USING TRANSFORMATION KNOWLEDGE FOR THE CLASSIFICATION OF RAMAN SPECTRA OF BIOLOGICAL SAMPLES

Klaus.-D. Peschke, Bernard Haasdonk,
Olaf Ronneberger and Hans Burkhardt
Lehrstuhl für Mustererkennung und Bildverarbeitung,
Institut für Informatik, Albert-Ludwigs-Universität Freiburg,
Georges-Köhler-Allee 52, D-79110 Freiburg, Germany
email: peschke@informatik.uni-freiburg.de

Petra Rösch, Michaela Harz and Jürgen Popp
Institut für Physikalische Chemie
Friedrich-Schiller-Universität Jena
Helmholtzweg 4, D-07743 Jena, Germany

ABSTRACT
For the classification of biological samples based on Raman spectra, a robust classifier is necessary. This requirement is met by using Support Vector Machines (SVMs) enhanced by incorporating a-priori knowledge about pattern variations. In the described approach *transformation knowledge* is included directly into the classification process by using regularized tangent distance kernels. This approach replaces the standard Euclidean distance in the kernel function by the distance of the linear approximation (tangent spaces) of known transformation manifolds. These transformations represent first a global scaling of the spectral values referring to intensity variations, and second a baseline shift by Lagrange polynomials. Experiments are carried out and reported in this paper. The results show, that incorporating a-priori knowledge by tangent distances improves the classification rates substantially, while a lossy baseline correction becomes superfluous.

KEY WORDS
biomedical computing, tangent distance, prior knowledge, SVM single cell classification, Raman spectra

1 Introduction

The increasing production times in bio-science as well as pharmaceutical industries raises the task of developing fast and robust methods for detection of pathogenic particles (e.g. bacteria). Most of the state of the art techniques are based on extracting features about the shape of the organisms, requiring different cultivation steps of the samples for the identification. This process needs several hours up to several days leading to a drop down of the production time. Therefore, applying a fast and reliable classification method is an important task. Within the research project *Online Monitoring and Identification of Bioaerosols (OMIB)* it was possible to provide a way to acquire Raman spectra of bacteria cells on a single cell level [1] [2] that are a characteristic finger print of the microorganisms. The classification of Raman spectra is usually carried out with methods like *k-nearest neighbour (k-nn) classification, principal component analysis (PCA)* or *soft independent modeling of class analogy (SIMCA)*. However, most of the methods require a correction of the baseline to gain a low classification error rate. In [1] it was shown that also support vector machines are a suitable tool for the classification of Raman spectra. While SVMs are already well known in pattern recognition for solving classification problems, they are not widespread in the area of (bio-)chemometrics and only few applications have been published yet (cf. [3] or [4]).

This paper successfully introduces an advanced method of enhancing SVMs by exploiting transformation knowledge to the field of classification of spectral data. By this approach known transformations like baseline shifts or scaling of intensity values are incorporated directly into the classification step. In that way, good recognition results can be achieved without carrying out a baseline correction as a preprocessing of the data. This is an enormous advantage compared to other classification techniques because the correction of the baseline is difficult and implies a loss of information. Moreover baseline corrections methods are usually based on heuristic approximations and until now only a few algorithms have been published, that achieve a reasonable baseline correction automatically.

2 Theory

The following section provides the theoretical background of the used methods. The main focus will be on prior transformation knowledge modeled by tangent distance kernels. Therefore the description of the support vector machines is deliberately limited to the parts that are relevant for understanding. A more detailed description of SVMs can be found in standard textbooks (e.g. [5] and [6]).

2.1 Classification based on SVM

SVMs are based on statistical learning theory that is described in [5] and [6]. Instead of modeling a distribution of different classes, SVMs try to find a separating boundary that is defined by the largest margin between two classes. This task can be reformulated into a quadratic optimization problem. The solution of this problem is a global optimum.

