

Wavelet Frame Accelerated Reduced Support Vector Machines

Matthias Rätsch, Gerd Teschke, Sami Romdhani, and Thomas Vetter *Member, IEEE*

Abstract—In this paper a novel method for reducing the runtime complexity of a Support Vector Machine classifier is presented. The new training algorithm is fast and simple. This is achieved by an Over-Complete Wavelet Transform that finds the optimal approximation of the Support Vectors. The presented derivation shows that the wavelet theory provides an upper bound on the distance between the decision function of the Support Vector Machine and our classifier. The obtained classifier is fast, since a Haar wavelet approximation of the Support Vectors is used, enabling efficient Integral Image based kernel evaluations. This provides a set of cascaded classifiers of increasing complexity for an early rejection of vectors easy to discriminate. This excellent runtime performance is achieved by using a hierarchical evaluation over the number of incorporated and additional over the approximation accuracy of the Reduced Set Vectors. Here this algorithm is applied to the problem of face detection, but it can also be used for other image based classifications. The algorithm presented, provides a 530 fold speed-up over the Support Vector Machine, enabling face detection at more than 25 fps on a standard PC.

Index Terms—Over Complete Wavelet Transform, Reduced Support Vector Machine, Coarse to Fine Classifier, Cascaded Evaluation, Face Detection, Machine Learning.

EDICS Category: SRE-MLRN, MRP-WAVL,
OTH-RCGN

I. INTRODUCTION

IMAGE based classification tasks are time consuming. For instance, detecting a specific object in an image, such as a face, is computationally expensive, as all the pixels of the image are potential object centers. Hence all the pixels must be classified.

Recently, more efficient methods have emerged based on a cascaded evaluation of hierarchical filters: image patches easy to discriminate are classified by a simple and fast filter, while patches that resemble the object of interest are classified by more involved and slower filters. In the area of face detection [13], cascaded based classification algorithms were introduced by Keren *et al.* [7], by Romdhani *et al.* [12] and by Viola and Jones [19]. With Kerens *et al.* [7] detector the negative examples (i.e. the non-faces) need to be modeled by a Boltzmann distribution and must be smooth. This assumption could increase the number of false positive in presence of a

M. Rätsch*, S. Romdhani and T. Vetter are with the University of Basel, Computer Science Department, Bernoullistrasse 16, CH-4057 Basel, Switzerland, Phone: + 41 61 267 0554, Fax: + 41 61 267 0559, E-mail: {matthias.raetsch,sami.romdhani,thomas.vetter}@unibas.ch

G. Teschke is with the Konrad Zuse Institute Berlin, Takustr. 7, D-14195 Berlin, Germany, Phone: + 49 30 84185 290, Fax: + 49 30 84185 107 E-mail: teschke@zib.de

cluttered background. Romdhani *et al.* [12] use a Cascaded Reduced Set Vectors (RSV) expansion of a Support Vector Machine [18]. The speed bottleneck of [12] is that at least one convolution of a 20×20 filter has to be carried out on the full image, resulting in a computationally expensive evaluation of the kernel with an image patch. Kienzle *et al.* [8] present an improvement of this method, where the first (and only the first) RSV is approximated by a separable filter. Viola & Jones [19] use Haar-like oriented edge filters having a block like structure enabling a very fast evaluation by use of an Integral Image. These filters are weak, in the sense that their discrimination power is low. They are selected, among a finite set, by the AdaBoost algorithm that yields the ones with the best discrimination. A drawback of their approach is that it is not clear that the cascade achieves optimal generalization performances. Practically, the training proceeds by trial and error, and often, the number of filters per stage must be manually selected so that the false positive rate decreases smoothly. Another drawback of the method is that the set of available filters is limited and has to be selected manually. The training for the classifier is "on the order of weeks" ([19], Section 5.2), as every filter (about 10^5) is evaluated on the whole set of training examples and this is done every time a filter is added to a stage of the cascade.

Taking the above mentioned problems into account, we developed a novel classification algorithm. The following features make the algorithm accurate and efficient:

- 1) **Support Vector Machine:** Use of an SVM classifier that is known to have optimal generalization capabilities.
- 2) **Reduced Support Vector Machine:** The RVM uses a reduced set of Support Vectors [12].
- 3) **Double Cascade:** For non-symmetric data (i.e. only few positives to many negatives) we achieve an early rejection of easy to discriminate vectors. It is obtained by the two following cascaded evaluations over coarse-to-fine Wavelet Approximated Reduced Set Vectors (W-RSV's): (i) **Cascade over the number of used W-RSV's** and (ii) **Cascade over the resolution levels of each W-RSV**. The Double Cascade constitutes one of the major novelties of our approach. The trade-off between accuracy and speed is very continuous.
- 4) **Integral Images:** As the RSV's are approximated by a Haar wavelet transform, the Integral Image method is used for their evaluation, similarly to [19].
- 5) **Wavelet Frame:** We use an over-complete wavelet system to find the best representation of the RSV's.

The learning stage of our proposed Wavelet Approximated

Reduced SVM (**W-RVM**) is fast, straightforward, automatic and does not require the manual selection of ad-hoc parameters. For example, the training time (Section III) is two hours which is a vast improvement over former detectors.

In our approach, we apply an Over-Complete Wavelet Transform (OCWT) to the Reduced Support Vector Machine itself, and not of the input space as a pre-processing like [6], [5].

This paper presents the coherent and complete frame work of our approach where we summarize and extend the conference papers [10], [11], [12]. The improvement of [10] compared to [11] are the features 3. (ii) and 5. (see above): The Simulated Annealing optimization using morphological filters was replaced by a sparse wavelet frame representation of the RSV's. Simulated Annealing does not provide the global optimum of the RVM approximation in all cases and it is difficult to adjust the resolution level.

In this paper we take advantage of recent progress in wavelet analysis: the optimality of sparse signal approximation (rectangular structure) in wavelet space. Moreover, we show the double cascade structure of the learning and detection process that is obtained by the proposed recursive refinement of the wavelet frame representation of the RSV's.

In addition we show in Section II-B.3 that the wavelet frame approach provides an upper bound of the hyperplane approximation error. Exploring this characteristic the training of the W-RVM works without heuristics and is fast. Also as an expansion, we show in Section II-B.3 the relation between the hyperplane approximation error of the decision functions and a training parameter to control the trade-off between sparsity and approximation. As demonstrated in Section II-C.1 the parameter for setting the approximation accuracy does not play a decisive role, opposite to former methods, using only one resolution level.

The paper is organized as follows: Section II details our novel training (Section II-B) and detection algorithm (Section II-C). It is shown in Section III that the new expansion yields a comparable accuracy to the SVM while providing a significant speed-up. In addition to the mentioned papers [10], [11] we carried out experiments on well known databases, like FERET [9] to provide the comparability to other approaches.

