BEYOND STRAIGHT LINES - OBJECT DETECTION USING CURVATURE

Antonio Monroy, Angela Eigenstetter and Björn Ommer

Interdisciplinary Center for Scientific Computing, University of Heidelberg, Germany
{amonroy,aeigenst,bommer}@iwr.uni-heidelberg.de

ABSTRACT

We present an approach that directly uses curvature cues in a discriminative way to perform object recognition. We show that integrating curvature information substantially improves detection results over descriptors that solely rely upon histograms of orientated gradients (HoG). The proposed approach is generic in that it can be easily integrated into state-of-the-art object detection systems. Results on two challenging datasets are presented: ETHZ Shape Dataset and INRIA horses Dataset, improving state-of the-art results using HoG by 7.6% and 12.3% in average precision (AP), respectively. In particular, we achieve higher recall at lower false positive rates.

1. INTRODUCTION

Visual object detection in cluttered scenes is one of the key problems of computer vision. Localizing all instances of an object category is highly challenging due to the large intra-class variability. Finding a common model for all the widely diverse class instances thus poses a major difficulty. To yield robust, powerful object representations, the vision community has now broadly adopted the theme of gradient histograms: Almost, all present approaches, ranging from semi-local descriptors such as SIFT [1] to holistic object representations [2, 3, 4], are based on histograms of local gradient orientation. In effect, this results in a straight line approximation of object boundaries since local regions are described by a histogram over a discrete set of edge orientations that they contain. In this framework, a smooth curve cannot be distinguished from one with sharp bends or from a set of differently oriented lines in arbitrary configuration as can be seen in Fig. 1. Moreover, natural objects are actually not existing in a blocks-world domain [5] and have not been designed with a ruler on a drawing table. Instead they do exhibit characteristically curved boundaries, e.g., consider the differences between apples and pears. Thus, we extend the widely used object representation based on gradient orientation histograms by incorporating a robust description of curvature. Furthermore, the importance of curvature for visual search tasks in human perception has been confirmed in different studies within the perception community. In his review [6], Wolfe follows the consensus that there appears to
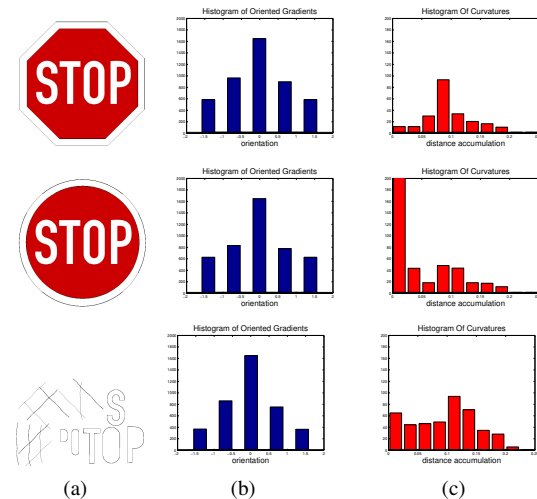
Fig. 1. (a) Original images, (b) Histograms of oriented gradients, (c) Histograms of Curvature. A smooth curve cannot be distinguished from one with corners or from a set of differently oriented lines in an arbitrary configuration based only on histograms of oriented gradients.

be about eight to ten basic features that play an important role for visual search tasks: color, orientation, motion, size, curvature, depth, vernier offset, gloss and, perhaps, intersection and spatial position/phase. Features like color and orientation, have reached maturity also within the computer vision community leading to powerful detectors (e.g [2, 7, 8]), while others like curvature have not received the same level of attention in the current work on object detection.

In this paper, we describe an object detection system that efficiently represents an object shape using both, orientation and curvature features. Our experiments on the two well known datasets, ETHZ Shape Dataset and INRIA horses Dataset, show that our combined descriptor improves performance significantly over state-of-the-art IKSVM ([7]) detector that solely uses histograms of oriented gradients (HoG). These results clearly confirm what our experience tells us: we live in a curved world and we need curvature to describe objects.

