Discriminative Learning of Contour Fragments for Object Detection

Peter Kontschieder
kontschieder@icg.tugraz.at
Hayko Riemenschneider
hayko@icg.tugraz.at
Michael Donoser
donoser@icg.tugraz.at
Horst Bischof
bischof@icg.tugraz.at

Institute for Computer Graphics and Vision
Graz University of Technology
Austria

Abstract

The goal of this work is to discriminatively learn contour fragment descriptors for the task of object detection. Unlike previous methods that incorporate learning techniques only for object model generation or for verification after detection, we present a holistic object detection system using solely shape as underlying cue. In the learning phase, we interrelate local shape descriptions (fragments) of the object contour with the corresponding spatial location of the object centroid. We introduce a novel shape fragment descriptor that abstracts spatially connected edge points into a matrix consisting of angular relations between the points. Our proposed descriptor fulfills important properties like distinctiveness, robustness and insensitivity to clutter. During detection, we hypothesize object locations in a generalized Hough voting scheme. The back-projected votes from the fragments allow to approximately delineate the object contour. We evaluate our method *e.g.* on the well-known ETHZ shape data base, where we achieve an average detection score of 87.5% at 1.0 FPPI only from Hough voting, outperforming the highest scoring Hough voting approaches by almost 8%.

1 Introduction

Object localization in cluttered images is a big challenge in computer vision. Typical methods in this field learn an object category model, *e.g.* from a set of labeled training images and use this model to localize previously unseen category instances in novel images. The two dominating approaches in this field are either sliding window [9] or generalized Hough-voting [3] based. The final detection results are mostly returned as bounding boxes, highlighting the instance locations but some methods also return accurate object outlines.

In general, the detection approaches can additionally be divided into appearance and contour based methods. Appearance-based approaches first detect interest points and then extract strong image patch descriptors from the local neighborhoods of the detected points using versatile features like color, texture or gradient information. In contrast, contour-based methods exhibit interesting properties like invariance to illumination changes or variations in

BMVC 2011 http://dx.doi.org/10.5244/C.25.4

color or texture. It is also well-established in visual perception theory [5] that most humans are able to identify specific objects even from a limited number of contour fragments without considering any appearance information.

In this work we propose that also for a machine vision system, it is sufficient to solely rely on shape cues for the task of object detection, *i.e*. we are deliberately neglecting appearance information. In particular, we determine discriminative contour fragments of an object category shape in a state-of-the-art machine learning algorithm. Additionally, we show that the integration of local shape fragment descriptors into a generalized Hough voting scheme enables to outperform previous methods on challenging reference data bases like the ETHZ shape data base, without considering additional appearance based information.

1.1 Related Work

In this section, we recapitulate the main approaches addressing contour-based object category localization. In [10], Fergus *et al*. showed a learning approach which incorporates shape (besides appearance information) as a joint spatial layout distribution in a Bayesian setting for a limited number of shape parts. Ferrari *et al*. [11] partition image edges of the object model into groups of adjacent contour segments. For matching, they find paths through the segments which resemble the outline of the modeled categories. In a later approach [13] they investigate how to define contour segments in groups of k approximately straight adjacent segments (kAS) as well as how to learn a codebook of pairs of adjacent contours (PAS) from cropped training images [12] in combination with Hough-based center voting and non-rigid thin-plate spline matching.

Ravishankar *et al*. [26] use short line fragments, favoring curved segments over straight lines allowing certain articulations by splitting edges at high curvature points. In [18], Leordeanu *et al*. introduce a recognition system that models an object category using pairwise geometrical interactions of simple gradient features. The resulting category shape model is represented by a fully connected graph with its edges being an abstraction of the pairwise relationships. The detection task is formulated as a quadratic assignment problem. Shotton *et al*. [28] and Opelt *et al*. [24] simultaneously introduced similar recognition frameworks based on boosting contour-based features and clutter sensitive chamfer matching.

