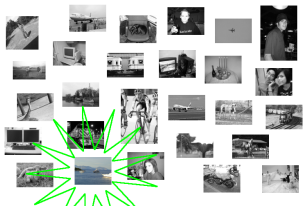


# Recherche d'images et indexation multimédia basées contenu

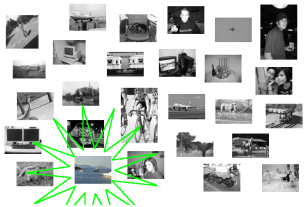
David Picard

Travaux réalisés aux LIP6 - UPMC



## Recherche de cible

# Paradigmes de la recherche d'images

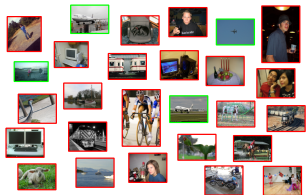


Recherche de cible

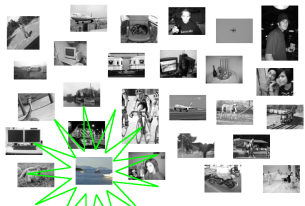


Exploration interactive

# Paradigmes de la recherche d'images



Classification



Recherche de cible

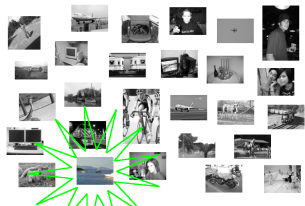
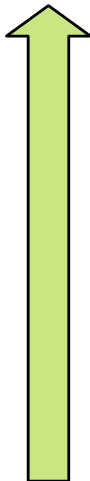


Exploration interactive

# Paradigmes de la recherche d'images



Classification



Recherche de cible



Exploration interactive  
Sémantique

## Recherche de quasi-copies



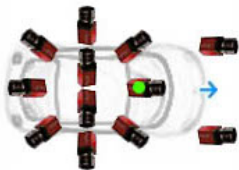
# Numérisation des rues de Paris en haute résolution



12 caméras

3 lasers

10 000 panoramiques  
pour le XIIIème



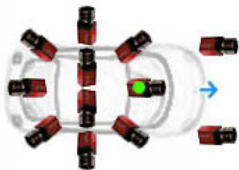
## Numérisation des rues de Paris en haute résolution



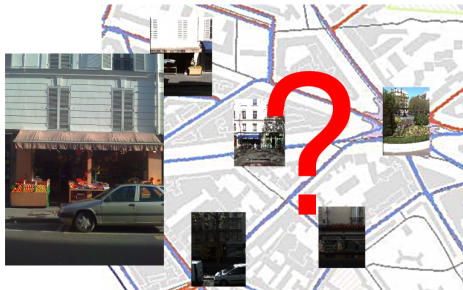
12 caméras

3 lasers

10 000 panoramiques  
pour le XIIIème

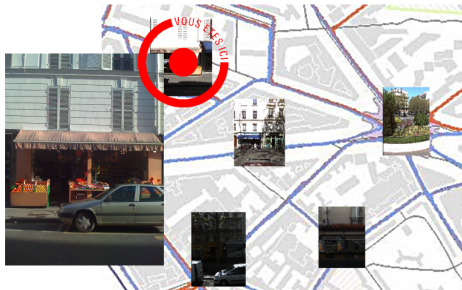


Moteur de recherche multimédia en environnement urbain



À partir d'une photo d'un  
usager (prise avec un  
smartphone, par ex.)

Trouver la localisation  
(panoramique  
correspondante) dans la  
base iTowns





Descripteurs locaux très discriminants

SIFT [Lowe04] : histogramme d'orientations de gradient

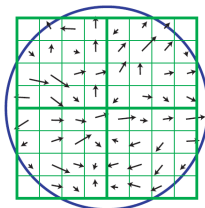
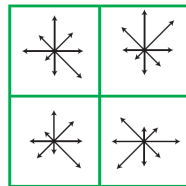


Image gradients



Keypoint descriptor

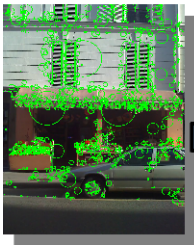
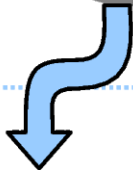
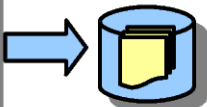
Apparier les points d'intérêt entre la requête et la cible

