Multi-camera system for traffic light detection: About camera setup and mapping of detections

Conference Paper · October 2017

DOI: 10.1109/ITSC.2017.8317946

3 authors, including:

Julian Müller
Daimler
8 PUBLICATIONS **51** CITATIONS

SEE PROFILE

Multi-Camera System for Traffic Light Detection: About Camera Setup and Mapping of Detections

Julian Müller
Department of Measurement, Control and Microtechnology
University of Ulm
Albert-Einstein-Allee 41
Ulm, Germany
Email: julian-2.mueller@uni-ulm.de

Andreas Fregin
Research and Development
Daimler AG
Wilhelm-Runge-Straße 11
Ulm, Germany
Email: andreas.fregin@daimler.com

Klaus Dietmayer
Department of Measurement, Control and Microtechnology
University of Ulm
Albert-Einstein-Allee 41
Ulm, Germany
Email: klaus.dietmayer@uni-ulm.de

***Abstract*—Traffic light recognition is an object detection task mainly investigated on monocular images from a single camera [1], [2], [3], [4], [5] usually one with a small field of view. They guarantee high object resolution and less distortion. Both properties support common state of the art detectors. However, a single-camera setup only covers a certain distance range. Traffic lights in close range disappear from the field of view. When stopped at the stop line of an intersection, the vehicle can-not recognize the toggle process of the traffic light. In this paper a two-camera setup for traffic light recognition is presented. A stereo camera is used for far and moderate range, while for close range a wide-angle camera is used. The fields of view of both cameras overlap. In particular, the transition of a detection in the stereo camera to the wide-angle camera is presented. Objects detected by the stereo camera are handed over to the wide-angle camera. Since the hand-over is erroneous, a template matching correction presented. The accuracy of the hand-over as well as a subsequent correction method is evaluated.**

I. Introduction

Research on traffic light recognition has increased strongly during recent years. Especially in context of autonomous driving a reliable detection of traffic lights is crucial. Previously introduced methods like the color detection [6], [2], [4], [5] or shape detection [6], [3] already reached satisfying results. However, most systems in the related work used camera setups incapable of providing a sufficient solution for autonomously driven vehicles. Traffic lights have to be detected up to the stop line. In a challenging situation, the vehicle has to brake near to the stop line. In this case a single-camera setup with a narrow field of view fails as the traffic lights are not visible in the camera images up to the vehicle coming to a complete stop. In the related work, the field of view of the camera is only rarely described. Fairfield et al. [7] mention their field of view of 30 degrees, which has advantages for distant traffic lights. Traffic light detection at close-range is not possible with such a camera. Section II describes the camera setup used in this paper. It consists of two cameras, which operate on different distance ranges. The camera setup and properties are chosen in order to fulfill this task. The main contribution of this paper is presented in Sections III and IV. The mapping of a detection from the stereo camera frame into the wide-angle camera frame and its correction is described.

Fig. 1. Identical scene in two camera frames. The stereo camera image is on the left, whereas the wide-angle camera image is on the right. Two detected traffic lights in the stereo camera image are mapped into the wide-angle camera system as they will move out of the image during the next frames. Hence, close range detection is performed by the wide-angle camera system.

Figure 1 illustrates an identical scene in both camera frames. The stereo camera image is given on the left, whereas the wide-angle camera image is given on the right. Both detected green traffic lights are bounded by a green bounding rectangle. Since they will move out of the image during the next frames, those objects must be handed over to the wide-angle camera. In this camera frame the objects are detected and tracked until the vehicle passes the objects in real-world. As already mentioned, previous publications in traffic light recognition introduced detection algorithms based only on one mono or stereo camera. Multi-view systems are not published. Publications with multi-camera systems are available in related research topics such as person or pedestrian detection and tracking. Hale et al. [8] present an indoor person tracking system based on two stereo cameras. Hamdoun et al. [9] presented a multi-camera system in order to re-identify persons in a video sequence. However, their algorithm is based on interest point descriptors as they have non-overlapping camera fields of view. Nummiaro et al. [10] use two monocular cameras. Thus, they can only calculate epipolar lines on the other image planes, on which they estimate a detected object from another camera. Most presented applications differentiate themselves from our problem because objects are not tracked in both cameras during longer periods. Since we assume the vehicle to either stop or move forward, a disappeared object from the stereo camera is not likely to appear again. Once mapped, an object is processed further by the second camera system. Since no

Fig. 2. Illustration of the research vehicle as well as two views from both cameras, which express the environment perception in our traffic light detection system. This visualization clearly shows the need for wide-angle cameras at close range. Two traffic lights are in the transition area between both frames. Four traffic lights are only visible in the wide-angle camera as they are too close to the vehicle and thus outside the field of view of the stereo camera.

public dataset provides stereo and multi-view data we present mapping results evaluated on our own dataset.