II. WAVELET FRAME APPROXIMATED SUPPORT VECTOR MACHINE

Support Vector Machines (SVM) are well-known for good generalization capabilities. Their decision function has the form: $y(\mathbf{x}) = \sum_{i=1}^{N_s} \alpha_i \cdot k(\mathbf{x}, \mathbf{x}_i) + b$, where $k(\cdot, \cdot)$ represents the kernel determining the feature space. In order to improve the runtime performance, it is proposed in [16] to approximate the SVM by a Reduced SVM (RVM) in combination with a cascaded evaluation as in [12]. The RVM aims to approximate the SVM by a *smaller* set of new Reduced Set Vectors (RSV's), \mathbf{z}_i instead of the Support Vectors, \mathbf{x}_i . The RVM approach provides a significant speedup over the SVM, but is still not fast enough, as the image has to be convolved in steps of full convolutions, e.g. by 20×20 RSV's. The algorithm

presented in this paper improves this method since it does not require to perform this convolution explicitly. Instead, it approximates the RSV's by Haar-like vectors and computes the evaluation of a patch using an Integral Image of the input image. They can be used to compute very efficiently the dot (or inner) product of an image patch with an image that has a block-like structure, i.e. rectangles of constant values.

A. Integral Images Based on Haar-like W-RSV's

During an RVM evaluation, most of the time is spent in kernel evaluations. In the case of the Gaussian kernel, $k(\mathbf{x}, \mathbf{z}_i) = \exp(-\|\mathbf{x} - \mathbf{z}_i\|^2 / (2\sigma^2))$, chosen here, the computational load is spent in evaluating the norm of the difference between a patch and a RSV. This norm can be expanded as follows: $\|\mathbf{x} - \mathbf{z}_i\|^2 = \mathbf{x}'\mathbf{x} - 2\mathbf{x}'\mathbf{z}_i + \mathbf{z}_i'\mathbf{z}_i$. As \mathbf{z}_i is independent of the input image, it can be pre-computed. The sum of squares of the pixels of a patch of the input image, $\mathbf{x}'\mathbf{x}$ is efficiently computed using the Integral Image ([3], [19]) of the squared pixel values of the input image. As a result, the computational load of this expression is determined by the term $2\mathbf{x}'\mathbf{z}_i$.

The novelty of our approach is the approximation of the RSV's, \mathbf{z}_i , by optimally wavelet frame approximated Reduced Set Vectors (W-RSV's), \mathbf{u}_i which have a block-like structure, as seen in Figure 1. Optimally approximated means here the usage of an optimally shifted wavelet basis that represents the image as sparse as possible. Then the term $2\mathbf{x}'\mathbf{u}_i$ can be evaluated very efficiently using the Integral Image. If \mathbf{u}_i is an image patch with rectangles of constant (and optionally different) gray levels then the dot product is evaluated in constant time by the addition of four pixels of the Integral Image of the input image per rectangle and one multiplication per gray level value.

B. Learning Process

In contrast to several approaches like [6], [5], we do not wavelet-transform the input images as a pre-processing at runtime. The novelty is that we apply the OCWT to the Reduced Support Vector Machine itself.

1) *Soft-Shrinkage to Build Rectangular Structured W-RSV's*: We are searching for an approximation of a given image \mathbf{z} by a piecewise block structured image \mathbf{u} which is as sparse as possible. This optimization problem can be casted in the following variational form

$$\min_{\hat{\mathbf{u}}} \left\{ \|\mathbf{z} - \hat{\mathbf{u}}\|_{L_2}^2 + 2\mu|\hat{\mathbf{u}}|_{B_1^1(L_1)} \right\}, \quad (1)$$

where $B_1^1(L_1)$ denotes a particular Besov semi-norm (for more details we refer the reader to [17], [14] and for a detailed discussion of the problem to [1]). It is known that the Besov (semi) norm of a given function can be expressed by means of its wavelet coefficients. In two spatial dimensions the Besov penalty is nothing else than a ℓ_1 constraint on the wavelet coefficients (promoting sparsity as required).

The minimization of (1) is easily obtained: Let $\{\psi_\lambda\}_{\lambda \in \Lambda}$ be the underlying wavelet basis, where Λ is the index set over all possible grid point, scaling and wavelet species. Then we

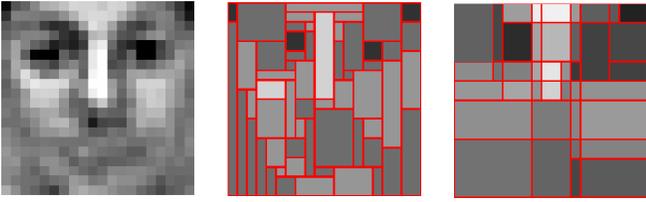


Fig. 1. Examples for Haar-like approximations: RSV (left) approximated using morphological filter (H-RSV [11], middle) and using an OCWT (W-RSV, right). The OCWT representation meets optimally the local image structure. The ratio of the decreasing of the hyper-plane distance to the used operations (see Section II-B.5) is more efficient for the W-RSV (0.73) than for the H-RSV (0.51).

may express \mathbf{z} and $\hat{\mathbf{u}}$ as follows: $\mathbf{z} = \sum_{\lambda \in \Lambda} z_\lambda \psi_\lambda$, $\hat{\mathbf{u}} = \sum_{\lambda \in \Lambda} \hat{u}_\lambda \psi_\lambda$, where $z_\lambda = \langle \mathbf{z}, \psi_\lambda \rangle$ and $\hat{u}_\lambda = \langle \hat{\mathbf{u}}, \psi_\lambda \rangle$. Thus we may completely rewrite (1) as

$$\mathbf{u} = \arg \min_{\mathbf{u}} \sum_{\lambda \in \Lambda} \{(z_\lambda - \hat{u}_\lambda)^2 + 2\mu|\hat{u}_\lambda|\}. \quad (2)$$

Minimizing summand-wise, we obtain the following explicit expression for the optimum u_λ , see, e.g. [4],

$$u_\lambda = S_\mu(z_\lambda) = \text{sgn}(z_\lambda) \max\{|z_\lambda| - \mu, 0\}, \quad (3)$$

where S_μ is the soft-shrinkage operation with threshold μ . Consequently, the optimum \mathbf{u} is simply obtained by soft-shrinking the wavelet coefficients of \mathbf{z} , i.e.

$$\mathbf{u} = \sum_{\lambda \in \Lambda} S_\mu(z_\lambda) \psi_\lambda = W^{-1} S_\mu(W\mathbf{z}), \quad (4)$$

where W stands for the wavelet transform operator.