Lu *et al*. showed a method for fragment grouping in a particle filter formulation [20] and obtained consistent models for detection. Bai *et al*. [2] captured intra-class shape variations within a certain bandwidth by a closed contour object model called shape band. In [33], Zhu *et al*. formulated object detection as a many-to-many fragment matching problem. They utilize a contour grouping method to obtain long, salient matching candidates which are then compared using standard shape context descriptors. The large number of possible matchings is handled by encoding the shape descriptor algebraically in a linear form, where optimization is done by linear programming. In a follow-up work [29], Srinivasan *et al*. showed promising results when models are automatically obtained from training data instead of using a single category shape prototype. Again, their method relies on the availability of long, salient contours and has high complexity with detection times in the range of minutes per image. In [27], Riemenschneider *et al*. perform object detection by partially matching detected edges to a prototype contour in the test images. In such a way, piecewise contour approximations and error-prone matches between coarse shape descriptions at local interest points are avoided. Another approach proposed by Yarlagadda *et al*. [32] aims on grouping mutually dependent object parts of uniformly sampled probabilistic edges. In their Hough voting stage, object detection is formulated as optimization problem which groups dependent parts, correspondences between parts and object models and votes from groups to object hy-

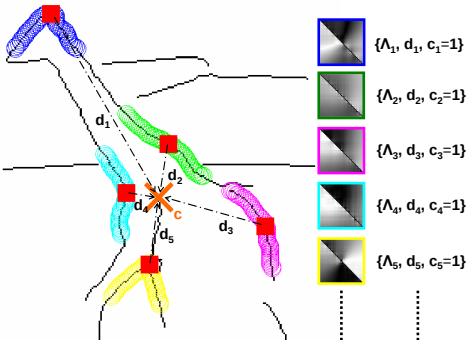

Figure 1: Illustration of our object category localization method. Local edge fragments are discriminatively trained in a Hough Forest analyzing a triple per fragment $\{\Lambda_i, \mathbf{d}_i, \mathbf{c}_i\}$, where Λ_i is our novel fragment descriptor matrix, $\mathbf{d}_i$ its corresponding center voting vector and $\mathbf{c}_i$ its class label. During testing descriptor matches vote for object centroid $\mathbf{c}$ to hypothesize object locations in the image. Best viewed in color.

potheses. Payet and Todorovic [25] presented an approach for object detection by mining repetitive spatial configurations of contours in unlabeled images. Their contours are represented as sequences of weighted beam angle histograms and are transferred into a graph of matching contours, whose maximum a posteriori multicoloring assignment is taken to represent the shapes of discovered objects. Finally, we refer to the approach of Amit and Geman in [1] where randomized trees are used to perform shape recognition for handwritten digits.

1.2 Contributions

As can be seen in the comprehensive summary of related work in Section 1.1, many approaches were proposed that exploit shape information for object detection. However, we observed that many researchers first put an enormous effort into learning a suitable object model from training data [12, 24, 26, 28] but then neglect to directly apply this gained knowledge in the matching or verification phase. Instead, techniques like error-prone chamfer matching are used to decide whether an object is present or not. We are convinced that a detection system benefits from an approach which inherently unifies the strengths of the individual methods instead of either solely relying on a particular technique or serially concatenating them. As a consequence, we propose to jointly learn a novel shape fragment description with its spatial location information about the object contour. For recognition, we can directly apply the learned knowledge in a generalized Hough voting manner.

A key issue of such an approach is to use a powerful local shape descriptor, which should fulfill various requirements like distinctiveness, robustness to clutter and noise, invariance and efficiency. Since common local shape descriptors are limited concerning these requirements, we propose a novel contour fragment descriptor which describes relative spatial arrangements of sampled points by means of angular information. As it is shown in the experiments in Section 3 our novel descriptor outperforms related methods like shape context in this scenario.

2 Discriminative Learning of Fragments

In this section we present our novel approach for object detection. As underlying representation we use connected and linked edges as it is described in Section 2.1. From the obtained edges we extract local fragment descriptions using a novel, discriminative descriptor that captures local angular information along edges (see Section 2.2). To learn a model from a set of labeled training images we train a Random Forest on the obtained descriptors storing the relative location and scale of the fragments with respect to the object centroid for the positive training samples, as it is explained in Section 2.3. At run-time, we cast probabilistic votes for possible center locations of the target objects in a generalized Hough voting manner. The resulting local maxima in the Hough space serve as detection hypotheses. In Figure 1 we illustrate our proposed method.

