Combining complementary kernels in complex visual categorization

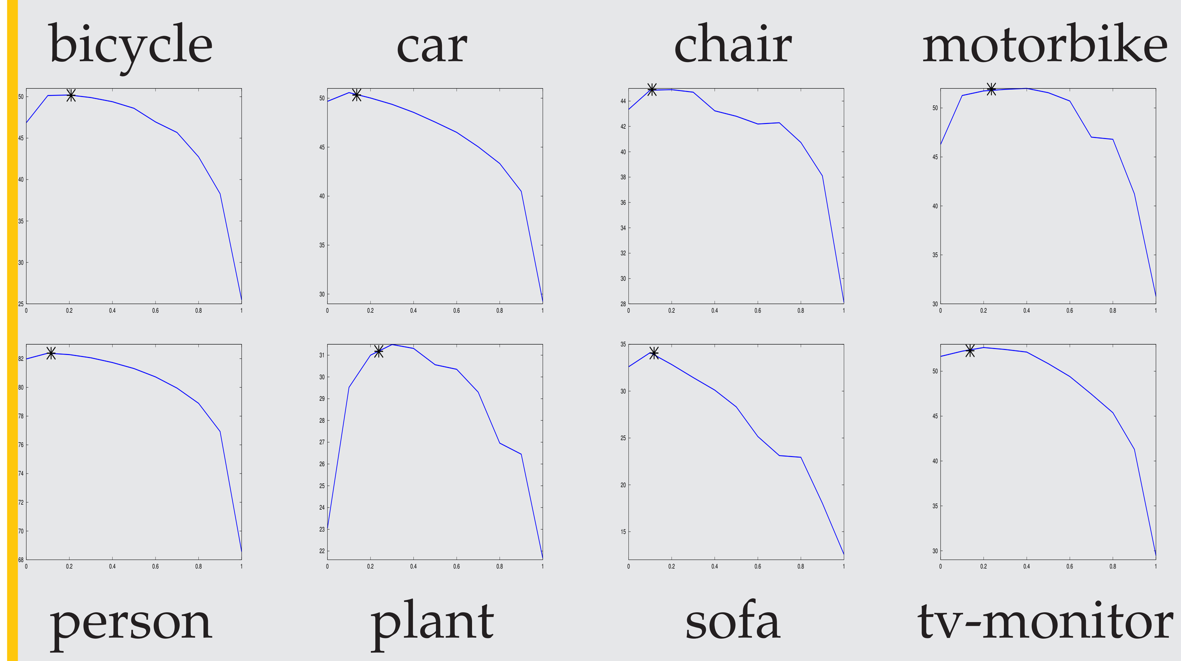
{nicolas.thome, matthieu.cord}@lip6.fr[1], picard@ensea.fr[2]

[1]LIP6, UPMC Paris 6, France
[2]ETIS, ENSEA/CNRS/Université de Cergy Pontoise, France

Contribution

Two kernel-learning proposed algorithms:

1. Hybrid strategy published in [1]: new MKL algorithm ⇒ non-sparse combination between different image modalities



2. Unpublished work: learning a powered product of kernels, denoted **Product Kernel Learning** (PKL).

Hybrid MKL-scheme

Non-sparse combination between ≠ image modalities, still using ℓ_1 optimisation scheme

Idea: Each descriptor ⇒ numerous kernels with varying parameters (*e.g.* σ for gaussian)

- Each channel c: set of M kernels $K_{c,\sigma}$
- ℓ_1 MKL strategy to select the relevant σ parameter (SimpleMKL [2])

Adapted MKL problem formulation:

$$f(\mathbf{x}) = \sum_{i=1}^{N_e} \alpha_i y_i \sum_{c=1}^{N_c} \sum_{\sigma=\sigma_1}^{\sigma_M} \beta_{c,\sigma} k_{c,\sigma}(\mathbf{x}, \mathbf{x}_i) - b$$

joint optimization performed on α_i (N_e parameters) and $\beta_{c,\sigma}$ ($N_c \times M$ parameters).

- Kernel parameter tuning & learning at the same time: option to cross-validation (≠ [3]).