An important component of SVMs are kernel functions. Those kernels implicitly calculate a transformation into a high dimensional feature space and compute the corresponding inner product. Besides linear kernels, that calculate a scalar product of two vectors, there are also several non-linear kernels available [5], for instance the Gaussian radial basis function (rbf) kernel

$$k^{\mathrm{rbf}}(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \tag{1}$$

or the negative distance kernel

$$k^{\mathrm{nd}}(\mathbf{x}, \mathbf{x}') = -\|\mathbf{x} - \mathbf{x}'\|, \tag{2}$$

where the vector $\mathbf{x}$ is given as the Raman spectrum and the parameter γ used to adjust the local support of the rbf. The kernels presented in this work are based on those two kernels. In general positive definiteness of the kernel functions is a desirable property. However, SVMs can also work with conditionally positive definite functions [6] or more general with indefinite kernels [7]. In particular k^{rbf} is a positive definite kernel whereas k^{nd} is a conditionally positive definite kernel. The regularized tangent distance kernels presented are in general indefinite.

2.2 Prior knowledge in Raman spectroscopy

In pattern recognition it is widely known, that incorporation of problem-specific prior knowledge is crucial for improving the accuracy of a recognition system. A frequent type of such prior knowledge is the knowledge about typical pattern variations $t(\mathbf{x})$, which do not change the inherent meaning of the patterns $\mathbf{x}$. This is exemplified for the case of Raman spectra, where two types of variations are realistic and will be used in the sequel.

Baseline shift: The measurements can differ by some non-constant offset due to changing measuring environments, e.g. varying background fluorescence. A typical assumption on this baseline shift is that it can be approximated by low-frequency curve, e.g. by polynomials [8]. As a polynomial basis Lagrange polynomials of degree d are chosen with equidistant base points [9, p. 261] that are illustrated in Fig. 1:

$$P_l(x) := \prod_{i \neq l} \frac{(x - \frac{i}{d})}{(\frac{l}{d} - \frac{i}{d})}, \quad l = 0, \ldots, d.$$

An equidistant sampling of these polynomials in $x \in [0, 1]$ yields vectors $\mathbf{l}_l(\mathbf{x})$, which can model the baseline shift with arbitrary polynomials by linear combinations.

Intensity scaling: The global intensity of a Raman spectrum measurement is proportional to the duration of the accumulating measuring-interval or the area of the sample, which is irradiated by the laser beam. Hence it is reasonable to model this effect on the measured spectra by a scaling of the intensity values.

These two pattern variations can be modelled by the following equation with real valued parameter vector $\mathbf{p} = (p_0, \ldots, p_d, p_{d+1})^T$, where p_l for $l = 0, \ldots, d$ are the coefficients for the Lagrange polynomials $\mathbf{l}_l$ and p_{d+1} is the parameter for the scaling operation:

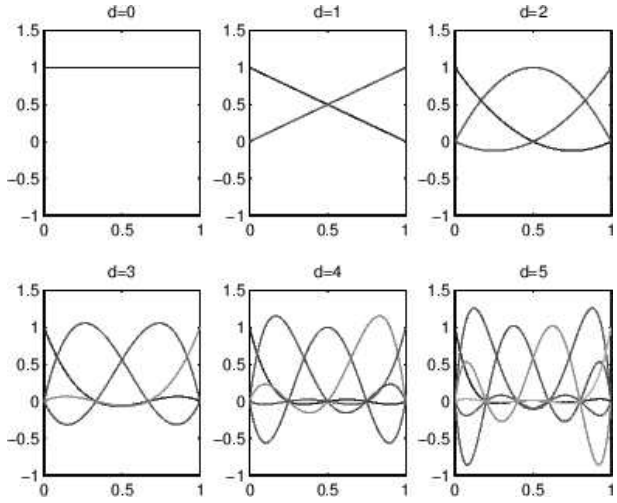

Figure 1. Polynomial Lagrange bases for increasing degree $d = 0, \ldots 5$. Note that the polynomials within each basis cover the unit interval symmetrically, have similar absolute values and curvature.

$$t_{\mathbf{p}}(\mathbf{x}) := \sum_{l=0}^{d} p_l \mathbf{l}_l(\mathbf{x}) + (1 + p_{d+1})\mathbf{x} \tag{3}$$

Note, that $\mathbf{p} = \mathbf{0}$ recovers the original sample $t_{\mathbf{0}}(\mathbf{x}) = \mathbf{x}$.