2.1 Edge Detection and Linking

As a first step we extract edges of the input image. We use the Berkeley edge detector [22] in all experiments and link the edges into oriented, connected point coordinate lists. We want to explicitly stress the importance of point linking which has great impact on the obtained fragments, since different splits of T-junctions or gap closing yields very different fragments. The obtained lists of connected points state the basis for all subsequent steps which is why we introduce the following terminology, used throughout the rest of this paper: We name all lists as *edges*, while we denote all extracted edge parts as *fragments*. In our method all analyzed fragments have the same length, consisting of exactly N points.

2.2 Shape Fragment Descriptor

To be able to discriminatively learn the shape of local, equal sized fragments, we need a powerful shape descriptor. Typical local fragment descriptors are shape context [4], turning angles [7], beam angles [25], partial contours [27] or contour flexibility [31]. We propose a novel descriptor which, as shown in the experiments in Section 3, outperforms related methods. Our shape descriptor is related to the recently proposed descriptor of Riemenschneider *et al*. [24] which also uses angles between fragment points. However, we want to stress crucial differences as follows: First, we employ another sampling strategy to define the angles and second, we differ in the selection of descriptor values. We analyze angles defined by lines connecting a reference point and the fragments' points. This is in contrast to [24] where only relative angles between points on the fragments are considered and in addition no fixed length of the fragment is assumed. Moreover, we are using a fragment-dependent reference point which is defined by the fragments' bounding box and therefore also contributes to a discriminative and local description of each considered fragment (see Fig. 3). Our sampling strategy was carefully designed for usage within a discriminative learning framework and is crucial for reasonable performance since otherwise we would not be able to distinguish between different locations on regularly shaped object parts as *e. g.* the semi-circles in Fig. 2.

We first outline the main definitions for extracting the fragment descriptors and then discuss the general properties of the obtained representation in detail. Let an individual *edge* B_i be defined as a sequence of linked points $B_i = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{M_i}\}$, where $M_i = |B_i|$ is the total number of connected edge points and $M_i \geq N$. We always analyze fragments of the same length N. Therefore, we compute $(M_i - N + 1)$ fragment descriptors for every individual edge B_i. For every *fragment* we define an $N \times N$ descriptor matrix Λ with $diag(\Lambda) = \mathbf{0}$. Every entry $\alpha_{ij} (i \neq j)$ is defined by the angle between a line connecting the points $\mathbf{b}_i$ and $\mathbf{b}_j$ and a line from $\mathbf{b}_j$ to a reference point $\mathbf{p}_0$, which is defined by the upper left corner of the

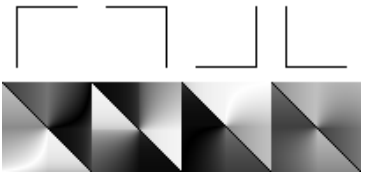

Figure 2: Selected contour fragment primitives (first row) and the corresponding, unique angular abstractions into the proposed fragment descriptor matrices (second row).

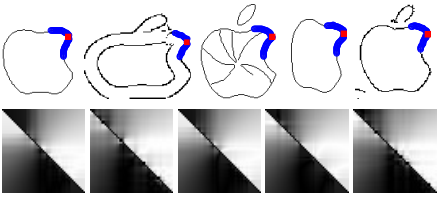

Figure 3: Mappings of intra-class variations for similar locations on object contours to descriptor matrices. Blue circles highlight the selected fragments and red squares indicate their center position.

fragments' surrounding bounding box. In such a way, we define

$$\alpha_{ij} = \sphericalangle(\overline{\mathbf{b}_i\mathbf{b}_j}, \overline{\mathbf{b}_j\mathbf{p}_0}) \quad \forall i,j = 1,\ldots,N \tag{1}$$

where $\mathbf{b}_i, \mathbf{b}_j$ are the i^{th} and j^{th} points on the respective fragment. Hence, we are mapping the fragment description onto the interval $[0,\pi]$. For a single fragment, the angles are calculated over all possible point combinations, yielding the descriptor matrix. Our proposed descriptor has a number of important properties which we discuss next.

