Kernel-Based Implicit Regularization of Structured Objects

François-Xavier Dupé, Sébastien Bougleux, Luc Brun, Olivier Lezoray, Abderrahim Elmoataz

Kernel-Based Implicit Regularization of Structured Objects

François-Xavier Dupé*, Sébastien Bougleux†, Luc Brun*, Olivier Lézoray† and Aberrahim Elmoataz*
ENSICAEN - Université de Caen - GREYC CNRS UMR 6072
6 Bd du Maréchal Juin
14050 Caen Cedex, France
*Email: *{fdupe,luc.brun,abder}@greyc.ensicaen.fr, †{sebastien.bougleux,olivier.lezoray}@unicaen.fr*

***Abstract*—Weighted graph regularization provides a rich framework that allows to regularize functions defined over the vertices of a weighted graph. Until now, such a framework has been only defined for real or multivalued functions hereby restricting the regularization framework to numerical data. On the other hand, several kernels have been defined so far on structured objects such as strings or graphs. Using definite positive kernels, each original object is associated, by the "kernel trick", to one element of a Hilbert space. As a consequence, this paper proposes to extend the weighted graph regularization framework to objects implicitly defined by their kernel hereby performing the regularization within the Hilbert space associated to the kernel. This work opens the door to the regularization of structured objects.**

***Keywords*-kernel; graph-based regularization; total variation; classification; discrete structures**

I. Introduction

Structured objects such as strings or graphs allow to encode objects made of several labels related by sequential or structural dependencies. Until recent years, the similarity or the distance between two structured objects was mainly computed from the two related notions of maximal common sub-graph and graph edit distance [1].

However, such measures of distance (or similarity) operate directly in the space of structured objects which lacks mathematical properties. This lack forbids the use of basic statistical tools such as mean or variance. Such a limitation explains at least partially the recent popularity of explicit graph embedding algorithms [2]. Kernels on structured objects provide an elegant alternative solution to this problem. By using a symmetric positive definite kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ on a set $\mathcal{X}$, any object of $\mathcal{X}$ can be mapped to a Hilbert space $\mathcal{F}$ (called feature space) by a function $f : \mathcal{X} \to \mathcal{F}$ such that

$$\forall x, y \in \mathcal{X}, \quad K(x, y) = \langle f(x), f(y) \rangle, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ represents the inner product in $\mathcal{F}$. Many learning algorithms use only an inner product between input data hereby avoiding the explicit computation of the feature map f. This computation scheme corresponds to the well-known "kernel trick". See [3] or [4] for a survey on kernels and kernel methods.

An ideal fonction f should map two close objects x and y of $\mathcal{X}$ onto two close points $f(x)$ and $f(y)$ of $\mathcal{F}$. Thus, the regularity of the function f should be defined according to a similarity criterion between objects. This property is usually achieved through the design of the associated kernel K. However, enforcing the regularity of a given feature map f, or regularizing f according to a similarity measure different from its kernel, are not straightforward if one wants to keep the positive definiteness of the kernel. Note that given a regularized version $g : \mathcal{X} \to \mathcal{F}$ of f, the kernel K^g defined by $K^g(x, y) = \langle g(x), g(y) \rangle$ is positive definite and provides a regularized version of the initial kernel K.

Several approaches can be used to achieve such a regularization. Among them, variational models based on energy minimization allow to find a function g which is regular on $\mathcal{X}$ and remains close to the initial feature map f. Such a problem can be formalized as the solution of the following optimization problem

$$\underset{g : \mathcal{X} \to \mathcal{F}}{\operatorname{argmin}} \; E(f, g, \alpha) = \alpha R(g) + (1 - \alpha) A(g, f), \tag{2}$$

where the functional $R(g)$ measures the regularity or smoothness of g, the functional $A(g, f)$ measures the approximation error between f and g, and the parameter $\alpha \in (0, 1)$ measures the trade-off between these two terms. The determination of a function g minimizing an energy defined by a regularization term and an approximation one is well-known and closely related to the definition of a kernel (see e.g. [3], [4], [5]). However, in the specific case of the regularization of mapping functions, we have to face to two main problems: Firstly, the mapping function f to be regularized is usually only implicitly defined through its associated kernel K. Secondly, even considering an explicit formulation of the mapping function f, its embedding space $\mathcal{F}$ may have an infinite dimension. These two drawbacks forbid usual regularization techniques (except approaches that rely on the graph Laplacian spectrum [6]).

