SIXTH FRAMEWORK PROGRAMME
MARIE CURIE ACTIONS
HUMAN RESOURCES AND MOBILITY
MARIE CURIE RESEARCH TRAINING NETWORKS (RTN)

COMPUTATIONAL AND COGNITIVE VISION SYSTEMS:
A TRAINING EUROPEAN NETWORK

VISIONTRAIN

Rezaul Karim

Optimizing Local Classifiers For Appearance Based Visual Object Detection

Contract number: **MRTN-CT-2004-005439**
Project start/end dates: **1/5/2005 – 30/4/2009**
Reporting period: **1/5/2005 – 30/4/2009**

Phd advisor: **Christoph Schnörr** (schnoerr@math.uni-heidelberg.de)
Host institution: University of Heidelberg (moved from University of Mannheim) (partner 11), Germany

Contents

1 Introduction: Phd Topic

Class-specific object recognition is challenging in view of the high intra-class variability of object instances (Figure 1). In this project, we aim to detect humans by concentrating on the structural approach in terms of local parts and their relationships. The overall detection process comprises the

Figure 1: Large variation in human pose renders human detection as a difficult task.

following two main steps : First, local classifiers find probable locations of different body parts in the input image, Second, possible objects and their poses in the image are detected by graphical inference (Figure 2).

Figure 2: **Left**: Input test image. **Middle**: Possible candidates (dots) for different body parts detected by local classifiers. **Right**: Detected human with pose found by using graphical inference.

This thesis report is focusing on the first step, that is, about optimizing the local classifiers, which is the author's PhD topic.
Notation. FG = Foreground or Target; BG = Background.

2 Brief review of the state of the art

Run time optimization of classifiers is a key issue for fast object detection. A prominent example is the work of Viola and Jones [12] on face detection based on a cascade of boosted weak classifiers that only require simple image convolutions for feature extraction and thresholding. This framework is not directly applicable to kernel classifiers like support vector machines (SVMs)[11], for instance, because boosting based on such strong classifiers as components is less effective. In many applications, however, the flexibility of kernel machines is a decisive advantage, as they can be applied to arbitrary features and pattern representations including histograms, sets, graphs, etc. This raises the question of how to design structured architectures for efficient classification using kernel machines as components.

Accordingly, this problem has spurred research recently. Related work can be roughly, but not disjointly, classified

- into approaches [2, 9, 8, 13, 4, 5] to the design of *Reduced Support Vector Machines (RSVMs)* that require less computational costs than the standard SVM for classifying a pattern, and
- into approaches [3, 9, 10, 1] that exploit SVMs (either reduced or not) as components of a structured architecture for classification.

Regarding the former class of approaches, RSVMs require only a *fraction of kernel evaluations* for classifying a pattern, either by computing a sparse subset of the support vectors of the full SVM [2, 4], or by computing a novel small set of vectors in order to replace the support vectors altogether [9, 13]. Additionally, wavelet approximations of these latter vectors have been investigated in [8] in order to *efficiently evaluate the arguments* (i.e. dot products between pattern vectors) to which the kernel function is applied.

The latter class of approaches, on the other hand, is focusing on *structured SVM-based classification* for face detection. Heisele et al. [3] studied a hierarchy of linear SVMs including a single nonlinear SVM as top node. Thresholds were tuned for optimizing classification performance and speed, followed by feature selection. Romdhani et al. [9] proposed a single chain of SVMs that is optimized also by threshold tuning, and by approximating a fully nonlinear SVM that has to be computed beforehand, whereas a decision tree with linear SVMs is suggested in [1]. Finally, Sahbi and Geman [10] recently presented a tree-structured hierarchy of SVMs that again is optimized by the reduced set technique used in [9] and threshold selection, and is operating on an application specific partitioning of the space of patterns (faces) according to different poses.