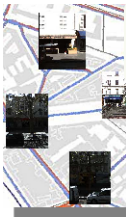


Apparier les points d'intérêt entre la requête et la cible

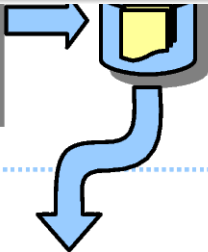


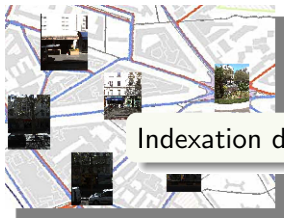
Recherche des  $k$  plus proches voisins



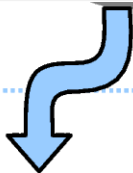


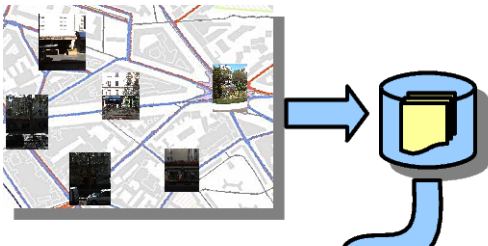
Extraction des descripteurs pour toutes les images de la base



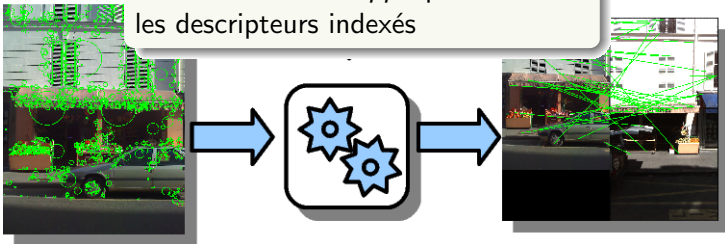


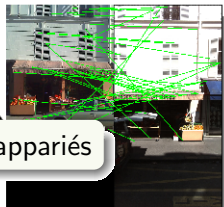
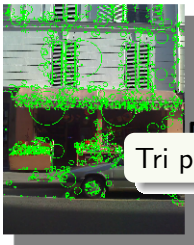
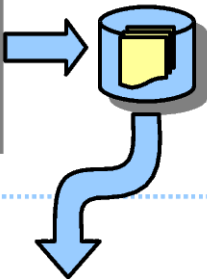
Indexation de tous les descripteurs





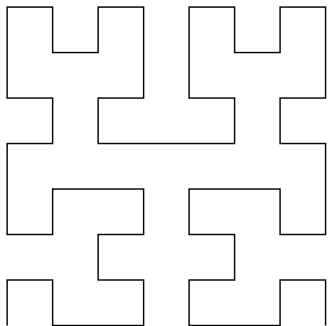
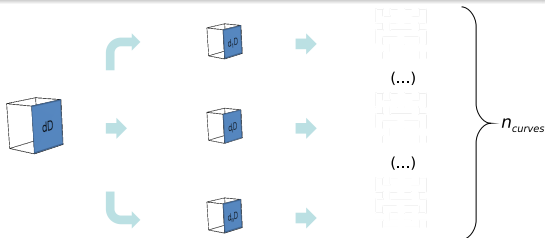
Recherche des  $k - ppv$  parmi tous les descripteurs indexés





Tri par vote des descripteurs appariés

Multicurves : Découpage de l'espace en blocs (sous espaces)

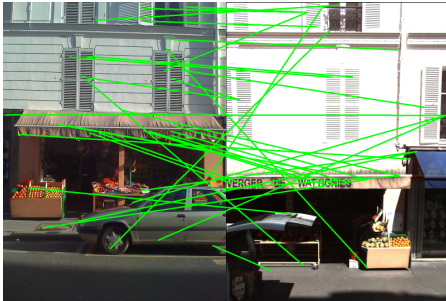


- Indexation de chaque bloc par une courbe de Hilbert
- Espace de dimension 1  $\Rightarrow$  relation d'ordre (indexation facile)
- Recherche de candidats sur chaque courbe
- Calcul de la vraie distance sur les candidats



Tri par vote de  $k - ppv$

## Post-traitements



Suppression des faux appariements par critères géométriques

- Estimation d'une transformation affine entre la requête et la cible
- Max dans l'histogramme des différences d'angles [Jegou08]



Tri par vote de  $k - ppv$  et nettoyage géométrique



Site du projet : <http://www.itowns.fr>

Quelques publications :

[1] Study of SIFT Descriptors for Image Matching based Localization in Urban Street View Context

Picard, D. and Cord, M. and Valle, E. In *CMRT09 - CityModels, Roads and Traffic*. Volume 38, Part 3 / W4.

