# A REAL WORLD DETECTION SYSTEM
## *Combining Color, Shape and Appearance to Enable Real-time Road Sign Detection*

Peng Wang, Jianmin Li and Bo Zhang
*State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, China*

Keywords: Object Detection, Road Sign, Haar-wavelet, Cascade Detector, HSV Color Space, Contour, RANSAC.

Abstract: Although specific object detection has undergone great advances in recent years, its application to critical real-time circumstances like those in automated vehicle controlling is still limited, especially when facing strict speed and precision requirements. This paper uses a hybrid of various computer vision techniques including color space analysis, Haar-wavelet cascade detector, contour analysis and RANSAC shape-fitting, to achieve a real-time detection speed while maintaining a reasonable precision and false-alarm level. The result is a practical system that out-performed most rivals in an automated vehicle contest and an indication of feasible CV application to speed critical areas.

# 1 INTRODUCTION

## 1.1 Motivation

The real time road sign detection system introduced by this paper was motivated as a crucial subsystem of an automatic driving system for vehicles. Three challenges arise in designing such a real time object detection application which cannot be fulfilled together by a single existing computer vision algorithm: the strict requirement of speed, of precision and the diversity of target objects.

## 1.2 Outline

To tackle all these three challenges, we designed a combination of various computer vision techniques to utilize color, shape and appearance information all together. The design principle of our system is to rely on a stable detection algorithm, which may be time consuming, to act as the main detector, while utilize various kinds of pre-processing and post-processing stages to shrink the area of regions performed on by this main detector. The reduction in regions of interests (ROIs) compensates the slow speed of the main detector. For orthogonality, these pre and post processing stages should exploit information different from that used by the main detector.

We designed a pipeline architecture to combine these pre and post processing stages and the main detector. The pipeline of our road sign detection system consists of 4 stages, as shown in Figure 1. At the first stage, color space analysis is used to spot approximate ROIs. After that, the potential regions are tested against contour analysis to shrink the number and size of candidate regions. At the third stage, a Haar-wavelet cascade detector plays the major role of detecting road signs. At the final stage, these detected signs are checked by a RANSAC shape-fitting post-validator.


Figure 1: The 4-stage pipeline.

# 2 RELATED WORK

Color-based methods (Broggi, 2007; Escalera, 1997; Escalera, 2003) use thresholds within certain color-space to pick up the pixels that comprise the target object. The biggest problem of color-based methods is the poor distinguishing power and weak robustness of sole color information, and hence the difficulty in distinguishing road signs from background noises in a similar color.

Wang P., Li J. and Zhang B..
A REAL WORLD DETECTION SYSTEM - Combining Color, Shape and Appearance to Enable Real-time Road Sign Detection.
DOI: 10.5220/0003814306750678
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 675-678
ISBN: 978-989-8565-03-7

Papageorgiou (2000) showed that compositions of simple features like Haar-wavelet turned out to have a great advantage in speed while not suffering much from precision drop. Viola (2004) introduced the canonical Cascade Classifier. The problem of shape-based methods is that they are all aimed for one specific kind of visual object. If we force the training samples to contain various kinds of objects, the output will be a detector poor in both hit rate and false-alarm rate. If we train a detector for each kind of objects, the computational resources required during detecting will be overwhelming.

# 3 PIPELINE

## 3.1 Color Space Analysis

Among the many color spaces that can be used for color space analysis, our system selects HSV because it is the most coherent with the intuition of human conception. Rendering a road image in H, S and V channels, we found that road signs stand out prominently in H and S channels, but not so much in V channel. Via setting upper and lower thresholds on H and S channels, we can approximately pick out pixels that belong to a road sign. These pixels will connect with each other to form irregular regions, which, after certain image processing techniques, can be used to calculate bounding boxes that most tightly contain them. These bounding boxes form the ROIs for the next pipeline stage.

## 3.2 Contour Analysis

Computer Vision toolset like OpenCV usually provides some contour analysis tools, which can be used to extract contours from a binary image, to match a contour against a template contour, etc. Contour analysis is based on the binary image output of color thresholding. Contour matching algorithms take as input two contours and output a real number indicating the extent to which they match. A threshold can be put on this number to rule out candidates whose contours are too far away from the wanted contour. Contour algorithms do not use sliding window, thus is much faster than algorithms that are performed in a sliding window manner.