3 Approach and methodology

Earlier, we use Support Vector Machine (SVM) [11] based classifiers as local classifiers to find possible candidate for different body parts. The whole human body is divided into several parts. In the training phase, we train a single classifier for each of these parts. In the test phase, these classifiers classify the test image data where FGs are the corresponding parts for each of the classifiers. BG images, without containing any human, are used against FGs to train the classifiers.
There are up to **1700 SVs for just one part**! And as due to kernel evaluations by SVs, classification time increases with the increasing of SVs, this makes part detection even more time consuming than the combinatorial graphical inference in our system. Thus, to make it real time applicable, at first, we need to use classifiers with reduced number of SVs (RSVMs). Figure 3 (on page 4) illustrates the basic idea underlying the design of RSVMs.

3.1 RSVM

We assess a state of the art direct greedy strategy [4] for designing reduced SVMs (RSVMs). The rationale behind our choice is as follows: Firstly, we focus on *direct* RSVM computation rather than

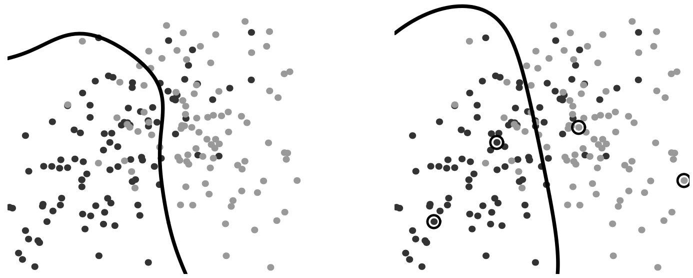

Figure 3: **Illustration of SVM & RSVM. Left:** Decision surface of standard SVM trained with all patterns, and optimized parameters. Number of SVs is 93. Error rate is $L_{full} \approx 17.005\%$. **Right:** An RSVM with 4 SVs indicated by circles. Error rate $L_{red} \approx 17.71\%$ is only slightly worse than that of the standard SVM, whereas the kernel evaluations have been reduced by a factor of about 23.

on approximations of fully nonlinear SVMs, in order to avoid the need to train the latter beforehand. Secondly, we refrain from the computation of novel representatives of support vectors as done in [9, 13] because this relies on complex optimization problems that are sensitive to initialization, step sizes, etc.

In this method [4] the basic idea is to perform a greedy search for an optimal subset of the training set; it trains directly an RSVM by minimizing the primal objective function of standard SVM, which is an explicit function of the classification performance on the training set. Thus, the classifier, produced by using this small subset demands much lower number of kernel evaluation than standard SVM while having similar performance.

Cascading of RSVMs: A natural way to reduce the number of (expensive) kernel evaluations is to reject a major part of the background patterns with a simple classifier as early as possible. Thus, using a simple classifier as a first stage to reject them and a comparatively powerful (with more SVs) classifier to work on the remaining patterns highly reduces the overall number of kernel evaluations. Therefore, instead of having a single complex classifier as detector, we prefer to design a detector by cascacading two RSVM classifiers in series in the following way:

As shown in figure 4 (on page 5), the whole test set is input of M1. Only patterns, which are detected as targets (FG) by M1, will go as input into M2. From these input patterns, those, which are classified as FG by M2, are FG patterns detected by the whole cascaded detector. Patterns that are classified as BG by any of M1 or M2 are discarded instantly and not processed further.

3.2 Kernel free detector for improving efficiency

Although sparse kernel classifiers are less computational expensive by one order of magnitude, they still require the evaluation of kernels, which is an intrinsically expensive operation by itself. This raises the problem of how to reduce further the computational costs for detection without significant loss of performance.

To this end, we utilize kernel machines in a preprocessing phase to determine an ensemble of weak learners, which subsequently is combined by standard boosting methods to produce a complex detector that runs much faster, however. The weak classifiers considered are more flexible than decision stumps but have similar low computational classification costs in comparison with kernel evaluations.

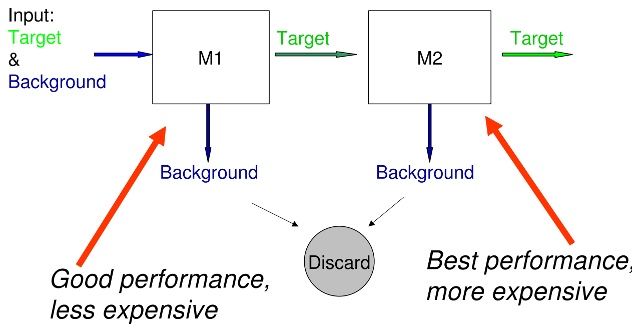

Figure 4: Architecture of a two-stage classifier for reducing the number of kernel evaluations.