2. PREVIOUS WORK

In the perception community exists an extensive body of work on the importance of curvature stimuli for visual search tasks [6]. Moreover, the estimation of curvature in images has been studied in depth and several methods have been proposed. They can be classified in three groups according to which

Fig. 2. Examples of local curvature approximation used by our descriptor on the ETHZ Shape Dataset.

definition of curvature they are using: tangent direction, osculating circle and derivation [9]. Recently, [10] proposed an efficient, new approach to approximate discrete curvature at a given point p by means of the accumulation of euclidean distances from different secant lines to the point p. This method proved to be more stable compared to the curvature-space method [11], where a boundary is represented as a parametric function of arc length, and inflection points are detected as stable zero-crossing points over convolution of the shape with Gaussian filters at different σ levels. Many methods using curvature information for finding interest points (e.g. high-curvature points) have been proposed in the literature, most recently in [12] and [13]. However, the direct use of curvature information for building object descriptors has seen comparably little progress. The early approach of [11] works for object recognition under the assumption of closed boundaries. In medical imaging curvature statistics are utilized to capture local changes away from the mean curvature on spherical 3D constructions of tumors to discriminate different types of cancer [14]. However, this approach is not suited for general object recognition, since the approach describes global curvature distribution on spherical objects, but does not capture the local distribution of curvature, that is needed to describe the general shape of an arbitrary object.

In contrast to this, modern descriptors like k-AS [15] explicitly decide not to take curvature into account and rather consider the segments as completely straight segments so as to capture only the relevant information of the geometric configuration they form. Moreover, most detection systems like [2] or more recent extensions of this approach [7, 16, 8] solely encode orientation of gradients in form of histograms.

Our approach directly encodes curvature and uses this shape cue together with orientation of gradients to perform object detection. The used curvature representation is able to capture the shape information of complex objects, by directly using curvature to build up local and global histograms of curvature, without making any assumptions about their shape. This work shows that (i) curvature information can be integrated effortlessly into all state-of-th-art object representations that are based on gradient histograms. Moreover, (ii) the representation has low computational cost and, most importantly, (iii) it provides complementary object information that significantly enriches the widely used orientation histograms.

3. ROBUST REPRESENTATION OF CURVATURE

In this section we describe a method to perform object detection based on curvature information from shapes and use this information directly as a discriminative feature together with
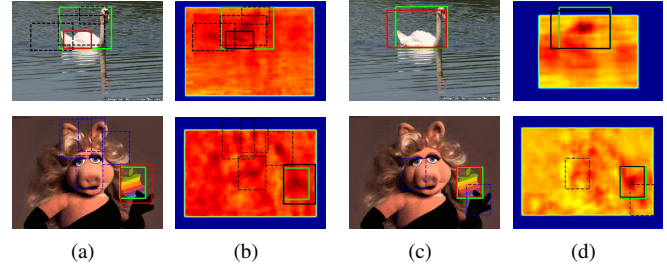

Fig. 3. Detection results using standard HoG (implementation of [7]) (first two columns) and results using HoGC (last two columns). First detection is shown in red and false positives in dashed black. These examples illustrate a general finding in this database that compared to the widely used HoG, our proposed representation yields a better localization of the maxima compared to ground-truth and generation of less false-positives.

histograms of oriented gradients (HoG) [2]. We abbreviate the joint descriptor with **HoGC**.

A very fast and stable way to approximate the curvature for planar boundaries is to use the chord-to-point distance accumulation (or distance accumulation) [10]. Let B be a set of N consecutive boundary points, $B := \{p_0, p_1, p_2, \cdots, p_{N-1}\}$. The set of points is obtained by following the edge contours of objects in clockwise direction. Each pair of points p_i and p_{i+l} define a line L_i, where $i + l$ is taken modulo N. L_i depends on the parameter l whose adjustment is explained later in this section. For each point p_k the perpendicular distance D_{ik} from the line L_i is computed using the euclidean distance. The distance accumulation for a point p_k and a chord length l is the sum

$$h_l(k) = \sum_{i=k-l}^{k} D_{ik}. \qquad (1)$$

[10] showed that equation (1) is more stable, regardless of different values of l, than in Gaussian smoothing curvature calculation methods, which give dislocation, broadening and flattening of the features ([19]). Furthermore, it was shown that the chord-to-point distance accumulation asymptotically approximates (up to a constant) the true curvature of the boundary.