Contributions and outline.

For a finite set $\mathcal{X}$ of objects and a positive definite kernel K on $\mathcal{X}$, we propose to compute explicitely a regularized kernel from the solution of the minimization problem (2) corresponding to the p-TV-L_2 regularization on weighted graphs (see e.g. [7], [8], [9]), adapted here to feature maps (Section II). This is similar to the regularization of multivalued functions defined on $\mathcal{X}$. Then, we show that the

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.525

regularization equations derived from the solution of (2) allow to compute the regularized kernel without explicitly computing the regularized feature map (Section III). The validation of the proposed kernel regularization is provided in Section IV using graph kernels measuring the similarity between shape's skeleton.

II. Discrete regularization of feature maps

Given a finite set $\mathcal{X}$, recent works on discrete regularization ([7], [8], [9]) consider the elements of $\mathcal{X}$ as the vertices of a weighted graph, that can be fully represented by a weight matrix $W=(w_{xy})_{x,y\in\mathcal{X}}$. The symmetric positive weight function $w:\mathcal{X}\times\mathcal{X}\to\mathbb{R}_+$ measures the similarity between two elements of $\mathcal{X}$ such that $x,y\in\mathcal{X}$ are connected by an edge (x,y) in the graph if $w_{xy}>0$, and not connected if $w_{xy}=0$. Also, the graph is supposed to have no self-loops, that is $w_{xx}=0$ for all $x\in\mathcal{X}$.

Then, the minimization of $E(f,g,\alpha)$ (see (2)) is defined for functions $f,g:\mathcal{X}\to\mathbb{R}^q$ with a least square functional as an approximation term $A(g,f)$, and a total variation-like regularization term $R(g)$.

Rewritten in the context of feature maps $f,g:\mathcal{X}\to\mathcal{F}$, the approximation term $A(g,f)$ corresponds to the quadratic functional given by

$$A(g,f)=\frac{1}{2}\sum_{x\in\mathcal{X}}\|f(x)-g(x)\|^2, \tag{3}$$

where $\|\cdot\|=\sqrt{\langle\cdot,\cdot\rangle}$ corresponds to the inner product in the feature space $\mathcal{F}$. Similarly, the regularization term $R(g)$ is the (isotropic) p-total variation of g over the graph given by

$$R(g)=\frac{1}{2p}\sum_{x\in\mathcal{X}}|\nabla_x^w g|^p,\quad p\geq 1, \tag{4}$$

where $\nabla_x^w g$ represents the pointwise graph-gradient of g defined by the vector operator

$$\forall x\in\mathcal{X},\quad \nabla_x^w g=\left(\sqrt{w_{xy}}\left(\frac{g(y)}{\sqrt{d_y}}-\frac{g(x)}{\sqrt{d_x}}\right)\right)_{y\in\mathcal{X}},$$

where $d_x=\sum_{y\sim x}w_{xy}$ is the degree of x, and $x\sim y$ denotes the set of elements $y\in\mathcal{X}$ connected to x.

Its magnitude is usually given by its L_2 norm over $\mathcal{F}^{|\mathcal{X}|}$

$$\begin{aligned}|\nabla_x^w g| &= \left(\sum_{y\sim x}w_{xy}\left\|\frac{g(y)}{\sqrt{d_y}}-\frac{g(x)}{\sqrt{d_x}}\right\|^2\right)^{\frac{1}{2}},\quad \forall x\in\mathcal{X}\\ &= \left(\sum_{y\sim x}w_{xy}\left\langle\frac{g(y)}{\sqrt{d_y}}-\frac{g(x)}{\sqrt{d_x}},\frac{g(y)}{\sqrt{d_y}}-\frac{g(x)}{\sqrt{d_x}}\right\rangle\right)^{\frac{1}{2}}.\end{aligned} \tag{5}$$

When $p=2$, it can be shown that $R(g)=\langle\mathcal{L}g,g\rangle$, where $\mathcal{L}$ denotes the normalized graph Laplacian, which is usually considered as a kernel on $\mathcal{X}$ [5]. In this case, (2) corresponds to the Tikhonov regularization. When $p=1$, $R(g)$ is the (isotropic) total variation of g and (2) corresponds to the TV-$L2$ regularization.

