ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

Detecting zebra crossings utilizing AdaBoost

Ludwig Lausser[1] and Friedhelm Schwenker[2] and Günther Palm[2]

1- University Hospital Ulm - Department of Internal Medicine I
Robert Koch Str. 8, 89069 Ulm - Germany

2- University of Ulm - Dept of Neural Information Processing
89069 Ulm - Germany

Abstract. This paper introduces a visual zebra crossing detector based on the Viola-Jones approach. The basic properties of this cascaded classifier and the use of integral images are explained. Additional pre- and postprocessing for this task are introduced and evaluated.

1 Introduction

In the last decades vehicle safety systems evolved from simple damage minimizing and preventing tools to complex multimodal sensory systems recognizing dangerous situations before they emerge. Many different sensors have been developed on particular parts of this tasks. One possibility of getting abstract data is to analyse visual information of cameras aligned to the driving direction. Different methodes of computer vision can be used therefore. While in many tasks better results could be achieved by combining other sensors with the visual input, the task of detecting markings on a street surface could only be solved by using visual information. Here a detector for zebra crossings, based on the approaches of Viola and Jones, is introduced. The detector works on grey-scaled pictures of 320×240 pixels recorded from a camera placed in front of the rear-view mirror inside of the test vehicle. The images were recorded with a sampling rate of 15 pictures per second.

2 Cascaded AdaBoost

Viola and Jones applied the AdaBoost-algorithm which was introduced by Freund and Schapire [2, 3, 1] to train a single stages of their cascaded architecture. AdaBoost is a meta algorithm that builds an ensemble of weak classifiers in order to get a final strong classifier. After selecting one weak classifier, the training examples get reweighted to intensify examples, which are not correctly classified yet. Trained on the reweighted training set, the next weak learner will minimize the error of the current ensemble. Viola and Jones restrict the weak classifiers to classification functions, which depend on one single feature. In this way the AdaBoost-algorithm produces a small set of features, which classifies the training set quite well. The cascaded structure used by Viola and Jones [7, 4] combines classifiers h_i trained by the AdaBoost algorithm to the cascaded classifier c:

$$c(x) = \bigwedge_{i=1}^{K} h_i(x)$$

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

In this context K is the resulting number of stages produced by the cascade's training. The great benefit of this structure comes from iterative evaluation of the single classifiers h_i. A sub-window, which was rejected at a certain stage, must not be evaluated at higher stages. The benefit is a gain in terms of computational time, because the most sub-windows will be rejected in early stages. The expected number of weak classifiers N which have to be evaluated per sub-window is reduced to

$$N = n_1 + \sum_{i=2}^{K} \left(n_i \prod_{j<i} p_j \right)$$

where n_i is the number of weak classifiers in the ith stage and p_i it's positive rate. Viola and Jones describe the positive rate as "*...the propotion of windows which are labelled as potentially containing the object of interest*" [7].

3 Feature extraction and Variance Filtering

In their very first approach Viola and Jones [7] use a set of very simply rectangular features influenced by the work of Papageorgiou [5] (Fig. 1). The sum of pixels, which lies within the black area, gets simply subtracted from the pixels, which lie in the white area. The resulting set consists of simply contrast-, edge- and corner- detectors, which are specialized for horizontal and vertical structures. For our base resolution of 24×24 the set is composed of 162,332 single rectangular features.

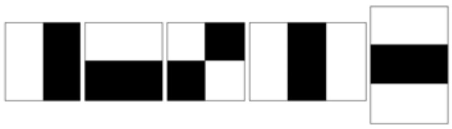


Fig. 1: Reactangular features used by Viola and Jones

In order to compute these features effictively an intermediate representation, the so called integral image [7], was applied. Variance normalization brings a lot of benefit to the classification task, e.g. images with low contrast will become exhanced. Some problems appeared on the task of detecting single stripes of a zebra crossing. In the street scenario there are a lot of other objects very similar to single stripes. For example grooves in the lane could look like markings if the sun shines from a particual position. Variance filters as used by Peters[6] take care of this problem. Peters tests the standard deviation σ of a sub-window against a treshold, and only if variance σ^2 is greater than this treshold, the sub-window will be scanned by the detector cascade. A small value for σ indicates a low contrast within the sub-window. Using this method only sub-windows with high contrasts are scanned. In our approach a bandpass filter implemented. This additionally rejects sub-windows with too high contrasts. This method