The effects of these transformations are depicted in Fig. 2, where the shifts of the baseline along degree 3 Lagrange polynomials and the intensity scaling are illustrated. The middle row of plots shows the original pattern, which is transformed by positive and negative parameters p_l columnwise. For the scaling in the rightmost column an extinction of the spectrum is obtained by shift with the $p_{d+1} = -1$ scaling, which must be prevented. The remaining columns of the figure demonstrate shifts along the Lagrange polynomials $\mathbf{l}_l(\mathbf{x})$, which models the baseline variation. Again, large absolute values of p_l for $l = 0, \ldots, d$ result in unrealistic patterns. So it will be crucial to limit the extent of those parameters p_l in the sequel.

The reason for the choice of these transformations is both the simple modelling and the (chemometric) relevance concerning the formulated problem. These two points are also a guideline for application of tangent distances in other fields. For instance in image recognition tasks, basic geometric transformations like rotations, translations or stretching are known to be meaningful and simple [10, 11, 12].

2.3 Tangent distance kernels

In the general case where transformations $t_{\mathbf{p}}(\mathbf{x})$ can be smoothly parameterized by a real valued vector $\mathbf{p}$, the set of

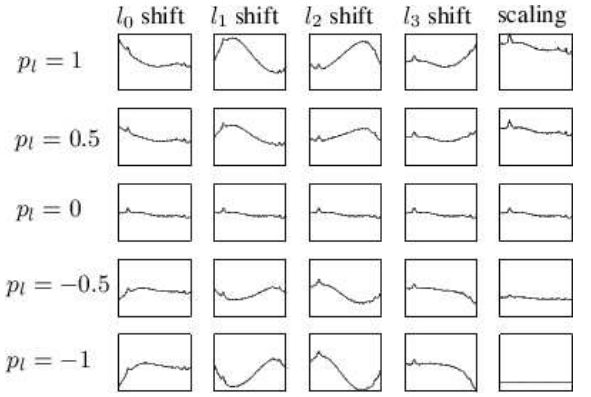

Figure 2. Raman spectrum transformations. The middle row shows the original sample $\mathbf{x}$ (intensity against wavenumber), above and below are increasing transformations $t_{\mathbf{p}}(\mathbf{x})$ with respect to the 4 Lagrange baseline shifts l_l of degree 3 ($l = 0, \ldots, 3$) and the scaling.

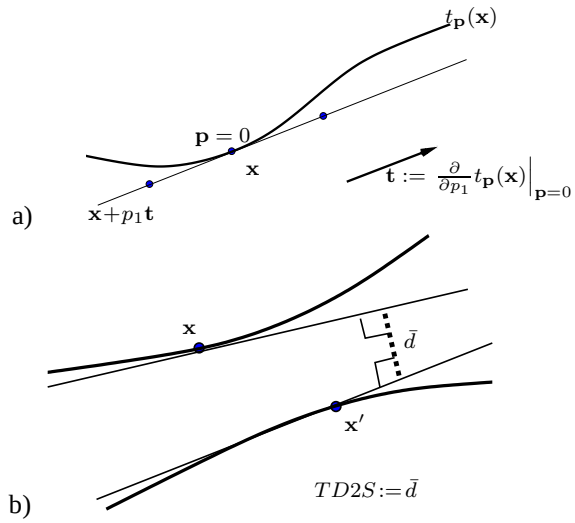

Figure 3. Illustration of tangent distance. a) Illustration of the tangent space approximating the true transformation manifold, b) unregularized tangent distance in 3D.

