Robust Matching and Recognition using Context-Dependent Kernels

Hichem SAHBI[*]
Col. Jean-Yves AUDIBERT[†,‡], Jaonary RABARISOA[†]
and Renaud KERIVEN[†]

[*] CNRS - LTCI, UMR 5141, TELECOM ParisTech, Paris.
[†] Certis - Ecole des Ponts, Paris.
[‡] Willow - ENS Ulm/ INRIA, Paris.

July 7, 2008

Outline

- I. Issues and Contribution.
- II. Subset Kernels.
- III. Context Dependent Kernel Design.
- IV. Experiments and Take-Home Message.

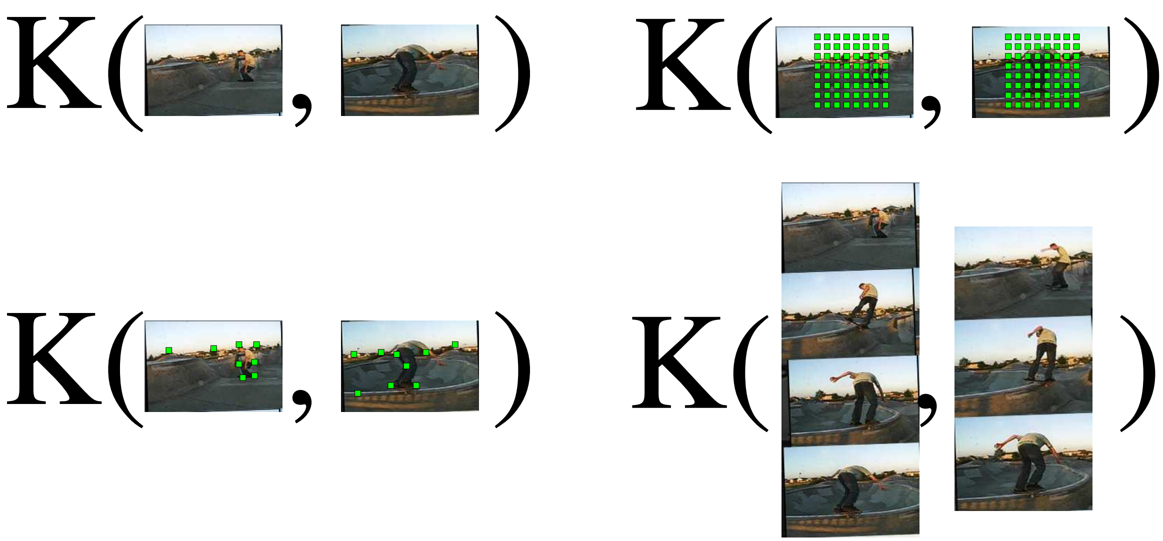
Outline

- I. Issues and Contribution.
- II. Subset Kernels.
- III. Context Dependent Kernel Design.
- IV. Experiments and Take-Home Message.

I. Holistic vs Subset Kernels

- Initially kernels were designed in order to handle fixed length and ordered data (i.e. as Holistic kernels).

- Define $\mathcal{X}$ as an input space and $X, X' \in \mathcal{X}$. $K : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}^+$.



- Unfixed length and unordered data (graphs, trees, interest points) require subset and/or Local kernels.

I. Methods

- Alignment kernels
  - Max kernel (Wallraven et al. 2003).
  - Circular shift kernel (Lyu 2005).
  - Intermediate kernel (Boughorbel 2005).
  - Pyramid match kernel (Grauman & Trevor 2007).
  - Dynamic programming kernel (Bahlmann et al. 2002).
- Order and length insensitive kernels
  - Kullback Leibler divergence kernels (Kondor & Jebara, 2003).
  - Battacharyya affinity (Moreno et al. 2003).
  - Principal angles (Wolf & Shashua, 2003).
  - Subset kernels (Shashua & Hazan, 2004).

I. Contribution

- Kernel Design
  - A new variational framework for kernel design.
  - Context dependent similarity kernel.
  - Fixed point of an energy.
  - Positive definite (Mercer) kernel.
- Applications
  - Interest point based object matching and recognition.
  - Extension to video retrieval.
  - Extension to machine translation.

Outline

- I. Issues and Contribution.