Boosting of an ensemble of such weak learners produces promising detectors. The overall method comprises the following steps:

- Train a full SVM and find its decision boundary.
- Determine a set of linear weak learners from this decision boundary.
- Run AdaBoost on this set of weak learners.

4 Scientific achievements

4.1 Design of fast low-level detectors

The main goal of the PhD project was to optimize the local classifiers, that is finding local classifiers having least computational cost while keeping classification performance reasonably high. As we are dealing with very complex images, a complex classifier is mandatory in order to obtain a desired classification accuracy. At the same time, due to the large size of data from many different bodyparts, an efficient classifier is really crucial. We realised a novel technique that produces a classifier with high complexity and less computational cost with significant accuracy. This classifier produces decision boundaries as complex as do SVMs. While this complexity induces high computational costs of SVMs in the classification phase, our classifier is only little more expensive to evaluate than a simple linear classifier. The evaluation on 13 different body parts of real human data set from our own database, showed that, compared to SVM, our classifier gives 1.02% more error rate while being 393.56 times faster!
Using such detectors as component of a larger structured detector can further enhance efficiency.

4.2 Graphical models for object recognition

Low-level detectors of features form an essential component of visual object recognition systems. In this context, our group studied the performance of the detectors outlined above for object recognition with graphical models. Based on the output of object part detectors, contextual inference with a graphical model computes the most probable configuration of parts in a given image.

So far, the preprocessing stage took most of the computation time, followed by graphical inference in few seconds. The novel detectors speed up the whole detection process dramatically without loss of overall detection performance.

5 Results

This section summarizes our experimental evaluation of RSVM-cascade and kernel free detector

5.1 Cascading reduces computational cost up to 99%

We applied a two-stages classifier combining a very sparse RSVM at stage 1, followed by the RSVM designed as reported in section 3.1, using the bechmark data[7], and averaged the results over the corresponding 100 training-test pairs of data sets. The effective number of support vectors is the sum of #SVs of the first machine plus #SVs of the second machine multiplied by the acceptance rate of the first machine.
Table 1 shows that in comparison to a single SVM based classifier [4], by cascading of RSVMs, classification cost can be reduced upto 99% without a significance loss of performance.

Dataset	Our Cascade		Keerthi *et al* [4]		SVM[4]	
	Effective #SVs	Error	#SVs	Error	#SVs	Error
Breast	4.67(0.66)	26.80(4.92)	12.1(5.6)	29.22(2.11)	185.8(16.44)	28.18(3.00)
Diabetis	6.48(0.54)	26.32(2.28)	13.8(5.6)	23.47(1.36)	426.3(26.91)	23.73(1.24)
German	4.41(0.63)	27.80(2.45)	14.0(7.3)	24.90(1.50)	630.4(22.48)	24.47(1.97)
Ringnorm	9.79(0.19)	2.04(0.28)	12.9(2.0)	1.97(0.57)	334.9(108.54)	1.68(0.24)
Thyroid	5.45(0.41)	5.61(2.41)	10.6(2.3)	5.47(0.78)	57.80(39.61)	4.93(2.18)
Waveform	9.16(0.40)	12.75(1.33)	14.4(3.3)	10.66(0.99)	246.9(57.80)	10.04(0.67)

Table 1: Benchmark evaluation of the two-stages sparse SVM. The effective number of SVs minimizes the classification costs upto 99% compared to a single SVM based classifier and yields comparable classification performance.

5.2 Kernel-free detectors: boosting computational efficiency

5.2.1 Decision boundary: SVM & Kernel free detector

Figure 5 (on page 7) shows decision boundaries by SVM and our kernel free AdaBoost detector on Banana data set. Comparing these boundaries, that our kernel free detector is able to approximate complex decision boundaries of SVMs. This SVM is construced with 80 SVs whereas our kernel free detector is constructed with 30 linear classifier. So, to classify a single novel pattern, SVM will demand minimum 80 kernel evaluations whereas our kernel free detector requires about 30 dot products only!

