathcal{SV}} \eta_k^j \exp\left(-\gamma\, d(\theta, \theta_j)^2\right)$$

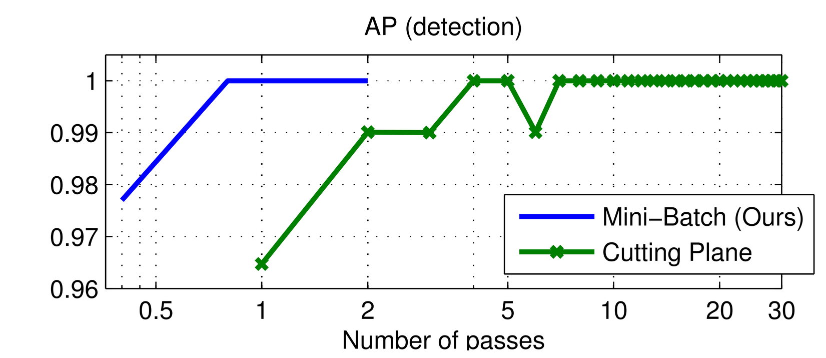
with gradient-based optimization, e.g. L-BFGS.

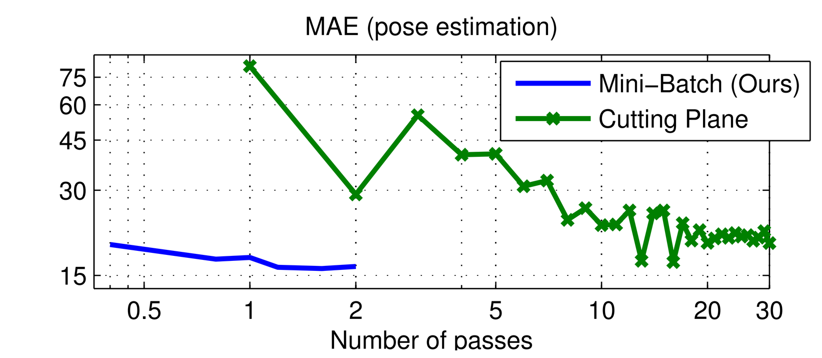
Online Structural SVM Learning

$$\min_{\mathbf{w},\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i$$
$$\forall i, \forall y: \langle \mathbf{w}, \Psi(x_i, y_i) \rangle - \langle \mathbf{w}, \Psi(x_i, y) \rangle \ge \Delta(y_i, y) - \xi_i$$
where $\Delta(y_i, y) = \beta \Delta_{loc}(B_i, B) + (1-\beta)\Delta_{pose}(\theta_i, \theta)$

- **Batch algorithm (cutting plane)**: in each step, find violated constraints in entire training set S.
- **Our online algorithm**: in each step, find violated constraints in a sampled subset S_t instead.

Comparisons on the EPFL Cars dataset

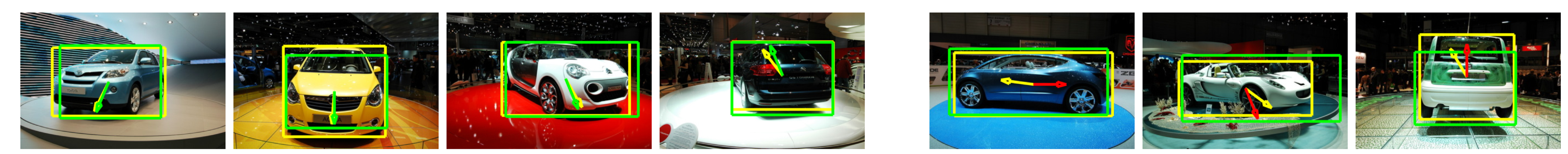


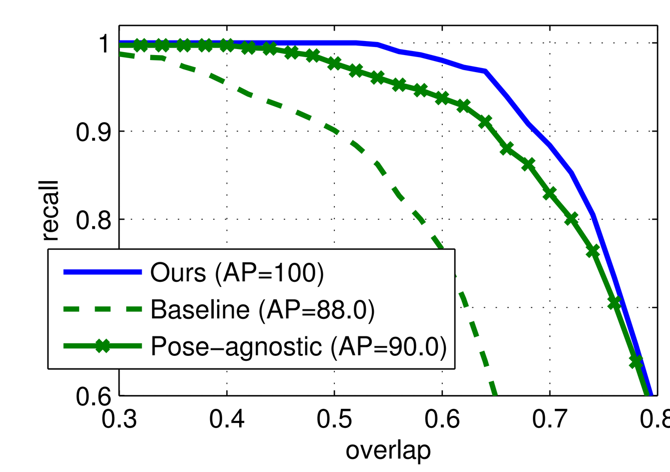


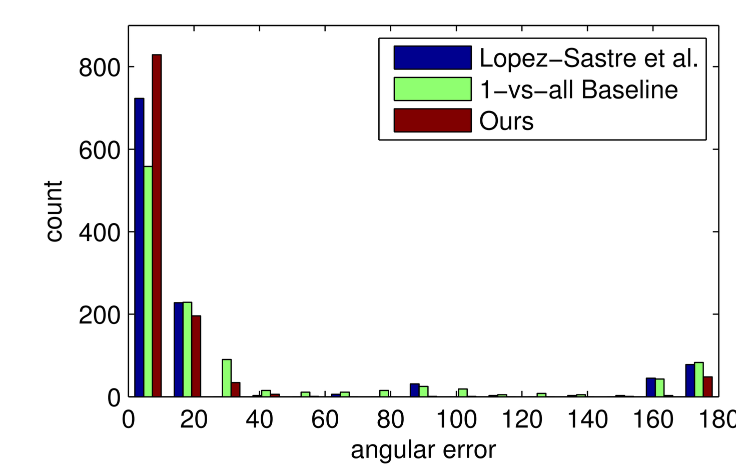
Experimental Results

Appearance model: single rectangular template (no mixtures/parts). Baseline: view-specific 1-vs-all SVMs.