Distinctiveness Most importantly, descriptors calculated from locations on the target object contour need to be discriminative to those generated from background data or clutter. Furthermore, we want the descriptors to be distinguishable from each other when they are calculated from the same object contour, but at different locations. This property is of particular interest when features are trained together with their spatial support. On the other hand, features that are computed at similar locations but from different training samples should result in similar representations. In other words, the descriptor should be able to capture intra-class variations and tolerate small perturbations in the training set. Our descriptor satisfies both requirements as can be seen in Figures 2 and 3.

Efficacy Another important property of our descriptor is the efficacy of data abstraction. Using a measure to describe relations between connected points has multiple advantages over considering each pixel independently. It assists to identify discriminative fragments during training and simultaneously reduces noise since information is encoded in a redundant way.

Invariance Features are often classified according to their level of invariance to certain geometrical transformations. For example, features invariant to Euclidean transforms keep unchanged after applying translation and rotation. Increasing the degree of invariance generally decreases the distinctiveness and thus weakens the distinctiveness property. Our proposed descriptor is invariant to translation and actively encodes orientational information, therefore relying on the so-called gravity assumption.

Efficiency

Features should be computable in an efficient manner. In our case, we can precompute the values for all possible pixel combinations, hence reducing actual feature composition to a lookup-operation which can be done in constant time for any fragment. Since the angular description in Equation (1) corresponds to a surjective mapping $f : I \subset \mathbb{R}^2 \rightarrow [0,\pi]$, we can provide a lookup-table (LUT) for all possible point pairs $(x,y) \in I$ as follows. The maximum Euclidean distance between the reference point $\mathbf{p}_0$ and any fragment point $\mathbf{b}_i$ is

bounded by the fragment length N. Hence, the area under the lower right quadrant with center $\mathbf{p}_0$ corresponds to the total number of possible fragment point locations and defines I. Consequently, we can precompute a square matrix of size $|I| \times |I|$ holding the angular descriptions according to Equation (1) for all tuples $(x,y) \in I$. This matrix then serves as LUT during feature calculation for arbitrary fragment points.

2.3 Discriminative Fragment Learning

For discriminatively learning our proposed contour fragment descriptor, we use the recently introduced Hough Forest [15]. Hough Forests are an extension of the standard random forest [6] which in addition to finding optimal splits of the training data also optimize grouping of class specific object center votes. In such a way, a test sample is classified according to its appearance as well as its localization properties.

Each tree is constructed on a set of training samples $\{\mathscr{P}_i = (\Lambda_i, c_i, \mathbf{d}_i)\}$, where Λ_i is the fragment descriptor matrix, c_i its corresponding class label and $\mathbf{d}_i$ the offset vector, describing the displacement to the center for positive training samples (see Fig. 1). Positive samples are selected by taking all edges within the bounding box annotations. The binary tests are chosen according to [15], randomly reducing either class label or centroid offset uncertainty.

Once the entire Forest is constructed, the detection process can be started on the test images. Edges are extracted from the test images and arranged into ordered, connected edge lists B_i with $|B_i| \geq N$. For each B_i, again a total number of $(|B_i| - N + 1)$ fragment descriptors $\{\Lambda_j\}$ are computed and then classified into tree-specific leaf nodes. Please note that the descriptors are only computed along edges, which significantly reduces the computational costs in comparison to a sliding window approach or dense sampling as used in [15].

Since the voting vectors D_L and class label probabilities C_L are known in every leaf node, we are able to cast the voting vectors into the Hough image V in an accumulative way. All pixel locations in V are incremented with the value $\frac{C_L}{|D_L|}$. Finally, the resulting Hough image is Gauss-filtered and its local maxima hypothesize the detected object centroids.