transformed patterns is a nonlinear manifold as illustrated in Fig. 3 a). For computationally dealing with these nonlinearities, a linear approximation of this manifold can be obtained by a series expansion

$$t_{\mathbf{p}}(\mathbf{x}) = t_{\mathbf{0}}(\mathbf{x}) + \sum_{l=0}^{d+1} p_l \cdot \frac{\partial}{\partial p_l} t_{\mathbf{p}}(\mathbf{x})\Big|_{\mathbf{p}=0} + \mathcal{O}(\|\mathbf{p}\|^2), \quad (4)$$

where $\frac{\partial}{\partial p_l} t_{\mathbf{p}}(\mathbf{x})\,|_{\mathbf{p}=0}$ are the tangents spanning the tangent space. In fact, for the transformations of the Raman spectra given in (3), the transformations already are linear spaces spanned by the tangents $\mathbf{x}$ and $\mathbf{l}_l(\mathbf{x})$, so no approximations are required here, the error-term $\mathcal{O}(\cdot)$ in (4) is zero.

Then, the main motivation for the so called *tangent distance* [11, 12] is that instead of the Euclidean distance, the distance of the tangent spaces is a more reasonable measure of dissimilarity. For instance in Fig. 3 b) the patterns $\mathbf{x}$ and $\mathbf{x}'$ are quite far away, however, their transformation manifolds' tangent spaces reveal that certain transformations of both patterns lie much closer (Their distance indicated as $\bar{d}$). This can be formalized in a very compact way by collecting the $d+2$ tangents ($d+1$ for baseline shift and 1 for scaling) of a point $\mathbf{x}$ as columns of the matrix $\mathbf{T}$, the tangents of a point $\mathbf{x}'$ into the columns of the matrix $\mathbf{T}'$. The *regularized two-sided tangent distance* which is involved in the experiments can be defined as

$$TD2S(\mathbf{x}, \mathbf{x}')^2 := \quad (5)$$
$$\min_{\mathbf{p},\mathbf{p}'} \left(\|\mathbf{x} + \mathbf{T}\mathbf{p} - \mathbf{x}' - \mathbf{T}'\mathbf{p}'\|^2 + \lambda \left(\|\mathbf{p}\|^2 + \|\mathbf{p}'\|^2 \right) \right).$$

The first term performs the minimization of the distance between the tangent spaces, while the second regularization term prevents, that the transformation parameters p_l become too large. The weight between the two terms can be adjusted by the parameter λ. In particular $\lambda = \infty$ recovers the standard Euclidean distance.

The solution of the optimization problem reduces to an unconstrained convex quadratic optimization problem. These problems can be explicitly solved by a linear equation system after setting the gradient of $TD2S^2$ to zero. The solution is obtained by

$$\begin{pmatrix} \mathbf{p} \\ -\mathbf{p}' \end{pmatrix} := \mathbf{S}\,(\mathbf{x}' - \mathbf{x}) \quad \text{with} \quad (6)$$
$$\mathbf{S} := \left(\begin{pmatrix} \mathbf{T}^T \\ \mathbf{T}'^T \end{pmatrix} (\mathbf{T}, \mathbf{T}') + \lambda \mathbf{I} \right)^{-1} \begin{pmatrix} \mathbf{T}^T \\ \mathbf{T}'^T \end{pmatrix},$$

where $\mathbf{I}$ denotes the identity matrix. The *regularized tangent distance kernels (TD-kernels)* are then obtained by replacing the Euclidean distance by $TD2S$, which leads to the kernels such as

$$k^{\text{rbf}}_{TD2S}(\mathbf{x}, \mathbf{x}') := e^{-\gamma TD2S(\mathbf{x},\mathbf{x}')^2} \quad \text{and}$$
$$k^{\text{nd}}_{TD2S}(\mathbf{x}, \mathbf{x}') := -TD2S(\mathbf{x}, \mathbf{x}').$$

More distance-based kernels but also inner-product-based kernels such as the polynomial kernel can be defined correspondingly, cf. [7]. These kernels are extensions of the TD-kernels presented in [10] by enabling regularization. This is necessary in the present application, as the unregularized kernels would be trivially constant due to the intersecting transformation spaces.