Given an image, we first extract edges using the Berkeley edge detector [20]. Connected components on the binarized edge map yield a set of segments B_j. On these segments we calculate the distance accumulation given in equation (1). To be robust against the choice of l we choose a bank of values $\{l_1, \cdots, l_n\}$ ranging between 5 and 40 pixels and take for every point p_i on segment B_j the median

$$c_j(p_i) := \text{median} \left\{ \frac{h_{l_s}(i)}{l_s^3} \ \middle| \ s = 1, \cdots, n \right\} \qquad (2)$$

as boundary feature. In Fig. 2 we show some examples of the curvature on natural images.

The idea behind the HoG descriptor of Dalal and Triggs [2] is that local statistics about intensity and orientation of gradients can encode the appearance and shape of objects. Curva-

		Average Precision			Recall @ 0.3/0.4/(1) FPPI					
		Curv.	HoG	HoGC	Curv.	HoG	IKSVM [7]	Voting [17]	DSM [18]	HoGC
ETHZ Shape	Applelogos	72.3	86.7	92.5	86.3/91.2	90.0/90.0	90.0/90.0	90.6±6.2/-	95.0/95.0	100/100
	Bottles	72.0	79.0	88.4	92.8/96.4	96.3/96.3	96.4/96.4	94.8±3.6/-	100/100	96.4/96.4
	Giraffes	31.0	56.0	60.1	43.0/43.0	72.3/78.7	79.1/83.3	79.8±1.8/-	87.2/89.6	74.4/85.1
	Mugs	34.1	71.2	82.2	54.8/54.8	87.1/87.1	83.9/83.9	83.2±5.5/-	93.6/93.6	90.3/93.5
	Swans	50.2	59.4	66.9	76.4/76.4	82.3/82.3	88.2/88.2	86.8±8.9/-	100/100	94.1/94.1
	Average	52.1	**70.4**	**78.0**	70.6/72.3	**85.6/86.8**	87.5/88.4	87.1±2.8/-	95.2/95.6	**91.0/93.8**
INRIA horses		52.2	**71.3**	**83.6**	53.2/56.5/72.8	**81.5/82.6/91.3**	-/-/86.0	-/-/-	-/-/-	**90.2/90.2/94.5**

Table 1. We compare the performance of the HoGC against the state-of-the-art detector IKSVM [7] for the **ETHZ Shape Dataset** and the **INRIA horses dataset**. We follow the standard setup of HoG and search over location and scale, but not over aspect ratios. Thus explains the performance gap between our HoG and IKSVM [7] on ETHZ. [18] deviate from HoG by adding a computationally costly part-based model .

ture information of shapes can be encoded in a similar way. We divide the image into connected cells and for each cell we build a 1D histogram of curvature information. For this, we discretized the values $c_j(p_i)$ from Eq.(2). Each pixel then casts a vote proportional to the gradient magnitude. Following a "soft binning" approach, it also contributes to the histograms in the four cells around it using bilinear interpolation. In practice, to calculate both, the histograms of oriented gradients and histograms of curvature, the image is divided into grids of increasing resolutions for 4 levels, and histograms from each level are weighted according to $w = 2^{l-1}$, where $l = 1$ is the coarsest scale and the histograms are concatenated together to form a feature vector that encodes local and global curvature statistics of the image. The range of values from Eq.(2) is subdivided into 10 equally sized bins.

Learning the model

Because of the histogram-nature of the feature vectors, we use a SVM with histogram intersection kernel ([4]) as classifier. [21] proposed an approximation method for Intersection Kernel SVM, which essentially reduces the runtime of the classifier to that of a linear SVM. We train our model with an initial, randomly picked subset of negative examples and then collect negative examples that are incorrectly classified by the initial model. A new model is trained using the new negative examples and the support vectors from the old model. We repeat this procedure three times. To detect an object instance the classifier is run in sliding window mode over different location and scale. Note, that using this setting, curvature has not to be scale invariant to be used as a descriptor since the curvature computation is performed for different sizes of the sliding window, i.e. curvature is computed on different scales during detection.