2.4 Ranking and Verification

The previous stages of our method provide object hypotheses in the test image and a corresponding score obtained from the Hough votes. Similar to related work [21, 23, 27] we additionally provide a ranking according to a pyramid matching kernel (PMK) [16] where histograms of oriented gradients (HOG) are used as features. The PMK classifier is trained on the same training examples as used for the Hough forest. We use the classifier for ranking and for verification where we additionally consider nearby locations and scales around the proposed hypotheses. Including the local search is still efficient since our hypotheses generation stage delivers only a few hypotheses per image, therefore an order of magnitude fewer candidates have to be considered as in a sliding window approach.

3 Experimental Evaluation

In order to demonstrate the quality of our proposed method, we performed several experiments: First, we demonstrate improved performance of our novel fragment descriptor in comparison to several related shape descriptors as shown in Section 3.1. Subsequently, we show the performance of our method on challenging data bases like the ETHZ shape data base [11] and the INRIA horses data base [17] (Section 3.2). We compare our results to several competitive contour-based recognition approaches while we deliberately ignore methods additionally using segmentations or appearance information as in [19, 30].

length	CA	BA	PC	TA	SC	Ours
l=51	41.67	41.61	41.13	40.49	37.13	**33.60**
l=41	43.15	42.67	42.68	42.14	37.51	**35.93**
l=31	43.43	43.48	42.99	42.81	**38.33**	38.58

Table 1: Evaluation of descriptor performances at several lengths, showing the per-pixel classification error in % (see text for definition) for each descriptor when learned in a random forest. Our descriptor yields to the classification error of 33.60% for a length of 51, significantly lower than all compared state-of-the-art shape descriptors.

3.1 Fragment Descriptor Evaluation

Due to the large variety of shape-based descriptors, it is vital to evaluate our proposed descriptor in a quantitative experiment. Therefore, we designed an experiment to evaluate different shape descriptors within different learning algorithms for a classification task. In particular, we compared our proposed descriptor to five types of descriptors, namely the chord angle (CA) [8], beam angle (BA) [25], partial contours (PC) [27], turning angle (TA) [7] and shape context (SC) [4]. The learning algorithms we used were random forest, linear SVM and a simple nearest neighbor classifier.

We define a test setup where we classify individual edge pixels on Berkeley edges, detected in the giraffe category test images of the ETHZ shape data base. Specifically, we compare the per-pixel classification results to the ground-truth edge annotations. Hence, we define a classification error, describing the ratio of false classifications to all edge pixels. The protocol for the experiment uses 50% of the images for learning a classifier and 50% for testing. We extracted fragments (or quadratic patches containing edges for SC) at varying sizes from the images, such that for all images a reasonable number of foreground edges remained in the test set.

In Table 1 we list some selected results for fragment lengths / patch sizes of 31, 41 and 51 pixels, when learning the individual descriptors in a random forest framework. As shown, our proposed descriptor outperforms all of the other descriptors at length $l = 51$ and is hence well suited for use in a discriminative setting. Please note that increasing the length even more may result in better classification scores, however, the number of edges belonging to the object category contour might decrease. Using linear SVM or nearest neighbors for classification results in approximately similar distributions of the scores. However, the mean error is on average about $5-10\%$ higher which suggests that random forest like classifiers are better suited for our task.

3.2 Object Detection

Object detection performance is evaluated on the ETHZ shape data base [11] and the INRIA horses data base [17]. For all our experiments, we use the following parameters. Our forest consists of 12 decision trees, each with a depth of 15. The fragment length is fixed to $N = 51$, as suggested from the classification task in the previous section. We randomly extract 10000 positive training samples from edges within the bounding box annotation. The positive training images are all scaled to the median height of the selected training data base and the aspect ratio is fixed. 10000 negative training samples are extracted from the same training images, but outside of the bounding boxes. Detection performance is evaluated using the strict PASCAL 50% criterion.