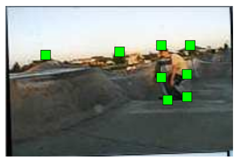
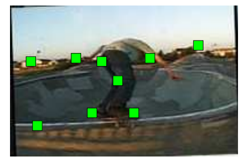
- II. Subset Kernels.

- III. Context Dependent Kernel Design.

- IV. Experiments and Take-Home Message.

II. Subset Kernels

Let's consider a database of images.

K( , )

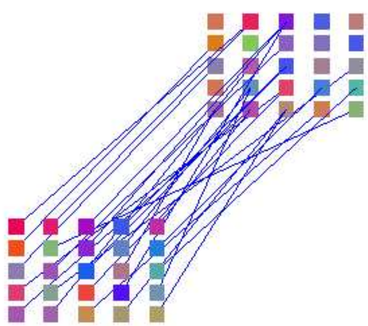
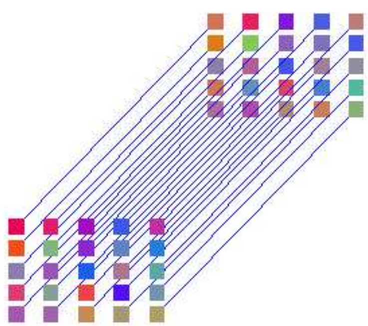
Definition 1. [Subset Kernels] *let* $\mathcal{X} = \cup_{p \in \mathbb{N}^+} \mathcal{S}_p$ *be an input space, and consider* $\mathcal{S}_p, \mathcal{S}_q \subseteq \mathcal{X}$ *as two finite subsets of* $\mathcal{X}$*. We define the similarity function or kernel* K *between* $\mathcal{S}_p = \{x_1^p, \dots, x_i^p, \dots, x_n^p\}$ *and* $\mathcal{S}_q = \{x_1^q, \dots, x_j^q, \dots, x_m^q\}$ *as*

$$K(\mathcal{S}_p, \mathcal{S}_q) = \sum_i^n \sum_j^m k\left(x_i^p, x_j^q\right), \tag{1}$$

Proposition 1. K *defined in (1) is also a Mercer kernel. i.e.,* $K(\mathcal{S}_p, \mathcal{S}_q) = \langle \Phi(\mathcal{S}_p), \Phi(\mathcal{S}_q) \rangle$*, for some* $\Phi : \mathcal{X} \longrightarrow \mathcal{H}$*.*

II. Subset Kernels (Minor Kernel Deficiency)

Naive matching	**'H'**	**'i'**		**'S'**	**'i'**	**'r'**
'S'	0	0	-	1	0	0
'i'	0	1	-	0	1	0
'r'	0	0	-	0	0	1
Context-dependent	-	-	-	-	-	-
'S'	0	0	-	.38	0	0
'i'	0	.36	-	0	.39	0
'r'	0	0	-	0	0	.38

Outline

- I. Issues and Contribution.
- II. Subset Kernels.
- III. Context Dependent Kernel Design.
- IV. Experiments and Take-Home Message.

III. Kernel Design

Minor kernel design: $k(x_i^p, x_j^q) = \mathbb{P}(X' = x_j^q, X = x_i^p)$.

$$
\begin{aligned}
\min_{\mu} \quad & \sum_{i \in \mathcal{I}_p, j \in \mathcal{I}_q} k(x_i^p, x_j^q)\ d(x_i^p, x_j^q) + \\
& \beta \sum_{i \in \mathcal{I}_p, j \in \mathcal{I}_q} k(x_i^p, x_j^q)\ \log(k(x_i^p, x_j^q)) + \\
& \alpha \sum_{i \in \mathcal{I}_p, j \in \mathcal{I}_q} k(x_i^p, x_j^q) \left(- \sum_{\substack{x_k^p \in \mathcal{N}_p(x_i^p), \\ x_\ell^q \in \mathcal{N}_q(x_j^q)}} k(x_k^p, x_\ell^q) \right) \qquad (2) \\
\text{s.t.} \quad & k(x_i^p, x_j^q) \in [0,1] \quad i \in \mathcal{I}_p, \quad j \in \mathcal{I}_q \\
& \sum_{i,j} k(x_i^p, x_j^q) = 1
\end{aligned}
$$