II. Camera Setup

The camera setup was selected to guarantee detection of all traffic lights at a defined distance range. Consequently, three main requirements arise: The required number of cameras, the specification of the field of view and a sufficient camera resolution.

A. Number of Cameras

As already mentioned, traffic lights have to be detected up to passing the stop line at an intersection. On streets in Europe it is common that traffic lights are directly installed at stop lines. While a human resolves this situation with turning the head, an autonomous vehicle needs to cover such cases with the sensor setup. Using cameras with a high field of view provide an opportunity with one crucial disadvantage. High field of views strongly decrease the resolution of objects in the camera image. Hence, high fields of views are coupled to high required camera resolutions. In application to autonomous driving very high resolutions are critical as real-time requirements have to be guaranteed. A multi-camera setup avoids the need for high resolution cameras and decreases camera resolution requirements as described just as it decreases the computation effort per image cycle. We decided for a two-camera setup consisting of a stereo and a wide-angle camera for close-range. The stereo camera providing 3D information is mandatory as is needed for the presented mapping.

B. Field of View Specification

Figure 2 illustrates the research vehicle as well as both camera views. A stereo camera with a field of view of 60 degrees covers the distance range from approximately 20 up to infinity. The camera sensor with a resolution of 2048x1024 pixels provides a sufficient resolution for this task without unnecessarily increasing the computational cost. The additional three-dimensional information which can be calculated is of great importance for the described mapping introduced in Section III. The point cloud is also illustrated in the figure. The bright image visualization shows the viewing area of the wide-angle camera. The wide-angle camera is used for close-range traffic light recognition because the stereo camera has a limited viewing area. The specifications of the wide-angle camera are defined mainly by the number of lanes n_l on a street which have to be monitored. The number of lanes leads to a total width w when assuming a fixed lane width w_l. Furthermore, the vehicle stops at a resulting minimum distance d_{stop} in front of the traffic light. Thereby the necessary horizontal field of view

$$\gamma_h = 2 \cdot \arctan\left(\frac{w}{d_{\mathrm{cam}} + d_{\mathrm{stop}}}\right) = 2 \cdot \arctan\left(\frac{n_{\mathrm{l}} \cdot w_{\mathrm{l}}}{d_{\mathrm{cam}} + d_{\mathrm{stop}}}\right) \quad (1)$$

can be calculated. The distance between vehicle front end and the camera mounting is denoted as d_{cam}. The vertical field of view γ_v is mainly defined by the height h_{TL} of the traffic light installed above the road. It can be calculated as

$$\gamma_v = 2 \cdot \arctan\left(\frac{h_{\mathrm{TL}} - h_{\mathrm{cam}}}{d_{\mathrm{cam}} + d_{\mathrm{stop}}}\right) \quad (2)$$

while still assuming a minimum stop distance d_{stop}. Please note, that this equation is valid for a front-looking camera without tilt angle. Requirements on the horizontal viewing angle decrease with rising tilt angle.
Figure 3 illustrates vertical and horizontal viewing angles as well as the important geometries.

Our datasets recorded on German streets mainly contain intersections, in which no or at most one lane must be monitored. However, at some intersections two lanes need

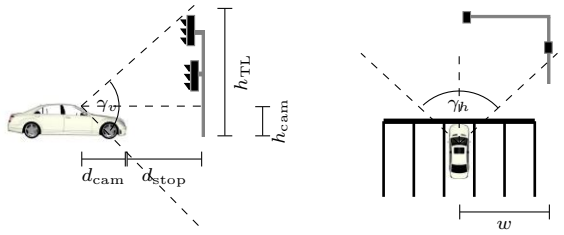

Fig. 3. Illustration of the geometries for calculating the necessary field of view. The left subimage shows the vertical viewing angle γ_v, which is mainly dependent on the traffic light height h_{TL} and the distance of the camera to the traffic light. The right subimage shows the vertical horizontal angle γ_h, which is dependent on the number of lanes to oversee and the distance to the traffic light.

to be monitored, which strongly increases the horizontal viewing angle. In order to also guarantee a detection of traffic lights at intersections with five lanes, the wide-angle camera has a horizontal viewing angle of $\gamma_h = 130°$. The minimum installation height of traffic lights in Germany amounts 2.1 meters above sidewalks and 4.5 meters above the road. Pedestrians and cyclists on sidewalks as well as trucks are protected against collisions. This installation height mainly defines the required vertical viewing angle γ_v.

C. Camera Resolutions

Besides the field of view a sufficient camera resolution is crucial for effective traffic light recognition.