Since our method is not implicitly scale invariant, we run the detector on multiple scales. However, this can be done efficiently since descriptor calculation takes constant time due to

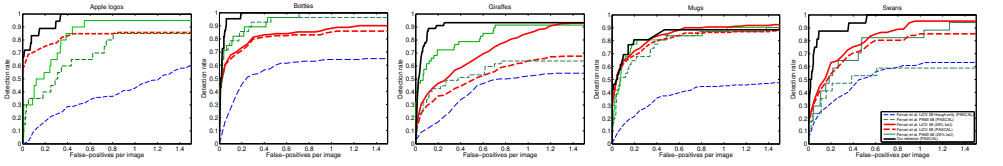

Figure 4: Object detection performance on ETHZ in comparison to [13, 14]. Each plot shows curves for the 50% Pascal criterion and the 20%-IoU criterion of the methods proposed in [13, 14]. Our results for 50% PASCAL criterion are shown in thick solid black. Note, that we mostly outperform results of [13, 14] consistently over all classes although evaluating under the stricter 50% PASCAL criterion.

the use of Look-Up-Tables and traversing the trees has logarithmic complexity. In practice, the average evaluation time is restricted to a few seconds per image for our *C* implementation.

ETHZ shape data base The ETHZ shape data base consists of five object classes and a total of 255 images. The images contain at least one and sometimes multiple instances of a class and have a large amount of background clutter. All classes contain significant intra-class variations and scale changes. Therefore, we run the detector on 15 different scales, equally distributed between factors of 0.2 and 1.6. We use the same test protocol as specified in [14] where a class model is learned by training on half of the positive examples from a class, while testing is done on all remaining images from the entire data base.

In Table 2 we list the results of our described object detector for each object class in comparison to current state of the art [21, 23, 27, 32] where divisions into voting, ranking and verification stages are applicable. However, due to the large number of competing methods, we only provide the scores of the initial Hough *voting* stage and the PMK *ranking* stage in tabular form. Recognition performance is evaluated by ranking the hypotheses according to their confidence scores. In the initial voting stage this confidence corresponds to the accumulated values in the Hough space. For ranking, the confidence scores are updated using the HOG-based verification as described in Section 2.4. Both *voting* and *ranking* are evaluated at 1.0 FPPI. Ranking is quite efficient since on average only 3.5(!) hypotheses are returned by our method. Finally, we also show results of our method for the full verification step (where also nearby locations and scales are tested around the returned hypotheses) at FPPI $= 0.3/0.4$, which is the standard measure for comparing results on ETHZ data base.

As can be seen in Table 2, we substantially outperform the currently best scoring methods after both, Hough-voting and ranking stage. We achieve a performance boost of **7.5%** over the previously best voting method in [32], 18.1% over [27], 26.6% over [23], 24.5% over [21] and even 34.2% over [14]. After applying the learned HOG models for ranking we are 11.0% better than the previously best scoring method [27] and 15.6% better than [23] ([32] has no scores for ranking). Finally, our method also shows high performance for the full verification system, providing an average recognition score of $93.3/96.1$ at $0.3/0.4$ FPPI, which is approximately on par with the highest scores reported from contour-based approaches in [29] (95.2/95.6). However, the authors in [29] do not provide scores for detection and refinement stages individually which makes direct comparison difficult. Furthermore, their method has a high computational complexity and detection takes minutes per image.

To further emphasize the contribution of our proposed descriptor with respect to the

	Voting Stage (FPPI=1.0)							*Ranking Stage (FPPI=1.0)*			*(FPPI=0.3/0.4)*
ETHZ Classes	Hough [14]	M^2HT [21]	w_{ac} [23]	PC [27]	Group [33]	Hough Forest [15]	**Our work**	[23] + PMK	[27] + PMK	**Ours + PMK**	**Our work Verification**
Apples	43.0	80.0	85.0	90.4	84.0	80.0	94.4	80.0	90.4	100.0	94.4/100
Bottles	64.4	92.4	67.0	84.4	93.1	70.8	90.9	89.3	96.4	95.5	100/100
Giraffes	52.2	36.2	55.0	50.0	79.5	60.5	86.7	80.9	78.8	93.3	91.1/93.3
Mugs	45.1	47.5	55.0	32.3	67.0	73.1	92.3	74.2	61.4	88.5	80.8/87.2
Swans	62.0	58.8	42.5	90.1	76.6	81.3	73.3	68.6	88.6	93.3	100/100
Average	53.3	63.0	60.9	69.4	80.0	73.1	**87.5**	78.6	83.2	**94.2**	**93.3/96.1**