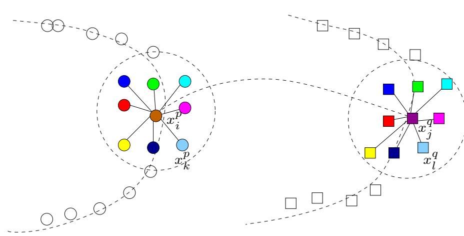
III. Kernel Design

Proposition 2. *(2) admits a solution in the form of a context-dependent kernel* $k_t(x_i^p, x_j^q) = v_t(x_i^p, x_j^q)/Z_t$*, with* $t \in \mathbb{N}^+$*,* $Z_t = \sum_{i,j} v_t(x_i^p, x_j^q)$ *and* $v_t(x_i^p, x_j^q)$ *defined as*

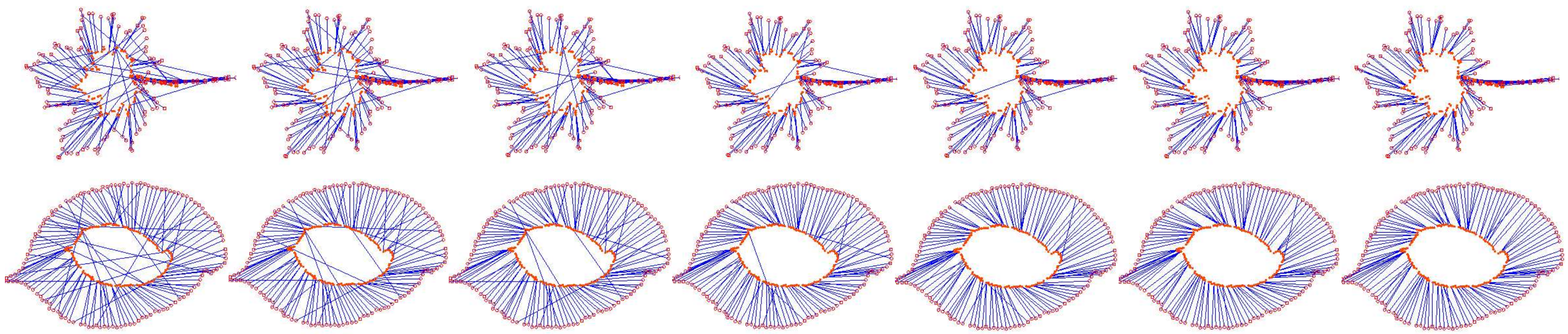
$$\exp\left(-\frac{d(x_i^p, x_j^q)}{\beta} - 1\right) \times \exp\left(\frac{2\alpha}{\beta} \sum_{k,\ell} \mathbb{V}(x_i^p, x_k^p, x_j^q, x_\ell^q) \;\; k_{t-1}(x_k^p, x_\ell^q)\right) \tag{3}$$

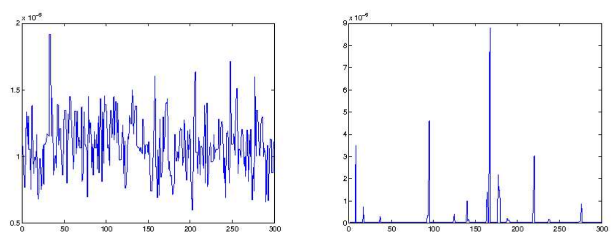
which is also a Gibbs distribution.

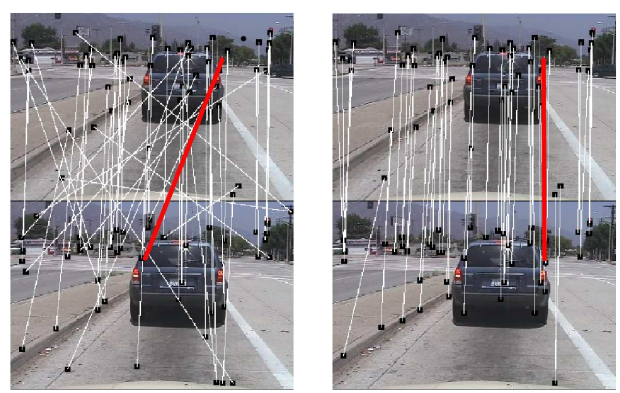
Proof 1. *see appendix.*



III. Neighborhood Parameter







III. Mercer Condition

Proposition 3. [Closure] *the sum and the product of any two Mercer kernels is a Mercer kernel. The exponential of any Mercer kernel is also a Mercer kernel.*

Proof 2. *see, for instance, (Cristianini and Taylor 2000).*

Proposition 4. *let* $\mathbb{V}(x_i^p, x_k^p, x_j^q, x_\ell^q) = g(x_i^p, x_k^p)g(x_j^q, x_\ell^q)$*, consider* $g : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ *and* k_0 *positive definite. The kernel* k_t *is then positive definite.*

III. Mercer Condition (Sketch of the Proof)

Proof 3.

