Model selection in pedestrian detection using multiple kernel learning

SUARD Frédéric, RAKOTOMAMONJY Alain and BENSRHAIR Abdelaziz

Laboratoire Informatique, Traitement de l'Information, Systèmes EA4051
Institut National des Sciences Appliquées (INSA) de Rouen, Université de Rouen,
Email: frederic.suard@insa-rouen.fr
http://asi.insa-rouen.fr/~fsuard/

Introduction

Context :

- Pedestrian detection using vision system,
- pattern recognition : which images contain a pedestrian ?

Problems :

- how to characterize efficiently an image ?
- how to tune efficiently the classifier and the descriptors ?

Multiple kernel approach :

- choosing automatically the best kernels,
- $\mathbf{k}(\mathbf{x},\mathbf{x}') = \sum_{k=1}^{K} \beta_k \mathbf{k}_k(\mathbf{x},\mathbf{x}')$

Application :

→ combining and selecting characterization,
→ tuning automatically descriptors and kernels,
→ selecting relevant features.

Multiple Kernel Framework

- $\mathbf{x}_i$: training examples vector, y_i : class label, $\mathbf{k}(.,.)$ a kernel,
- training data set : $\{\mathbf{x}_i, y_i\} \in \mathcal{X} \times \{-1, 1\}$, $i = 1 : N$,
- Decision function [6] :

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i \mathbf{k}(\mathbf{x}_i, \mathbf{x}) + b\right)$$

- Lanckriet et al. [2] : $\mathbf{k}$ is convex linear combination of kernels

$$\mathbf{k}(\mathbf{x},\mathbf{x}') = \sum_{k=1}^{K} \beta_k \mathbf{k}_k(\mathbf{x},\mathbf{x}')$$

with $\beta_k \geq 0$, $\sum_k \beta_k = 1$

→ idea : computing automatically value of β instead of selecting and combining the descriptors manually

- optimization problem :

$$\left\{\begin{array}{ll} \min\limits_{\mathbf{w},\beta,b,\xi} & \frac{1}{2}\left(\sum_{k=1}^{K} \beta_k \|\mathbf{w}_k\|_2\right) + C \sum_{i=1}^{N} \xi_i \\ \text{s.t.} & y_i f(\mathbf{x}_i) \geq 1 - \xi_i \quad \forall i = 1, \ldots, N \\ \text{and} & \sum_{k=1}^{K} \beta_k = 1 \end{array}\right.$$

- semi-infinite linear program (SILP), Sonnenburg et al. [5] :

$$\left\{\begin{array}{ll} \max\limits_{\theta,\beta} & \theta \\ \text{s.t.} & \sum_{k=1}^{K} \beta_k = 1 \\ \text{and} & \sum_{k=1}^{K} \beta_k S_k(\boldsymbol{\alpha}) \geq \theta \end{array}\right.$$

with $S_k(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j \mathbf{k}_k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^{N} \alpha_i$.

Image representation

- Pixel value : value of the luminosity for each pixel,
- Gradient norm : value of the gradient magnitude for each pixel,
- Wavelet [3] : Haar wavelet transform at a scale n of a neighboring of size 2^n pixels, with a distance of $\frac{1}{4}2^n$ pixels between two neighboring.
- Histograms of gradient : compute local histogram of gradient orientation
  - Shashua [4] : image splitting according to human morphology,
  - Dalal [1] : dense image splitting.

Evaluation process :

- 310 urban scenes : night and day, different weather and ligthning conditions.
- 1240 pedestrians, 6220 non-pedestrians,
- images rescaled at 128×64 pixels, pedestrians centered,
- learning and test set : 500 pedestrians and 500 non-pedestrians,
- learning and test sets randomly chosen, 20 iterations.

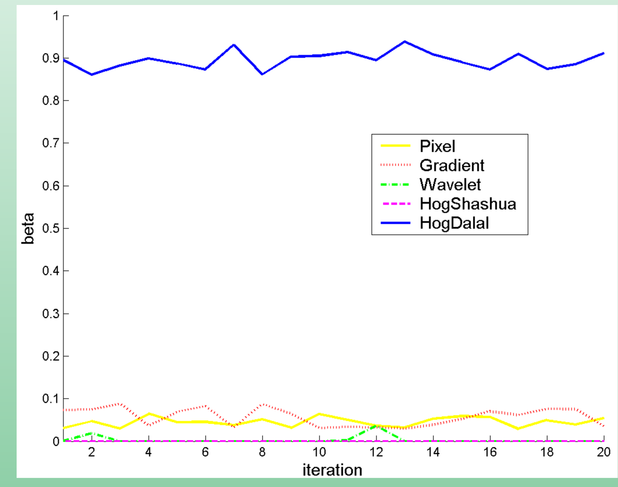
Combination of representations

- all descriptors : pixel, gradient, wavelet, HoGShashua, HoGDalal,
- one kernel for each descriptor,

→ select and combine best image characterization.