3 Experimental

The experiments are carried out with a micro-Raman setup of a standard Raman spectroscope (HR LabRam invers, Jobin-Yvon-Horiba). The excitation wavelength of the used laser (Nd:YAG, frequency doubled) is centered at 532 nm. On the whole, there are 2545 spectra for 20 different strains available (cf. Tab. 1). However, as it can be seen in the table, the population sizes differ and range from 805

samples for the largest class (*S. epidermidis*) to 20 samples for the smallest class (*E. coli*). This difference is due to the chosen measurement setup and does not represent the a-priori probability of the strains. The microorganisms were cultivated on a standard or nutrition agar (Micrococcus, Escherichia and Bacillus) as well as on CA and CASO agar (Staphylococcus) for different growing ages. In order to simulate samples from clean rooms the Raman measurements were directly performed on single cells from smears on fused silica plates. The spectra are preprocessed in two steps. First, a running median filter is used to filter out spikes, that are caused by cosmic radiation. Second, the grid points are re-sampled by a bilinear interpolation. This was necessary, due to the fact that different measuring setups lead to slightly different sample points on the grid axis.

As a reference method a k-nearest neighbour classification method is used. This method is a robust classifier, however, is rather costly to calculate. Instead of using training and test set for validation, a leave-one-out (LOO) test is chosen. In that setup the classifier is trained on all available data except one sample. This sample is then identified by the trained classifier. Afterwards, the training and classification is repeated until all of the available data samples have been left out and classified exactly once. By using a LOO test in each step almost all data can be used for training. This is an essential advantage, when the amount of training data is limited for some classes. The results are presented by the classification error (LOO-error) as well as the average classification error (av-LOO-error). The latter is the average over the error rate of each strain class and is chosen because of the unrepresentative class prior probabilities.

The computation of the tangent distances is carried out by programs that are based on the scripting language MatLab 6.5 (The MathWorks, Inc.). The SVM classification uses the library *LIBSVMTL* [13], that is built on a C++ template architecture and specially suited for the classification of large data sets.

4 Classification results and discussion

As mentioned above the k-nn classifier is a common classifier for spectral data. So in Fig. 4 the classification error for the described dataset is presented where the number of neighbors k is increased. The continuous line gives the av-LOO-error while the dashed line describes the LOO-error. For the 1-nearest neighbour case the av-LOO-error is 42.68% and the LOO-error is 21.06%. Those are also the lowest error rates that have been achieved, as the error curves clearly increase for the cases considering more than one neighbours. This tendency reveals a glance on how the data is distributed in the sample space. Since the error increases when more neighbours are involved, the data seems to be sparsely distributed.

For comparison, the error rates for the k-nn classifier based on tangent distances are given in Fig. 5 (the parameters for the calculation of the distances were those that gave the best result in the SVM classification, described below). The lowest error is 19.44% for the av-LOO-error (continuous line) and 9.00% for LOO-error (dashed line). As in the k-nn classifier results before, the lowest error rate occurs for the 1-nearest neighbour case, confirming the assumption about the sparse distribution of the data. Obviously, the results improve when considering pattern variations by tangent distances, as can be seen in Fig. 5 and Fig. 4. This shows that baseline variation and scaling is crucial for classification.