Figure 4 illustrates the resolution of a traffic light in pixels with respect to its real-world distance to the camera origins. Wide fields of view have the disadvantage of lower resolutions per degree. This can be clearly seen by comparing the curves of the stereo and wide-angle cameras. As a result, higher resolutions must be used which are critical in autonomous driving application. We use a frame rate of 15 Hz, i.e. the maximum computing time for one frame encompasses approximately 67 ms. As already mentioned, the wide-angle camera detects traffic lights at distances up to approximately 20 m. For this reason the characteristic curve of the stereo

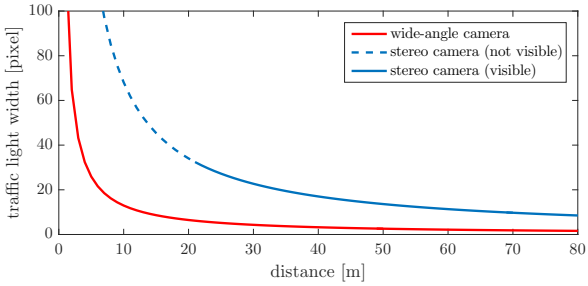

Fig. 4. Comparison of the resolution of a traffic light in both cameras with respect to the real-world distance in meters. The stereo camera has a significantly higher resolution because of less horizontal viewing angle. Since the objects leave the field of view of the camera (dashed line), a wide-angle camera is required.

camera is shown as a dashed line. It shows the theoretical size of the traffic light object at distances below 20 m. In practice, the real-world horizontal offset (traffic light is installed few meters besides the road) causes leaving the field of view earlier.

III. Mapping of Detections

The principle of our traffic light recognition system is as follows:

1) Detection in the stereo camera frame
2) Transition to the wide-angle camera frame at a defined distance range
3) Further detection in the wide-angle camera frame up to passing

In this paper we mainly want to present the mapping step.

A. Detection System

For the purpose of completeness we briefly introduce the detection system in the stereo camera frame. Figure 5 illustrates our system together with the input data. The left camera image as well as the disparity image are required. The first step consists of three proposal generators. We use three different methods as a single proposal generator does not reach sufficient detection results. The color proposal generator classifies each pixel as red, green, yellow or background. Connected components are combined to one color hypothesis. The spotlight detector is implemented similar to [4]. A top-hat operation followed by edge detection and contour finding algorithm delivers blob hypotheses. The third proposal approach uses hough transform [11] to detect circular objects in the image. All hypotheses are combined and transformed from blob to object hypotheses. Object hypotheses do circumscribe the traffic light housing. Therefore we use the disparity-based method described in [12] as it increases the detection accuracy for imperfect proposals. In a subsequent step those hypotheses are classified. We propose to add background around the traffic light hypotheses as the edges to the background can be strong features depending on the type of feature and classifier. A tracking module can improve both, detection rates as well as the false positive rate. Please note, that implementation details are not stated as this is not the scope of this paper.

B. Mapping

In our system traffic lights are initially detected in the left stereo camera image. This means, that traffic light detection in the whole camera image is only done in the stereo camera image. After mapping those detections to the wide-angle camera, only existing detections are processed further. The mapping of image points describing the traffic light detection can be formulated as

$$\underline{x}_1^p \rightarrow \underline{x}_2^p, \tag{3}$$

where $\underline{x}_1^p$ describe the detection in the stereo camera frame and $\underline{x}_2^p$ the detection in the wide-angle camera frame. In a

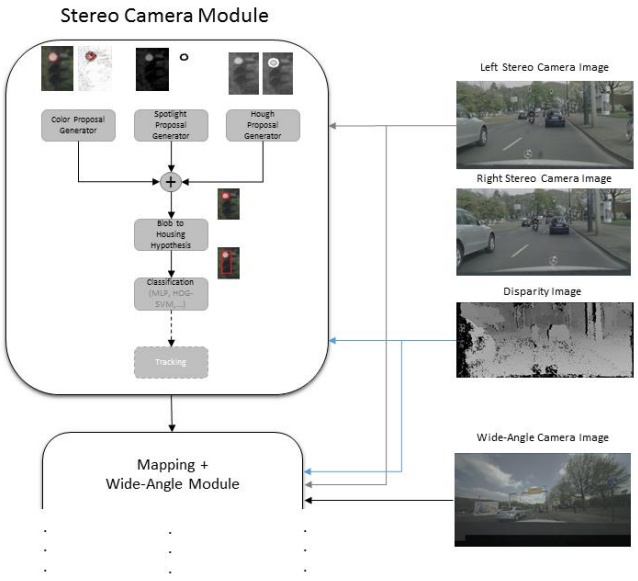

Fig. 5. Detection system before the mapping and wide-angle unit. Three proposal generators, which detect candidates based on color, shape and intensity are used. We use blob shape, geometry and disparity information to create hypotheses circumscribing the traffic light housing according to [12]. A classification module (e.g. support vector machine with HOG features, multilayer perceptron) verifies each hypothesis as traffic light or no traffic light. The remaining candidates are tracked.

