Computer Vision Paper Critique

Kshitij Srivastava (MT18099)

Paper Title: Object Detection using a Max-Margin Hough Transform
Authors: Subhransu Maji, Jitendra Malik
Venue: CVPR 2009
Link: `http://www.cse.psu.edu/~rtc12/CSE586/papers/match_smajiMaxMarginHough.pdf`

I. SUMMARY

The prevalent techniques for object detection include sliding window classifiers, pictorial structures, constellation models and implicit shape models. Each suitable for different use-cases. Some of the techniques to tackle complexity issues associated with the above techniques include looking at salient regions, coarse to fine grain search, branch and bound. The Hough transform has also been used widely for this. The implicit shape model, a probabilistic formulation of the Hough transform, is of much importance because of its probabilistic voting scheme. Making the local parts to vote for possible transformations such as translation, scale and aspect variation, allows for importance sampling of windows by using the peaks of voting space.

The main contribution of the paper is in placing the Hough transform in a discriminative framework where the object center is found using a local part voting scheme. It takes into account the appearance of the part as well as the spatial distribution of its position with respect to the object center. This results in a convex formulation, which can be solved using off the shelf optimization packages. The authors have called their approach Max-Margin Hough Transform or M^2HT. They have conducted their experiments on three data-sets; the ETHZ shape data-set, UIUC cars data-set, and INRIA horse data-set and achieved significant improvements over state-of-the-art techniques with much lower complexities.

A. Max-margin Hough Transform

The procedure of probabilistic Hough transform can be viewed as a weighted vote for code-book entries Ci (local parts). Learning these weights in a discriminative manner can optimize the classification performance. The idea here is to realize that the score of an object at a location x is a linear function of $p(O \mid C_i)$.

$$S(O,x) \propto \sum_{i,j} p(x \mid O,C_i,l_j)p(C_i \mid f_i)p(O \mid C_i,l_j)$$

$$= \sum_{i,j} p(x \mid O,C_i,l_j)p(C_i \mid f_i)p(O \mid C_i)$$

$$= \sum_i p(O \mid C_i) \sum_j p(x \mid O,C_i,l_j)p(C_i \mid f_i)$$

$$= \sum_i w_i xa_i(x) = w^T A(x)$$

The above equation proves the previous point of linearity of the score function.

1) Discriminative Training: For a training set, we are given the positive instances and pick hard negative instances by finding the peaks in the voting space of negative training images. Weights are learned by maximizing the score obtained by summing up the results of the voting process ($w^T A_i$).

$$\min_{w,b,\mathscr{E}} \ \frac{1}{2} w^T w + C \sum_{i=1}^{T} \mathscr{E}$$

$$\begin{aligned} s.t.: \quad & y_i(w^T A_i + b) \geq 1 - \mathscr{E} \\ & w \geq 0, \mathscr{E} \geq 0, \forall_i = 1,2,...,N \end{aligned}$$

The above optimization is very similar to that of a linear SVM and has been solved using the CVX package.

B. Overall Detection Strategy

The detector works in two stages:

1) The M2HT detector is run on the input image to find a small set consisting of regions of interest.
2) A verification classifier based on a SVM is used to find the true location and score of the object by doing a local search.

The Hough transform step helps in narrowing down the search to a smaller subset of regions.

1) $M^2HTDetector$: Code-books generated using k-means clustering of Geometric Blurring (GB) features sampled uniformly along the edges in an image are used to learn the weights. It was found that voting over scales is not as reliable, so instead the authors have run the detector over a small set of scales and vote for the rest of the pose parameters. Negative instances (not containing the category of interest) are obtained by running the Hough transform on negative images and finding the peaks of the voting space above a threshold.

2) Verification Classifier: The SVM classifier has been training using the pyramid match kernel on histograms of oriented gradients. The image is divided into grids of increasing resolution and histograms from each level are weighted according to the given equation.

$$w_l = 2^{l-1}, \text{l = 1 being the coarsest scale}$$

Training on all data-sets has been done by resizing the positive instances of the category to the median aspect ratio and a number of windows sampled from negative training images. Detection is done by running the classifier at various locations and aspect and scales, at fixed aspect ratio. The speedup method for IKSVM classification has been used to make the runtime of the classifier equivalent to a linear SVM.

II. PROS

- This algorithm proposed in the paper is a significant improvement over existing state-of-the-art methods used for object detection, both in terms of improved results as well as reduced runtime.
- The algorithm has been explained quite well by the authors and the results are shown in an effective way.
- The authors have used very simple techniques such as Hough transform and SVM classifier to achieve sizable results.

III. CONS

- The data-sets used are of good quality images and nothing can be said about the robustness of the algorithm.
- Hough transform is still naive in the sense that it works efficiently for lower number of parameters. Hence, the algorithm cannot be used for image sets with higher number of transformations been done to images.

IV. SUGGESTIONS

- Using a boosted version of SVM might help in increasing the accuracy of verification.
- I would like to try an enhanced version of the Hough transform for detecting important regions of the image like Agent-Based Hough transform.
- Instead of SVM as the verification classifier, using neural networks might help. However, it might add more computational overhead.