2) *Optimal Match by Translated Wavelet Bases*: Typically, orthogonal or so-called non-redundant representations and filtering very often creates artifacts in terms of undesirable oscillations or non-optimally represented details, which manifest themselves as ringing and edge blurring. For our purpose it is essential to pick a representation that optimally meets the local image structure (see Figure 1). The most promising method for adequately solving this kind of problem has its origin in translation invariance (the method of cycle spinning, see, e.g. [2]), i.e. representing the image by all possible shifted versions of the underlying (Haar) wavelet basis. But contrary to the idea of introducing redundancy by averaging over all possible representations of \mathbf{z} , we aim to pick only that one that optimally meets the given image structure.

In order to give a rough sketch of this technique, assume that we are given an RSV \mathbf{z} with $2^M \times 2^M$ pixel. Following the cycle-spinning approach, see again [2], we have to compute $2^{2(M+1-j_0)}$ different representations of \mathbf{z} with respect to the $2^{2(M+1-j_0)}$ translates, s of the underlying wavelet basis. The scale j_0 denotes the coarsest resolution level of \mathbf{z} . The family $\{\mathbf{z}^s\}_s$ generated this way serves now as our reservoir of possible wavelet representations of one single \mathbf{z} . The best shift s^* is that one for which we have a minimal discrepancy to the SVM hyper-plane per operations for the kernel-evaluation. We evaluate all possible local shifts (in our case $s = 64$), hence the global optimum shift is guaranteed (see Section II-B.5).

3) *Hyper-plane Approximation*: The first reduction step was computing the reduced SVM by the means of [11] and [12]. This yields $\Psi_{\text{RVM}} = \sum_{i=1}^{N_z} \beta_i \Phi(\mathbf{z}_i)$. As outlined above, an essential improvement can be achieved by accelerating the numerical integration. To this end, we have suggested the use of Haar-like sparse approximations \mathbf{u}_i of \mathbf{z}_i that generates rectangular representations of the images and fits thus well with the concept of Integral Images. Replacing \mathbf{z}_i by \mathbf{u}_i amounts to $\sum_{i=1}^{N_z} \beta_i \Phi(\mathbf{u}_i)$. The change of the supporting vectors might likely require a slight adjustment of the β_i 's which is done iteratively (see below), i.e. the approximation we are proposing finally reads as

$$\Psi_{\text{W-RVM}} = \sum_{i=1}^{N_z} \gamma_i \Phi(\mathbf{u}_i). \quad (5)$$

The natural question that arises is how well approximates the reduced and Haar-like designed $\Psi_{\text{W-RVM}}$ the original SVM Ψ_{SVM} , i.e. we have to consider the quantity

$$\|\Psi_{\text{SVM}} - \Psi_{\text{W-RVM}}\| \leq \|\Psi_{\text{SVM}} - \Psi_{\text{RVM}}\| + \|\Psi_{\text{RVM}} - \Psi_{\text{W-RVM}}\|,$$

where the first misfit term on the right hand side is minimized through the iterative method in [11] and [12]. It remains to analyze the second discrepancy between Ψ_{RVM} and $\Psi_{\text{W-RVM}}$.

By making use of kernel-based evaluations of the inner products (and using $k(\mathbf{z}_i, \mathbf{z}_i) = 1$) and Cauchy-Schwarz we obtain

$$\begin{aligned} \|\Psi_{\text{RVM}} - \Psi_{\text{W-RVM}}\|^2 &\leq \left(\sum_{i=1}^{N_z} \|\beta_i \Phi(\mathbf{z}_i) - \gamma_i \Phi(\mathbf{u}_i)\| \right)^2 \\ &= \langle \mathbf{I}_{N_z \times 1}, (\|\beta_1 \Phi(\mathbf{z}_1) - \gamma_1 \Phi(\mathbf{u}_1)\|, \dots, \\ &\quad \|\beta_{N_z} \Phi(\mathbf{z}_{N_z}) - \gamma_{N_z} \Phi(\mathbf{u}_{N_z})\|) \rangle^2 \\ &\leq N_z \sum_{i=1}^{N_z} \|\beta_i \Phi(\mathbf{z}_i) - \gamma_i \Phi(\mathbf{u}_i)\|^2 \\ &= N_z \sum_{i=1}^{N_z} \{\beta_i^2 + \gamma_i^2 - 2\gamma_i \beta_i k(\mathbf{z}_i, \mathbf{u}_i)\} \\ &= N_z \left\{ \sum_{i=1}^{N_z} (\beta_i - \gamma_i)^2 + 2 \sum_{i=1}^{N_z} \gamma_i \beta_i (1 - k(\mathbf{z}_i, \mathbf{u}_i)) \right\} \\ &= N_z \left\{ \|\beta - \gamma\|^2 + 2 \sum_{i=1}^{N_z} \gamma_i \beta_i (1 - k(\mathbf{z}_i, \mathbf{u}_i)) \right\}. \quad (6) \end{aligned}$$

Now, when choosing the Gaussian kernel we may approximate k as follows

$$\begin{aligned} 1 - k(\mathbf{z}_i, \mathbf{u}_i) &= 1 - \exp\left(\frac{-\|\mathbf{z}_i - \mathbf{u}_i\|^2}{2\sigma^2}\right) \\ &= \frac{\|\mathbf{z}_i - \mathbf{u}_i\|^2}{2\sigma^2} + \mathcal{O}(\|\cdot\|^4). \quad (7) \end{aligned}$$

Thus the data misfit discrepancy is directly controlled by the ℓ_2 distance of the sparse approximation \mathbf{u}_i of \mathbf{z}_i (which is minimized under sparsity constraints) and the distance $\|\beta - \gamma\|$.

Thus, up to higher order terms, we achieve

$$\|\Psi_{\text{RVM}} - \Psi_{\text{W-RVM}}\|^2 \approx N_z \left\{ \|\beta - \gamma\|^2 + \sigma^{-2} \sum_{i=1}^{N_z} \gamma_i \beta_i \|\mathbf{z}_i - \mathbf{u}_i\|^2 \right\}, \quad (8)$$

where the relation between the set vector approximation error and the threshold parameter μ needs to be made. This is important to control the tradeoff between sparsity (i.e. computational cost) and the approximation (classification preciseness).