[2] Geometric Consistency Checking for Local-Descriptor Based Document Retrieval

Valle, Eduardo, Picard, David and Cord, Matthieu. In *Proceedings of the 9th ACM Symposium on Document Engineering*

Certaines données disponibles : mail [david.picard@lip6.fr](mailto:david.picard@lip6.fr)

## Classification - Challenge PASCAL VOC 2009/2010



Objectif :

Pour une catégorie d'objets donnée, quelles sont les images dans lesquelles cet objet apparaît ?

Objectif :

Pour une catégorie d'objets donnée, quelles sont les images dans lesquelles cet objet apparaît ?

- Classification binaire (pertinent/non-pertinent)

Objectif :

Pour un catégorie d'objets donnée, quelles sont les images dans lesquelles cet objet apparaît ?

- Classification binaire (pertinent/non-pertinent)
- Pas de localisation (pas de bounding-box)

Objectif :

Pour une catégorie d'objets donnée, quelles sont les images dans lesquelles cet objet apparaît ?

- Classification binaire (pertinent/non-pertinent)
- Pas de localisation (pas de bounding-box)
- Catégories très variées (méthode générique)

Objectif :

Pour un catégorie d'objets donnée, quelles sont les images dans lesquelles cet objet apparaît ?

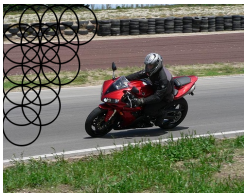
- Classification binaire (pertinent/non-pertinent)
- Pas de localisation (pas de bounding-box)
- Catégories très variées (méthode générique)
- Variance intra classe très grande (changement d'échelle, de point de vue, etc.)

Objectif :

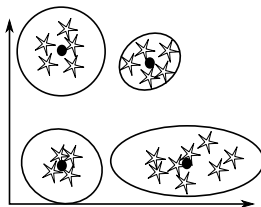
Pour un catégorie d'objets donnée, quelles sont les images dans lesquelles cet objet apparaît ?

- Classification binaire (pertinent/non-pertinent)
- Pas de localisation (pas de bounding-box)
- Catégories très variées (méthode générique)
- Variance intra classe très grande (changement d'échelle, de point de vue, etc.)

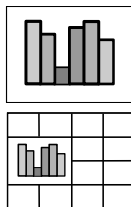
Challenge difficile



extraction

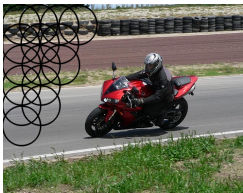


prototypes

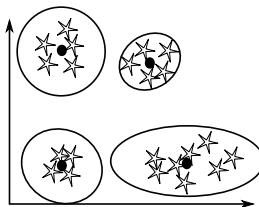


local histograms

- 1 Extraction de descripteurs locaux (SIFT par exemple)
- 2 Dictionnaire de prototypes ( "*dictionnaire visuel*" obtenu par clustering de tous les descripteurs de toutes les images)
- 3 Histogramme des occurrences des prototypes dans un région de l'image (combien de fois chaque "mot visuel" apparaît dans l'image)



extraction



prototypes



local histograms

- ① Extraction de descripteurs locaux (SIFT par exemple)
- ② Dictionnaire de prototypes ( "*dictionnaire visuel*" obtenu par clustering de tous les descripteurs de toutes les images)
- ③ Histogramme des occurrences des prototypes dans un région de l'image (combien de fois chaque "mot visuel" apparaît dans l'image)

Classification : SVM sur les histogrammes

Support Vector Machines : fonction de classification de la forme

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle - b \quad (1)$$

optimisée selon

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{\mathbf{x}_i \in \mathcal{A}} \ell(y_i f(\mathbf{x}_i)) \quad (2)$$

mapping non-linéaire :