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

	members	gradient	distance to regression line
zebra crossing	≥ 35	$[-0.15; +0.15]$	≤ 7 pixels
single line	≥ 20	$\geq \lvert 0.15 \rvert$	≤ 2 pixels

Table 1: Group parameters

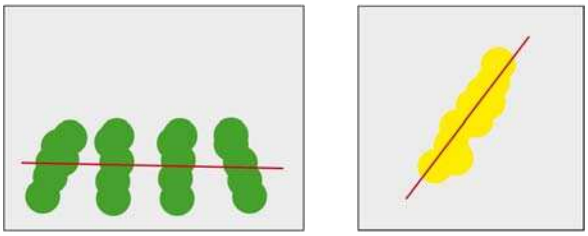
was chosen after some tests with the normal highpass. The single stripes of a zebra crossing didn't pass an additionally set treshold on their part. So sub-windows with a too high contrast don't show zebra crossings neither. In this concrete setting only the middle part of an image, which shows the lane, is of interest (Fig. 3(a)). A single stripe of a zebra crossing will never appear bigger than 40×40 pixels. So bigger scales won't be used neither. In this way the number of sub-windows, which have to be scanned and analyzed, are reduced to approximately 31,000 per image (for more details see [8]).

4 Grouping

Without some postprocessing the AdaBoost cascade showed a somewhat different behaviour than that which is known from tasks like face-recognition. A cascade trained for face-recognition allways detects the presented face in more than on sub-window. The single sub-windows are arranged concentric around the real face. Each of this sub-windows shows approximately the entire face. The cascade trained for single stripes not only reacts on whole stripes but also on parts of them. The single sub-windows will therefore be ordered in a line. The center points of the sub-windows will lie on the single stripes. In this application a single hit was represented by a rectangle of 24×36 pixels. On this way a single stripe will be covered by small rectangels. The difference could be seen on diagonal stripes. Here a representation of an hit through it's sub-window would include wide aeras, where no stripe could be seen. Two hits belong to the same group, if their boundings overlap. A simple bounding box represents the final group. The width of 36 pixels was chosen to create connections to neighboured stripes. Now features had to be found to characterize those groups, which really show a zebra crossing. Within some experiments three features had been chosen: The number of hits within a group, the gradient of the regression line through this points and the average distance between the points and the regression line. The values mentioned here were chosen for pictures of 320×240 pixels (Tab. 1). Once a zebra crossing was found, there must be 100 single hits in the next picture, which lie in the boundings of the zebra crossing to prolong the detection.

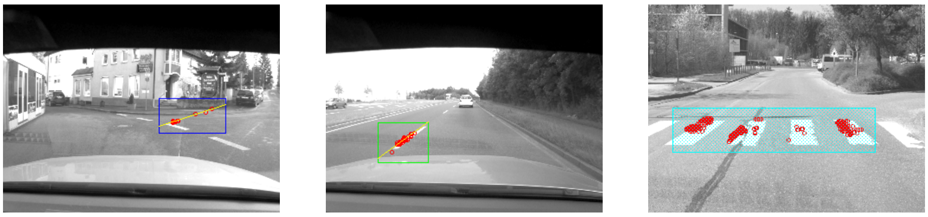
Scanning a whole zebra crossing the system will detect a group consisting of some more or less parallel lines (Figure 2(a)). A regression line of this group will therefore be quite horizontal. The average distance of the single points to the regression line will overcome a certain minimum because of the stripe structure within it.

Normally other groups don't satisfy this requirements, but their characteris-



(a) (b)

Fig. 2: Schematic Groupings of single hits : 2(a) grouping caused by a zebra crossing, 2(b) grouping caused by a single line



(a) (b) (c)

Fig. 3: Examples for some different groups: 3(a) normal group, 3(b) group declared as single strip, 3(c) some groups merged as zebra crossing

tics can be disturbed by some random single hits. On this way groups can be declared to zebra crossings falsely. If for example some random single hits were detected, attaching the left and right side of a group, the group's regression line will be flattened. In order to minimize such effects, it's from benefit to declare another group. In our task it is quite impossible to differ between the single white stripes of a zebra crossing and other white lines on the lane. Therefore a lot of false detections may be found on this lines. Even if the groups could be distinguished from a zebra crossing, they are likely to be interpreted falsely. Because of the large number of single detections on the line, only a few single random detections are needed to achieve the 35 hits, which are needed to interprete the group as a zebra crossing. Therefore if a group is labelled as a single line (Figure 2(b)), its corresponding hits in the number of subsequent images can be rejected.