At first, we consider the difference of the set vectors and express them by means of the corresponding wavelet coefficients, i.e.

$$\|\mathbf{z}_i - \mathbf{u}_i\|^2 = \sum_{\lambda \in \Lambda} (z_{i,\lambda} - S_\mu(z_{i,\lambda}))^2.$$

Assuming further that \mathbf{z} consists of $2^M \times 2^M$ pixel, we have

$$1 - k(\mathbf{z}_i, \mathbf{u}_i) \leq 1 - \exp\left(\frac{-2^{2M}\mu^2}{2\sigma^2}\right).$$

Consequently, an upper bound E for the worst case error is then given by

$$\begin{aligned} \|\Psi_{\text{RVM}} - \Psi_{\text{W-RVM}}\|^2 &\leq N_z \left\{ \|\beta - \gamma\|^2 + \right. \\ &2 \left(1 - \exp\left(\frac{-2^{2M}\mu^2}{2\sigma^2}\right) \right) \sum_{i=1}^{N_z} \beta_i \gamma_i \left. \right\} \\ &=: E(\mu). \end{aligned}$$

Neglecting higher order terms of the exp series, we may write

$$E(\mu) \approx N_z \left(\sigma^{-2} 2^{2M} \mu^2 \sum_{i=1}^{N_z} \beta_i \gamma_i + \|\beta - \gamma\|^2 \right). \quad (9)$$

From the last formula we see that the influence of μ is of quadratic nature which assures a rapid error decay of the left hand summand. The quantity $\|\beta - \gamma\|^2$ will be studied below when we have given the rule for deriving the vector γ . In the limit case, $\mu \rightarrow 0$, we then achieve $\lim_{\mu \rightarrow 0} E(\mu) = 0$, which shows that the proposed scheme acts in the limit case as the RVM. For the case in which we really achieve complexity reduction by sparsity and thus a significant gain in computational time and cost, we refer to section III.

4) *Hierarchical Evaluation via Resolution Levels:* The early rejection of easy to discriminate vectors is achieved by a double cascade. The inner cascade is a hierarchy over the number $i = 1, \dots, N_z$ of incorporated W-RSV's, \mathbf{u}_i^l . After incorporating a certain number of W-RSV's with a constant resolution level l it is more efficient to improve the approximation accuracy of the first (already incorporated) vectors. Hence we train in Section II-B.5 $l = 0, \dots, L$ sets of W-RSV's for the outer cascade of coarse-to-fine resolution levels. The trade-off between the two cascades is determined in Section II-C. To exploit these cascades is the superior way to reject most image points by only few operations. Moreover this novel method is robust since the adjustment of only one optimal resolution level was sensitive in [11]. The proposed

evaluation selects the most efficient approximation accuracy automatically at detection time based on the image patch to be classified. In contrast to former methods the trade-off between accuracy and speed is smooth.

5) *Algorithm to Generate Hierarchically Refined W-RSV's:* The algorithm is based on residual Haar wavelet approximations of the RSV's \mathbf{z}_i which are pre-computed by minimizing $\|\Psi_{\text{SVM}} - \Psi_{\text{RVM}}\|^2$ via the algorithm suggested in [12].

Before presenting the algorithm, we introduce the basic quantities. Starting with computing 2^{2J} different initial Haar-like approximations $\mathbf{r}_i^{0,s}$ by (4), where $s \in \{1, \dots, 2^J\}^2$ is the shift of the underlying Haar wavelet basis, we recursively define for $l = 0, \dots, L$ and $i = 1, \dots, N_z$

$$\mathbf{u}_i^l = \sum_{j=0}^l \mathbf{r}_i^{j,s^*}, \quad (10)$$

$$\mathbf{r}_i^{l+1,s} = (W^s)^{-1} S_\mu (W^s (\mathbf{z}_i - \mathbf{u}_i^l)),$$

where the shift s^* denotes the best shift (selected by an optimally criterion introduced below) of the residual at resolution level l , see Figure 2. Note that s^* may differ for each $\mathbf{r}_i^{l,s}$. Within this setting each set vector \mathbf{z}_i is then approximated at level l by \mathbf{u}_i^l . The benefit of the residual structure is that (i) \mathbf{u}_i^l converge to \mathbf{z}_i , if $l \rightarrow \infty$, (ii) we can store all the residuals and thus they do not need to be recomputed in the cascade step when tuning the resolution (i.e. the accuracy of the set vector representation) from coarse to fine, and (iii) the evaluation of the kernel at run-time is more efficient (detailed later at (16) in Section II-C). Beside the computational cost, the discrepancy to the original SVM is of importance. Such a discrepancy depends on the resolution level l and the number i of used set vectors,

$$\begin{aligned} \delta_i^l(s) &= \left\| \Psi_{\text{SVM}} - \sum_{k=1}^{i-1} \gamma_k^{l,i} \Phi(\mathbf{u}_k^l) - \gamma_i^{l,i} \Phi(\mathbf{u}_i^{l-1} + \mathbf{r}_i^{l,s}) \right. \\ &\quad \left. - \sum_{k=i+1}^{N_z} \gamma_k^{l,i} \Phi(\mathbf{u}_k^{l-1}) \right\|_F^2, \quad (11) \end{aligned}$$

where we set $\mathbf{u}_i^{-1} = 0$. The cascade structure is thus achieved when adding residuals $i \rightarrow i + 1$ and then, after reaching $i = N_z$, passing to the next level $l \rightarrow l + 1$, i.e. subsequently adding $\mathbf{r}_i^{l,s}$. Note that for each added residual $\mathbf{r}_i^{l,s}$ we have to compute a new vector $\gamma^{l,i} = (\gamma_1^{l,i}, \dots, \gamma_{N_z}^{l,i})'$. Since we are searching for the best shift s for $\mathbf{r}_i^{l,s}$ and the optimal $\gamma^{l,i}$, we have to minimize $\delta_i^l(s)$. The optimal vector $\gamma^{l,i}$ can be computed explicitly. Introducing the $N_x \times N_z$ matrix

$$\Phi_{\mathbf{x}, \mathbf{u}}^{l,i,s} = \begin{pmatrix} \langle \Phi(\mathbf{x}_1), \Phi(\mathbf{u}_1^l) \rangle & \dots & \langle \Phi(\mathbf{x}_{N_x}), \Phi(\mathbf{u}_1^l) \rangle \\ \vdots & \ddots & \vdots \\ \langle \Phi(\mathbf{x}_1), \Phi(\mathbf{u}_{i-1}^l) \rangle & \dots & \langle \Phi(\mathbf{x}_{N_x}), \Phi(\mathbf{u}_{i-1}^l) \rangle \\ \mathbf{v}^s & & \\ \langle \Phi(\mathbf{x}_1), \Phi(\mathbf{u}_{i+1}^{l-1}) \rangle & \dots & \langle \Phi(\mathbf{x}_{N_x}), \Phi(\mathbf{u}_{i+1}^{l-1}) \rangle \\ \vdots & \ddots & \vdots \\ \langle \Phi(\mathbf{x}_1), \Phi(\mathbf{u}_{N_z}^{l-1}) \rangle & \dots & \langle \Phi(\mathbf{x}_{N_x}), \Phi(\mathbf{u}_{N_z}^{l-1}) \rangle \end{pmatrix}$$

