Color transformation for improved traffic sign detection

DOI:
10.1109/ICIP.2012.6466896

Document status and date:
Published: 01/01/2012

Document Version:
Accepted manuscript including changes made at the peer-review stage

COLOR TRANSFORMATION FOR IMPROVED TRAFFIC SIGN DETECTION

I.M. Creusen[1,2], *L. Hazelhoff*[1,2], *P.H.N. de With*[1,2]

[1]Cyclomedia Technology
Waardenburg, The Netherlands

[2]Eindhoven University of Technology
Eindhoven, The Netherlands

ABSTRACT

This paper considers large scale traffic sign detection on a dataset consisting of high-resolution street-level panoramic photographs. Traffic signs are automatically detected and classified with a set of state-of-the-art algorithms. We introduce a color transformation to extend a Histogram of Oriented Gradients (HOG) based detection algorithm to further improve the performance. This transformation uses a specific set of reference colors that aligns with traffic sign characteristics, and measures the distance of each pixel to these reference colors. This results in an improved consistency on the gradients at the outer edge of the traffic sign. In an experiment with $33,400$ panoramic images, the number of misdetections decreased by 53.6% and 51.4% for red/blue circular signs, and by 19.6% and 28.4% for yellow speed bump signs, measured at a realistic detector operating point.

***Index Terms*—** Traffic Sign Detection, Object Detection, Object Recognition, Color Transformation

1. INTRODUCTION

In recent years, country-wide datasets consisting of panoramic photographs have been recorded. By using these datasets, surveying of real-world objects can be enabled with image analysis. By using computer vision algorithms, this surveying process can be partially automated, leading to a further increase in efficiency.

Traffic signs are examples of objects that require occasional maintenance, because they can become damaged, dirty, vandalized or can get replaced or lost. For the governmental agencies and their contractors that have to perform this maintenance, a regularly updated survey is a useful tool.

In this paper, the focus is on traffic sign detection algorithms. This work is part of a larger automatic traffic sign surveying system described recently [1, 2]. In addition to detection, the system needs to perform classification of the specific traffic sign, and the 3D-position of the sign needs to be determined with triangulation techniques.

Previously we have shown that by using color information in the detector algorithm, its performance is substantially improved [3]. Moreover, it was shown that the selection of the color space has a large influence on the performance of the detector. We introduce a novel color transformation that solves one of the issues that normally occurs when extracting gradient features from multiple color channels. The image gradient that occurs at the outer edge of the traffic sign is regularly poorly defined, because it depends to a large degree on the contents of the background. By utilizing the knowledge of the colors that are specifically used for traffic signs, the image can be transformed in a way such that this gradient is less dependent on the background contents. It will be shown that this transformation leads to an improved detection performance.

The remainder of this paper has the following structure. In Section 2, an overview of related literature is presented. In Section 3, the dataset and recording conditions are introduced. Section 4 provides an overview of the traffic sign surveying system in which the detector is utilized. In Section 5, the novel color transformation is presented and explained in more detail. Section 6 contains experimental results and finally the conclusions are given in Section 7.

2. RELATED WORK

There is a large amount of existing research on the subject of object detection. One of the first successful object detectors is the face detector by Viola and Jones [4]. The detector described in this paper is based on the popular *Histogram of Oriented Gradients* (HOG) algorithm by Dalal and Triggs [5]. Another popular approach is the use of parts-based detection algorithms, as described for example in the work of Crandall *et al.* [6] and Leibe *et al.* [7]. Since traffic signs are very rigid and static, we do not expect a parts-based approach to offer significant advantages in our application.

In addition, there is much prior work on traffic sign detection. A large portion of this research is focused on other applications, such as driver assistance systems or autonomous vehicles. These alternative applications have different constraints from our surveying application. The alternative systems often work in real-time on video streams, while supporting a small subset of traffic signs. In contrast, our system does not need to work in real-time, it processes high-resolution panoramic images taken every 5 meters and supports most types of traffic signs. The real-time constraints of the systems related to vehicles often lead to algorithms based on the Vi-

This work was funded by the ITEA2 program under the ViCoMo project.

Fig. 1. Example of a typical spherical panoramic image from our dataset mapped onto a flat 2D surface.

ola and Jones detector, or highly specific algorithms created for this purpose. Some example systems can be found in the work of De la Escalera *et al.* [8] and Bahlmann *et al.* [9].