As for $p\geq 1$, $E(f,g,\alpha)$ is a strickly convex functional, one can find a solution of the minimization of $E(f,g,\alpha)$ by computing its critical points in the direction of any element $u\in\mathcal{F}$, e.g. functions g satisfying

$$\langle\partial_{g(x)}E(f,g,\alpha),u\rangle=0,\ \forall x\in\mathcal{X}. \tag{6}$$

It is easy to show that for all $x\in\mathcal{X}$

$$\begin{aligned}\langle\partial_{g(x)}R(g),u\rangle &= \sum_{y\sim x}\gamma_{xy}(g)\left\langle\frac{g(x)}{d_x}-\frac{g(y)}{\sqrt{d_xd_y}},u\right\rangle\\ &= \langle(\Delta_p g)(x),u\rangle,\end{aligned} \tag{7}$$

where $\gamma_{xy}(g)=\frac{1}{2}w_{xy}\left(|\nabla_x^w g|^{p-2}+|\nabla_y^w g|^{p-2}\right)$, and $\Delta_p g$ is the gradient-based p-Laplacian of g over W. Then from (6) and (7) we obtain for all $x\in\mathcal{X}$ and $u\in\mathcal{F}$

$$\alpha\langle(\Delta_p g)(x),u\rangle+(1-\alpha)\langle g(x)-f(x),u\rangle=0. \tag{8}$$

By the positive-definiteness property of the kernel $\langle\cdot,\cdot\rangle$, if the function g is a minimizer of $E(f,g,\alpha)$ for any direction $u\in\mathcal{F}$, we have for all $x\in\mathcal{X}$

$$\alpha(\Delta_p g)(x)+(1-\alpha)(g(x)-f(x))=0. \tag{9}$$

This latter system of nonlinear equations, expressed here for mapping functions from $\mathcal{X}$ to $\mathcal{F}$, can be approximated by several methods such as the nonlinear Jacobi method or the steepest descent one (see for instance [7],[8] or [9], and Section III). In this paper, we consider the nonlinear Jacobi method given by the following iterative algorithm

$$\begin{cases} g^0=f\\ g^{t+1}(x)=L_{xx}^t f(x)+\sum_{y\sim x}L_{xy}^t g^t(y),\ \forall x\in\mathcal{X}\end{cases} \tag{10}$$

where g^t is the mapping function g at the time t, and

$$\begin{aligned}L_{xx}^t &= \frac{1-\alpha}{1-\alpha+\frac{\alpha}{d_x}\sum_{z\sim x}\gamma_{xz}^t}\\ L_{xy}^t &= \frac{\alpha\gamma_{xy}^t}{\sqrt{d_xd_y}\left(1-\alpha+\frac{\alpha}{d_x}\sum_{z\sim x}\gamma_{xz}^t\right)},\quad \forall x\neq y.\end{aligned} \tag{11}$$

When the initial mapping function f is unknown, or when $\mathcal{F}$ is infinite-dimensional, the regularized version g of f cannot be computed explicitely with (10), or with any other method solving (6). Nevertheless, (10) can be used to compute a regularized kernel K^g of K.

III. Kernel regularization

The minimizer g of problem (2) can be used to define a regularized kernel $K_{xy}^g=\langle g(x),g(y)\rangle,\ \forall(x,y)\in\mathcal{X}\times\mathcal{X}$, which can be seen as a regularized version of the initial kernel K defined by (1). By using the iterative algorithm (10) to compute g, at convergence the inner product $\langle g^{t+1}(x),g^{t+1}(y)\rangle$ tends to K_{xy}^g. By recursion on both

$g^{t+1}(x)$ and $g^{t+1}(y)$, we show that this allows to compute the regularized kernel without computing explicitely the mapping function g.