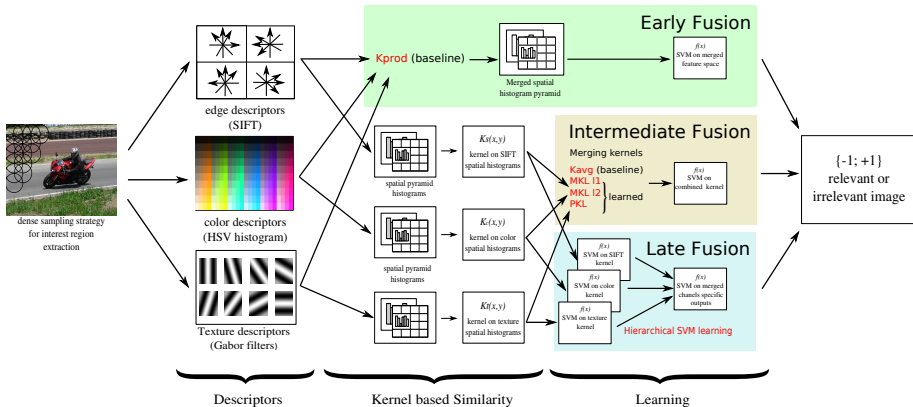
$$f(\mathbf{x}) = \langle \phi(\mathbf{w}), \phi(\mathbf{x}) \rangle - b \quad (3)$$

avec  $\langle \phi(\mathbf{x}_1), \phi(\mathbf{x}_2) \rangle = k(\mathbf{x}_1, \mathbf{x}_2)$ , pas besoin d'explicitier  $\phi$  (kernel trick), alors

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in \mathcal{A}} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) - b \quad (4)$$

Comment combiner plusieurs types de descripteurs (couleur, texture, contour, etc.) ?

# Comment combiner plusieurs types de descripteurs (couleur, texture, contour, etc.) ?



Différentes configurations possibles pour la combinaison de descripteurs dans le contexte SVM

SVM :

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle - b \quad (5)$$

$$\text{solution de} \quad \min_w \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{\mathbf{x}_i \in \mathcal{A}} \ell(y_i f(\mathbf{x}_i)) \quad (6)$$

Problème dual :

$$f(\mathbf{x}) = \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) - b \quad (7)$$

$$\text{solution de} \quad \max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \quad (8)$$

$$\text{s.t.} \quad \forall i, 0 \leq \alpha_i \leq C, \quad \sum_i \alpha_i y_i = 0 \quad (9)$$

SVM :

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle - b \quad (5)$$

$$\text{solution de} \quad \min_w \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{\mathbf{x}_i \in \mathcal{A}} \ell(y_i f(\mathbf{x}_i)) \quad (6)$$

Problème dual :

$$f(\mathbf{x}) = \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) - b \quad (7)$$

$$\text{solution de} \quad \max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \quad (8)$$

$$\text{s.t.} \quad \forall i, 0 \leq \alpha_i \leq C, \quad \sum_i \alpha_i y_i = 0 \quad (9)$$

Fusion intermédiaire : combinaison de noyaux

$$k(\mathbf{x}, \mathbf{x}') = (h(k_1, \dots, k_N))(\mathbf{x}, \mathbf{x}') \quad (10)$$

Combinaisons “*Baseline*” :

Average :  $k(\mathbf{x}, \mathbf{x}') = \sum_n k_n(\mathbf{x}, \mathbf{x}')$

Product :  $k(\mathbf{x}, \mathbf{x}') = \prod_n k_n(\mathbf{x}, \mathbf{x}')$

Combinaisons “*Baseline*” :

Average :  $k(\mathbf{x}, \mathbf{x}') = \sum_n k_n(\mathbf{x}, \mathbf{x}')$

Product :  $k(\mathbf{x}, \mathbf{x}') = \prod_n k_n(\mathbf{x}, \mathbf{x}')$

Peut on optimiser la combinaison par apprentissage ?