Table 1. Different bacteria strains that are included in the dataset (the DSM number refers to an identification number of the Deutsche Sammlung von Mikroorganismen und Zellkulturen, the ATCC number refers to the American Type Culture Collection)

name	number of samples
B. pumilus DSM 27	57
B. pumilus DSM 361	43
B. sphaericus DSM 28	53
B. sphaericus DSM 369	42
B. subtilis DSM 10	306
B. subtilis DSM 347	42
E. coli DSM 423	51
E. coli DSM 498	21
E. coli DSM 499	20
M. luteus DSM 348	619
M. luteus DSM 20030	48
M. lylae DSM 20315	20
M. lylae DSM 20318	20
S. cohnii DSM 6669	67
S. cohnii DSM 20260	65
S. cohnii DSM 6718	65
S. cohnii DSM 6719	63
S. epidermidis ATCC 35984	805
S. warneri DSM 20036	67
S. warneri DSM 20316	71

Scaling tangent: The error rates can be reduced, when using SVMs with tangent distance kernels. Even TD-kernels, which apply only the intensity scaling tangent, already improve the recognition results as listed in Table 2. The base results with the standard kernels are given in the first row. The results with the regularized $TD2S$-kernels are given in the remainder of the table, where the regularization parameter λ is chosen exponentially decreasing. It can be seen, that in the case $\lambda = \infty$, which corresponds to the usual Euclidean distance, the rbf-kernel and the negative distance perform identically with respect to the LOO-error. For the av-LOO-error, the negative distance kernel is slightly better. With decreasing regularization parameter the error rates are first higher than the base approach, but then reach a minimum around $\lambda = 0.000010$. For $\lambda = 0$ the distance will be constant 0, as all tangent spaces intersect. So, the distances are

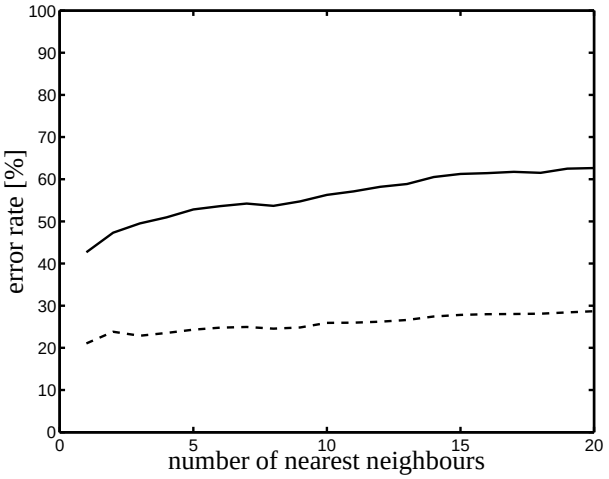

Figure 4. Error curves for the av-LOO-error (continuous line) and LOO-error (dashed line) for the k-nn classifier.

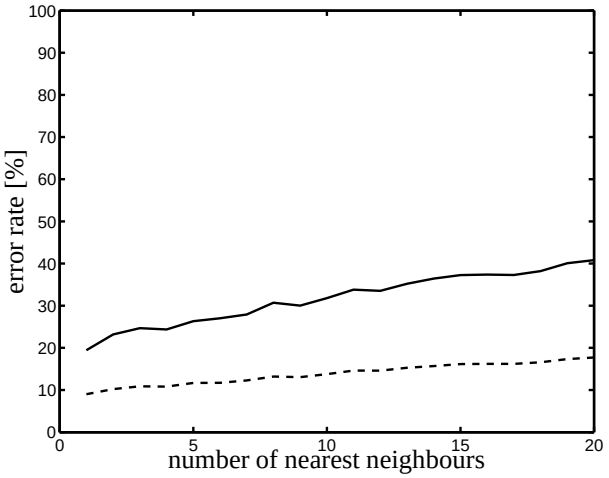

Figure 5. Error curves for the av-LOO-error (continuous line) and LOO-error (dashed line) for the k-nn classifier, while considering a baseline modeling and intensity scaling by tangent distances

non-informative, which is reflected by the increasing error rates for $\lambda \leq 0.000010$ at the end of the table and beyond. The optimal rates for each column are highlighted by bold numbers, which demonstrates that for the rbf-kernel, both the LOO- and the av-LOO-errors indeed decrease by 0.19% and 0.82%, respectively, compared to the standard SVMs. The negative distance kernel, however, does only improve the errors marginally.