with the i th row

$$\mathbf{v}^s = (\langle \Phi(\mathbf{x}_1), \Phi(\mathbf{u}_i^{l-1} + \mathbf{r}_i^{l,s}) \rangle, \dots, \langle \Phi(\mathbf{x}_{N_x}), \Phi(\mathbf{u}_i^{l-1} + \mathbf{r}_i^{l,s}) \rangle)$$

and the same way the $N_z \times N_z$ matrix $\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s}$ with entries $\langle \Phi(\mathbf{u}_i^l), \Phi(\mathbf{u}_i^{l'}) \rangle$ but where the i th row is replaced with

$$\mathbf{w}^s = (\langle \Phi(\mathbf{u}_1^l), \Phi(\mathbf{u}_i^{l-1} + \mathbf{r}_i^{l,s}) \rangle, \dots, \langle \Phi(\mathbf{u}_i^{l-1} + \mathbf{r}_i^{l,s}), \Phi(\mathbf{u}_i^{l-1} + \mathbf{r}_i^{l,s}) \rangle, \dots, \langle \Phi(\mathbf{u}_{N_z}^{l-1}), \Phi(\mathbf{u}_i^{l-1} + \mathbf{r}_i^{l,s}) \rangle)$$

and i th column with $(\mathbf{w}^s)'$, we recast the discrepancy $\delta_i^l(s)$ as follows,

$$\delta_i^l(s) = \|\Psi_{\text{SVM}}\|_F^2 - 2(\gamma^{l,i})' \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s} \alpha + (\gamma^{l,i})' \Phi_{\mathbf{u},\mathbf{u}}^{l,i,s} \gamma^{l,i}.$$

Evaluating the derivative of the discrepancy and setting it to 0, the optimal $\gamma^{l,i}$ is then obtained by

$$\gamma^{l,i}(s) = (\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s})^{-1} \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s} \alpha \quad (12)$$

and depends thus on s . With the explicit expression (12), the discrepancy becomes

$$\delta_i^l(s) = \|\Psi_{\text{SVM}}\|_F^2 - \alpha' (\Phi_{\mathbf{x},\mathbf{u}}^{l,i,s})' (\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s})^{-1} \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s} \alpha. \quad (13)$$

This of course requires the existence of $(\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s})^{-1}$ what clearly means then linear independency of all involved $\Phi(\cdot)$'s. If this cannot be assured, we have to consider a regularized version of $\delta_i^l(s)$, namely

$$\delta_i^l(s) = \|\Psi_{\text{SVM}}\|_F^2 - 2(\gamma^{l,i})' \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s} \alpha + (\gamma^{l,i})' (\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s} + \rho) \gamma^{l,i}.$$

This yields

$$\gamma^{l,i}(s) = (\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s} + \rho)^{-1} \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s} \alpha \quad (14)$$

and thus

$$\delta_i^l(s) = \|\Psi_{\text{SVM}}\|_F^2 - \alpha' (\Phi_{\mathbf{x},\mathbf{u}}^{l,i,s})' (\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s} + \rho)^{-1} \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s} \alpha. \quad (15)$$

With the matrix notation, the double cascade structure becomes now more visible: beside the residual cascade with respect to l in the approximation of each \mathbf{z}_i by \mathbf{u}_i^l , there is for each l a matrix cascade structure with respect to i that allows to store the entries up to the i th row in $\Phi_{\mathbf{x},\mathbf{u}}^{l,i,s}$ and up to i th row and i th column in $\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s}$. The remaining entries $(\Phi_{\mathbf{x},\mathbf{u}}^{l,i,s})_{n,m}$ for $m > i$ and $(\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s})_{n,m}$ for $n, m > i$ can be taken from the previous level $l-1$.

We summarize our findings and design the algorithm for the generation of the W-RVM:

- 1) Set $\Psi_{\text{SVM}} = \sum_{i=1}^{N_x} \alpha_i \Phi(\mathbf{x}_i)$, $\mathbf{u}_i^{-1} = 0$ and set $l = 0$
- 2) Start with $i = 1$
- 3) Compute for $s \in \{1, \dots, 2^J\}^2$ (here $J = 3$)

$$\mathbf{u}_i^{l-1} = \sum_{j=0}^{l-1} \mathbf{r}_i^{j,s^*}$$

$$\mathbf{r}_i^{l,s} = (W^s)^{-1} S_\mu (W^s (\mathbf{z}_i - \mathbf{u}_i^{l-1}))$$

- 4) Compute $\forall s \in \{1, \dots, 2^J\}^2$ the decrement of the discrepancy

$$\begin{aligned} \text{if } i = 1, l = 0 : \delta \Delta_i^l(s) &= \|\Psi_{\text{SVM}}\|_F^2 - \delta_1^0(s) \\ \text{if } i = 1, l > 0 : \delta \Delta_i^l(s) &= \delta_{N_z}^{l-1}(s^*) - \delta_1^l(s) \\ \text{else : } \delta \Delta_i^l(s) &= \delta_{i-1}^l(s^*) - \delta_i^l(s) \end{aligned}$$

and the number of operations

$$\omega \Delta_i^l(s) = 4 * \#[\mathbf{r}_i^{l,s}] + v(\mathbf{r}_i^{l,s}),$$

where $\#[\mathbf{r}_i^{l,s}]$ is the number of piecewise constant rectangles and $v(\mathbf{r}_i^{l,s})$ the number of gray values of $\mathbf{r}_i^{l,s}$.

- 5) Select the best shift s^* out of $\{1, 2, \dots, 2^J\}^2$ by

$$s^* = \arg \max_s \frac{\delta \Delta_i^{l+1}(s)}{\omega \Delta_i^{l+1}(s)}$$

- 6) Save the rectangle structure of \mathbf{r}_i^{l,s^*} and the coefficient vector

$$\hat{\gamma}^{l,i} = \gamma^{l,i}(s^*) = (\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s^*})^{-1} \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s^*} \alpha$$

- 7) If $i < N_z$, increment i and proceed to step 3. If $i = N_z$ and $l < L$, increment l and proceed to step 2; else, stop.