To do this, let $S^t_{xy} = \langle f(x), g^t(y)\rangle = \langle g^t(y), f(x)\rangle$ be the similarity between the initial function f and the regularized one at step t. By applying (10) in S^{t+1}_{xy}, we obtain for all $(x,y) \in \mathcal{X}\times\mathcal{X}$

$$\begin{aligned} S^{t+1}_{xy} &\overset{(10)}{=} L^t_{yy}\langle f(x), f(y)\rangle + \sum_{z\sim y} L^t_{yz}\langle f(x), g^t(z)\rangle, \\ &= L^t_{yy}K_{xy} + \sum_{z\sim y} L^t_{yz}S^t_{xz}. \end{aligned} \tag{12}$$

Similarly, we have for all $(x,y) \in \mathcal{X}\times\mathcal{X}$

$$\begin{aligned} K^{g^{t+1}}_{xy} &= \langle g^{t+1}(x), g^{t+1}(y)\rangle \\ &\overset{(10)}{=} L^t_{xx}\langle f(x), g^{t+1}(y)\rangle + \sum_{z\sim x} L^t_{xz}\langle g^t(z), g^{t+1}(y)\rangle \\ &\overset{(10)}{=} L^t_{xx}S^{t+1}_{xy} + L^t_{yy}\sum_{z\sim x} L^t_{xz}\langle g^t(z), f(y)\rangle \\ &\quad + \sum_{z\sim x} L^t_{xz}\sum_{v\sim y} L^t_{yv}\langle g^t(z), g^t(v)\rangle, \\ &= L^t_{xx}S^{t+1}_{xy} + L^t_{yy}\sum_{z\sim x} L^t_{xz}S^t_{yz} + \sum_{z\sim x} L^t_{xz}\sum_{v\sim y} L^t_{yv}K^{g^t}_{zv} \end{aligned} \tag{13}$$

which depends on the initial kernel K, the regularized kernel K^{g^t} at the step t, the similarities S^t and S^{t+1}, and L^t defined by (11). One can observe that L^t depends on the gradient magnitude (5) of g^t, which can be explicitly computed for all $x \in \mathcal{X}$ from K^{g^t} by

$$|\nabla^w_x g^t| = \left(\sum_{y\sim x} w_{xy}\left(\frac{K^{g^t}_{xx}}{d_x} + \frac{K^{g^t}_{yy}}{d_y} - \frac{2K^{g^t}_{yx}}{\sqrt{d_x d_y}}\right)\right)^{\frac{1}{2}}. \tag{14}$$

By recursion, this shows that the kernel $K^{g^{t+1}}$ only depends on the initial kernel K.

The proposed computation of the regularized kernel K^g is summarized by the following algorithm

$$\begin{cases} (a) \;\; K^{g^0} \leftarrow K,\; S^0 \leftarrow K,\; t \leftarrow 0 \\ (b) \;\; |\nabla_x g^t| \leftarrow (14), \quad \forall x \in \mathcal{X} \\ (c) \;\; L^t_{xy} \leftarrow (11), \quad \forall (x,y) \in \mathcal{X}\times\mathcal{X} \\ (d) \;\; S^{t+1}_{xy} \leftarrow (12), \quad \forall (x,y) \in \mathcal{X}\times\mathcal{X} \\ (e) \;\; K^{g^{t+1}}_{xy} \leftarrow (13),\; \forall (x,y) \in \mathcal{X}\times\mathcal{X} \\ (f) \;\; \text{if not converged, } t \leftarrow t+1 \text{ and goto } (b). \end{cases} \tag{15}$$

The relative error on the regularized kernel can be used as a stopping criterion in step (e)

$$\frac{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}}(K^{g^{t+1}}_{xy} - K^{g^t}_{xy})^2}{\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}}(K^{g^t}_{xy})^2} \leq \epsilon^2,$$

where $\epsilon = 1e^{-8}$ in our experiments provided in the following section.