first step, the detected traffic lights two-dimensional position is projected into the camera coordinate frame. Hence, a three-dimensional point is received by

$$\underline{X}_1^c = \mathbf{K}_1^{-1} \underline{x}_1^p, \tag{4}$$

where $\mathbf{K}_1$ is the intrinsic camera matrix of the stereo camera. Since the stereo camera and wide-angle camera position and orientation in real-world differs, those coordinates are mapped into the wide-angle camera system. The 3D point is multiplied by an extrinsic camera matrix $[\mathbf{R}|\underline{t}]$ containing orientation and translation between both frames. This leads to

$$\underline{X}_2^c = [\mathbf{R}|\underline{t}]\mathbf{K}_1^{-1} \underline{x}_1^p \tag{5}$$

and gives the 3D point in the other camera (wide-angle camera system). Multiplying this three-dimensional point with the intrinsic camera matrix $\mathbf{K}_2$ of the wide-angle camera leads to the equation

$$\underline{x}_2^p = \mathbf{K}_2^{-1}[\mathbf{R}|\underline{t}]\mathbf{K}_1^{-1} \underline{x}_1^p, \tag{6}$$

which describes the mapping of a point $\underline{x}_1^p$ in the stereo camera frame into the wide-angle camera system. All equations are based on the common pinhole camera model [13], [14]. Mapping two points (i.e. lower left and upper right corner) of a traffic light is sufficient to describe a bounding rectangle in both frames. Nevertheless all four corner points were used because distortions may result in non-rectangular traffic lights in the camera image.

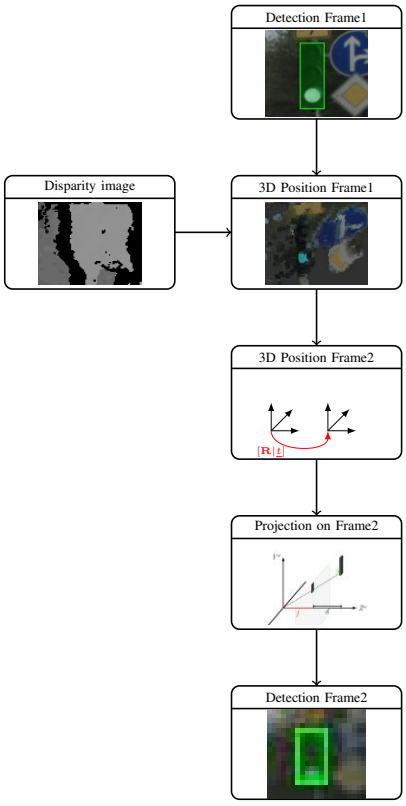

Fig. 6. Scheme of mapping a detection between two camera systems. The primary detection continues on the stereo camera images (frame 1). The green bounding box illustrates a detected traffic light. By using the disparity image the detection can be projected into 3D space. Because of the available camera calibration the 3D position can be calculated in wide-angle camera coordinates. Remapping this position into the wide-angle camera image gives the resulting detection in the wide-angle frame (frame 2).

Figure 6 illustrates the complete mapping step. The detection system operating on the stereo camera image delivers an established detection to be mapped and further processed by the wide-angle camera. By using the disparity image, the three-dimensional position of the detection can be calculated. The disparity data are received by the SGM algorithm [15], [16]. Since the cameras are extrinsically calibrated, this 3D position can also be calculated with respect to the second, wide-angle camera. Remapping this 3D point onto the image plane gives the resulting 2D position in the wide-angle frame. Please note, that the corner points of the original detection should be used for mapping to fully describe the detection in position and size. However, disparity values at corner points (transition areas to other objects) are critical. Thus we propose to either use the center pixel of the detection or calculate a median disparity to determine the depth. A comparison of different disparity methods can be seen in [12].

Although the previous figure shows a perfect mapping, in some cases the mapping contains errors.

Figure 7 illustrates different cases of perfect and imperfect mappings. The detections are illustrated on the left and the mapped detections on the right, respectively. The first three examples show perfect mappings, whereas the last three ex-

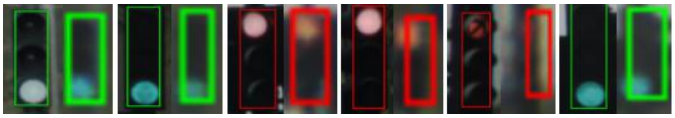

Fig. 7. Six examples of mapping hypotheses from the stereo camera (left) to the wide-angle camera frame (right). The first three examples show a successful mapping with a perfect result. The last three examples show an erroneous mapping. Please note, that the bounding box color indicates the state, not the quality of mapping. Although both cameras have the same image sensor the color of the lamps differ. This is mainly caused by different camera lenses.

amples show erroneous mappings.