- *Initially* $(t = 0)$, k_0 *is per definition a positive definite kernel.*
- *By induction, let us assume* k_{t-1} *a Mercer kernel i.e.,* $\exists \Phi_{t-1} : k_{t-1}(x, x') = \big\langle \Phi_{t-1}(x), \Phi_{t-1}(x') \big\rangle$, $\forall x, x' \in \mathcal{X}$.
- *Now, the sufficient condition will be to show that* $\Big(\sum_{y,y'} \mathbb{V}(x, y, x', y')\ k_{t-1}(y, y') \Big)$ *is also a Mercer kernel.*
- *Then, by the closure of the exponential and the product (see proposition 3),* k_t *will then be Mercer.*

III. Convergence

Let us assume $0 \leq g \leq 1$, and remind $\mu^{(t)} \in \mathbb{R}^{n \times m}$ be the vector of components $\mu_{i,j}^{(t)} = k_t(x_i^p, x_j^q)$. Introduce the mapping $f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}$ defined by its component

$$f_{i,j}(v) = \exp\left(-1 - \frac{d(x_i^p, x_j^q)}{\beta} + \frac{2\alpha}{\beta} \sum_{k,\ell} g(x_i^p, x_k^p) g(x_j^q, x_\ell^q) v_{k,\ell} \right) \tag{4}$$

By construction of the kernel k_t, we have $\mu^{(t)} = f\big(\mu^{(t-1)}\big)$. Let A and B satisfy

$$\sup_{1 \leq i \leq n,\ 1 \leq j \leq m} \sum_{k,\ell} g(x_i^p, x_k^p) g(x_j^q, x_\ell^q) \leq A \tag{5}$$

$$\sum_{i,j} \exp\left(-1 - \frac{d(x_i^p, x_j^q)}{\beta} \right) \leq B \tag{6}$$

Consider $L = \frac{2B\alpha}{\beta}\exp\left(\frac{2\alpha A}{\beta}\right)$, and let $\mathcal{B} = \left\{v \in \mathbb{R}^{n\times m} : \forall\, 1 \leq i \leq n,\ 1 \leq j \leq m,\ |v_{i,j}| \leq 1\right\}$ be the $\|\cdot\|_\infty$-ball of radius 1. Finally, let $\|\cdot\|_1$ denote the 1-norm on $\mathbb{R}^{n\times m}$: $\|u\|_1 = \sum_{1\leq i\leq n,\ 1\leq j\leq n} |u_{i,j}|$.

Proposition 5. *If $\|\mu^{(0)}\|_\infty \leq 1$ and $2\alpha A \leq \beta$, then we have $f(\mathcal{B}) \subset \mathcal{B}$, and on $\mathcal{B}$, f is L-Lipschitz for the norm $\|\cdot\|_1$.*

In particular, if $L < 1$, then there exists a unique $\tilde{v} \in \mathcal{B}$ such that $f(\tilde{v}) = \tilde{v}$, and the sequence $(\mu^{(t)})$ satisfies

$$\|\mu^{(t)} - \tilde{v}\|_1 \leq L^t \|\mu^{(0)} - \tilde{v}\|_1 \underset{t\rightarrow+\infty}{\longrightarrow} 0. \tag{7}$$