Remark. We can derive an equivalent scheme considering the steepest descent method for solving (9), that is

$$\frac{dg^t(x)}{dt} = -\alpha(\Delta_p g)(x) - (1-\alpha)(g(x) - f(x)),\; \forall x \in \mathcal{X},$$

with $g^0 = f$ as initial value. By using the Euler method

$$\begin{aligned} g^{t+1}(x) =& g^t(x) - \tau\alpha(\Delta_p g^t)(x), \\ & - \tau(1-\alpha)(g^t(x) - f(x)),\; \forall x \in \mathcal{X} \end{aligned}$$

where $\tau > 0$ is the marching step size, the kernel associated to the regularized version of the mapping function f can be computed by the following equation

$$\begin{aligned} K^{g^{t+1}}_{xy} =& T^{t,t+1}_{xy} + \tau(1-\alpha)(S^{t+1}_{xy} - T^{t,t+1}_{xy}) \\ & - \tau\alpha\sum_{z\sim x}\gamma_{xz}\left(\frac{T^{t,t+1}_{xy}}{d_x} - \frac{T^{t,t+1}_{zy}}{\sqrt{d_x d_z}}\right), \end{aligned}$$

where $S^{t+1}_{xy} = \langle f(x), g^{t+1}(y)\rangle$ is given by

$$\begin{aligned} S^{t+1}_{xy} =& S^t_{xy} + \tau(1-\alpha)(K_{xy} - S^t_{xy}) \\ & - \tau\alpha\sum_{z\sim y}\gamma_{yz}\left(\frac{S^t_{xy}}{d_y} - \frac{S^t_{xz}}{\sqrt{d_y d_z}}\right), \end{aligned}$$

and $T^{t,t+1}_{xy} = \langle g^t(x), g^{t+1}(y)\rangle$ is given by

$$\begin{aligned} T^{t,t+1}_{xy} =& K^{g^t}_{xy} + \tau(1-\alpha)(S^t_{xy} - K^{g^t}_{xy}) \\ & - \tau\alpha\sum_{z\sim y}\gamma^t_{yz}\left(\frac{K^{g^t}_{xy}}{d_y} - \frac{K^{g^t}_{xz}}{\sqrt{d_y d_z}}\right). \end{aligned}$$

IV. Application to shape classification

The proposed kernel regularization framework is validated in the context of shape classification. Given a set $\mathcal{X}$ of 2D shapes, each shape can be described by its skeleton, represented by a graph which encodes the main shocks along the skeleton [10]. Each pair of shapes is then compared using the graph kernel proposed by [10]: each graph is associated to a bag of trails which covers it and contains the most important information about the shape. The bag of trails associated to the two graphs are then compared using a convolution kernel which weights each trail according to its relevance within its bag. The resulting kernel K is positive definite within the space of graphs.

To classify the shapes in $\mathcal{X}$, a training set $\mathcal{X}' \subset \mathcal{X}$ is selected and structured by a weighted graph $W = (w_{xy})_{x,y\in\mathcal{X}'}$. An intuitive weight function between two shapes x and y is given by the graph edit distance [11], which compares the graph representation of x and y. Unfortunatly, defining a positive definite kernel from this distance is not straightforward, for instance the Gaussian kernel based on this edit distance is not definite positive.

In this paper, we choose to correct the metric provided by [10] by regularizing the initial graph kernel K with (15) on the graph W with weights defined by

$$w_{xy} = exp\left(\frac{-d^2(G_x, G_y)}{2\sigma^2}\right), \tag{16}$$

Table I
CONFUSION MATRIX WITH REGULARIZATION

$$\begin{pmatrix} 10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 8 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 9 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 11 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 9 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 10 \end{pmatrix}$$

Table II
CONFUSION MATRIX WITHOUT REGULARIZATION

$$\begin{pmatrix} 10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 10 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 9 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 9 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 9 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 11 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 10 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 8 \end{pmatrix}$$

where G_x and G_y describe the graphs associated to the shapes $x, y \in \mathcal{X}'$, and d is the edit distance. The resulting kernel K^g is definite positive (Section 3) and is associated to a regularized mapping function according to the weight matrix W. Roughly speaking this kernel thus maps two graphs with a small edit distance to two close points in the feature space. This process can be seen as an alternative to [12], which defines a non positive definite kernel from edit distances.