As already stated the mapping is erroneous in many cases. Possible reasons are

- Errors in calibration (intrinsic and extrinsic calibration)
- Errors in 3D position caused by the disparity image
- Asynchronous image acquisitions,

and each of these reasons may cause situation-dependent errors. Errors in camera calibration (particularly angle errors) cause position-dependent mapping errors which rise with increasing distance to the camera center.

IV. Mapping Correction

Even small detection errors are critical for traffic light recognition systems. If the light is not part of the mapped detection, a determination of the state is not possible. Furthermore, traffic light tracks are in peril of being lost as many classifiers are vulnerable to detection inaccuracy caused by shifting or scaling.

For this reason, a method is presented to correct the mapped traffic lights.

A. Template Matching Correction

The template matching correction is based on two principal components.

In a first step, a search area must be determined. The search window is calculated based on assumed errors in yaw, pitch and roll (calibration and vehicle orientation errors) and errors in disparity.

The second step slides the traffic light template, i.e. the detection cutout of the stereo camera image, over the predetermined search window. The position with the highest matching score displays the position of the traffic light.

Since the resolution of the detection in the stereo camera image is much higher than in the wide-angle camera image, the template requires resizing. The size of the detection in the wide-angle camera is known from the previous mapping step. The original detection is resized according to this size.

Figure 8 illustrates the detection in the stereo camera image (left), resized detection (middle) and for the purpose of comparison the detection in the wide-angle camera frame. Both, resized detection and detection in the wide-angle frame have the same resolution.

The resized template is slid over the search window in the wide-angle camera image. The images are compared against

Fig. 8. Traffic light detection coutout in the stereo camera image (left) and wide-angle camera image (right). Since the resolution in the stereo camera is much higher, the template is resized to the same resolution as expected in the wide-angle camera frame. The resized template is shown in the middle.

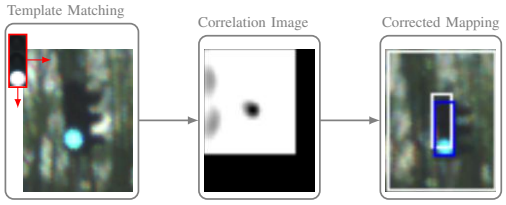

Fig. 9. Scheme of the template matching correction. The resized template received from the stereo camera is shifted over the wide-angle region of interest. Hence, a binary correlation image results. High correlation values are illustrated in black. The border areas are black as the template leaves the region of interest. The right image shows the comparison between initially mapped detection (white) and corrected detection (blue).

each other and the matching scores are saved in a result matrix R containing scores at each position in the search window. Two different methods are used for score calculation [17], [18]. The sum of squared differences is calculated as

$$R_{SQ}(x^p, y^p) = \sum_{x^{p'}, y^{p'}} \left[T(x^{p'}, y^{p'}) - I(x^p + x^{p'}, y^p + y^{p'}) \right]^2, \quad (7)$$

where T denotes the template and I the search window the template is slid over. Smaller correlation values express a better similarity between both image patches. Another method used is to calculate the cross correlation given by

$$R_{CO}(x^p, y^p) = \sum_{x^{p'}, y^{p'}} T(x^{p'}, y^{p'}) \cdot I(x^p + x^{p'}, y^p + y^{p'}) \quad (8)$$

where higher values express a better similarity.

Figure 9 illustrates the correction of a detection with template matching. The resized stereo detection template is slid over the wide-angle camera image. The second image shows the resulting correlation image in which higher similarity is plotted with dark pixels. The resulting blue bounding box is clearly more accurate than the original white bounding box, resulting from mapping without correction.

V. Evaluation

For the evaluation of the proposed method we do not use published traffic lights datasets such as LISA traffic sign dataset [19], LARA traffic light benchmark [20] or the recent Bosch Small Traffic Light Dataset [21] as they do not provide stereo camera images and multi-views.

A. Dataset

Our dataset recorded in 11 german cities includes stereo camera images as well as wide-angle camera images. For the evaluation of this method, we use the dataset recorded in Düsseldorf. A major part of the data is already annotated, whereas the stereo camera images and wide-angle camera images were initially annotated independently. Since we only annotate every tenth frame (changes from frame to frame are marginal), only a small percentage of the annotated data can be used for this evaluation. The majority of the annotations are not synchronous, this means for a given stereo camera annotation no appropriate wide-angle annotation exists and vice versa. When we started to evaluate the precision of the camera-to-camera mapping we started to annotate sequences with corresponding annotations in both camera frames. Up to now, 30 sequences (at 30 intersections) with approximately 300 images are finished. Only frames, in which traffic lights are close to leave the field of view of the stereo cameras are annotated. In comparison with our huge database 300 images is a low value. Nevertheless, for evaluating the proposed mapping system the number of annotations is not as crucial as for completely image processing based approaches. The mapping itself is mainly dependent on camera calibration quality, which is independent of the current vehicle environment. Only the mapping correction can show dependencies. The dataset is planned to be published.