MKL [Bach04] : apprendre une combinaison linéaire de noyaux

$$k(\mathbf{x}, \mathbf{x}') = \sum_n \beta_n k_n(\mathbf{x}, \mathbf{x}') \quad (11)$$

classifieur :

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in \mathcal{A}} \alpha_i y_i \sum_n \beta_n k_n(\mathbf{x}_i, \mathbf{x}) - b \quad (12)$$

$$\text{solution de } \max_{\alpha, \beta} \sum_i \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j \sum_n \beta_n k_n(\mathbf{x}_i, \mathbf{x}_j) \quad (13)$$

$$\text{s.t. } \forall i, 0 \leq \alpha_i \leq C, \sum_i \alpha_i y_i = 0$$

$$\sum_n \beta_n = 1, \forall i, \beta_i \geq 0$$

MKL [Bach04] : apprendre une combinaison linéaire de noyaux

$$k(\mathbf{x}, \mathbf{x}') = \sum_n \beta_n k_n(\mathbf{x}, \mathbf{x}') \quad (11)$$

classifieur :

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in \mathcal{A}} \alpha_i y_i \sum_n \beta_n k_n(\mathbf{x}_i, \mathbf{x}) - b \quad (12)$$

$$\text{solution de } \max_{\alpha, \beta} \sum_i \alpha_i - \frac{1}{2} \sum_{ij} \alpha_i \alpha_j y_i y_j \sum_n \beta_n k_n(\mathbf{x}_i, \mathbf{x}_j) \quad (13)$$

$$s.t. \quad \forall i, 0 \leq \alpha_i \leq C, \sum_i \alpha_i y_i = 0$$

$$\sum_n \beta_n = 1, \forall i, \beta_i \geq 0$$

Algorithmes rapides : SimpleMKL[Rakoto08], SILP [Sonnenburg06]

Sparcity ?

Contraintes  $\sum_n \beta_n = 1, \forall i, \beta_i \geq 0$  imposent solution clairsemée en  $\beta$

Idéal pour faire de la sélection de noyau

Sparcity ?

Contraintes  $\sum_n \beta_n = 1, \forall i, \beta_i \geq 0$  imposent solution clairsemée en  $\beta$

Idéal pour faire de la sélection de noyau

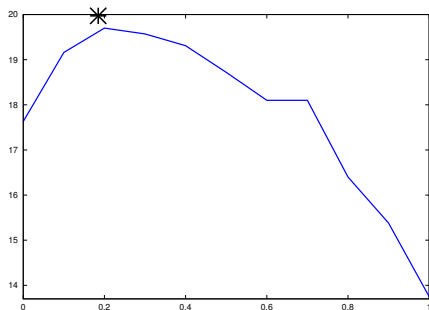
Fusion de descripteurs : Peu de noyaux, tous informatifs - pas de sélection  
!

Idee : noyau paramétrique pour chaque descripteur (ex. gaussien), utiliser la combinaison pour apprendre les paramètres

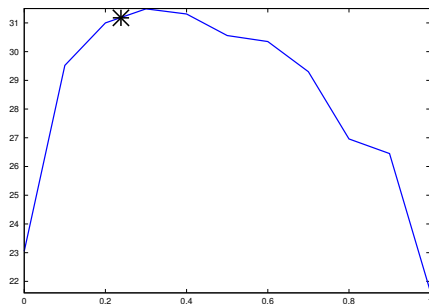
EX : 3 types de descripteurs (contours, couleurs, textures)  $\Rightarrow$  10 noyaux gaussien pour chaque descripteur (10  $\sigma$  différents)

30 noyaux au total et nécessité de faire de la sélection : MKL bien adapté

## Comparaison entre MKL adapté et somme *ad hoc* VOC 2009



bottle



potted-plant

2 descripteurs (SIFT et couleur)  $\Rightarrow$  MKL avec 20 noyaux (10 SIFT et 10 couleur)

Noyau global :  $\lambda k_{SIFT} + (1 - \lambda) k_{couleur}$

Performances en fonction de  $\lambda$  : combinaison manuelle (courbe), MKL (\*)

Produit de noyaux ?

Noyau :

$$K(\mathbf{x}, \mathbf{x}') = \prod_j k_j(\mathbf{x}, \mathbf{x}')^{\beta_j} \quad (14)$$

Produit de noyaux ?

