Evidential combination of pedestrian detectors

Philippe Xu[1]
https://www.hds.utc.fr/~xuphilip
Franck Davoine[12]
franck.davoine@gmail.com
Thierry Denœux[1]
https://www.hds.utc.fr/~tdenoeux

[1] UMR CNRS 7253, Heudiasyc
Université de Technologie de Compiègne
Compiègne, France
[2] CNRS, LIAMA
Beijing, P. R. China

The importance of pedestrian detection in many applications has led to the development of many algorithms. In this paper, we address the problem of combining the outputs of several detectors. A pre-trained pedestrian detector is seen as a black box returning a set of bounding boxes (BB) with associated scores. We conducted our experiments using the Caltech Pedestrian Detection Benchmark [2]. More than 30 state-of-the-art detectors were tested on this dataset and their outputs are publicly available.

To illustrate the potential gain from combining multiple detectors, we show in Fig. 1 (a) some detection statistics for the Caltech dataset. We can see that, at one False Positive Per Image (FPPI), more than 95% of the pedestrians in the "Reasonable" scenario were detected by at least one detector. The "Reasonable" scenario corresponds to pedestrians over 50 pixels tall and with an occlusion rate lower than 35%. As a comparison, the currently best performing algorithm has a recall rate of about 80% at 1 FPPI. Similarly, in the "Overall" scenario where all the pedestrians were considered, about 60% of the pedestrians were detected by at least one detector. The currently best algorithm hardly reached a 40% recall rate. The potential gain of combining in a proper way all those detectors is thus fairly significant.

In order to combine the outputs of different detectors, the BBs returned by the detectors need to be associated. In a sliding windows approach, a single pedestrian is often detected at several nearby positions and scales. A non-maximal suppression (NMS) step is often needed in order to select only one BB per pedestrian. In our context, the same issue occurs but, instead of having multiple detections from a single detector, they are returned by several ones. We formulated the NMS problem as a simple hierarchical clustering where the distance between two BBs is defined as their area of overlap. The clustering was done greedily by defining the distance between two clusters as the distance between their respective highest-scored BBs. This implies that the outputs from the different detectors are comparable.

To handle this issue, a calibration step was used to transform the scores into calibrated probabilities. Fig. 1 (b) illustrates the calibration results obtained from a logistic regression and an isotonic one. One particularity of object detection is the relatively high false positive rate. For example with the 'HOG' algorithm [1], more than 99% of the detections have a score less than 0.1 and less than 0.1% of these detections are true positives. As a result, most detections have an associated probability lower than 0.1. From a Bayesian perspective, multiple sources of information returning low probabilities would actually lead to an even lower one. This would go against the idea that multiple detections should lead to increased confidence.

The theory of belief function [3] was used to handle this issue. We interpreted the output $q \in [0,1]$ of a calibration function as a simple mass function $m = \{1\}^{1-q}$ defined over the frame of discernment $\Omega = \{0,1\}$ as

$$m(\{1\}) = q, \qquad m(\{0,1\}) = 1-q. \tag{1}$$

To combine two mass functions $\{1\}^{\alpha_1}$ and $\{1\}^{\alpha_2}$, three combination rules were considered in our experiments: Dempster's rule, the cautious rule and a triangular norm-based rule. They are defined, respectively, as

$$\{1\}^{\alpha_1} \oplus \{1\}^{\alpha_2} = \{1\}^{\alpha_1\alpha_2}, \tag{2}$$

$$\{1\}^{\alpha_1} \owedge \{1\}^{\alpha_2} = \{1\}^{\min(\alpha_1,\alpha_2)}, \tag{3}$$

$$\{1\}^{\alpha_1} \circledT_p \{1\}^{\alpha_2} = \{1\}^{\alpha_1 \top_p \alpha_2}, \tag{4}$$

where

$$\alpha_1 \top_p \alpha_2 = \begin{cases} \alpha_1 \wedge \alpha_2 & \text{if } p = 0, \\ \alpha_1 \alpha_2 & \text{if } p = 1, \\ \log_p\left(1 + \frac{(p^{\alpha_1}-1)(p^{\alpha_2}-1)}{p-1}\right) & \text{otherwise.} \end{cases} \tag{5}$$

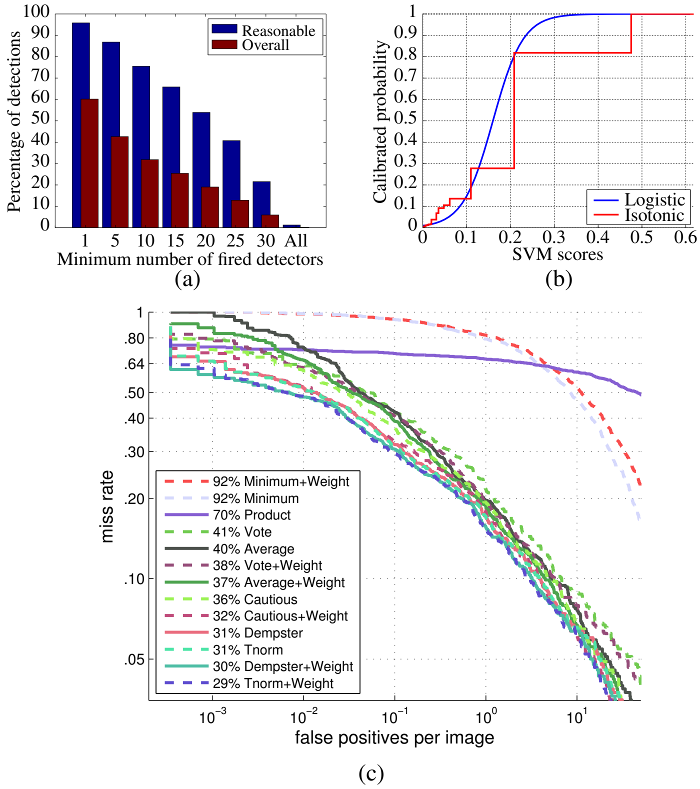

Figure 1: (a) Percentage of detected pedestrians by at least $k \in \{1,5,\ldots,34\}$ detectors at 1 FPPI. (b) Logistic and isotonic calibration of the scores from the 'HOG' pedestrian detector [1]. (c) Results of different combination strategies using a logistic regression calibration on the "Reasonable" scenario.

The triangular norm-based rule was used to better handled the dependencies among detectors. The detectors were first grouped using a hierarchical clustering and the parameter $p \in [0,1]$ of the triangular norm was then optimized for each pairwise combination.

In our experiments, we compared probabilistic combination rules (product, average, min and max) to evidential ones. Figure 1 (c) shows the results obtained from a logistic calibration on the "Reasonable" case scenario. We can see that the product and minimum rules performed very poorly. The average rule performed better than the majority vote. The cautious rule, which is equivalent to the maximum rule, performed better than all the other probabilistic rules but worse than Dempster's rule and the t-norm based rule. Using an additional weight led to better results for all combination methods except the minimum combination rule. Similar conclusions were reached by using an isotonic calibration. Compared to the best single pedestrian detector, the logistic weighted t-norm led to an improvement of 9% in terms of log-average miss rate and 6% for the isotonic one. The weighted average only led to 1% improvement. All the other probabilistic combination rules led to a decrease in performance.

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, San Diego, USA, 2005.

[2] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):743–761, 2012.

[3] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.