Table 2: Hypothesis voting and ranking showing detection rates (using PASCAL50 criterion) for the ETHZ shape data base [11]. For the voting stage our coverage score increases the performance by **7.5%** [32], 18.1% [27], 24.5% [21], 26.6% [23] and 34.2% [14]. After ranking we achieve an improvement of **11.0%** over [27] and 15.6% over [23].

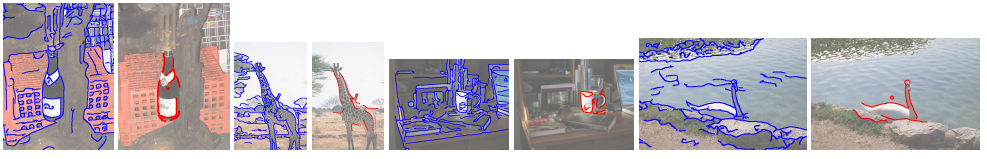

Figure 5: Examples for successful object localizations for some classes of ETHZ. The highly cluttered edge responses (in blue) and the reprojected fragments (in red) for the object hypothesis with the highest confidence per image are shown. Best viewed in color.

Hough Forest learning environment, we trained and evaluated on the same number of trees using all 32 provided features of [15] (mostly appearance-based). To have a fair comparison and accomplish each of the 15 considered scales in reasonable time, we evaluated on the same locations as we did for our descriptor. As shown in Table 2, our single feature descriptor clearly outperforms the standard Hough Forest using all 32 features (14.4% better).

In Figure 4 we show the detection rate vs. FPPI plots for all ETHZ classes in comparison to all results provided by Ferrari *et al.* [13, 14]. Figure 5 illustrates some results for different classes. We show the highly cluttered edge responses of the test images used for localization and the corresponding reprojections of the classified fragments for an object hypothesis. In addition, they can be used to approximately delineate the object contour. The inpainted circles indicate the voting centers.

INRIA horses data base The INRIA horses data base [17] contains a total number of 340 images where 170 images belong to the positive class showing at least one horse in side-view at several scales and 170 images without horses. The experimental setup is chosen as in [14] where the first 50 positive examples are used for training and the rest of the images are used for evaluation $(120+170)$. We run the detector on 8 different scale factors between 0.5 and 1.5. We achieve a competitive detection performance of **85.50%** at 1.0 FPPI, compared to recently presented scores in [21] (85.3%) and [32] (87.3%). Moreover, we outperform the methods in [27] (83.72%), [13] (80.77%) and [17] (73.75%).

4 Conclusion

In this paper we investigated the use of contour fragment descriptors for the task of object detection. Our novel method discriminatively learns a number of local contour fragment descriptors in combination with their spatial location relative to object centroid in a Random

Forest classifier. We designed a fragment descriptor that abstracts spatially connected edge points into angular relations in a matrix form and demonstrated that our proposed descriptor shows distinctive patterns for differently shaped fragment primitives, while tolerating small perturbations and intra-class variabilities. Experiments demonstrated that the proposed descriptor significantly outperforms related shape descriptors. We further showed excellent results on the well-known ETHZ and INRIA horses data bases. For example, our method outperforms the currently highest scoring contour based methods by approximately 8% at 1.0 FPPI after the Hough voting stage. In addition, we demonstrated that back-projections of the fragments voting for a hypothesis allows delineating the object outline. Our future work will focus on turning this information into concise image segmentations.

Acknowledgements We acknowledge the financial support of the Austrian Science Fund (FWF) from project Fibermorph (P22261-N22), the Research Studios Austria Project μSTRUCSCOP (818651) and the Austrian Research Promotion Agency (FFG) project FIT-IT CityFit (815971/14472-GLE/ROD).