## 3.3 Haar-wavelet Cascade Detector

The design of the cascade detector is the same as Viola (2004). Because cascade detector is used in a sliding window manner, it is the biggest time consumer of the whole system, and the major way of speeding up is thus to reduce of area of regions this sliding window is performed on.

## 3.4 RANSAC Shape-fitting

When we get some points that are believed to be generated from the edge of certain shape, in principle we can recover the generating shape (i.e. its parameters) from the information provided by these points. Typically we will have much more points than theoretically needed. RANSAC (RANdom SAmple Consensus) (Fischler, 1981) is a method that exploits this redundant information to improve the precision and stability of shape fitting. It randomly selects points that are mathematically sufficient to calculate the shape parameters, and repeat this procedure certain times to get multiply sets of calculated parameters. The final values of parameters are decided by a voting among these calculated parameter sets.

In addition to gaining the values of shape parameters, RANSAC can also be used to check our presumption of shape model. For example, if we assume the generating shape is a circle, but the calculated parameter sets differ too much from each other, i.e. the statistical deviation exceeds some criteria, then we should forgo our previous assumption and claim that the shape would not be a circle. RANSAC is used in this way as a shape validator in our system.

# 4 OPTIMIZATION

To stabilize the sizes of ROIs, we implemented a tracking mechanism. When a sign is detected and ensured (for example by a consecutive series of appearances), subsequent detection will be performed only on its neighborhood, with a thorough detection every several frames to allow for new signs.

There are 25 kinds of signs as shown in Figure 2. Theoretically we should train one cascade detector for each sign, leading to 25 cascade detectors in total. The speed requirement cannot afford such an amount of computing, thus we grouped the signs into 9 groups and trained a detector for each group, as shown in Figure 2. Multiple detectors also open the possibility of parallelization.

For trade-off between hit rate and false-alarm, we prefer lowering false-alarm rate to lifting hit rate during parameter adjusting, because a low false-alarm rate can serve both to precision and speed, and

a moderate hit rate is somewhat tolerable in object detection, since the detection is measured object-wise, that is to say, the detection of an object should be considered succeeded as long as one of its many appearances is detected.

# 5 EXPERIMENT RESULTS

To measure our system's precision and speed, we conducted a suite of experiment on a set of videos captured from real road. The set consists of 33 videos with resolution 1280*960, containing totally 11345 frames. About 20% of the frames contain one or more road signs. A road sign is one of the 25 target road signs used in this experiment, as shown in Figure 2. We thoroughly annotated the bounding boxes of the signs on all frames, and used them as ground truth for testing. The training samples were extracted from a different set of videos than the testing set, containing 12,454 patches as positive samples and 384 full images as the source of negative patch samples.

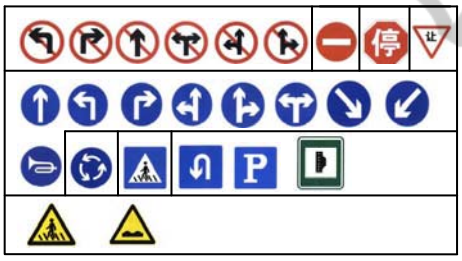

Figure 2: Road signs used in this experiment. Each cell contains a group sharing the same cascade detector.

For measurement, we defined several criteria. A detect (the bounding box of a sign) is treated as 'correct' or 'hit' is it intersects with a ground truth rectangle, and the area of intersection is larger than both 80% of the area of the detect and 80% of the area of the ground truth. Hit rate is defined as the proportion of the number of the ground truth bounding boxes been hit over the number of all ground truth bounding boxes. The number of false-alarms is the number of detects that do not hit any ground truth. We reported it in the form of false-alarms per frame. We reported the speed of the system as Frames Per Second (FPS). The speed is measured on a quad-core Intel i7 CPU with 2.8GHz main frequency.

## 5.1 Full System Performance

Table 1 shows the performance of turning on all four stages of the pipeline. A FPS of about 8 is a conservative estimate, which doesn't take advantage of the optimization techniques like tracking described in Section 4, whose results will be reported in Subsection 5.5. A false-alarm rate of about 0.13 is good enough to be based by further process such as consecutive appearing validation during tracking and validation during recognition.