Noyau :

$$K(\mathbf{x}, \mathbf{x}') = \prod_j k_j(\mathbf{x}, \mathbf{x}')^{\beta_j} \quad (14)$$

$K$  est-il un noyau ? (vrai si  $j \in \mathbb{N}$ , ? cas général)

Cas particulier :  $k_j$  gaussien  $e^{-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\sigma^2}} \Rightarrow k_j^{\beta_j}$  revient à changer  $\sigma$

Produit de noyaux ?

Noyau :

$$K(\mathbf{x}, \mathbf{x}') = \prod_j k_j(\mathbf{x}, \mathbf{x}')^{\beta_j} \quad (14)$$

$K$  est-il un noyau ? (vrai si  $j \in \mathbb{N}$ , ? cas général)

Cas particulier :  $k_j$  gaussien  $e^{-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\sigma^2}} \Rightarrow k_j^{\beta_j}$  revient à changer  $\sigma$

$$\begin{aligned} \prod_j k_j(\mathbf{x}_1, \mathbf{x}_2)^{\beta_j} &= \prod_j \langle \phi_j(\mathbf{x}_1), \phi_j(\mathbf{x}_2) \rangle \\ &= \langle \phi_1(\mathbf{x}_1) \otimes \cdots \otimes \phi_M(\mathbf{x}_1), \phi_1(\mathbf{x}_2) \otimes \cdots \otimes \phi_M(\mathbf{x}_2) \rangle \end{aligned} \quad (15)$$

$\Rightarrow$  Optimiser les  $\alpha_i$  et les  $\phi_j$  conjointement

$$f(\mathbf{x}) = \langle w, \phi_1(\mathbf{x}) \otimes \cdots \otimes \phi_M(\mathbf{x}) \rangle - b \quad (16)$$

Trouver la solution de :

$$\begin{aligned} \max_{\alpha_i, \beta_m} \quad & -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \prod_m k_m(x_i, x_j)^{\beta_m} + \sum_i \alpha_i \\ \text{s.t.} \quad & \sum_i \alpha_i y_i = 0 \quad \forall i, \quad 0 \leq \alpha_i \leq C \quad \forall m, \quad \beta_m \geq 0 \end{aligned} \quad (17)$$

Optimisation alternée :

- $\alpha$  fixés, descente de gradient du second ordre sur les  $\beta$
- $\beta$  fixés, optimisation des  $\alpha$  à l'aide un solveur SVM

## Test sur VOC 2009 :

| category  | SIFT | Color | Gabor | Prod | Avg  | $\ell_1$ -MKL | $CV_{\ell_1}$ MKL | PKL         |
|-----------|------|-------|-------|------|------|---------------|-------------------|-------------|
| bicycle   | 46.9 | 25.5  | 32.5  | 45.9 | 46.0 | 47.8          | <b>48.3</b>       | 47.0        |
| bird      | 55.9 | 30.1  | 39.2  | 53.0 | 54.4 | 56.5          | <b>57.4</b>       | 57.0        |
| bottle    | 17.6 | 13.7  | 13.2  | 18.7 | 19.1 | 19.5          | <b>20.1</b>       | 19.2        |
| car       | 49.7 | 29.2  | 32.4  | 49.5 | 49.1 | 50.4          | 51.2              | <b>51.7</b> |
| cow       | 21.1 | 10.2  | 11.7  | 24.3 | 24.7 | 21.7          | 24.9              | <b>26.5</b> |
| dog       | 39.1 | 24.4  | 24.2  | 35.8 | 35.2 | 37.4          | 38.2              | <b>39.4</b> |
| motorbike | 46.3 | 30.8  | 43.2  | 54.9 | 55.3 | 53.2          | <b>55.8</b>       | 55.0        |
| sofa      | 32.6 | 12.6  | 13.0  | 25.9 | 25.3 | 32.5          | 32.0              | <b>33.2</b> |

Table: VOC 2009: (M)AP for different combination strategies.

## Conclusions

- $CV_{\ell_1}$ -MKL est un bon compromis entre sélection et combinaison (meilleures performances car paramétrage plus précis)
- MKL et PKL complémentaires

[1] An efficient system for combining complementary kernels in complex visual categorization tasks

Picard, D., Thome, N., and Cord, M. (2010a). *International Conference on Image Processing (ICIP'10)*.