Product Kernel Learning: PKL

Geometric combination of kernels

$$K(\mathbf{x}_1, \mathbf{x}_2) = \prod_c k_c(\mathbf{x}_1, \mathbf{x}_2)^{\beta_c}$$

Adapted PKL problem formulation:

$$f(x) = \sum_i \alpha_i y_i \prod_c k_c(\mathbf{x}_i, \mathbf{x})^{\beta_c} - b$$

As in MKL: jointly learning α_i and β_c

- Algorithm for exponential kernels: $k_c(\mathbf{x}_1, \mathbf{x}_2) = e^{-\beta_c d_c^2(\mathbf{x}_1, \mathbf{x}_2)}$
- Alternate optimization scheme:
  1. Classic SVM solver on α
  2. Approximate second order gradient descent on β
- Step 1 convex, Step 2 not ⇒ overall problem not convex.

Results

UCI Toys like datasets for algorithm validation. Combination of Gaussian kernels on each axis.

Data set	ℓ_1-MKL (%)	PKL (%)
Inonosphere	89.0 ± 2.1	**94.2** ± 1.4
Sonar	83.8 ± 3.8	**86.2** ± 4.5

⇒ **PKL is competitive to existing MKL algorithms: more accurate, sparser, faster**

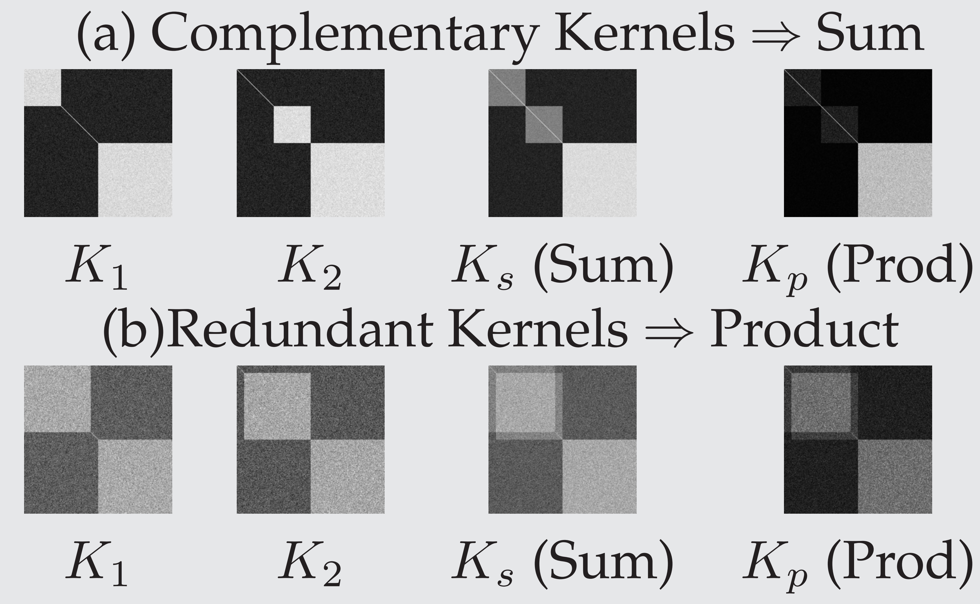
VOC 2009 Categorization with multiple visual features (15 kernels, 150 for hybrid strategy).