Table 1: Performance of the full system. HR stands for hit rate. FA stands for false-alarms per frame. FPS stands for frames per second.

| HR | FA | FPS |
|---|---|---|
| **0.586** | **0.132** | **7.86** |

## 5.2 Effect of Contour Analysis

Turning off the contour analysis stage and thus letting all the ROIs output by the previous stage to reach the next stage, we got results shown in Table 2. Note the increase of hit rate, but also the increase of false-alarm and the drop of FPS. As the FPS was nearly halved as a result of the removal of contour analysis, and dropped to an intolerable level, we can prove that the contour analysis stage is really a crucial component of the whole system, especially when it is intended to be used in a real-time circumstance. It also justifies our bias stated in Section 4 that in object detection tasks, we should better prefer a low false-alarm rate to a high hit rate.

Table 2: Performance with Contour Analysis stage turned off.

| HR | FA | FPS |
|---|---|---|
| **0.782** | **0.244** | **3.834** |

## 5.3 Effect of Color Space Analysis

The effect of removing color space analysis is very obvious as shown in Table 3. (Because the contour analysis is based on the output mask image of color thresholding, that stage have to be also removed.) Putting the nearly unchanged hit rate and the dramatic boost of false-alarm rate aside, the FPS alone would make the system unworkable. This simple result is enough to show that color space analysis is a fundamental part of our system.

Table 3: Performance with Color Space Analysis stage turned off.

| HR | FA | FPS |
|---|---|---|
| **0.572** | **1.446** | **0.694** |

### 5.4 Effect of RANSAC

Turning off the RANSAC stage, we got results shown in Table 4. Turning off RANSAC stage doesn't affect hit rate much, but increases false-alarm rate by about 15%. Equally saying, introducing RANSAC can lower false-alarm rate by about 9% without sacrificing hit rate. The speed is not affected much either. This proves that RANSAC is a safe and feasible post processing stage for the system, though the influence is not as dramatic as color and contours.

Table 4: Performance with RANSAC stage turned off.

| HR | FA | FPS |
|---|---|---|
| **0.583** | **0.154** | **7.93** |

### 5.5 Effect of Tracking

In addition to the full system tested in Subsection 5.1, we can add simple tracking mechanism described in Section 4 to stabilize the performance and further boost the speed. The results of this addition are shown in Table 5. Surprisingly and happily, both precision and speed enjoy a significant enhancement. The precision is enhanced in that the false-alarm rate drops by more than 60% without the hit rate suffering much. The FPS is increased by about 50%. Both benefits are due to the dramatically reduced ROI sizes. Some detects and false-alarms are shown as Figure 3.

Table 5: Performance with tracking mechanism added.

| HR | FA | FPS |
|---|---|---|
| **0.535** | **0.081** | **11.750** |

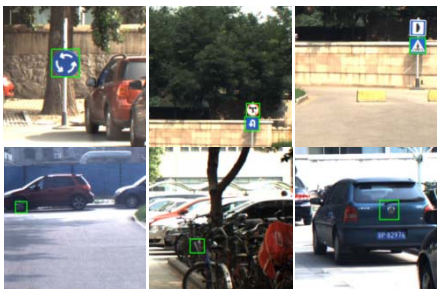

Figure 3: Some detects and false-alarms.

## 6 CONCLUSIONS

In this paper we presented a system that can fulfill the task of road sign detection in real-time and real world circumstances. By effectively utilizing various computer vision techniques, we proved that though there may not be a single algorithm that can tackle all the challenges in a real world task, a wise selection and hybrid of existing techniques can still produce a feasible and robust application.

## ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China under Grant No. 90820305, National Basic Research Program (973 Program) of China under Grant No.2012CB316301, and Basic Research Foundation of Tsinghua National Laboratory for Information Science and Technology (TNList).

## REFERENCES

Broggi, A., Cerri, P., et al., 2007. Real time road signs recognition. In *Intelligent Vehicles Symposium.* IEEE.

Escalera, A., Luis, E., Moreno, M., et al., 1997. Road traffic sign detection and classification. In *Transactions on Industrial Electronics*. IEEE.

Escalera, A., Armingol, J., et al., 2003. Traffic sign recognition and analysis for intelligent vehicles. In *Image and Vision Computing.*

Fischler, M., Bolles, R., et al., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *Communications of the ACM.*

Papageorgiou, C., Poggio, T., 2000. A trainable system for object detection. In *International Journal of Computer Vision.*

Viola, P., Jones, M.,et al., 2004. Real-time face detection. In *International Journal of Computer Vision.*