There is some existing work on large-scale traffic sign surveying systems. Timofte *et al.* [10] describe a similar system. The detector is based on an ROI selection, followed by a detector based on the Viola and Jones algorithm. After the detection stage, a multiview algorithm is employed to perform 3D localization of the signs. Their dataset has a limited size and consists of photographs taken every meter, making the problem considerably easier compared to our application, where the dataset is country-wide and the capturing interval is larger to keep it manageable.

3. DATASET

Our experimental dataset consists of $33,409$ street-level panoramic images taken at 5-m intervals in a rural region. The resolution of the images is 4800×2400 pixels, of which an example is shown in Figure 1. Initially, traffic signs seem easy objects for detection, however, the large variety in recording conditions, viewpoints and sign appearances make the reliable detection and classification quite challenging. More specifically, the panoramic images are captured from a moving vehicle, resulting in motion blur and a large variety in lighting conditions. Moreover, various objects can partially occlude the traffic signs and the signs themselves become damaged by collisions, dirty and they fade with age. Additionally, there is poor standardization of the traffic signs, leading to subtle different variations of signs. In addition to these variations, there is a broad variety of background objects with similar characteristics as traffic signs, such as a red character "O" in advertisements.

The set contains $73,754$ traffic signs of various classes, and each traffic sign is visible within several different panoramas. In this paper, we concentrate on the detection of red/blue circular and priority-way traffic signs as they appear in numerous cases. The training images are taken from a separate set of panoramic images. The number of training samples used in the experiments is 348 and 125 for red/blue circular signs and priority-way signs, respectively.

Fig. 2. Overview of the main algorithmic stages of the automatic traffic sign survey system.

4. SYSTEM OVERVIEW

The traffic sign surveying system consists of several components. The detection algorithm identifies the location of traffic signs in the panoramic images. The positioning algorithm calculates the 3D position of the traffic signs by combining detections from adjacent panoramas and the corresponding GPS coordinates. The classification algorithm determines to which class the identified traffic signs belong. The system architecture overview is shown in Figure 2. We will now briefly explain these components in more detail.

The used detection algorithm is based on the *Histogram of Oriented Gradients* (HOG) algorithm by Dalal and Triggs [5]. In previous work, we have described the extraction of HOG features from the color channels of various color spaces [3]. This was shown to significantly improve the performance of traffic sign detection.

The classification algorithm uses the *Bag of Words* (BoW) approach, pioneered by Csurka *et al.* [11] and Sivic *et al.* [12]. To handle the highly unbalanced data of traffic signs, the standard BoW algorithm has been modified to obtain an equal number of visual words for each class. For more information, we refer to Hazelhoff *et al.* [13].

The 3D positioning algorithm calculates the position of the traffic signs based on the detection results of several adjacent panoramas. For each 2D detection, it is known that the corresponding traffic sign should lie on the line in 3D space, extending from the location of the photograph in the direction of the 2D detection. By calculating (near) intersections between these lines, possible traffic sign candidates are created. By applying the mean shift clustering algorithm from Fukunaga and Hostetler [14] to these intersections, we can calculate the likely positions of the considered traffic signs.

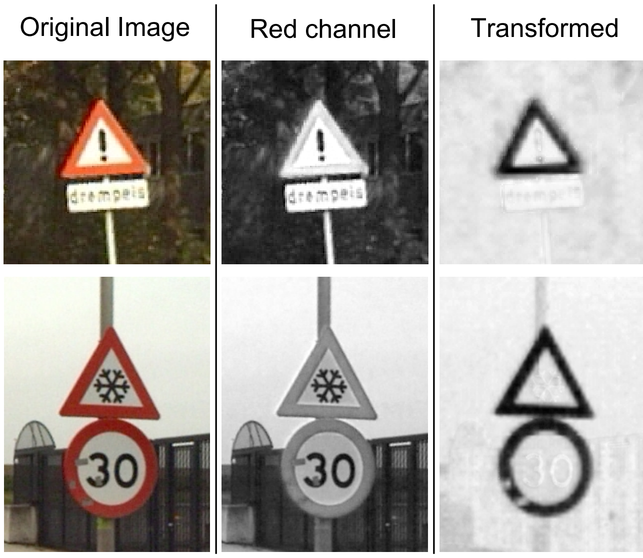

Fig. 3. Example of a gradient appearance problem solved by the color transformation. At the left: original images. The middle column contains the corresponding red channel of the image. At the right: the converted image using a red reference color and polar coordinates. The result is an identical appearance of the sign with a consistent gradient for both cases.