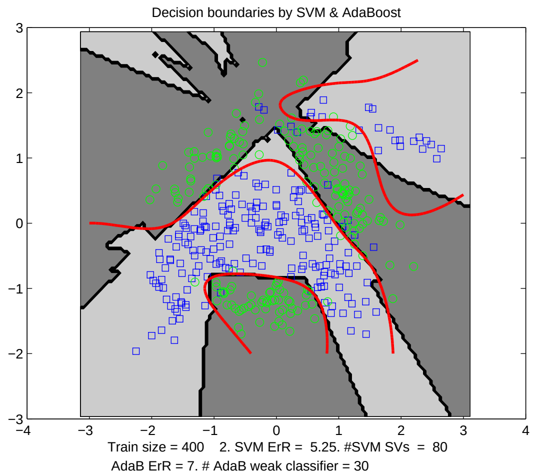

Figure 5: Decision boundary by SVM(solid red curve) and our Kernel free detector.

5.2.2 Benchmark: Kernel based detector & Our Kernel free detector

We evaluated our Kernel free detector on the benchmark data[7], and averaged the results over corresponding 100 training-test pairs of data sets. '
Table 2 (on page 7) and 3 (on page 8) show that in comparison to some reported standard kernel based detectors [4, 6] the classification cost can be reduced up to several hundred times without a significance loss of performance. In this table, '#LinClas' means the number of linear classifiers (selected weak learners) used to construct the Kernel free detector.

Dataset	Datatype	Kernel Free		AB*[6]	SVM [4]		RSVM [4]	
	TrPat TPat Dim Sets	#LinClas(SD)	ER (SD)	ER(SD)	# SV(SD)	ER(SD)	#SV(SD)	ER(SD)
Banana	400 4900 2 100	35(0)	12.3(1.1)	12.3(0.7)	221.7(67.0)	10.5(0.7)	17.3(7.3)	10.9(1.7)
Breast	200 77 9 100	6(0)	28.3(4.3)	30.4(4.7)	185.8(16.4)	28.2(3.0)	12.1(5.6)	29.2(2.1)
Diabetis	468 300 8 100	4(0)	23.8(2.1)	26.5(2.3)	426.3(26.9)	23.7(1.2)	13.8(5.6)	23.5(1.4)
Flare	666 400 9 100	3(0)	32.6(1.7)	35.7(1.8)	629.4(29.4)	34.0(1.3)	8.4(1.2)	33.9(1.1)
German	700 300 20 100	11(0)	25.4(2.3)	27.5(2.5)	630.4(22.5)	24.5(2.0)	14.0(7.3)	24.9(1.5)

*200 (RBF net) base classifiers were combined to build a single Boosted machine.

Table 2: Benchmark evaluation of our kernel free Boosted detector. Evaluating a Gaussian kernel (poses by a single SV on single pattern) is at least 3 times more expensive than evaluating a dot product (poses by a single linear classifier on single pattern).The Kernel free detector is up to several hundred times compuationally more efficient compared to a kernel based detector (SVM; AdaBoost, boosted with many RBF net) and yields comparable classification performance.

Dataset	Datatype	Kernel Free	AB*[6]	SVM [4]	RSVM [4]
Heart	170 100 13 100	3(0) 17.1(0)	20.3(3.4)	166.6(8.7) 15.8(2.2)	4.3(2.6) 15.5 (1.1)
Thyroid	140 75 5 100	17.0(0) 7.5(3.0)	4.4(2.2)	57.8(39.6) 4.9(2.2)	10.6(2.3) 5.5(0.8)
Titanic	150 2051 3 100	6(0) 23.0(1.2)	22.6(1.2)	150.0(0.0) 22.3(0.7)	3.3(0.9) 22.7(1.9)
Twonorm	400 7000 20 100	5(0) 3.2(0.4)	–(–)	330.3(137.0) 2.4(0.2)	8.7(3.7) 3.0(0.8)
Waveform	400 4600 21 100	22(0) 11.2(0.7)	10.8(0.6)	246.9(57.8) 10.0(0.7)	14.4(3.3) 10.7(1.0)

*200 (RBF net) base classifiers were combined to build a single Boosted machine.