References

[1] Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural Computation*, 9, 1996.

[2] X. Bai, Q. Li, L. J. Latecki, W. Liu, and Z. Tu. Shape band: A deformable object detection approach. In *(CVPR)*, 2009.

[3] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2), 1981.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *(PAMI)*, 2002.

[5] I. Biederman and G. Ju. Surface vs. edge-based determinants of visual recognition. *Cognitive Psychology*, 20:38–64, 1988.

[6] L. Breiman. Random forests. In *Machine Learning*, pages 5–32, 2001.

[7] L.B. Chen, R.S. Feris, and M. Turk. Efficient partial shape matching using smith-waterman algorithm. In *Proc. of NORDIA workshop at CVPR*, 2008.

[8] M. Donoser, H. Riemenschneider, and H. Bischof. Efficient partial shape matching of outer contours. In *(ACCV)*, 2009.

[9] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *(PAMI)*, 2010.

[10] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *(CVPR)*, 2003.

[11] V. Ferrari, T. Tuytelaars, and L. Van Gool. Object detection by contour segment networks. In *(ECCV)*, 2006.

[12] V. Ferrari, F. Jurie, and C. Schmid. Accurate object detections with deformable shape models learnt from images. In *(CVPR)*, 2007.

[13] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. *(PAMI)*, 2008.

[14] V. Ferrari, F. Jurie, and C. Schmid. From images to shape models for object detection. *(IJCV)*, 2009.

[15] J. Gall and V. Lempitsky. Class-specific hough forests for object detection. In *(CVPR)*, 2009.

[16] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In *(ICCV)*, 2005.

[17] F. Jurie and C. Schmid. Scale-invariant shape features for recognition of object categories. In *(CVPR)*, 2004.

[18] M. Leordeanu, M. Hebert, and R. Sukthankar. Beyond local appearance: Category recognition from pairwise interactions of simple features. In *(CVPR)*, 2007.

[19] F. Li, J. Carreira, and C. Sminchisescu. Object recognition as ranking holistic figure-ground hypotheses. In *(CVPR)*, 2010.

[20] C. Lu, L. J. Latecki, N. Adluru, X. Yang, and H. Ling. Shape guided contour grouping with particle filters. In *(ICCV)*, 2009.

[21] S. Maji and J. Malik. Object detection using a max-margin hough transform. In *(CVPR)*, 2009.

[22] D. R. Martin, Ch. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *(PAMI)*, 2004.

[23] B. Ommer and J. Malik. Multi-scale object detection by clustering lines. In *(ICCV)*, 2009.

[24] A. Opelt, A. Pinz, and A. Zisserman. A boundary-fragment-model for object detection. In *(ECCV)*, 2006.

[25] N. Payet and S. Todorovic. From a set of shapes to object discovery. In *(ECCV)*, 2010.

[26] S. Ravishankar, A. Jain, and A. Mittal. Multi-stage contour based detection of deformable objects. In *(ECCV)*, 2008.

[27] H. Riemenschneider, M. Donoser, and H. Bischof. Using partial edge contour matches for efficient object category localization. In *(ECCV)*, 2010.

[28] J. Shotton, A. Blake, and R. Cipolla. Contour-based learning for object detection. In *(ICCV)*, 2005.

[29] P. Srinivasan, Q. Zhu, and J. Shi. Many-to-one contour matching for describing and discriminating object shape. In *(CVPR)*, 2010.

[30] A. Toshev, B. Taskar, and K. Daniilidis. Object detection via boundary structure segmentation. In *(CVPR)*, 2010.

[31] C. Xu, J. Liu, and X. Tang. 2D shape matching by contour flexibility. *(PAMI)*, 2009.

[32] P. Yarlagadda, A. Monroy, and B. Ommer. Voting by grouping dependent parts. In *(ECCV)*, 2010.

[33] Q. Zhu, L. Wang, Y. Wu, and J. Shi. Contour context selection for object detection: A set-to-set contour matching approach. In *(ECCV)*, 2008.