This regularization scheme has been tested on 99 shapes of the well-known Kimia database, which is composed of 9 classes of 11 shapes. The training set was composed of 5 shapes of each class. The parameters of the kernel K [10] were optimized on this set using a 5-fold cross-validation combined with a grid-search.

The classification is performed from the regularized kernel using the quadratic discriminant analysis proposed by [13]. Table 1 and Table 2 show the confusion matrices obtained from the classification respectivley based the regularized and non-regularized Gram matrices defined by the kernel K on our trianing set. The regularization parameters used for these experiments are $p = 1$ and $\alpha = 0.8$. This experiments show improvements for the last class (8 to 10 shapes correctly classified), and the second one. This improvements are partially counter-balanced by small loss on classes 5 and 8, where one additional shape is missclassified in each case. The overall classification rate is improved from 0.868 to 0.878. This low improvement may be explained by the already good results obtained by [12], without the proposed regularization, and by the small size of the training set.

V. CONCLUSION

This paper describes a new regularization framework for mapping functions implicitely defined by positive definite kernels. This applies both on numerical and symbolic data such strings or graphs. The regularization steps allow to combine different metrics while preserving the postive definiteness of the initial kernel. Our current work [14] investigates other regularization approaches, such as the one based on unnormalized p-Laplacians. We also plan to provide a deep study of the different parameters involved in these regularization schemes.

REFERENCES

[1] H. Bunke, "On a relation between graph edit distance and maximum common subgraph," *Pattern Recogn. Lett.*, vol. 18, no. 9, pp. 689–694, 1997.

[2] A. Robles-Kelly and E. R. Hancock, "A riemannian approach to graph embedding," *Pattern Recognition*, vol. 40, no. 3, pp. 1042–1056, March 2007.

[3] J. P. Vert, K. Tsuda, and B. Schölkopf, "A primer on kernel methods," in *Kernel Methods in Computational Biology*. MIT Press, 2004, pp. 35–70.

[4] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*. Cambridge University Press, 2004.

[5] F. Steinke and B. Schölkopf, "Kernels, regularization and differential equations," *Pattern Recognition*, vol. 41, pp. 3271–3286, 2008.

[6] A. J. Smola and R. I. Kondor, "Kernels and regularization on graphs," in *COLT*, 2003, pp. 144–158.

[7] S. Osher and J. Shen, "Digitized PDE method for data restoration," in *Analytical-computational methods in applied mathematics*, E. G. A. Anastassiou, Ed. Chapman & Hall/CRC, 2000, pp. 751–771.

[8] D. Zhou and B. Schölkopf, "Regularization on discrete spaces," in *27th DAGM Symp.*, ser. LNCS, vol. 3663. Springer, 2005, pp. 361–368.

[9] A. Elmoataz, O. Lezoray, and S. Bougleux, "Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1047–1060, 2008.

[10] F. X. Dupé and L. Brun, "Tree covering within a graph kernel framework for shape classification." in *ICIAP*, 2009, pp. 278–287.

[11] K. Riesen, M. Neuhaus, and H. Bunke, "Bipartite graph matching for computing the edit distance of graphs," in *GbRPR*, 2007, pp. 1–12.

[12] M. Neuhaus and H. Bunke, "Edit distance based kernel functions for structural pattern classification," *Pattern Recognition*, vol. 39, no. 10, pp. 1852–1863, 2006.

[13] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, "Kernel quadratic discriminant analysis for small sample size problem," *Pattern Recognition*, vol. 41, no. 5, pp. 1528–1538, 2008.

[14] F.-X. Dupé, S. Bougleux, L. Brun, O. Lézoray, and A. Elmoataz, "Kernel-based implicit regularization of structured objects," GREYC CNRS UMR 6072, Tech. Rep., 2010.