Table 3: Benchmark evaluation of our kernel free Boosted detector. Evaluating a Gaussian kernel (poses by a single SV on single pattern) is at least 3 times more expensive than evaluating a dot product (poses by a single linear classifier on single pattern).The Kernel free detector is up to several hundred times compuationally more efficient compared to a kernel based detector (SVM; AdaBoost, boosted with many RBF net) and yields comparable classification performance.

5.2.3 Real data: SVM vs kernel free detector

We evaluated our kernel free detector on 13 different body parts of real human data set from our own database.
Table 4 (on page 9)shows that in comparison to SVM, our kernel free detector requires up to several hundred times less computational cost without a significance loss of performance.

6 Contribution to Visiontrain's workpackages

6.1 WP1: Computational theories and methods for low-level vision

We contributed to this topic along two lines of research.

- A *novel design methodology* was developed for fast local detection of object parts. The resulting architecture combines the classification performance of SVMs with the computational efficiency of boosted weakly learners, omitting any expensive kernel evaluations. Although these detectors were studied in connection with detection parts of the human body, the range of applications of this methodology extends to numerous other problems of low-level vision where fast computation and localization of cues in a pre-processing phase are essential.

- We studied novel method for graphical inference with complete graphs (see also the section below). The method is based on the *theoretical derivation* of novel bounds that are utilized in a A^*-search algorithm. Tightness of these bounds leads to sparse branching of the search algorithm an enables efficient inference with highly contextual models.

6.2 WP3: Learning and recognition of shapes, objects, and categories

We thoroughly studied the detection of *object categories* with completely connected graphical models. Vertices index a moderate number of object parts whose configuration and localization is to be inferred

Dataset	Datatype			SVM		Kernel Free Detector	
	TrPat	TPat	Dim	ToER	ClasifTime(sec)	ToER	ClasifTime(sec)
Head	6284	9714	640	3.6	77.1	3.9	0.2
L-Sho	6254	9667	640	5.8	85.3	6.7	0.2
R-Sho	6241	9646	640	4.7	82.1	5.5	0.2
L-Elb	6139	9449	640	6.6	131.6	8.2	0.4
R-Elb	6120	9418	640	6.0	52.6	7.0	0.2
L-Han	6092	9358	640	6.0	51.7	6.4	0.2
R-Han	6091	9340	640	3.5	163.9	6.6	0.2
L-Hip	6161	9463	640	7.0	141.9	7.6	0.6
R-Hip	6155	9468	640	7.2	91.8	8.2	0.2
L-Kne	6086	9238	640	5.9	88.4	6.1	0.3
R-Kne	6076	9234	640	5.8	81.5	6.2	0.2
L-Foo	6464	9582	640	6.2	117.8	7.7	0.2
R-Foo	6446	9614	640	6.2	172.2	7.8	0.3

Table 4: Real data evaluation of our Kernel free Boosted detector and SVM. Compared to SVM, Kernel free detector poses upto several hundred times less computational cost without a significance loss of performance.

in a given image. Candidate locations are computed in a pre-processing step as described above.

Probabilistic *representation of the shape* of instances of an object category is learned in terms of relations between parts. This not only concerns relative geometrical relations, but also the visual appearance of single parts and joint appearance of pairs of parts.

These relations are conveniently encoded with a *complete* graphical model. The high connectivity of the model makes object detection robust against failure of local part detection in the pre-processing stage. It also enables to detect and localize occluded parts.

Our approach can be generally applied to various object categories. It was comprehensively evaluated for the detection of faces and humans in single images, for the detection of humans in images taken simultaneously from multiple views, and for the extraction of the human spine in 3D medical image data.

7 Future research objectives

Concerning the local part detectors, there two major open problems.

1. A better theoretical insight as to why our kernel free classifier performs so well, and if this holds in general, is desirable. In particular, performance bounds should be derived and related to the performance of state-of-the-art kernel based classifiers, along with a characterization of data sets where these bound can be achieved.

2. Our analysis was confined to two-stage classifiers, so far. An extension to structural combinations of $n > 2$ classifiers will be subject of our further work.

8 Career plans

I have a great desire to go on with research follwoing an optimum environment, preferably in the academy and regardless of the place(country).