category	SIFT	Prod	Avg	ℓ_1-MKL [2]	ℓ_2-MKL [3]	Hybrid-MKL [1]	PKL
aeroplane	79.5	78.3	77.9	79.7	79.3	**79.7**	**79.7**
bicycle	46.9	45.9	46.0	47.8	47.9	**48.3**	47.0
bird	55.9	53.0	54.4	56.5	**57.5**	57.4	57.0
boat	61.4	56.9	56.4	62.3	60.1	**62.8**	62.2
bottle	17.6	18.7	19.1	19.5	19.8	**20.1**	19.2
bus	71.4	69.2	69.8	**72.3**	72.0	**72.3**	71.5
car	49.7	49.5	49.1	50.4	50.2	51.2	**51.7**
cat	54.8	54.4	54.1	56.8	**57.2**	57.0	56.8
chair	43.3	41.2	41.5	42.3	42.8	**43.6**	43.4
cow	21.1	24.3	24.7	21.7	25.1	24.9	**26.5**
dining-table	35.9	30.1	31.2	35.5	34.4	35.6	**36.0**
dog	39.1	35.8	35.2	37.4	37.4	38.2	**39.4**
horse	**47.5**	40.1	40.8	46.0	43.8	45.1	47.3
motorbike	46.3	54.9	55.3	53.2	**56.0**	55.8	55.0
person	82.0	81.8	81.7	82.5	82.8	**82.9**	82.8
potted-plant	23.0	29.9	30.9	30.7	**31.8**	31.3	29.4
sheep	33.0	24.8	26.7	30.1	31.7	30.7	**32.9**
sofa	32.6	25.9	25.3	32.5	29.9	32.0	**33.2**
train	68.2	67.1	67.5	69.9	69.5	**69.8**	69.4
tv-monitor	51.6	51.0	50.4	54.0	**53.6**	53.5	52.5
mean	48.0	46.7	46.9	49.0	49.1	**49.6**	**49.6**

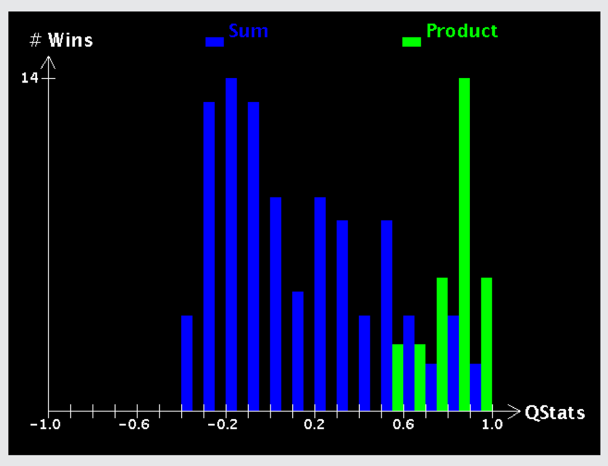
- Learned kernel combinations outperform best performing kernel (SIFT)
  - False for uniform weighting (averaging-product) ≠ [4]
  - Uniform weighting sub-optimal as soon as large performance variation between kernels
- Sparse *v.s.* dense combination: task-dependent (Learning ℓ_p norm *c.f.* [5])
  - Experimentally, Hybrid ℓ_1-MKL: good compromise between ℓ_1 and ℓ_2
- Globally, hybrid ℓ_1-MKL and PKL offer best MAP, but slight improvement

Discussion

- Unsucessful experiment: PKL for discriminative dictionnary learning, see [6]
- Unsucessful experiment: PKL for detector/descriptor combination, see [7]
- Sum or Product Kernel Learning ?



- Complementarity/Redundancy: metric ? Kernel correlation, Q-Stat, ρ-Stat ?



- Not effective in real image databases (VOC)

References

[1] David Picard, Nicolas Thome, and Matthieu Cord, "An efficient system for combining complementary kernels in complex visual categorization tasks," in *ICIP*, 2010, pp. 3877–3880.

[2] Alain Rakotomamonjy, Francis Bach, Stephane Canu, and Yves Grandvalet, "SimpleMKL," *JMLR*, vol. 9, pp. 2491–2521, 2008.

[3] Marius Kloft, Ulf Brefeld, Soeren Sonnenburg, Pavel Laskov, Klaus-Robert Müller, and Alexander Zien, "Efficient and accurate lp-norm multiple kernel learning," in *NIPS*, pp. 997–1005. 2009.

[4] Peter V. Gehler and Sebastian Nowozin, "On feature combination for multiclass object classification," in *IEEE ICCV*, 2009.

[5] Fei Yan, Krystian Mikolajczyk, Mark Barnard, Hongping Cai, and Josef Kittler, "Lp norm multiple kernel fisher discriminant analysis for object and image categorisation," in *IEEE CVPR*, 2010.

[6] H. Cai, F. Yan, and K. Mikolajczyk, "Learning weights for codebook in image classification and retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[7] Marcin Marszałek, Cordelia Schmid, Hedi Harzallah, and Joost van de Weijer, "Learning object representations for visual object class recognition," oct 2007, Visual Recognition Challange workshop, in conjunction with ICCV.