

Contour-based Object Detection

Joseph Schlecht
schlecht@uni-heidelberg.de
Björn Ommer
ommer@uni-heidelberg.de

Interdisciplinary Center for Scientific Computing
University of Heidelberg
Germany

Contour-based representations have a long history in object recognition and computer vision. Considerable effort was spent in the past matching geometric shape models of objects to image contours [4, 5, 8]. Although these approaches enjoyed some success, it is clear that finding contours exactly belonging to the shape of an object is a hard problem. This insight has given rise to an emphasis on local texture descriptors, the dominant approach today. These appearance-based descriptors summarize texture information in the form of histograms of gradients [6], shape context [2], geometric blur [3], and many others. While prominent edges are roughly encoded, exact shape location is replaced by a representation of texture. It is well known, however, that curvature of contours and junctions provide crucial shape information [1, 4]. Thus we believe it is time to investigate a contour representation alongside appearance-based descriptors.

We show that local contour descriptors produce significant gains in object detection and complement texture descriptors. We propose a contour representation that flexibly encodes junction information and curvature. The representation discretizes contour orientation at an interest point and records contour intensity at each angle as feature elements. Examples are shown in Figure 1. In contrast, recent work has parameterized adjacent contour segments with a fixed junction degree and a strict ordering of segments [7]. We also make a distinction with appearance-based descriptors, such as HoG [6], that integrate over orientations in image regions and produce a texture summary that marginalizes contour detail.

Our contour representation encodes curvature as a non-parametric distribution over oriented bars, or segments. Understanding contour curvature is important because shape cues concentrate there. This is illustrated in Figure 2, where we sampled interest points uniformly over detected edges and computed oriented bar features. We then rendered the features in order of their contour strength and curvature, and observe that the object is identifiable after a relatively small number are displayed.

We complement the oriented bar representation of contours with a texture descriptor collected at each interest point. The oriented bars are a set of filters F_1, \dots, F_D that have an oriented line segment with one endpoint in the filter center. An edge map E is convolved with each of the filters, creating D channels of contour orientation responses

$$B_d(x) = \sum_x E(x-x)F_d(x). \quad (1)$$

The channels are sampled at an interest point x_i giving the unnormalized feature $\hat{b}_i = B_1(x_i), \dots, B_D(x_i)$.

To describe local texture in the vicinity of interest points, we compute the geometric blur descriptor [3]. The geometric blur summarizes a signal under all affine transformations at a point. The descriptor centered at location x is a convolution with a spatially varying Gaussian kernel

$$G_x(y) = \sum_x E(x+y-x) \eta(x; \alpha \|x-y\| + \beta). \quad (2)$$

We sample the descriptor at C locations in concentric rings about the i -th interest point, giving the unnormalized feature $\hat{g}_i = G_i(y_1), \dots, G_i(y_C)$. After normalization, we concatenate the oriented bar and geometric blur features. Further details can be found in the paper.

We investigate the capabilities of our representation with the common Hough transform for object detection. The approach is kept general to remain widely relevant. Features located at x_i are matched to a training

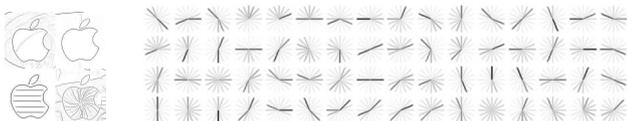


Figure 1: Oriented bar codebook learned from edge maps of Apples in the ETHZ dataset.



Figure 2: Shape information is concentrated at points of high curvature. Oriented bar features are ordered and rendered by contour strength and curvature using 10, 30, 50, 100 and 200 features.

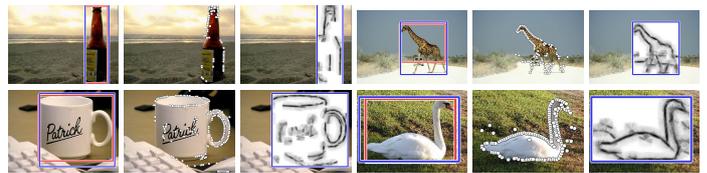


Figure 3: Example detections demonstrating the accuracy and contour localization ability of our representation. Ground-truth boxes are red and detections are blue.

set and associated with object shift vectors v_i . The vectors combine to cast votes for object position x through a set of discrete scales σ . The Hough accumulator is then

$$H(x, \sigma) = \sum_{i=1}^N w_i \delta(\|x_i + \sigma v_i - x\|), \quad (3)$$

where δ is the Dirac delta function. This equation has been used numerous times and the weights have been estimated in many ways. Surprisingly, we found in our experiments that uniform weights exceeds many state-of-the-art voting methods (excluding post-processing verification).

The proposed object representation is evaluated in the context of multi-scale object detection in cluttered scenes. We perform a rigorous evaluation using the ETHZ shape dataset and INRIA horses. We separately evaluate our approach on these datasets against several comparable state-of-the-art approaches. We further show a comparison of our combined representation against two baselines of only oriented bars or geometric blur. This isolates any performance gain to our approach. Figure 3 shows a few detection results. Further qualitative and quantitative results can be found in the paper. Taken together, they show our approach is an effective and computationally efficient representation of contours and junctions that accurately localizes and describes the local shape of contours.

- [1] F. Attneave. Some informational aspects of visual perception. *Psych. Review*, 61(3), 1954.
- [2] S. Belongie and J. Malik. Matching shapes. In *ICCV*, pages 454–461, 2001.
- [3] A. Berg and J. Malik. Geometric blur for template matching. In *CVPR*, pages 607–615, 2001.
- [4] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psych. Review*, 94(2):115–147, April 1987.
- [5] M. Clowes. On seeing things. *Artificial Intelligence*, 2:79–116, 1971.
- [6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, June 2005.
- [7] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contour segments for object detection. *PAMI*, 30:36–51, 2008.
- [8] D. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31(3):355–395, 1987.