5. COLOR SPACE AND TRANSFORMATION

From literature [3], it was already explored that a suitable choice of the color space is helpful for enhancing the traffic sign detection performance. Hence, a proper choice of the color space is the first step of our complete color transformation. However, a simple change of color representation is not sufficient for a consistent robust performance of the detection. The problem is that the color-based detection is sensitive to the background of the image giving a decreased performance and missed detections. This is due to the changing gradient as a function of the brightness of the background. With a dark background, the direction of the gradient to the red border of the traffic sign is different than with a bright background, although the traffic sign is identical (see Figure 3). By applying (1) a color coordinate transformation and (2) taking the red border of the traffic sign as a reference in the origin, this gradient difference can be avoided. The complete color transformation involves the following two steps, employing the above three elements.

Step 1: Choice of color space. The choice of the color space in which to apply this transformation is relevant, because it determines the distance metric as specified in the next step. In this paper, we have adopted the CIElab and CIEluv color spaces in conjunction with the following Step 2, as these have shown superior performance in our previous work.

Step 2: Color Transformation. The deviation to the reference color is specified as an L2 norm, leading to

$$p_t = \|\overline{p} - \overline{p_r}\| \,. \tag{1}$$

Here, $\overline{p}$ represents the color vector of a single pixel in a certain color space, $\overline{p_r}$ is a reference color in the same color space as $\overline{p}$. Since only the difference is computed, it effectively becomes a transformation to polar coordinates with a reference color at the origin (e.g. red in Figure 3), without considering the angle. For the reference colors, we take the well-defined standardized colors for traffic signs: red, blue, white and yellow. Figure 3 shows at the right that the combination of polar coordinates and a reference color in the origin, the background becomes bright and the traffic sign appears dark and *identically* in both cases. This leads to a consistent detection and thus a higher performance.

6. EXPERIMENTS

The detection algorithm extracts HOG-like features from each of the color channels. The detectors use the CIElab or CIEluv color spaces, and perform the transform based on the previously described set of reference colors. For comparison, detectors are operated in the original setting. A HOG cell size of 8 pixels is used in these experiments. Each image is processed on 45 scales with a downsampling factor of 0.95. Embedded in the HOG detector we use a linear Support Vector Machine (SVM) classifier with a C factor of 0.01 to indicate the presence of a sign. Each cell is only normalized once with respect to a 3×3 neighborhood of cells using the L1-sqrt metric described by Dalal and Triggs [5].

To test the effect of the previously introduced color transformation, an experiment has been performed on our dataset of $33,409$ panoramic photographs, described in Section 3. The resulting precision/recall curves are shown in Figure 4 and Figure 5 for red/blue circular signs and yellow speed-bump traffic signs, respectively. The performance gain at 90% recall can be seen in Table 1, where the performance gain of the proposed transformation becomes clear.

7. CONCLUSIONS

We have introduced a color transformation that measures the distance in color space of each pixel to a set of reference colors to improve the detection of traffic signs in panoramic images. These reference colors are chosen to be the standardized colors for traffic signs, thereby measuring how well each pixel matches these colors. It is shown that this transformation improves the consistency of gradients at the outer edge of traffic signs. Our optimized color transformation improves the detector performance in an experiment using $33,409$ panoramic images. In the CIElab and CIEluv color spaces, the number of misdetections in a realistic detector operating point (0.9 recall) decreased by 53.6% and 51.4% for red/blue circular signs, and by 19.6% and 28.4% for yellow speed bump signs.

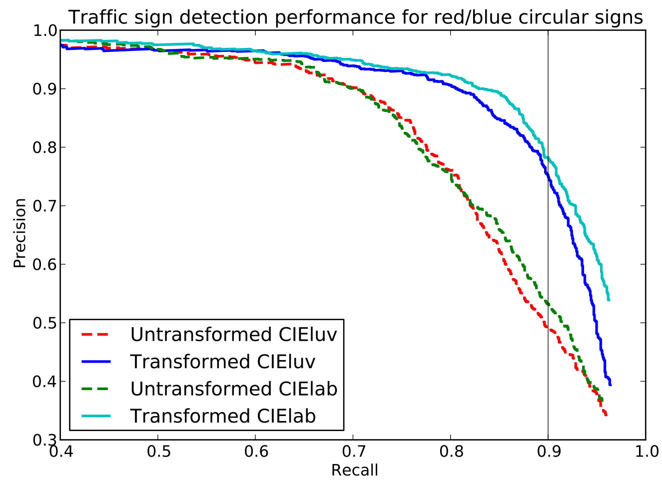

Fig. 4. Detection results for red/blue circular traffic signs, presented as an ROC curve. The vertical line at 90% recall indicates where we measure the detection improvement.

