MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

Combined generation of road marking and road sign databases applied to consistency checking of pedestrian crossings

Lykele Hazelhoff[a], Ivo Creusen[a], Thomas Woudsma[a], Xinfeng Bao[b] and Peter H.N. de With[b]

[a] CycloMedia Technology B.V.
Zaltbommel, The Netherlands
lhazelhoff@cyclomedia.com
ivocreusen@gmail.com

[b] Eindhoven University of Technology
Eindhoven, The Netherlands
thomaswoudsma@gmail.com
p.h.n.de.with@tue.nl

Abstract

Combined road marking and traffic sign databases are beneficial for both road maintenance and for usage within navigation devices and autonomous driving vehicles. The combination of both markings and signs completely provides all instructions and legislation for drivers. This paper presents a conceptual system for the automated creation of such combined databases and investigates the benefit of this combination for the specific case of pedestrian crossings. Evaluations on 62 km of road have shown that individual detection of road signs and markings indicating pedestrian crossings is very accurate ($\geq 95\%$), enabling selective safety analysis towards these specific locations. Combining both approaches enables very accurate identification of crosswalks and additionally leads to the retrieval of crossings with undetectable markings or signs, such that maintenance can be directed specifically towards these potentially dangerous crosswalks.

1 Introduction

Accurate mapping of road signs and road markings is essential for maintaining a high road safety, as the resulting databases allow for checking discrepancies between the desired and observed sign and marking placements. Such databases can also be embedded in navigation devices, for example to warn road users for upcoming alerting situations such as give-way locations or pedestrian crossings. These databases also may contribute to a higher navigation quality, e.g. by selecting the most efficient route that features the least amount of stops or give-way situations. These databases are also beneficial for usage within autonomous driving vehicles, i.e. by already decreasing speed when a give-way situation is coming up along its route, even when the sign is not yet visible. This also contributes to a smooth ride and a reduction in fuel consumption. This paper describes a combined approach for the recognition of road markings and road signs from street-level images, aiming at the reliable mapping of pedestrain crossings to enable novel safety and consistency checks.

Generation of databases of road signs and road markings from street-level images is an popular research topic. For example, several systems are described for creating surveys of road signs from street-level images [1, 2, 3, 4]. Most of such systems focus on the recognition of signs in single images, while only few also identify their 3D positions [2, 3]. Road-marking recognition systems can be based on LiDAR [5] or can exploit street-level images [6, 7, 8, 9]. Combined generation of such databases is less commonly explored,

Figure 1: Example of the road markings and road sign indicating a pedestrian crossing.

although Choi *et al.* [10] describe the combination of crosswalk and traffic light detection.

Although most legal information is denoted by road signs, combined databases of traffic signs and road markings allow for more detailed assessments of the actual road situation. Therefore, this paper explores the combined recognition of road signs and markings, which is attractive for the following reasons.

- *Accuracy*: focusing on two different types of information (markings and signs) containing partially redundant information which can be exploited to attain a higher surveying accuracy.
- *Consistency checking*: the co-occurrence and colocation of signs and markings that contain redundant information can be validated. For example, each pedestrian crossing should be indicated by both road markings and signs.
- *Information content*: road signs and road markings comprise both redundant and complementary information. In some situations, they provide redundant information (i.e. in case of a give-way situation), while in other situations only signs or markings are present. Combined databases will therefore provide a larger amount of information.
- *Localization*: road signs are placed at a certain distance in advance of the situation they apply to, while road markings indicate the exact location. For example, a stop sign can be placed up to 10 meters before the stop line.

This paper describes a conceptual system for the combined recognition of traffic signs and road markings, aiming at the generation of a combined database of road markings and signs. Additionally, we will present a case study for the surveying of pedestrian crossings, which is important for various tasks. First, a database of pedestrian crossing locations allows for safety inspections at those locations, which are necessary to maintain a high road safety, as missing or poorly visible signs or road markings cause potentially dangerous situations. Such databases not only allow for efficient checking based on the street-level images, but additionally feature automated consistency valida-

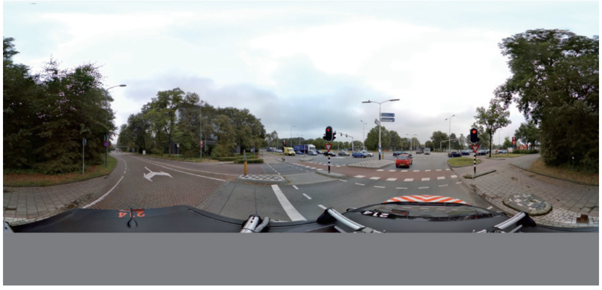

Figure 2: Example of an equirectangular input image.