1. **EPFL Cars** : detection & continuous pose estimation (Red: best. Green: second best.)

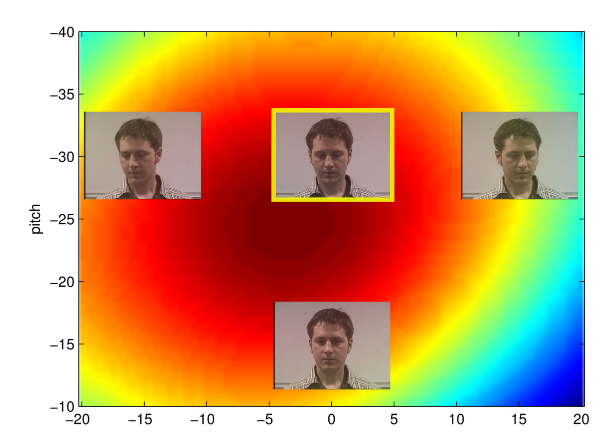






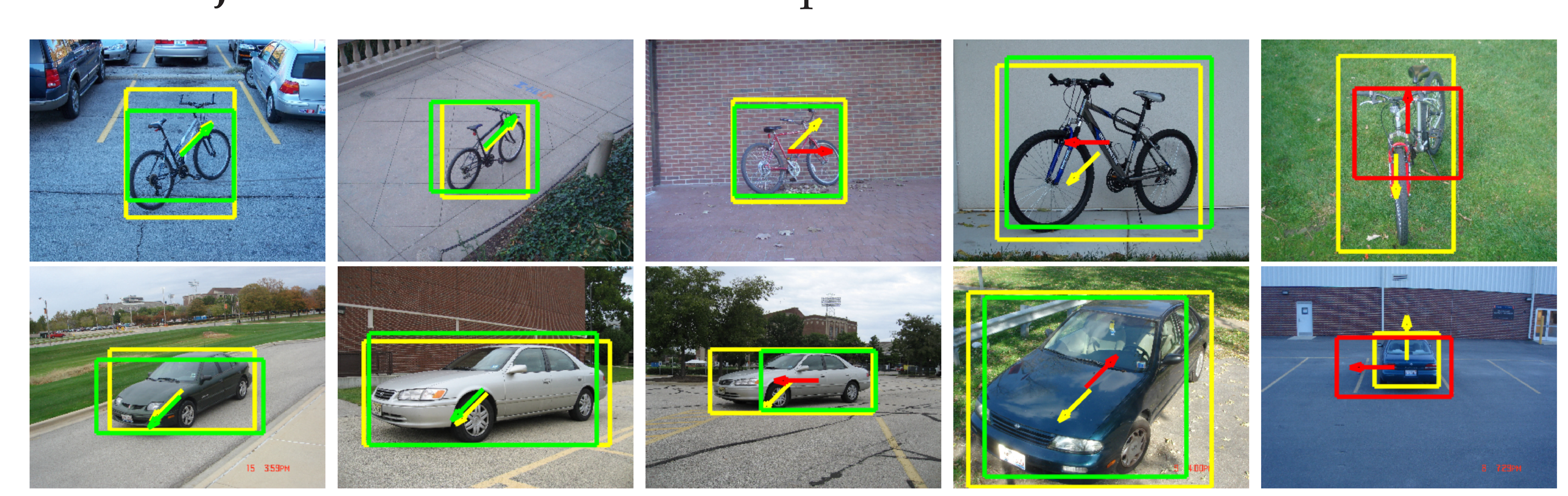
Method	AP	MAE/median		MPPE
Baseline	88.0	36.7	12.2	46.8
Ours	100	15.8	6.2	64.0
Pepik ECCV'12	97.5	–	6.9	69.0
Lopez ICCVW'11	97	27.2	–	66.1
Hara ECCV'14	(GT)	24.2	–	–

2. **Pointing'04 Faces** : continuous pose estimation



Method	pitch	yaw	avg
Baseline	6.37	7.14	6.76
Ours (avg)	4.30	5.36	4.83
Ours (best)	4.01	5.20	4.61
Hara ECCV'14	2.51	5.29	3.90
Fenzi CVPR'13	6.73	5.94	6.34
Haj CVPR'12	6.61	6.56	6.59

3. **3D Objects** : detection & discrete pose estimation



Method	bike: AP/MPPE		car: AP/MPPE	
Baseline	78.2	98.7	85.4	97.7
Ours (avg)	95.1	94.0	98.2	87.9
Ours (best)	96.8	97.6	97.8	93.0
Pepik ECCV'12	97.6	98.9	99.9	97.9
Schels CVPR'12	87.0	87.7	94.9	82.6
Lopez ICCVW'11	91	90	96	89

References

[1] Quan Yuan, Ashwin Thangali, Vitaly Ablavsky, and Stan Sclaroff. Learning a family of detectors via multiplicative kernels. *IEEE TPAMI*, 33(3):514–530, 2011.

[2] Christoph H. Lampert, Matthew B. Blaschko, and Thomas Hofmann. Efficient subwindow search: A branch and bound framework for object localization. *IEEE TPAMI*, 31(12):2129–2142, 2009.