Finally, as a byproduct of this section and as a contribution to section II-B.3, we are now able to quantify $\|\beta - \gamma\|$. Assume, the SVM is given by N_x set vectors \mathbf{x}_i and the RVM by N_z set vectors \mathbf{z}_i , then with $(\Phi_{\mathbf{z},\mathbf{z}})_{i,j} = \langle \Phi(\mathbf{z}_i), \Phi(\mathbf{z}_j) \rangle$ and $(\Phi_{\mathbf{x},\mathbf{z}})_{i,j} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{z}_j) \rangle$ it is common that $\beta = \Phi_{\mathbf{z},\mathbf{z}}^{-1} \Phi_{\mathbf{z},\mathbf{x}} \alpha$, see [12]. Consequently,

$$\|\beta - \hat{\gamma}^{l,i}\| \leq \|\Phi_{\mathbf{z},\mathbf{z}}^{-1} \Phi_{\mathbf{x},\mathbf{z}} - (\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s^*})^{-1} \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s^*}\| \|\alpha\|$$

and since we have $\|\mathbf{u}_i^l - \mathbf{z}_i\| \leq C_{\mu,l}$, by perturbation arguments we also have an entry-wise perturbation estimate for the full matrices which in turn yield an estimate for $\|\beta - \hat{\gamma}^{l,i}\|$ in dependence on μ and l (we omit a detailed examination here). Moreover, as the approximations \mathbf{u}_i^l at resolution level l tend to \mathbf{z}_i as μ tends to 0, we have an entry-wise convergence

$$\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s^*} \rightarrow \Phi_{\mathbf{z},\mathbf{z}}, \quad \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s^*} \rightarrow \Phi_{\mathbf{x},\mathbf{z}}$$

and hence

$$\|\Phi_{\mathbf{z},\mathbf{z}}^{-1} \Phi_{\mathbf{x},\mathbf{z}} - (\Phi_{\mathbf{u},\mathbf{u}}^{l,i,s^*})^{-1} \Phi_{\mathbf{x},\mathbf{u}}^{l,i,s^*}\| \xrightarrow{\mu \rightarrow 0} 0$$

C. Detection Process

The classification function of the W-RVM, denoted by $y_i^l(\mathbf{x})$ of the input patch \mathbf{x} , using N_z W-RSV's at the levels $0, \dots, l-1$ and i W-RSV's at the level l is as follows:

$$\begin{aligned} y_i^l(\mathbf{x}) &= \text{sgn} \left(\sum_{k=1}^i \hat{\gamma}_k^{l,i} k(\mathbf{x}, \mathbf{u}_i^l) \right. \\ &\quad \left. + \sum_{k=i+1}^{N_z} \hat{\gamma}_k^{l,i} k(\mathbf{x}, \mathbf{u}_i^{l-1}) + b_i^l \right) \quad (16) \\ k(\mathbf{x}, \mathbf{u}_i^l) &= \exp \left(-1/2\sigma^2 (\mathbf{x}'\mathbf{x} - 2\mathbf{x}'\mathbf{u}_i^l + \mathbf{u}_i^l \mathbf{u}_i^l) \right), \end{aligned}$$

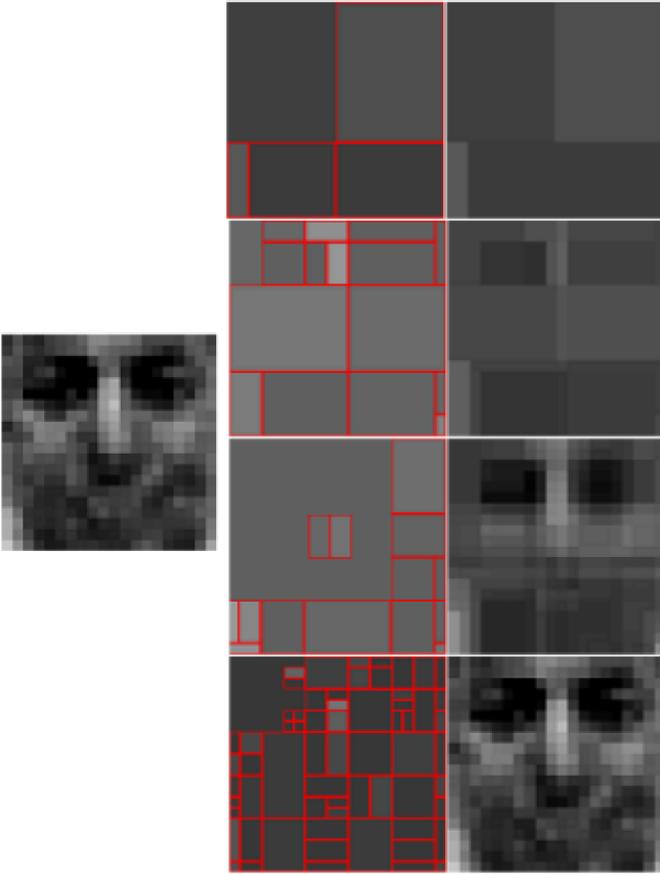


Fig. 2. Example of recursively approximating a RSV. *Left*: a RSV \mathbf{z}_i , *right*: W-RSV \mathbf{u}_i^l at different resolution levels (top to bottom: $l = 0, 1, 9, 18$), *middle*: related residuals $\mathbf{r}_i^{l,s*}$ (top to bottom: $l = 0, 1, 9, 18$).

where the kernel k is efficiently evaluated using Integral Images (Section II-A). For the term $2\mathbf{x}'\mathbf{u}_i^l = 2\mathbf{x}'\mathbf{u}_i^{l-1} + 2\mathbf{x}'\mathbf{r}_i^{l,s*}$ only $2\mathbf{x}'\mathbf{r}_i^{l,s*}$ has to be computed, since $2\mathbf{x}'\mathbf{u}_i^{l-1}$ can be stored at the previous level. The thresholds b_i^l are obtained automatically from an R.O.C. for a given accuracy. These thresholds are set to yield a given False Rejection Rate (FRR) so that the accuracy of the W-RVM is the same as the one of the full SVM (see [12] for details). The trade-off between FRR and FAR is the only parameter of our algorithm to be set by the user.

Realizing our double cascade algorithm (Section II-B.4) the detection process goes as follows:

- 1) Start at the first resolution level $l = 0$
- 2) Start with the first W-RSV, \mathbf{u}_1^l at the level l
- 3) Evaluate $y_i^l(\mathbf{x})$ for the input patch \mathbf{x} using (16)
- 4) If $y_i^l < 0$ then the patch is classified as not being the object of interest, the evaluation stops
- 5) If $i < N_z$, i is incremented and the algorithm proceeds to step 3; else if $l < L$, l is incremented and the

algorithm proceeds to step 2; otherwise the full SVM is used to classify the patch.

1) *Adjustment of Resolution Levels and Number of W-RSV's per Level*: When computing an approximation of an SVM, it is not clear how many approximation vectors N_z should be computed (see [12]). This number of vectors may vary depending on the level l of the approximation. To this end, it may be useful to let N_z depend on l . The reason is that at a certain point of the evaluation algorithm it is more efficient to increment l (and reset i), rather than to increment i . The best value of $N_z(l)$ is computed in an offline process using a validation dataset: $N_z(l)$ is set to the smallest i for which empirically

$$\frac{\text{Nops}(y_{i+1}^l)}{\text{Nrecs}(y_{i+1}^l)} > \frac{\text{Nops}(y_1^{l+1})}{\text{Nrecs}(y_1^{l+1})},$$

where Nops stands for the number of operations and Nrecs stands for the number of rejections of the negative examples.