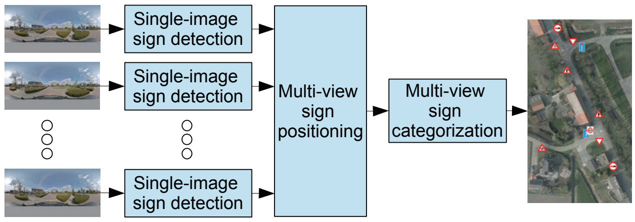

Figure 3: Overview of road-sign recognition system.

tions, such that manual checks can be reduced, where the remaining checks are focused on the potentially unsafe crosswalks. Second, such databases aid both navigation (as routes with lots of pedestrian crossings may cause large delays) and autonomous driving vehicles (since they can anticipate them, even if they are not yet visible from the car itself). Figure 1 shows an example of a pedestrian crossing and its road sign.

Although we concentrate on surveying pedestrian crossings, it should be noted that the proposed methodology is generic towards other traffic signs and road markings, as learning-based approaches are used.

The remainder of this paper starts with a description of the employed source data in Sect. 2. We then provide a brief overview of the road-sign and road-marking recognition systems in Sect. 3 and Sect. 4, respectively. The performed experiments and results are described in Sect. 5, followed by the conclusions in Sect. 6.

2 Source data description

The combined recognition systems described in this paper operate on street-level panoramic images, which provide a recent and accurate overview of the road infrastructure. These images are acquired at a large scale and are recorded at all public roads within the target area, using a capturing interval of 5 m. The recording vehicles drive along with regular traffic at normal speeds. The cars are utilized in an efficient way by capturing during daytime during all kinds of weather conditions, including sunny, cloudy and foggy weather, and directly after (but not during) rain or snow.

The panoramic images have a resolution of $2,400 \times 4,800$ pixels. The capturing location is also accurately known for each image, based on a high-quality positioning system featuring both GPS and IMU devices.

The employed capturing systems are calibrated precisely, resulting in panoramic images that are mapped to a sphere, on which angular distances can be measured. The resulting images are stored as equirectangular images, which have a linear relationship between the pixel coordinates within the image and the viewing directions in horizontal and vertical directions. This allows for the calculation of the real-world 3D positions based on triangulation. The position of an object can be retrieved in case multiple points (≥ 2) corresponding with the considered object are found in multiple images, using straightforward geometrical computations. Figure 2 displays an example image.

3 Road-sign recognition system

The road-sign recognition system processes all panoramic images captured within a region of interest, and is described extensively in [3]. Figure 3 displays the system overview. Each main system component is briefly summarized below, and more extensively in [3].

3.1 Single-image sign detection

At first, each panoramic image is analyzed and the present signs are detected by multiple, independent detectors, each focusing at a specific class of signs (i.e. *red circular* restriction signs). These detectors are based on the popular Histogram of Oriented Gradients [11] technique. We have extended the original approach with the use of color gradient information, in order to exploit the color transitions, both within the signs and between the sign and surroundings.

The road-sign recognition system currently employs detectors for 11 different sign appearance classes.

3.2 Multi-view sign positioning

The next step is computing the real-world coordinates of the road signs by combining all detections representing the same physical sign. This process exploits the geometrical properties of the used images, which directly results in a 3D location when ≥ 2 points corresponding to the same object are known in different images. Each detection is combined with all other detections of the same sign appearance class found in closeby images, and each combination results in an hypothesis of the sign position, which are clustered around the real sign position. These clusters are retrieved using the Mean Shift Algorithm [12], with the constraint that each sign contains at least 3 detections. This process operates independently per sign appearance class, such that for each positioned 3D sign both its location and sign class are available.

3.3 Multi-view sign categorization

Each positioned 3D sign is assigned a sign category. This categorization task is based on Bag of Words [13], which aims at the identification of the sign category, based on the occurrence counting of specific key patterns. Our approach exploits densely extracted SIFT descriptors as key patterns and compares these patterns to a large dictionary ($\sim 10,000$ entries, different for each sign appearance class) and then determines the sign category using a linear SVM. Each detection used during positioning of the sign is categorized independently, after which weighted voting is employed to retrieve the category of the positioned sign. This process also involves estimation of the sign orientation angle, indicating the road lane a sign applies to. This processing stage is extensively described in [14].

All found signs have now a location and sign type. Currently, the road-sign recognition system discriminates 176 different sign types. As this paper focuses at pedestrian crossings, we will further ignore all non-pedestrian crossings signs.

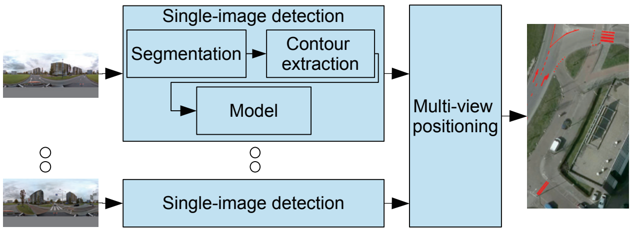

Figure 4: System architecture of the road-marking recognition system.