β value :

Kernel	β	variance
Pixel	0.0453	0.0115
Gradient	0.0571	0.0215
Wavelet	0.0029	0.0090
HogShashua	0	0
HogDalal	0.8947	0.0211



Independant evaluation of descriptors :

Method	Pixel	Gradient	Wavelet	HogShashua	HogDalal
AUC	0.8806	0.8963	0.7623	0.9592	0.9827
%	0.8259	0.8171	0.7237	0.8919	0.9399

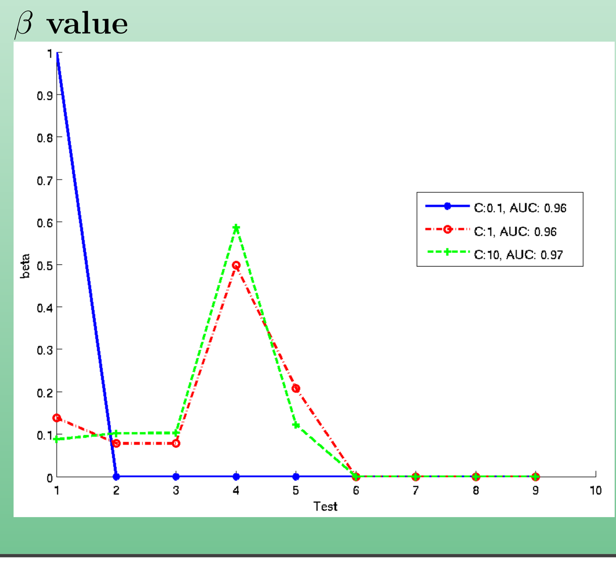
Parameter selection

- HOG descriptor,
- which kernel is well adapted ?

→ compare linear and gaussian kernel,
→ test different parameters value : bandwidth, hyperparameter,

- set of kernels : one kernel for each parameter, same dataset.

β value



Independant evaluation of each parameter

Kernel	**AUC**	**%**
MKL	0.9645	0.8976
Linear	0.9580	0.8758
Gaussian, $\sigma = 0.1$	0.8926	0.8520
Gaussian, $\sigma = 0.5$	0.8937	0.8432
Gaussian, $\sigma = 1$	0.9188	0.8646
Gaussian, $\sigma = 2$	0.9499	0.8888
Gaussian, $\sigma = 5$	0.9598	0.8438
Gaussian, $\sigma = 10$	0.9580	0.5972
Gaussian, $\sigma = 20$	0.9515	0.5120
Gaussian, $\sigma = 50$	0.8664	0.5682

Feature selection

- Hog Descriptor,
- descriptor size depends on parameter set,
- set of kernels : one kernel for each cell,
- select histograms more relevant → feature selection.

Parameter set	A	B	C	D	E	F
size of cell (pixels)	8	8	16	16	32	32
Size of Block (cells)	1	2	1	2	1	2
Overlapping (cells)	0	1	0	1	0	1
number of kernels	128	420	32	84	8	12

	MKL			no feat. sel.	
	AUC	%	$\frac{\#(\beta > 0)}{nbkernel}$	AUC	%
A	0.9726	0.9132	0.7891	0.9105	0.7327
B	0.9810	0.9298	0.7469	0.9107	0.7272
C	0.9622	0.8917	0.8203	0.9511	0.8804
D	0.9716	0.9095	0.7679	0.9313	0.8565
E	0.9479	0.8751	0.8375	0.9564	0.8891
F	0.9277	0.8499	0.8333	0.9472	0.8819

Conclusion

Conclusions :

- multiple kernel for automatic combination and selection,
- application : pedestrian detection
  - automatic parameter tuning,
  - combining descriptors,
  - feature selection.

Perspectives :

- Solve the problem of large kernel set,
- try new descriptors and combination.

References

[1] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 886–893, June 2005.

[2] Gert R. G. Lanckriet, Nello Cristianini, Peter Bartlett, Laurent El Ghaoui, and Michael I. Jordan. Learning the kernel matrix with semidefinite programming. *J. Mach. Learn. Res.*, 5:27–72, 2004.

[3] Constantine Papageorgiou and Tomaso Poggio. Trainable pedestrian detection. In *Proceedings of the 1999 International Conference on Image Processing*, pages 35–39, 1999.

[4] Amnon Shashua, Yoram Gdalyahu, and Gaby Hayon. Pedestrian detection for driving assistance systems: Single-frame classification and system level performance. In *Proceedings of IEEE Intelligent Vehicles Symposium*, 2004.

[5] Sören Sonnenburg, Gunnar Raetsch, and Christin Schaefer. A general and efficient multiple kernel learning algorithm. In *Advances in Neural Information Processing Systems 18*, pages 1273–1280, 2005.

[6] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer, N.Y, 1995.