By a similar evaluation the last used resolution level L can be achieved. For this L it is more efficient to classify the last few remaining patches by the SVM, instead of incrementing l . L depends also on the sparsity parameter μ . The smaller μ , the closer is \mathbf{u}_i^l to \mathbf{z}_i and the less resolution levels are required. However, the number of levels does not play a decisive role as the higher L , the sooner the evaluation process selects the next level, i.e. the less $N_z(l)$. Therefore our proposed approach is not very sensitive to the parameter for setting the approximation accuracy (e.g. μ), opposite to former methods using only one resolution level.

III. EXPERIMENTAL RESULTS

We applied our novel Wavelet Approximated Reduced SVM to the task of face detection. For the training and validation of the classifier we used two databases. The first set was crawled from the WWW (see Acknowledgment) and as second face database we used the grayscale version of FERET [9]. We chose this well-known dataset to provide the comparability to other approaches.

The training set includes 3500, 20×20 , face patches and 20000 non-face patches from the first dataset. The SVM computed on the training set yielded about 8000 Support Vectors that we approximated by $N_z = 90$ W-RSV's at $L = 5$ resolution levels by the method detailed in the previous section.

As first validation set (set I) we used 1000 face patches, and 100,000 non-face patches randomly chosen also from WWW images, but disjoint from the training examples. The first graph on Figure 3 plots the residual distance of the RVM (dashed line) and of the W-RVM (plain line) to the SVM (in terms of the distance $\Psi_{\text{SVM}} - \Psi_{\text{RVM}}$ and $\Psi_{\text{SVM}} - \Psi_{\text{W-RVM}}$) as a function of the number of vectors used. It can be seen that for a given accuracy more Wavelet Approximated Set Vectors are needed to approximate the SVM than for the RVM. However, as shown on the second plot, for a given computational load, the W-RVM rejects much more non-face patches from the validation set I than the RVM. This explains the improved run-time performances of the W-RVM. Additionally, it can be

seen that the curve is more smooth for the W-RVM, hence a better trade-off between accuracy and speed can be obtained by the W-RVM.

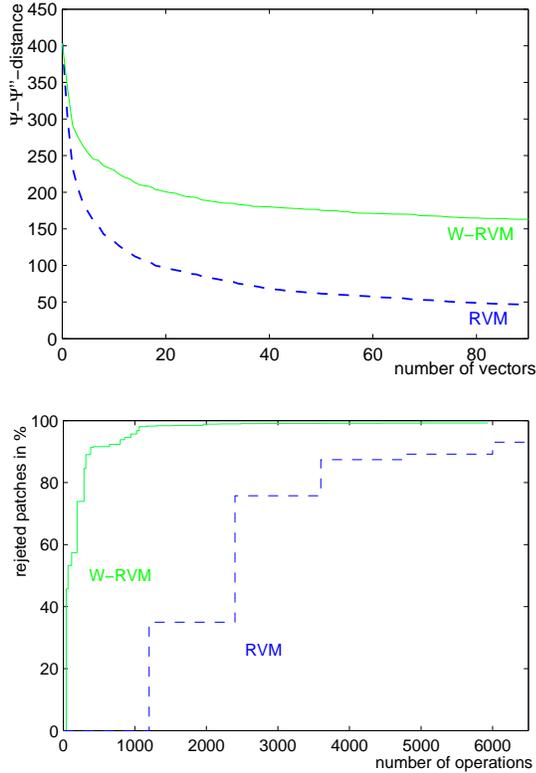


Fig. 3. Top: $\Psi_{SVM} - \Psi_{W-RVM}$ distance as function of the number of vectors for the RVM (dashed line), and the W-RVM (solid line). Bottom: Percentage of rejected non-face patches as a function of the number of operations required.

Figure 4 shows the R.O.C.'s, computed on the validation set I, of the SVM, the RVM and the W-RVM. It can be seen that the accuracies of the three classifiers are similar without (top plot) and almost equal with the final SVM classification for the remaining patches (bottom plot), see step 5. of the evaluation algorithm.

Table I compares the accuracy and the average time required to evaluate the patches of the validation set I. The speed-up over the former approach [11] is about a factor 2.5 ($3.85\mu s$). The novel W-RVM algorithms provides a significant speed-up (530-fold over the SVM and more than 15-fold over the RVM), for no substantial loss of accuracy.

TABLE I

COMPARISON OF ACCURACY AND SPEED IMPROVEMENT OF THE W-RVM TO THE RVM AND SVM.

method	FRR	FAR	time per patch
SVM	1.4%	0.002%	$787.34\mu s$
RVM	1.5%	0.001%	$22.51\mu s$
W-RVM	1.4%	0.002%	$1.48\mu s$

The validation set II contains 500 frontal and half profile images from the FERET database [9]. We compared our approach with the Viola & Jones method [19] implemented

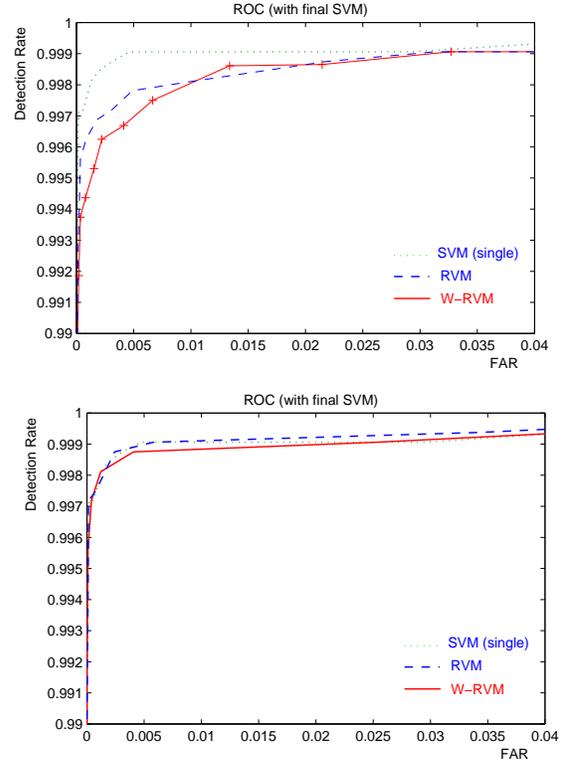


Fig. 4. R.O.C.'s for the SVM, the RVM and the W-RVM (top) without and (bottom) with the final SVM classification for the remainin patches. The FAR is related to non-face patches.

in OpenCV (version b5a). The Viola & Jones detector yields on set II a detection rate of 90.9% by 0.32 false acceptances (FA) and 0.29 sec per image (on a Pentium M Centrino 1600 CPU). Compared to the results given in [19] the processing time is slower since the image size of the FERET images is larger. The results on FERET are more accurate because of the higher quality of the images. With the W-RVM we obtained on the same PC and set II a detection rate of 90.1% by 0.25 FA and 0.15 sec processing time per image.