4. EXPERIMENTAL RESULTS

Objective of our experiments is to show that the direct use of curvature as a feature yields orthogonal shape information that helps to improve object detection results. Quantitatively this means that the use of our combined object descriptor should yield a higher average precision and a lower false positive rate for the same recall over the HoG descriptor using the same implementation.

We report our results on two challenging datasets: the ETHZ Shape Dataset and INRIA horses. The ETHZ Shape Dataset contains 255 images belonging to five different classes. We

follow the standard experimental protocol for creating training and test sets. The INRIA horses dataset consists of 170 images containing one or more side-viewed horses and 170 images without horses. 50 horse images and 50 negative images are used for training and the remaining 120 horse images plus 120 negative images are used for testing. In our experiments we are following the standard PASCAL setting for counting true positives and false positives among the predicted bounding boxes. In table 1 we compare the performance of our approach with several state-of-the-art detector systems [7, 17, 18] at 0.3, 0.4 and (for INRIA horses) 1 FPPI. Our HoG baseline implementation uses HoG and IKSVM, like the currently best reported results of a HoG based detection system on ETHZ [7]. Note, that [7] searched over different aspect ratios for some categories in the ETHZ Shape Dataset (e.g. Giraffes and Mugs). This explains the differences in the baseline results (HoG vs. IKSVM). Our final detector HoGC clearly improves performance over the baseline HoG detection system on both datasets. Furthermore, our approach outperforms the voting approach suggested in [17]. In addition we compared our detection system with the descriptive shape model (DSM) suggested in [18]. This approach performs slightly better than our HoGC descriptor on the ETHZ Shape dataset since it adds also a deformable part model to the holistic approach. As reported in [22] the average performance improves about 8% on PASCAL VOC 2007 when adding part-based HoG descriptors. However, we decided, for a fair comparison with HoG implementations, to use the standard setting without parts. Furthermore, detection takes several minutes per image using the descriptive shape model, whereas using HoGC is one order of magnitude faster. Figures 4 and 5 compare our approach with the state-of-the-art HoG detector. We remark that [7] did not include FPPI or precision-recall curves for his IKSVM + HoG detector for the ETHZ Shape Dataset. By incorporating curvature information, our combined HoGC representation outperforms HoG results in all categories of the ETHZ Shape Dataset and on INRIA horses. We achieve an average gain of 7.6% in AP on the ETHZ Shape Dataset and of 12.3% on INRIA horses. For the ETHZ Shape Dataset we get in average 5.4% higher detection rate at 0.3 FPPI and at 0.4 FPPI an improvement of 7%. On INRIA horses we improved the recall by 8.7% at 0.3
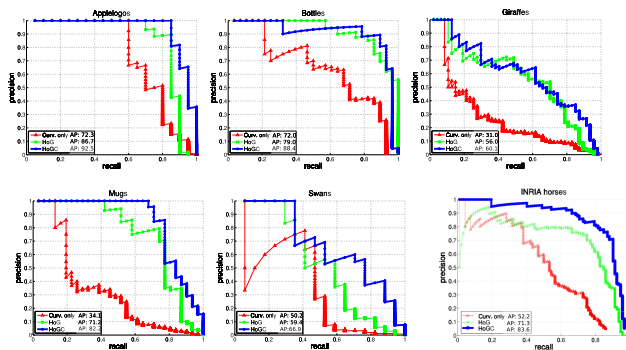
Fig. 4. Precision Recall Curves for ETHZ Shape Dataset and IN-RIA horses comparing curvature only (red), HoG (green) and HoGC (blue).
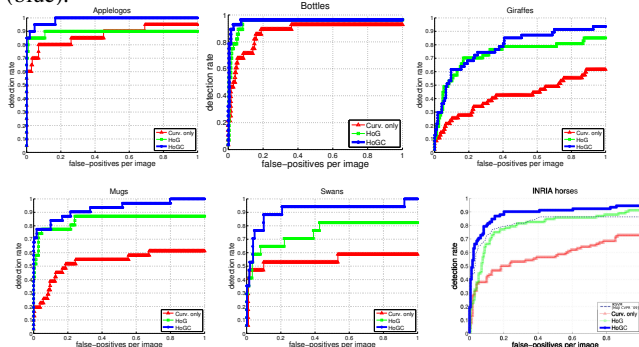

Fig. 5. Detection performance against FPPI for the ETHZ Shape Dataset and INRIA horses comparing curvature only (red), HoG (green) and HoGC (blue).

