High Precision Pedestrian Recognition in Low Resolution Images

Zs.T. Kardkovács, Cs. Gáspár-Papanek, G. Szabó, E. Sipőcz and G. Pécsi

Proper recognition of objects in images is one of the most important tasks of next generation intelligent systems. For example, identification of pedestrians plays a key role in building intelligent vehicles, surveillance robots, etc. To this end, NiSIS (Nature Inspired Smart Information Systems) announced a competition in 2007 which was aimed at boosting researches in this specific area. During the competition a large amount of low resolution (36x18 pixels) 8-bits grayscale still images were available based on DaimlerChrysler (DC) pedestrian classification benchmark [1]. The problem consists in the detection of pedestrians against a background or other pedestrian-like objects. Our solution proved to be the best among 17 participants from 14 countries.

DC pedestrian benchmark contains five sets of non-correlated still images. That is, neither pedestrians nor objects appear in more than one set, however, pedestrians may appear in several images mirrored or shifted by 2-4 pixels both vertically and horizontally within the same set. NiSIS made the benchmark more interesting by clearing (replacing adequate color codes by 205) a part of some images as if they were broken. In order to fulfill criteria of the benchmark our solution consists of four main modules: data preprocessing algorithms, image preprocessing techniques, classifiers, and ensemble methods.

Data preprocessing algorithms (DPA) are responsible for standardizing images as much as possible, i.e. all next processing techniques should not depend on mirroring, shifting, and other possible transformations. DPA discovered that images show pedestrians are common in having similar color values near the head and „shoe" regions. Nevertheless, the difference between common pedestrian and non-pedestrian images indicated that the real distinction between images can be captured near the sky region. We assume this was produced by the lens and auto-focusing methods of the camera since certain pixels have the same (reference) color codes in almost all images.

Image preprocessing techniques are responsible to identify common characteristics of object and filter foreground object from background. We assumed four approaches can help in solving data analysis discovered subproblems: we used an edge detection algorithm, a normalized skeletonization method, image based Z-normalization, and singular value decomposition (SVD). We used edge detection to find silhouettes of objects in images. We hoped skeletons determine basic shape of an object, and provide us hints how to transform them into a well-identified pedestrian's skeleton. Skeletons were also inputs of well-known neural network classifiers. Since neural networks are usually sensitive for input shifting, we normalized the skeletons in the sense that the skeleton element p closest to the pixel $p_0 = \langle 9;\ 18 \rangle$ is shifted to p_0, Z-normalization is widely applied for time series where intensities, trends, vertical shifting should be avoided. We found a 5–10% improvement on Z-normalized data results. SVD is used as a geometric transformation which determines proper orthonormal basis for each of the images. SVD is invariant for horizontal (and vertical) shifting, thus feature matrices of images were the most important input of classifiers.

Three classifiers were applied our solution: k nearest neighbor (kNN, with $k = 11$), a feed-forward multi layer Adaline network (two hidden layers, 5–5 sigmoid neurons in each), and an extreme learning machine (ELM) with Gaussian weights (a single layer network with 300 hidden neurons).

Our classifiers provide confidence like metrics between 0 and 1. We enhanced our model to evaluate both possibilities independently by each of the classifiers, and classifiers must predict confidences of both evaluations. One of the ensemble method is firmly used the sum of confidence weighted prediction for images. Since all classifiers perform very similarly, additional weighting was not necessary. While this voting strategy is very effective confidence values cannot be compared in general hence they have no common absolute scale or common background. Therefore, we also applied a feed-forward neural network to evaluate classifier confidence values, judgment and some other parameters. Since judgments of neural network based ensemble overperformed the other by 3-5%, we agreed on its judgment for pedestrians.

Our method was trained on the train data set given by the downloaded data. We used labeled test data for both validating and testing our model by splitting the set into two equal size parts. That is, we had a train data set denoted by S_1, a test set S_2 and a validation S_3. Each performance of the applied methods is seen in Table 1 where performance was calculated by the formula:

$$P(c) = \frac{\text{correct judgements by classifiers}}{\text{the number of images}}.$$

Our solution was found to be the most accurate (96.04%) on the NiSIS competition therefore we received a special award.

	SVD + kNN	Adaline	ELM	Simple ensemble	Neural ensemble
no modification	84%	84%	83%	91%	93%
edge detection	74%	74%	72%	78%	81%
skeletonization	76%	75%	73%	80%	83%

Table 1: Performances of classifiers trained on $S_1 + S_2$, tested on S_3

Acknowledgements

Our work was sponsored by Mobile Innovation Centre through the Asbóth Oszkár Programme of the Hungarian State.

References

[1] S. Munder, D. M. Gavrila. An Experimental Study on Pedestrian Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, November 2006.