9 References

- R. Karim, M. Bergtholdt, J. Kappes, and C. Schnörr. Greedy-Based Design of Sparse Two-Stage SVMs for Fast Classification. In Pattern Recognition – 29th DAGM Symposium, volume 4713 of LCNS, pages 395-404, 2007. Springer.
- M. Bergtholdt, J. Kappes, S. Schmidt, and C. Schnörr. A Study of Parts-Based Object Class Detection Using Complete Graphs. *International Journal of Computer Vision*, in press, published electr. January 28, 2009. [doi:10.1007/s11263-009-0209-1]

A Date and location of the Phd's final examination

A provisional date for my defense: Feb 2010

B List of attended events

- First VISIONTRAIN thematic school, 12 - 17 March 2006, Les Houches, France
- DAGM 2006 (28th Annual Symposium of the German Association for Pattern Recognition. Berlin, Germany September 12th-14th, 2007)
- Second VISIONTRAIN thematic school, 25 - 30 March 2007, Les Houches France
- DAGM 2007 (29th Annual Symposium of the German Association for Pattern Recognition. Heidelberg, Germany September 12th-14th, 2007)
- The Analysis of Patterns (Bertinoro International Centre for Informatics, Bertinoro, Italy. October 21st - October 27th, 2007)
- Third VISIONTRAIN thematic school, 9 - 14 March 2008, Les Houches France

C Links to publications

- "Greedy-Based Design of Sparse Two-Stage SVMs for Fast Classification.":
  `http://www.springerlink.com/content/g261j85277208159/`
- "A Study of Parts-Based Object Class Detection Using Complete Graphs.":
  `http://www.springerlink.com/content/665n8x3k270137q2/?p=3328a8d70a654c11886e56b76f9fef4c&pi=0`
  `http://ipa.iwr.uni-heidelberg.de/ipabib/Papers/PartBasedObjectClassDetection08.pdf`

References

[1] Karina Zapien Arreola, Janis Fehr, and Hans Burkhardt. Fast support vector machine classification using linear svms. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, pages 366–369, Washington, DC, USA, 2006. IEEE Computer Society.

[2] Vojtěch Franc and Václav Hlaváč. Greedy algorithm for a training set reduction in the kernel methods. In Nikolai Petkov and Michel A. Westenberg, editors, *CAIP 2003: Computer Analysis of Images and Patterns*, pages 426–433, Berlin, Germany, August 2003. Springer.

[3] Bernd Heisele, Thomas Serre, Sam Prentice, and Tomaso Poggio. Hierarchical classification and feature reduction for fast face detection with support vector machines. *Pattern Recognition*, 36(9):2007–2017, 2003.

[4] S. Sathiya Keerthi, Olivier Chapelle, and Dennis DeCoste. Building support vector machines with reduced classifier complexity. *Journal of Machine Learning Research*, 7:1493–1515, 2006.

[5] E. Osuna and F. Girosi. Reducing the run-time complexity of support vector machines, 1998.

[6] Gunnar Raetsch. Robust boosting via convex optimization: Theory and applications.

[7] G. Rätsch. Benchmark data sets.
http://ida.first.fraunhofer.de/projects/bench/benchmarks.htm.

[8] Matthias Rätsch, Sami Romdhani, Gerd Teschke, and Thomas Vetter. Over-complete wavelet approximation of a support vector machine for efficient classification. In *DAGM-Symposium*, pages 351–360, 2005.

[9] S. Romdhani, P. Torr, B. Schölkopf, and A. Blake. Efficient face detection by a cascaded support-vector machine expansion. *Proceedings of The Royal Society A*, 460(2501):3283–3297, 2004.

[10] Hichem Sahbi and Donald Geman. A hierarchy of support vector machines for pattern detection. *Journal of Machine Learning Research*, 7:2087–2123, 2006.

[11] V. N. Vapnik. *Statistical Learning Theory.* John_Wiley, September 1998.

[12] Paul A. Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[13] M. Wu, B. Schölkopf, and G. Bakir. A direct method for building sparse kernel learning algorithms. *Journal of Machine Learning Research*, 7:603 – 624, 04 2006.