Our proposed classifier is more efficient at detection, but mainly at training time than the AdaBoost method [19] and classifies about 25 times faster than the Rowley-Baluja-Kanade detector [13] and about 1000 times faster than the Schneiderman-Kanade detector [15].

We also proved the performance and detection accuracy under real-life conditions in the "Institut für Techno- und Wirtschaftsmathematik" (ITWM) in Kaiserslautern.

To demonstrate the efficient and accurate detection algorithm, we implemented an application using a standard webcam. Accurate face detection one obtained at real-time by 25 fps (on a Intel Pentium M Centrino 1600 CPU, at a resolution of 320x240, step size 1 pixel, 5 scales).

IV. CONCLUSION

In this paper, we presented a novel efficient method for SVM classifications on image based vectors. The essential ingredient was an recursively applied optimally matched Wavelet

Transform of the Reduced Set Vectors. It was demonstrated on the task of face detection.

As opposed to the RVM, the sparseness of operations required for classification is not only controlled by the number of Reduced Set Vectors but also by the number of wavelets basis functions used to approximate a Reduced Set Vector. Hence, negative examples can be rejected with much fewer number of operations, making the run-time algorithm very efficient. Moreover, as the Haar wavelets are used, the SVM kernel may be evaluated extremely efficient using Integral Images. The main advantage of this algorithm compared to other algorithms based on boosting, such as the Viola & Jones detector [19], is the fact that the training is much faster and does not require manual intervention.

ACKNOWLEDGMENT

The authors would like to thank the group of Hans Burkhardt of the University of Freiburg and particularly Ralf Jüngling for providing the face images used in the experiments. Thanks to Jamie Sherrah for his insights on the Viola & Jones detector.

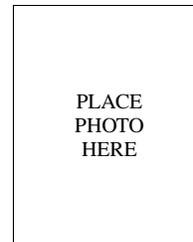
REFERENCES

- [1] A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear Approximation and the Space $BV(\mathbb{R}^2)$. *American Journal of Mathematics*, (121):587–628, 1999.
- [2] R.R. Coifman and D. Donoho. Translation-invariant de-noising. in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, New York, pages 125–150, 1995.
- [3] F. Crow. Summed-area tables for texture mapping. In *Proc. of SIGGRAPH*, 18(3):207 – 212, 1984.
- [4] I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising. *Applied and Computational Harmonic Analysis*, 19(1):1–16, 2005.
- [5] C. Garcia, G. Zikos, and G. Tziritis. Face detection in color images using wavelet packet analysis. *IEEE Int. Conf. on Multimedia Computing and Systems*, 1999.
- [6] D.A. Karras. Improved defect detection in textile visual inspection using wavelet analysis and support vector machines. *ICGST International Journal on Graphics, Vision and Image Processing*, 2005.
- [7] D. Keren, M. Osadchy, and C. Gotsman. Antifaces: a novel, fast method for image detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:747–761, July 2001.
- [8] W. Kienzle, G. H. Bakir, M. O. Franz, and B. Schölkopf. Efficient approximations for support vector machines in object detection. *Proc. DAGM'04*, pages 54 – 61, 2005.
- [9] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The feret evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, October 2000.
- [10] M. Rätsch, S. Romdhani, G. Teschke, and T. Vetter. Over-complete wavelet approximation of a support vector machine for efficient classification. *Proc. DAGM'05: 27th Pattern Recognition Symposium*, Vienna, 2005.
- [11] M. Rätsch, S. Romdhani, and T. Vetter. Efficient face detection by a cascaded support vector machine using haar-like features. *Proc. DAGM'04: 26th Pattern Recognition Symposium*, pages 62 – 70, 2004.
- [12] S. Romdhani, P. Torr, B. Schölkopf, and A. Blake. Computationally efficient face detection. In *Proceedings of the 8th International Conference on Computer Vision*, July 2001.
- [13] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *PAMI*, 20:23–38, 1998.
- [14] H.-J. Schmeisser and H. Triebel. *Topics in Fourier Analysis and Function Spaces*. John Wiley and Sons, New York, 1987.
- [15] H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied to faces and cars. *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1:746 – 751, 2000.
- [16] B. Schölkopf, S. Mika, C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch, and A. Smola. Input space vs. feature space in kernel-based methods. *IEEE Transactions on Neural Networks*, 10(5):1000 – 1017, 1999.
- [17] H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. Verlag der Wissenschaften, Berlin, 1978.
- [18] V. Vapnik. *Statistical Learning Theory*. Wiley, N.Y., 1998.
- [19] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.



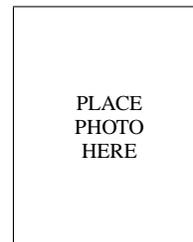
Matthias Rätsch studied mathematics, physics and computer science at the University of Potsdam, Germany. Then he worked on the fields of Optical and Intelligent Character Recognition at the University of Bremen, Germany. In 2002 he received the diploma-degree from the University of Potsdam, for his thesis on segmenting Touching Characters (Patent No: 195 33 585). Since 2003 he joined as PhD student the Graphics and Vision Research Group (GraVis) at the University of Basel, Switzerland. His research interests are in the fields of Face

and Facial Feature Detection, Machine Learning, and Human Computer Interface.



Gerd Teschke received the Diploma degree in mathematics in 1998 from the University of Potsdam. In 2001 he received the Dr. rer. nat. degree from the University of Bremen awarded with the Bremer Studienpreis (Bruker prize). From 2002–2003 he was with the PACM at Princeton University. In 2002 he became assistant professor at the University of Bremen where he received the Habilitation degree in 2006. In 2005 he moved to Berlin and became the Head of the research group Inverse Problems in Science and Technology at the Konrad Zuse

Institute. His current research centers around inverse problems, frame theory and sparse signal recovery.



Sami Romdhani will add biography text and photo when paper is accepted.



Thomas Vetter studied mathematics and physics and received the PhD degree in biophysics from the University of Ulm, Germany. As a postdoctoral researcher at the Center for Biological and Computational Learning at MIT, he started his research on computer vision. In 1993, he moved to the Max-Planck-Institut in Tübingen and, in 1999, he became a professor of computer graphics at the University of Freiburg. Since 2002, he has been a professor of applied computer science at the University of Basel in Switzerland. His current research is on image

understanding, graphics, and automated model building. He is a member of the IEEE and the IEEE Computer Society.