4 Road-marking recognition system

The road-marking recognition system processes all panoramic images captured within a region of interest and follows a similar architecture as the recognition systems described in literature [6, 7, 8, 9]. Figure 4 displays the system overview, of which each main system component is summarized below.

4.1 Single-image detection

Prior to determining the positions of the road markings, each panoramic image is remapped to a top-down view using Inverse Perspective Mapping (IPM), which is then processed as follows.

Image segmentation: In general, road markings are brighter than the road, so image regions that have a high local intensity are extracted. This involves calculation of the intensity difference between the grayscale pixel values and the average graycale intensity in a rectangular window around each considered pixel. A binary segmentation is then obtained by applying Otsu's threshold method on the found differences. Next, all connected components (groups of neighboring pixels) are extracted.

Contour classification: The contour of each connected component is extracted and represented by a feature vector to determine the road-marking type. First, all contours are translated to their centroid (geometric center) to attain translation invariance. Using Principal Component Analysis, each contour is then rotated to align its primary axis to be rotation invariant. Scale invariance is not used here as it is relevant for determining the marking type. Secondly, we extract the distance from the contour centroid to the edge at set angular intervals as feature values, since road markings have highly regular and convex shapes. This is illustrated in Fig. 5. Finally, the feature vector of each contour is classified by a set of SVMs, each trained to recognize different road-marking types, e.g. stripes, blocks and arrows.

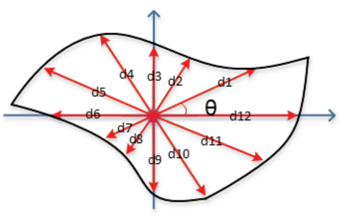

Figure 5: Illustration of the shape descriptor, which consists of a set of samples of the measured distances between origin and border at fixed angular intervals.

Model evaluation: Road markings often occur in specific patterns, e.g. a pedestrian crossing has equally-sized rectangles at regular intervals with equal orientations. These patterns can be modeled to link the individual markings together into e.g. a crosswalk, and to reduce false detections. The model evaluates all detected road markings and grouped detections that adhere to its adjacency and orientation rules, and discards detections not satisfying the model. The model evaluation improves the recognition results compared to the individual contour classification and pursues the recognition of high-level elements (i.e. crosswalks).

4.2 Multi-view positioning

Afterwards, the 3D positions of each detection are retrieved and multiple detections corresponding to the same marking, are merged. The image positions of the recognized markings can be converted to relative positions w.r.t. the capturing locations. Based on the known capturing locations, the 3D world coordinates of the markings can be computed. Afterwards, detections are merged using the Mean Shift Algorithm [12], where only clusters containing at least 2 detections in closeby images are accepted.

5 Experiments and results

5.1 Dataset description

Evaluations are conducted on a dataset containing all street-level panoramic images captured within a number of geographical regions. Each region contains both a number of pedestrian crossings, as well as numerous roads without such crossings. The combination of various geographical regions results in a dataset with a large number of different traffic situations, varying from roads in residential areas to main roads and parking lots of shopping malls. Also, because of its size, the dataset also involves large variations in included weather conditions, as each region is captured on a different day. The total dataset contains $12,373$ panoramic images (corresponding to 61.8 km of road). All enclosed pedestrian crossings and road signs denoting pedestrian crosswalks are manually annotated as ground truth by browsing all images and marking their 3D positions.

5.2 Experimental setup

Experiments are conducted at two different levels.

First, we independently assess the performance of the road-sign and road-marking systems by comparing the found signs and markings with the corresponding ground truth, based on a matching distance of 4 m.

Second, we evaluate the performance of the combined recognition approach, where we focus on consistency checking to retrieve a list of pedestrian crossings from which either the road markings or signs are not detected. During this experiment, connected pedestrian crossings (as shown in Fig. 6a) are grouped together, clarifying their lowered number w.r.t the first experiment. We consider a crossing as *consistent* when all connected marking detections have at least 6 stripes in total (needed to cover both road sides) and the crossing is accompanied by signs that are visible from all driving directions, which is evaluated based on the sign orientation angles.

Approach	Correctly det.		False det.
Marking recognition	166	94.9%	3
Sign recognition	163	98.8%	5

Table 1: Results of individual recognition.

Approach	Correctly det.		False det.
Combination	104	98.1%	6
-inconsistent	46	44.2%	
signs not ok	37	35.6%	
markings not ok	18	17.3%	

Table 2: Results of consistency checking.