Combined scaling and Lagrange tangents: Due to the previous experiments, the Gaussian rbf kernel turned out to perform best, so the focus is set on this kernel. The additional modeling of the baseline shift by the Lagrange tangents leads to additionally improved recognition rates. The results of the LOO- and av-LOO-error are given in Tab. 3 for increasing the polynomial degree of the polynomials from 0 to 4 and again decreasing the regularization parameter. The results indicate, that for $\lambda = 4 \cdot 10^{-6}$ both error-measures remarkably improve compared to Tab. 2 to an LOO-error of 2.91% and an av-LOO-error of 6.58%. The role of the polynomial degree, however, is not very explicit. The low polynomial degrees 1-2 seem to be favourable. Overall, these recognition errors are a considerable improvement from the base results in Tab. 2.

Table 2. LOO-errors [%] on the Raman spectra when using TD2S with scaling tangent.

	k^{rbf}_{TD2S}		k^{nd}_{TD2S}	
reg-par λ	LOO	av-LOO	LOO	av-LOO
∞ (=ED)	4.20	9.58	4.20	9.09
0.000100	4.40	9.99	4.48	10.42
0.000080	4.36	10.08	4.36	10.17
0.000040	4.17	9.41	4.32	9.63
0.000020	**4.01**	**8.76**	4.36	9.57
0.000010	**4.01**	8.99	**4.17**	**9.02**
0.000008	4.13	9.13	**4.17**	9.21
0.000004	4.20	9.36	4.24	9.44
0.000002	4.28	9.57	4.36	9.79

Table 3. LOO-errors [%] on the Raman spectra when using k^{rbf}_{TD2S} with combinedscaling and Lagrange tangents of certain degree l.

	LOO-error [%], varying degree l				
reg-par λ	0	1	2	3	4
0.000100	**3.38**	3.22	3.10	3.54	4.01
0.000080	**3.38**	3.30	3.06	3.61	4.01
0.000040	3.46	3.22	**2.99**	3.46	3.61
0.000020	3.50	3.14	3.14	3.34	3.30
0.000010	3.65	3.10	3.22	3.42	3.30
0.000008	3.61	3.06	3.14	3.30	3.14
0.000004	3.54	**2.91**	3.06	3.34	**3.14**
0.000002	3.77	2.99	3.14	**3.18**	**3.14**

	av-LOO-error [%], varying degree l				
reg-par λ	0	1	2	3	4
0.000100	**7.73**	7.31	7.18	8.01	9.19
0.000080	7.88	7.39	7.15	8.27	9.22
0.000040	8.06	7.27	**6.96**	7.99	8.33
0.000020	8.35	6.93	7.16	7.99	7.54
0.000010	8.59	6.87	7.56	7.77	7.66
0.000008	8.61	6.71	7.35	7.54	7.37
0.000004	8.41	**6.58**	7.16	7.63	**7.31**
0.000002	8.91	6.79	7.34	**7.51**	7.39

5 Conclusions

The described experiments represent a new application of tangent distances besides the known fields of optical character recognition [11, 12] and speech recognition [14]. Compared to the previous applications of tangent distance kernels [10] a regularization term was introduced to prevent trivial solutions of intersecting transformation spaces. Hence, the regularized tangent distance kernels can be used to model very meaningful variations of Raman spectra, i.e. scaling transformations and baseline shifts. This paves the way to a new field of classification techniques of spectral data without the need of an explicit baseline correction.