Intuitively the gradient of a single line seems higher than bounds defined in Table 1. The gradient bounds were chosen in this way because of curves. Once a single line was found, single hits of the subsequent image were rejected, if they lie less than ten pixels from the estimated regression line. The procedure is repeated in the next three images. If more than thirty single hits are rejected

	highway	city	crossroad
false-positive-rate	0,27%	0,03%	0,19%

Table 2: false-positive-rates of the internal Viola-Jones-Cascade

	highway	city	crossroad
rejected false detections	88,92%	14,78%	53,96%

Table 3: Rejection of single detection by single-line-groups

in this sequence, the procedure continues for further three images.

Sometimes parts on the outside of a zebra crossing build seperate groups. Those of them were seperated, which were worthy recognizing them. If such a group is found next to a detected zebra crossing, they will get merged to one zebra crossing detection. A selected group has to contain 10 single detections, an gradient lower than -0,15 or higher than 0,15 and the average distance between regression line an hits had to be lower than 7 (See Fig. 3).

5 Results and Conclusion

Because it's quite hard to find one set of representative samples, sequences of standard situations were tested instead of. Three situations without any zebra crossing were chosen: Driving on a highway, driving through the city and driving on a crossroad. For each situation a sequence of 100 pictures was evaluated to get results for the false-positive-rate (table 2) of internal Viola-Jones-Cascade. The concept of the single-line-group was also tested in this way (table 3). The fourth situation was crossing a zebra crossing. Therefore 75 Pictures had been evaluated. Sensitivity and specifity of the internal Viola-Jones-Cascade were computed for this situation(table 4). The whole grouping system was tested in the situations driving on a highway, driving through the city and crossing a zebra-crossing. The two first situations were therefore enlarged to 2700 and 2900 pictures (table 5).

The false-positive-rate differs from situation to situation. This is an effect, which comes from the different environments. Whereas there are no white lines at the city-situation, there are continuous white lines marking the border of the lane at the highway-situation. Therefore much more single false detections could be rejected in the highway-situation than in the city-situation.

The sensitifity and specifity evaluated at the zebra crossing situation show

	zebra crossing
Sensifity	0,013
Specificity	0,999

Table 4: Sensifity and Specificity

	highway	city	zebra crossing
containing zebra crossing groups	0,3%	0,21%	97,33%

Table 5: Percentage of pictures containing zebra crossing groups

that only a small part of sub-windows, which show a single white stripe, is recognized as it. But most of the sub-windows, which don't show a single white stripe were classified correctly. The correctly detected sub-windows are sufficient to detect $97,33\%$ of the zerba crossings in the test set. In $0,3\%$ of all pictures of the highway situation a zebra crossing was detected. In the city test set only $0,21\%$ were labelled incorrectly.

A zebra crossing detector based on the approach of Viola and Jones was introduced. It was shown how a variance filter could be used to reduce the time needed to scan a single picture. The internal cascade was trained to detect the single stripes of a zebra crossing. This caused high false positiv rates, because other white lines on the lane couldn't be distinguished from them. This problem was solved by grouping the single sub-windows in postprocessing. Therefore some characteristics for a zebra crossing group and a single line group were introduced. The single line group could reject up to 88,92% of all false positives detected by the internal cascade.

References

[1] Freund Y, Schapire RE, A short introduction to boosting. In: *Journal of Japanese Society for Artifical Intelligence* 14(5):771 – 780, 1999

[2] Freund Y, Schapire RE, A decision-theoretic generalization of on-line learning and an application to boosting. In: *Journal of Computer and System Sciences*, 55(1):119-139, 1997

[3] Freund Y , Schapire RE, Experiments with a New Boosting Algorithm. In: *International Conference on Machine Learning*, 1996, p. 148–156

[4] Jones M , Viola P: Fast Multi-view Face Detection / MITSUBISHI ELECTRIC RESEARCH LABORATORIES. 2003. – Tech Report

[5] Papageorgiou CP, Oren M, Poggio T, A General Framework for Object Detection. In: *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision* (1998), p. 555 IEEE Computer Society 1998

[6] Peters, V *Effizientes Training ansichtsbasierter Gesichtsdetektoren*, University of Bielefeld, Master Thesis, 2006

[7] Viola P, Jones M, Robust Real-time Object Detection. In: *Proc. of IEEE workshop on Statistical and Computational Theories of Vision*, Vancouver, CA, 2001

[8] Lausser, L *Erkennung von Fahrbahnmarkierungen mit maschinellen Lernverfahren*, University of Ulm, Master Thesis, 2007