FPPI, 7.6% at 0.4 FPPI and 3.2% at 1 FPPI. For the sake of completeness we also included detection results of our system solely using curvature information. However, the suggested curvature feature was never intended to be used in solitude and for that reason does not contain redundant information to the HoG descriptor, like the orientation of curvature. That explains the drop in performance when using curvature without HoG while the combination of both significantly improves state-of-the-art HoG object detection methods. These results approve our initial hypothesis that curvature is a complimentary feature to HoG.

5. CONCLUSION

The main contribution of this work is to provide quantitative evidence that curvature information of objects can be discriminatively used in a robust and reliably manner for object recognition. Our results show that the use of curvature information yields orthogonal information to the state-of-the-art theme of histograms of oriented gradients for visual search tasks. Combining both leads to better accuracy and performance on standard datasets and significantly improves state-of-the-art detection system solely based on HoG. The proposed curvature-based object representation is generic, efficient to compute, and it can be effortlessly integrated into all current object models that utilize histograms of gradients. Thus a wide applicability is automatically granted.

6. REFERENCES

[1] D.G. Lowe, "Object recognition from local scale-invariant features," *ICCV*, 1999.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *CVPR*, 2005.

[3] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," *CVPR*, 2006.

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

[5] J. Slaney and S. Thiébaux, "Blocks world revisted," *Artif. Intell.*, vol. 125, 2001.

[6] J. M. Wolfe, "Visual search," *Attention*, 1998.

[7] S. Maji and J. Malik, "Object detection using a max-margin hough transform," *CVPR*, 2009.

[8] P. Felzenszwalb, D. McAllester, and D. Ramaman, "A discriminatively trained, multiscale, deformable part model," *CVPR*, 2008.

[9] T. Lewiner, J. D. Gomes Jr., H. Lopes, and M. Craizer, "Arc-length based curvature estimator," *SIBGRAPI*, 2004.

[10] J.H. Han and T. Poston, "Chord-to-point distance accumulation and planar curvature: a new approach to discrete curvature," *Pattern Recogn. Lett.*, vol. 22, 2001.

[11] F. Mockhtarian and A. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *PAMI*, vol. 8, 1986.

[12] X. He and N. Yung, "Corner detector based on global and local curvature properties," *Optical Engineering*, vol. 47, 2008.

[13] M. Awrangjeb and G. Lu, "Corner detection based on the chord-to-point distance accumulation technique," *Trans. on Mult.*, 2008.

[14] M. G. Linguraru, S. Wang, F. Shah, R. Gautam, J. Peterson, W. M. Linehan, and R. M. Summers, "Computer-aided renal cancer quantification and classification from contrast-enhanced CT via histograms of curvature-related features," *EMBC*, 2009.

[15] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of adjacent contour segments for object detection," *PAMI*, vol. 30, 2008.

[16] P. Yarlagadda, A. Monroy, and B. Ommer, "Voting by grouping dependent parts," *ECCV*, 2010.

[17] C. Gu, J. J. Lim, P. Arbeláez, and J. Malik, "Recognition using regions," *CVPR*, 2009.

[18] P. Srinivasan, Q. Zhu, and J. Shi, "Manty-to-one contour matching for describing and discriminating object shape," *CVPR*, 2010.

[19] A.P. Witkin, "Scale-space filtering," *IJCAI*, 1983.

[20] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *PAMI*, vol. 26, 2004.

[21] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," *CVPR*, 2008.

[22] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object ?," *CVPR*, 2010.