TABLE I
DATABASE FOR THE DATASET RECORDED IN DUESSELDORF. A HIGH NUMBER OF ANNOTATIONS IN BOTH CAMERA FRAMES ARE AVAILABLE. ONLY A SMALL NUMBER OF IMAGES ARE SYNCHRONOUSLY ANNOTATED, I.E. AN ANNOTATION IN BOTH CAMERAS FRAMES AT THE SAME TIME EXISTS, WITH WHICH THE MAPPING CAN BE EVALUATED. WE USE 30 SEQUENCES WITH 410 ANNOTATIONS FOR EVALUATION.

number of	sequences	annotations
stereo camera labels	261	26468
wide-angle camera labels	34	69160
synchronous labels	30	410

B. Results

To express the accuracy of the mapping, the metric intersection over union is used defined as

$$IoU = \frac{|D_m \cap L|}{|D_m \cup L|} = \frac{|D_m \cap L|}{|D_m| + |L| - |D_m \cap L|}, \quad (9)$$

where D_m is the mapped detection and L are annotations in the wide-angle frame. Both rectangles are compared with another. Goal is to reach IoU values greater than 0.5. This is a typical threshold for detected objects in literature. Furthermore, our classifiers show good classification results for higher overlaps. Figure 10 illustrates a boxplot, which compares the mapping without correction against the correction by cross correlation and sum of squared differences, respectively. Both correction methods increase the median intersection over union by 0.15. The lower quantile is increased by 0.2. The difference between both correction methods is negligible, whereas the cross correlation correction performs best.

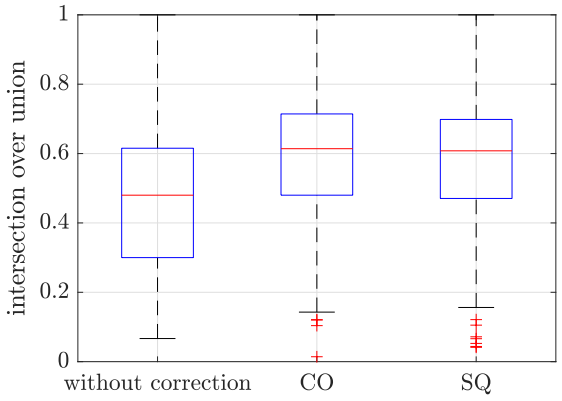

Fig. 10. Comparison of the mapping results without and with correction. The correction methods increase the median intersection over union by 0.15. The difference between both correction method is negligible.

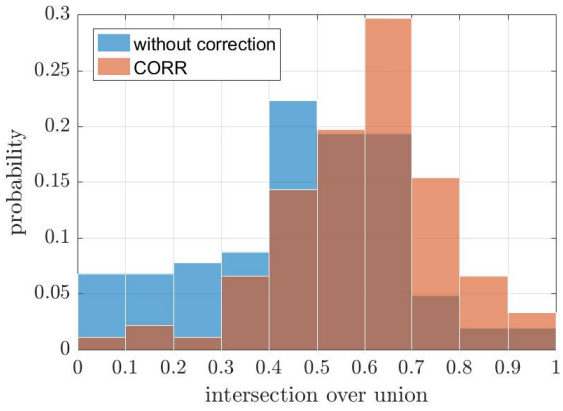

Fig. 11. Histograms comparing the mapping without and with cross correlation correction. Especially low intersection over union values are increased by correction.

Figure 11 illustrates all mapped detections as a probability histogram. The mapping without correction is plotted in blue, the mapping with cross correlation correction in red. A clear shift of the intersection over unions towards higher values is reached with correction. In particular, the low IoU values frequently appearing without correction are enhanced by the cross correlation. Still remaining low IoU values are caused mainly by the low resolution of the wide-angle camera images.

Please note, that a median intersection over union of 0.6 is a particularly high result with respect to the resolution. As seen in Figure 4, the resolution in the wide-angle camera is between 4 and 8 pixels (hand-over at 20-30m). Even positioning errors of 1-2 pixels cause high IoU decreases at that resolution.

Figure 12 illustrates further mapping results at different intersections.

