Implicit Shape Kernel for Discriminative Learning of the Hough Transform Detector

Yimeng Zhang
yz457@cornell.edu
Tsuhan Chen
tsuhan@ece.cornell.edu

School of Electronic and Computer Enginerring
Cornell University
Ithaca, NY, USA

The sliding window approach has been widely used for object detection because it provides a simple way to apply object recognition techniques to the detection task. Despite its effectiveness, though, the exhaustive search makes the approach inefficient in the case of a non-trivial classifier. The branch and bound techniques [3, 4] have been proposed in recent years to avoid the exhaustive search and still find the global optimal.

The Hough transform [1, 2, 5, 6, 7, 8] provides an alternative way to perform the detection task in a much more efficient manner. The Hough transform based detector has three main steps, as illustrated in Figure 1: 1) Each local patch in a test image is assigned to a codeword. 2) According to the spatial information of the codebook learned from the training images, each patch will cast weighted votes to the object locations and scales, and obtain the initial Hough image. 3) In order to tolerate shape deformation, kernel density estimation, such as Gaussian filtering or Mean-shift modes estimation is applied to the Hough image. This process gives us the final Hough image, and the peaks in the Hough image are extracted as the detection hypotheses. Figure 1 also shows the three weights we need to learn. The implicit shape model [6] puts the Hough transform into a probabilistic formulation by learning the locations of the codewords (weight 2) based on their spatial distribution in the training images.

Despite the success of the implicit shape model, it has two drawbacks. First is its discrimination power. There have been several works these years [2, 7, 8] that dealt with this issue. All of these methods have significantly improved the implicit shape model. However, they only make discrimination on the codewords (weight 1 or 3 in Figure 1), while the spatial weights are learned generatively as the spatial distribution of the codewords in positive training examples (weight 2). The second drawback is that it is difficult to interpret the scores in the final Hough image, especially after the kernel density estimation, so it is difficult to tell what function the learning process is optimizing.

In this paper, we propose a novel approach for learning the Hough transform. The approach puts the *whole* of the Hough transform into a maximum margin formulation by connecting the Hough transform with the SVM through the kernel methods. We design a kernel particularly for the Hough transform detector and call the kernel "Implicit Shape Kernel". During training, we use the kernel to train a SVM classifier, which determines the presence of the object of interest in a subwindow. During testing, we can follow the standard Hough transform process for the kernel calculations, and the final Hough image will provide the exact the output scores of the SVM at every location and scale.

We briefly describe the implicit shape kernel. The kernel is illustrated in Figure 2. For each pair of regions from two examples, if they are matched to the same codeword, we calculate the similarity value of their locations relative to the object centers. The similarity value is calculated using the window function K_w in the kernel density estimation . The window function can be either a Gaussian function or an Epanechnikov function. Different functions also define different processes at the runtime of detection with the Hough transform. The kernel value K of the implicit shape kernel is the summation of the similarities of all such pairs. Let I denote an image, C_i denote a codeword entry, $C(f_k)$ denote the codeword assignment of feature f_k, and x_k denote the location of the feature.

$$K(I_x, I_t) = \sum_i \sum_{k:C(f_k)=C_i} \sum_{k':C(f_{k'})=C_i} K_w\left(\frac{x_k - x_{k'}}{b}\right) \quad (1)$$

where b is the bandwidth of the window function. This kernel captures the deformation between the shapes of the two examples. If the two examples are the same image, the locations of the codewords will match exactly. Based on the property of the window function K_w, we get the highest similarity value for each pair of words. As the shape deforms, the value will decrease in a way defined by the window function.

The proposed kernel provides several good properties: 1) During testing, we can apply the standard Hough transform process to obtain the decision scores of the learned SVM for the subwindows by transferring the coefficients of the support vectors to the weights used for the Hough transform. The detailed implementation is described in the paper. we show that the scores in the final Hough image give the exact decision scores of the classifier for every location of the object in a testing image. Therefore, the method benefits from the detection efficiency of the Hough transform. Moreover, we are directly optimizing the classification scores during training. 2) By using the proposed kernel to train a maximum margin classifier (i.e. SVM), our method provides full discriminative learning of the Hough transform, including both appearances of the local patches and their spatial weights. The learning also takes the kernel density estimation into account. 3) Through the window function K_w, the implicit shape kernel avoids hard quantization of the image space when modeling the spatial information of the codewords, and therefore retains the flexibility of the implicit shape model [6].

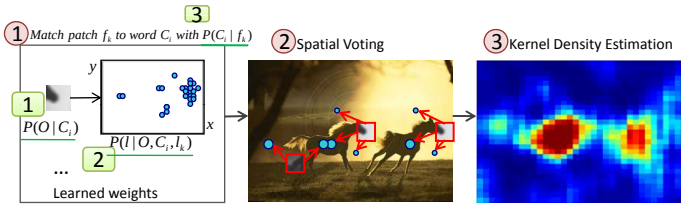

Figure 1: The illustration for the Hough transform. The three steps for testing an image are indexed with red circles. The three weights we need to learn are indexed with green circles.

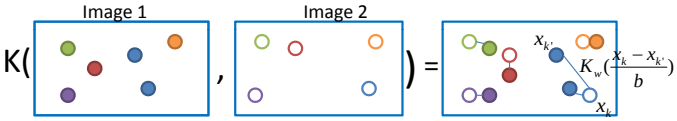

Figure 2: The illustration of the proposed kernel. The circles represent local patches. Different colors represent different codeword assignments.

We evaluate the proposed approach on Pascal VOC 2006 and 2007 datasets, INRIA horse and UIUC car datasets.The experiment results show that our approach significantly improves the detection performance over previous methods of learning the Hough transform.

[1] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. People-tracking-by-detection and people-detection-by-tracking. In *CVPR*, 2008.

[2] Juergen Gall and Victor Lempitsky. Class-specific hough forests for object detection. In *CVPR*, 2009.

[3] Christoph H. Lampert, Matthew B. Blaschko, and Thomas Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *CVPR*, 2008.

[4] Alain Lehmann, Bastian Leibe, and Luc van Gool. Feature-centric efficient subwindow search. In *ICCV*, 2009.

[5] Alain Lehmann, Bastian Leibe, and Luc van Gool. Prism: Principled implicit shape model. In *BMVC*, 2009.

[6] Bastian Leibe, Aleš Leonardis, and Bernt Schiele. Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, 77(1-3):259–289, 2008.

[7] Subhransu Maji and Jitendra Malik. Object detection using a max-margin hough transform. In *CVPR*, 2009.

[8] Ryuzo Okada. Discriminative generalized hough transform for object detection. In *ICCV*, 2009.