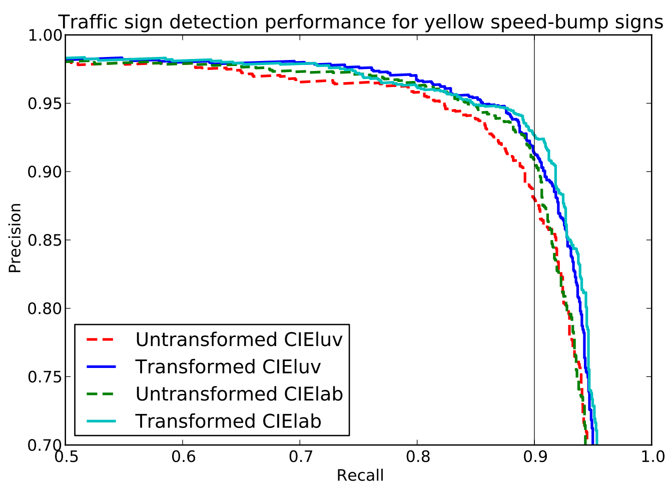

Fig. 5. Results for yellow speed-bump traffic signs, presented as an ROC curve. The vertical line at 90% recall indicates where we measure the detection improvement (see Table 1).

	Untransformed	Transformed	Error reduct.
RBC CIElab	0.532	0.783	53.6%
RBC CIEluv	0.491	0.753	51.4%
YSB CIElab	0.909	0.927	19.6%
YSB CIEluv	0.881	0.915	28.4%

Table 1. Results for red/blue circular (RBC) and yellow speed bump (YSB) traffic signs. All performance figures are precision scores at 0.9 recall. This represents a realistic operating point for the detector.

8. REFERENCES

[1] I. M. Creusen, L. Hazelhoff, and P. H. N. de With, "A semi-automatic traffic sign detection, classification and positioning system," in *Proc. of the IS&T/SPIE Electronic Imaging*, San Francisco, USA, Jan. 2012.

[2] L. Hazelhoff, I. M. Creusen, and P. H. N. de With, "Robust detection, classification and positioning of traffic signs from street-level panoramic images for inventory purposes," in *Proc. of the Workshop on Applications of Computer Vision (WACV)*, Breckenridge, USA, Jan. 2012.

[3] I.M. Creusen, R.G.J. Wijnhoven, E. Herbschleb, and P.H.N. de With, "Color exploitation in hog-based traffic sign detection," in *Proc. of the IEEE International Conference on Image Processing (ICIP)*, Sept. 2010, pp. 2669–2672.

[4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2001, vol. 1, pp. 511–518.

[5] N. Dalal and B. Triggs, "Histogram of oriented gradients for human detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2005, vol. 1, pp. 886–893.

[6] D. Crandall, P. Felzenszwalb, and D. Huttenlocher, "Spatial priors for part-based recognition using statistical models," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2005, vol. 1, pp. 10–17.

[7] B. Leibe, A. Leonardis, and B. Schiele, "Combined object categorization and segmentation with an implicit shape model," in *Proc. of the ECCV workshop on statistical learning in computer vision*, pp. 17–32.

[8] A. de la Escalera, L.E. Moreno, M.A. Salichs, and J.M. Armingol, "Road traffic sign detection and classification," *IEEE Transactions on Industrial Electronics*, vol. 44, no. 6, pp. 848–859, Dec. 1997.

[9] C. Bahlmann, Y. Zhu, and V. Ramesh, "A system for traffic sign detection, tracking, and recognition using color, shape and motion information," in *Proc. of the IEEE Symposium on Intelligent Vehicles*, 2005, pp. 255–260.

[10] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3d localisation," in *Proc. of the Workshop on Applications of Computer Vision (WACV)*, Dec. 2009, pp. 1–8.

[11] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. of the European Conference on Computer Vision (ECCV)*, May 2004.

[12] J. Sivic, B.C. Russell, A.A. Efros, A. Zisserman, and W.T. Freeman, "Discovering objects and their location in images," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, Oct. 2005, vol. 1, pp. 370–377.

[13] L. Hazelhoff, I. M. Creusen, D. W. J. M. van de Wouw, and P. H. N. de With, "Large-scale classification of traffic signs under real-world conditions," in *Proc. of the IS&T/SPIE Electronic Imaging*, San Francisco, USA, Jan. 2012.

[14] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, Jan. 1975.