VI. Conclusion

In this paper we presented a camera setup for traffic light recognition, which makes traffic light recognition possible from the perspective of an overall system. Regarding this property, it differs from all previously published methods as these only considered traffic light detection in their given database. We showed, that high fields of view are necessary to detect traffic lights at close range. Conventional field of views used in many publications drop detections at certain distances to the real-world object.

Traffic light detections in our stereo camera system were succesfully mapped into the wide-angle camera system by their calculated 3D object position and accurate camera calibrations for both cameras. However, the mapping step is very sensitive for errors in calibration or 3D position. For this reason we presented a correction method based on template matching. Many erroneous mappings were corrected by this matching step.

By presenting the mapping between different cameras, it is now possible to perform traffic light detection at a distance range from 0 to 100 meters. Thus, existing detections are now further detected and tracked in the second camera frame.

We plan to use a higher resolution in our next datasets. The current resolution of a traffic light in the image is 4-8 pixels in the distance range in which the hand-over is processed. Both, detecting and classification is complicated at that low resolutions. A higher wide-angle camera resolution does not cause computational problems as the pipeline on the wide-angle camera works on regions of interest created by previous detections from the stereo camera.

References

[1] M. Diaz-Cabrera, P. Cerri, and P. Medici, "Robust Real-time Traffic Light Detection and Distance Estimation Using a Single Camera," *Expert Syst. Appl.*, vol. 42, pp. 3911–3923, May 2015.

[2] M. Omachi and S. Omachi, "Traffic light Detection with Color and Edge Information," in *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, pp. 284–287, Aug 2009.

[3] M. Omachi and S. Omachi, "Detection of Traffic Light Using Structural Information," *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, vol. 2, pp. 809–812, 2010.

[4] R. De Charette and F. Nashashibi, "Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates," *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 358–363, 2009.

[5] J. Gong, Y. Jiang, G. Xiong, C. Guan, G. Tao, and H. Chen, "The recognition and tracking of traffic lights based on color segmentation and CAMSHIFT for intelligent vehicles," *2010 IEEE Intelligent Vehicles Symposium*, pp. 431–435, 2010.

[6] F. Lindner, U. Kressel, and S. Kaelberer, "Robust Recognition of Traffic Signals," *IEEE Intelligent Vehicles Symposium, 2004*, pp. 49–53, 2004.

[7] N. Fairfield, "Traffic Light Mapping and Detection," pp. 5421–5426, 2011.

[8] M. Hale, B. Brumitt, B. Meyers, S. Harris, J. Krumm, and S. Shafer, "Multi-Camera Multi-Person Tracking for EasyLiving," *Visual Surveillance, IEEE Workshop on*, vol. 00, p. 3, 2000.

[9] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences," in *2008 Second ACM/IEEE International Conference on Distributed Smart Cameras, Stanford, CA, USA, September 7-11, 2008*, pp. 1–6, 2008.

[10] K. Nummiaro, E. Koller-Meier, T. Svoboda, D. Roth, and L. van Gool, "Color-Based Object Tracking in Multi-Camera Environments," in *Proceedings of the DAGM'03, Springer LNCS 2781, pp. 591-599, Sep 2003*, pp. 591–599, 2003.

[11] J. Illingworth and J. Kittler, "The Adaptive Hough Transform," *IEEE Transaction on Pattern Analysis Machine Intelligence*, vol. 9, pp. 690–698, May 1987.

[12] A. Fregin, J. Mueller, and K. Dietmayer, "Three Ways of using Stereo Vision for Traffic Light Recognition," *Intelligent Vehicles (IV) 2017*, 2017.

[13] N. Navab, "3D Computer Vision II - Camera Models," 2009.

[14] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[15] H. H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, no. 2, pp. 807–814, 2005.

[16] S. K. Gehrig, H. Badino, and U. Franke, "Improving sub-pixel accuracy for long range stereo," *Computer Vision and Image Understanding*, vol. 116, no. 1, pp. 16–24, 2012.

[17] W. Ouyang, F. Tombari, S. Mattoccia, L. di Stefano, and W. Cham, "Performance Evaluation of Full Search Equivalent Pattern Matching Algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 127–143, 2012.

[18] F. Tombari, W. Ouyang, L. di Stefano, and W. Cham, "Adaptive Low Resolution Pruning for fast Full Search-equivalent pattern matching," *Pattern Recognition Letters*, vol. 32, no. 15, pp. 2119–2127, 2011.

[19] M. B. Jensen, M. P. Philipsen, A. Mogelmose, T. B. Moeslund, , and M. M. Trivedi, "Vision for Looking at Traffic Lights: Issues, Survey, and Perspectives," *IEEE Transactions on Intelligent Transportation Systems*, 2015.

[20] M. P. Philipsen, M. B. Jensen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Traffic light detection: A learning algorithm and evaluations on challenging dataset," in *IEEE 18th International Conference on Intelligent Transportation Systems, ITSC 2015, Gran Canaria, Spain, September 15-18, 2015*, pp. 2341–2345, 2015.