5.3 Individual recognition results

The recognition results are shown in Table 1. As follows, both approaches identify the vast majority of the signs and markings, where most missed markings are caused by worn out markings, as shown in Fig. 6b.

With respect to processing time, the traffic sign recognition system takes about 88.5 s per image on average [3]. The road marking recognition system requires 10 s per image on average. Both systems use single core implementations using MatLab and C++.

5.4 Combined results and consistency checking

Table 2 shows the combined recognition results of the road-markings and road-sign detectors for consistency checking. It should be noted that connected crosswalks are grouped together during this experiment, such that the results cannot be compared with the previous experiment. Our approach missed only 2 of the pedestrian crossings, which clearly indicates that safety checking based on the generated database is both accurate and efficient compared to manual inspections of all images. When assessing the consistency of the crossings, which enables specific manual safety checking of crosswalks with undetected markings or signs, we have found that over half of the detected crossings satisfy the set criteria (wide enough crosswalk to cover a two-way road and signs with correct orientation angles). Interestingly, we have found that inconsistencies in signs and markings seem correlated, i.e. undetectable markings and signs often coincide. This can be partly explained by aging, as crosswalks that are placed longer ago are more likely to be worn-out and have dirty, skewed or damaged signs.

6 Conclusions and future work

This paper has presented an approach for the generation of traffic signage databases from street-level panoramic images, based on a combination of road-sign and road-marking recognition. Such a combination leads to higher surveying accuracies, enables extraction of complementary information from both sources, and allows for consistency checking between markings and signs. In this paper, we have explored this combination for the surveying of pedestrian crossings.

(a) (b)

Figure 6: (a): Connected crossing, counted as single crossing. (b): Worn-out markings.

Experiments have shown that the individual recognition of markings and signs is very accurate ($\geq\sim 95\%$). The explicit use of combined recognition allows for a higher level of road signage detection by indicating the consistency of markings and signs. This paves the way for a reliable safety indication, and enables specific checkings of inconsistent crosswalks. In our experiments, more than 50% of the crosswalks are detected with consistent markings and signs.

Our combined recognition approach for traffic signs and road markings can be extended to other combinations denoted by both road markings and signs, such as priority situations and lane configurations.

References

[1] A. Ruta, F. Porikli, S. Watanabe, and Y. Li, "In-vehicle camera traffic sign detection and recognition," *Mach. Vision Appl.*, vol. 22, no. 2, pp. 359–375, 2011.

[2] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3d localisation," *Machine Vision and Applications*, dec. 2011.

[3] L. Hazelhoff, I. Creusen, and P.H.N. de With, "Exploiting street-level panoramic images for large-scale automated surveying of traffic signs," *Machine Vision and Applications*, vol. 25, no. 7, pp. 1893–1911, 2014.

[4] S. Maldonado-Bascon, S. Lafuente-Arroyo, and P. Gil-Jimenez *et al.*, "Road-sign detection and recognition based on support vector machines," *Intelligent Transportation Systems*, vol. 8, no. 2, pp. 264 –278, 2007.

[5] Y. Yu, J. Li, and H. Guan *et al.*, "Learning hierarchical features for automated extraction of road markings from 3-d mobile lidar point clouds," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. PP, no. 99, pp. 1–18, 2014.

[6] C. Li, I. Creusen, L. Hazelhoff, and P.H.N. de With, "Detection and recognition of road markings in panoramic images," in *ACCV Workshop on My Car Has Eyes - Intelligent Vehicles with Vision Technology*, 2014.

[7] B. Qin, W. Liu, and X. Shen *et al.*, "A general framework for road marking detection and analysis," in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*, pp. 619–625, 2013.

[8] H. Li, M. Feng, and X. Wang, "Inverse perspective mapping based urban road markings detection," in *Cloud Computing and Intelligent Systems (CCIS), International Conference on*, vol. 03, pp. 1178–1182, 2012.

[9] P. Foucher, Y. Sebsadji, and J.-P. Tarel *et al.*, "Detection and recognition of urban road markings using images," in *Intelligent Transportation Systems (ITSC), International IEEE Conference on*, pp. 1747–1752, 2011.

[10] J. Choi, B. T. Ahn, and I. S. Kweon, "Crosswalk and traffic light detection via integral framework," in *Frontiers of Computer Vision, (FCV)*, pp. 309–312, 2013.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005.

[12] D. Comanicu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603–619, 2002.

[13] G. Csurka, C. R. Dance, and L. F. *et al.*, "Visual categorization with bags of keypoints," in *Proc. European Conference on Computer Vision (ECCV)*, 2004.

[14] L. Hazelhoff, I. Creusen, and P.H.N. de With, "Robust classification of traffic signs using multi-view cues," in *Image Processing (ICIP), Int. Conf. on*, 2012.