In detail the following two research areas are envisaged. First, up to now the choice of the transformations is still rather general. More meaningful transformations seem possible, when applying further knowledge about the patterns. Effects like photo-bleaching of objects for instance lead to intensity reduction in certain sections of the spectra instead of scaling the complete patterns. This could additionally be modeled by transformations. Further, the baselines are not only random polynomials, so other parametric models than polynomials, e.g. Fourier-bases, might also be promising to represent these variations. Second, another research direction is the application of transformation knowledge via the regularized TD-distances for further pattern analysis with kernel methods. Future fields of application can be the detection of outliers by means of SVMs, kernel based principal component analysis or kernel based linear discriminant analysis.

6 Acknowledgement

The research project FKZ 13N8365 and FKZ 13N8369 was funded within the framework 'Biophotonik' from the Federal Ministry of Education and Research, Germany (BMBF). All financial support is gratefully acknowledged.

References

[1] P. Rösch, M. Harz, M. Schmitt, K.-D. Peschke, O. Ronneberger, H. Burkhardt, H.-W. Motzkus, M. Lankers, S. Hofer, H. Thiele, & J. Popp, Chemotaxonomic identification of single bacteria by micro-Raman spectroscopy: Application to clean-room-relevant biological contaminations, *Applied and Environmental Microbiology*, *71*(3), 2005, 1626–1637.

[2] M. Harz, P. Rösch, K.-D. Peschke, O. Ronneberger, H. Burkhardt, & J. Popp, Micro-raman spectroscopical identification of bacterial cells of the genus staphylococcus in dependence on their cultivation conditions, *The Analyst*, *130*(11), 2005, 1543–1550.

[3] A. I. Belousov, S. A. Verzakov, & J. von Frese, A flexible classification approach with optimal generalization performance: Support vector machines, *Chemometrics and Intelligent Laboratory Systems*, *64*, 2002, 15–25.

[4] A. Demiriz, K. P. Bennett, C. M. Breneman, & M. J. Embrechts, Support vector machine regression in chemometrics, *Computing Science and Statistics: Proceedings of Interface*, 2001.

[5] V. Vapnik, *The nature of statistical learning theory* (New York: Springer, 1995).

[6] B. Schölkopf, & A. J. Smola, *Learning with kernels: Support vector machines, regularization, optimization and beyond* (Cambridge: The MIT Press, 2002).

[7] B. Haasdonk, *Transformation Knowledge in Pattern Analysis with Kernel Methods - Distance and Integration Kernels*. PhD thesis, Computer Science Department, University of Freiburg, Germany, 2005.

[8] K. R. Beebe, R. J. Pell, & M. B. Seasholtz, *Chemometrics: A practical guide*, (New York: John Wiley & Sons, 1998).

[9] H. Jeffreys, & B. S. Jeffreys, *Methods of mathematical physics* (Cambridge: Cambridge University Press, 1950).

[10] B. Haasdonk, & D. Keysers, Tangent distance kernels for support vector machines, *Proceedings of the 16th International Conference on Pattern Recognition*, 2002, 864–868.

[11] P. Y. Simard, Y. A. LeCun, & J. S. Denker, Efficient pattern recognition using a new transformation distance, *Advances in Neural Information Processing Systems 5*, Morgan Kaufmann, 1993, 50–58.

[12] P. Y. Simard, Y. A. LeCun, J. S. Denker, & B. Victorri, Transformation invariance in pattern recognition — tangent distance and tangent propagation, *Neural Networks: Tricks of the Trade*, Springer, 1998, 239–274.

[13] O. Ronneberger, & F. Pigorsch, LIBSVMTL: a support vector machine template library, 2004. Software available at http://lmb.informatik.uni-freiburg.de/lmbsoft/libsvmtl/.

[14] D. Keysers, W. Macherey, J. Dahmen, & H. Ney. Learning of variability for invariant statistical pattern recognition, *Proceedings of the 12th European Conference on Machine Learning*, Springer, 2001, 263–275.