III. Convergence

Proof. The first assertion is proved by induction by checking that for $\|v\|_\infty \leq 1$, we have

$$f_{i,j}(v) \leq \exp\Big(-1 + \frac{2\alpha}{\beta} \sum_{k,\ell} g(x_i^p, x_k^p) g(x_j^q, x_\ell^q) v_{k,\ell} \Big) \quad (8)$$

$$\leq \exp\Big(-1 + \frac{2\alpha}{\beta} A \Big) \leq 1. \quad (9)$$

For the second assertion, note that for any v in $\mathcal{B}$, we have $|\frac{\partial f_{i,j}}{\partial v_{k,\ell}}(v)| \leq \exp\big(-1 - \frac{d(x_i^p, x_j^q)}{\beta}\big)$ For any v, v' in $\mathcal{B}$, we have

$$\|f(v) - f(v')\|_1 = \sum_{i,j} |f_{i,j}(v) - f_{i,j}(v')| = (**)$$

$$
\begin{aligned}
(**) \quad &\leq \sum_{i,j} \exp\big(-1 - \frac{d(x_i^p, x_j^q)}{\beta}\big) \frac{2\alpha}{\beta} \exp\big(\frac{2\alpha}{\beta} A\big) && (10)\\
&\times \big| \sum_{k,\ell} g(x_i^p, x_k^p) g(x_j^q, x_\ell^q) v_{k,\ell} && (11)\\
&- \sum_{k,\ell} g(x_i^p, x_k^p) g(x_j^q, x_\ell^q) v'_{k,\ell} \big| && (12)\\
&\leq \sum_{i,j} \exp\big(-1 - \frac{d(x_i^p, x_j^q)}{\beta}\big) \frac{2\alpha}{\beta} \exp\big(\frac{2\alpha}{\beta} A\big) \|v - v'\|_1 && (13)\\
&\leq L \|v - v'\|_1 && (14)
\end{aligned}
$$

which proves the second assertion. The last assertion directly comes from the fixed-point theorem $\Box$.

Outline

- I. Issues and Contribution.
- II. Subset Kernels.
- III. Context Dependent Kernel Design.
- IV. Experiments and Take-Home Message.

IV. Results

Experiments conducted on MNIST $(10 \times 100 + 100)$ and Swedish $(15 \times 25 + 50)$.
$K(\mathcal{S}_p, \mathcal{S}_q) = \sum_i \sum_j k_0(x_i^p, x_j^q)$ vs. $K(\mathcal{S}_p, \mathcal{S}_q) = \sum_i \sum_j k_t(x_i^p, x_j^q)$.

Initialization	Linear	Polynomial
Iterations (Mnist)		
k_0	11.4 ± 4.42	9.15 ± 4.63
k_1	8.80 ± 4.77	5.6 ± 2.72
k_2	6.90 ± 3.55	5.8 ± 2.36
k_3	6.90 ± 3.41	5.2 ± 2.07
k_4	**6.90± 3.41**	**5.2 ± 2.07**

Initialization	Linear	Polynomial
Iterations (Swedish)		
k_0	11.7 ± 2.88	6.53 ± 6.34
k_1	6.00 ± 2.30	3.33 ± 2.73
k_2	**3.06± 1.88**	**3.33± 2.73**

Conclusion and Extensions

- Take Home Message
  - Similarity in kernel design is not only intrinsic, it also depends on the context.
  - Context dependent kernels show better matching results with respect to context free kernels.
  - Model free approach (contrast to RANSAC, Hough, etc.)
  - CDKs are also (rotation, translation and scale) invariant and tolerant to deformations.

- Extensions
  - Video matching and retrieval.
  - Machine translation.

Machine Translation

- S_1=**he may go**, S_2=**il mai partir**, S_3=**il peut partir**.

- $K(S_1, S_2|S_3) = k(\textbf{he,il}) + k(\textbf{may,mai} \mid \textbf{peut}) + k(\textbf{go,partir})$.

- Context free or word based translation: if $k(\textbf{may,mai}) > k(\textbf{may,peut})$ then wrong french translation.

- Context or phrase based translation:

$$\begin{aligned} k_t(\textbf{may,mai}|\textbf{peut}) &= \exp\left(-\mathbb{C}(\textbf{may,mai}|\textbf{peut})\right) \\ & \exp(g(\textbf{he,may})\ g(\textbf{il,mai}|\textbf{peut})\ k_{t-1}(\textbf{he,il})) \\ & \exp(g(\textbf{go,may})\ g(\textbf{partir,mai}|\textbf{peut})\ k_{t-1}(\textbf{go,partir})) \end{aligned}$$

- If $k(\textbf{may,mai}) < k(\textbf{may,peut})$, then correct french translation.

Video Retrieval

- Similarity between images depends on their context.