[21] K. Behrendt and L. Novak, "A Deep Learning Approach to Traffic Lights: Detection, Tracking, and Classification," in *Robotics and Automation (ICRA)*, IEEE.

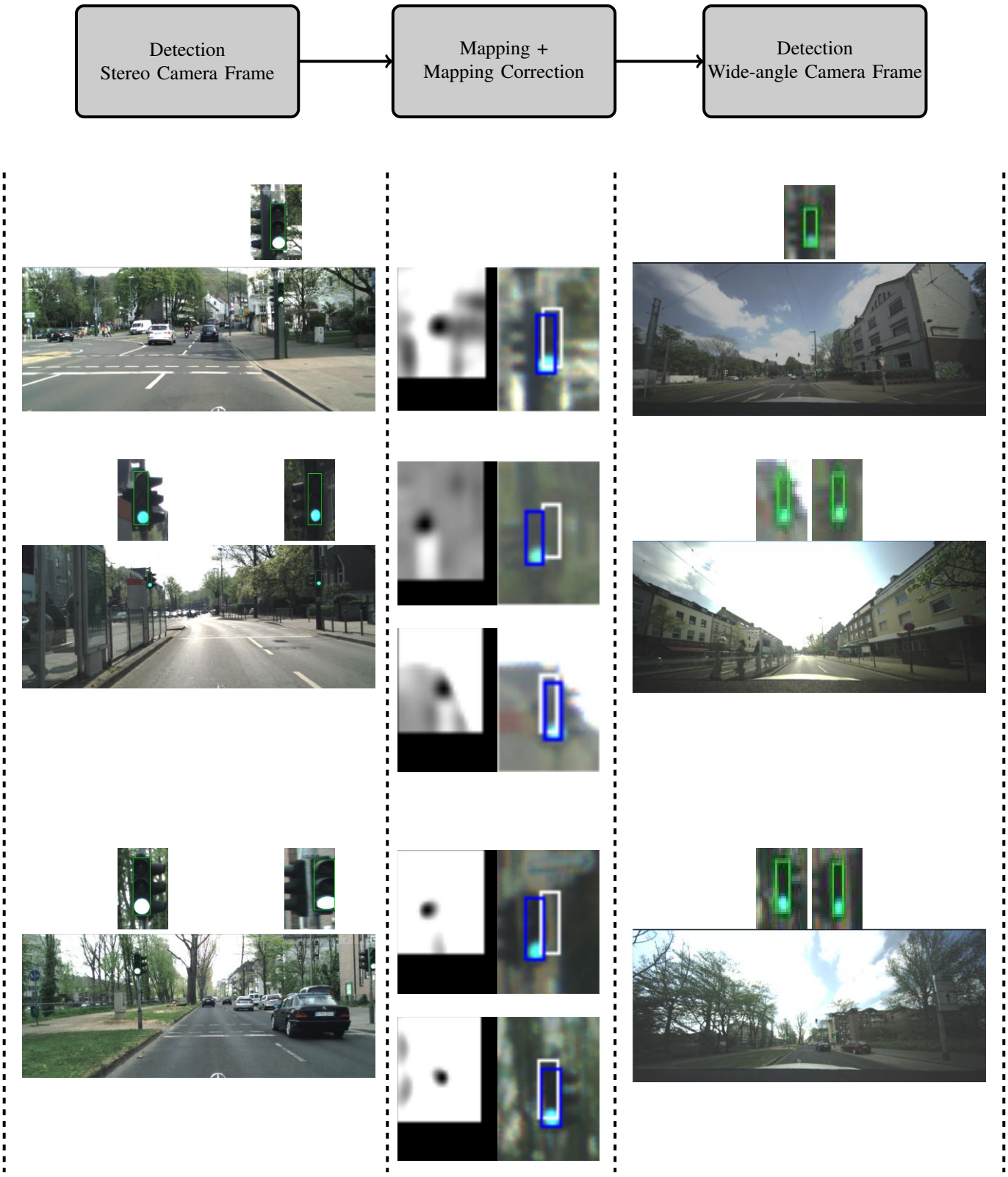

Fig. 12. Examples for a successful mapping and mapping correction. The left column shows example detections in the left stereo camera image. They are marked with bounding boxes. Those detections are mapped as described in Section III, which results in the white bounding box in the middle column. Since those mappings contain errors, the template matching correction step of Section IV is applied. A grayscale correlation image results, where dark pixels express a high correlation. A corrected rectangle can be determined, which is shown as a blue rectangle. The right column shows the wide-angle camera frame with the resulting detections. For better visibility the detections are shown as increased cutouts.