Color-based Free-Space Segmentation using Online Disparity-supervised Learning

Willem P. Sanberg, Gijs Dubbelman and Peter H.N. de With

***Abstract*— This work contributes to vision processing for Advanced Driver Assist Systems (ADAS) and intelligent vehicle applications. We propose a color-only stixel segmentation framework to segment traffic scenes into free, drivable space and obstacles, which has a reduced latency to improve the real-time processing capabilities. Our system learns color appearance models for free-space and obstacle classes in an online and self-supervised fashion. To this end, it applies a disparity-based segmentation, which can run in the background of the critical system path, either with a time delay of several frames or at a frame rate that is only a third of that of the color-based algorithm. In parallel, the most recent video frame is analyzed solely with these learned color appearance models, without an actual disparity estimate and the corresponding latency. This translates into a reduced response time from data acquisition to data analysis, which is a critical property for high-speed ADAS. Our evaluation on two publicly available datasets, one of which we introduce as part of this work, shows that the color-only analysis can achieve similar or even better results in difficult imaging conditions, compared to the disparity-only method. Our system improves the quality of the free-space analysis, while simultaneously lowering the latency and the computational load.**

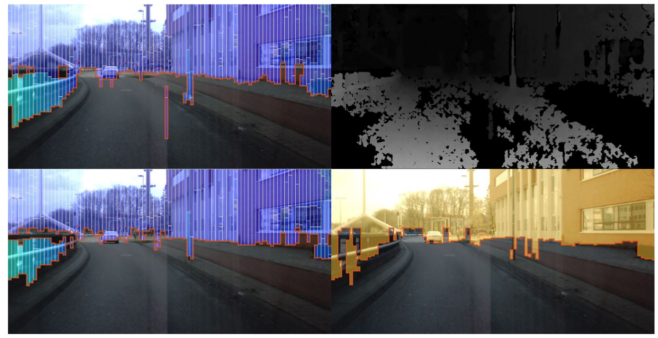

Fig. 1. Stixel segmentation results comparing disparity-only (top left), fused disparity and color (bottom left), and our new color-only (bottom right) analysis. The orange overlay with a dark border depicts space occupied by obstacles according to the detection algorithm. At the top right, the disparity signal is shown, which has several artifacts due to low texture in the road region and sun reflections in the windshield (middle). These artifacts cause false obstacle detections in the original disparity-based algorithm [1]. These errors have been resolved in the fused method presented in [2]. Here, we show that it is possible to obtain similar results as the fusion method, while requiring less data in the process.

I. INTRODUCTION

In recent years, vehicles are becoming increasingly intelligent with so-called Advanced Driver Assistance Systems (ADAS). This development is expected to significantly reduce traffic accidents, traffic congestion and fuel consumption simultaneously. To ensure traffic safety, ADAS can e.g. indicate the location of potentially hazardous obstacles to the driver and the position of safely drivable road. On the longer term, ADAS and related technologies will allow the development of fully autonomous vehicles. In this work, we improve a state-of-the-art vision-based free-space detection system by exploiting multiple image modalities in an efficient way.

To robustly facilitate situational awareness at a moving platform, several complementary sensor modalities should be employed. These modalities can include RADAR, LIDAR, ultrasound, and (thermal) imaging. The benefit of using vision-based systems is that they provide dense scene information in a cost-effective way. Image data is also a rich source of information, since it comprises of several informative properties. For stereo-based video imaging, these informative aspects include not only the usual *texture, color* and *shape* features, but also *optical flow* motion analysis and *disparity* estimation. All these elements can contribute to a robust situational analysis, such as e.g. the detection of partially occluded pedestrians who are about to cross the street. Although LIDAR, RADAR or ultrasound provide valuable complementary information, in this work we solely focus on vision-based detection systems.

Willem Sanberg, Gijs Dubbelman and Peter de With are with the Department of Electrical Engineering, Video Coding and Architectures Research Group, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands w.p.sanberg@tue.nl

Related work already indicates the use of vision-based ADAS in practice. These are mainly monocular systems that detect pedestrians [3], lane markings [4], or traffic signs [5]. Fully autonomous vehicles require more dense, accurate, and reliable sources of scene information, in particular 3D information. The current challenge in vision-based ADAS is to deliver comparable semantic 3D scene information at a more affordable price point than that of typically used high-end laser-based systems (accompanied by RTK-GPS) [6][7].

Multi-view image processing, in particular stereo vision, has the potential to fulfill these requirements. In stereo vision, the disparity, which is analogous to depth, can be estimated densely and in real-time [8]. This gives a direct description of the geometry of the scene and it facilitates, for example, a separation of flat, drivable surfaces from erect obstacles [9][10]. The Stixel World method [1] is a state-of-the-art approach for such a geometry description of the scene. It is a fully probabilistic framework to distinguish free-space from obstacles in the disparity signal, which can be implemented efficiently given several assumptions. This framework is generally more flexible and more robust than its predecessors.

A pitfall of the original Stixel World framework is that it requires a disparity signal of a certain quality. However, the quality of disparity estimation often degrades in cases of occlusions, reflections or image regions with too little texture information, which are common in traffic scenery. As a result of this degraded signal, the original Stixel World framework detects many false obstacles, rendering the results useless for a practical system under adverse conditions. An example of this is shown at the top-left image of Fig. 1. In our recent work [2], we show that the performance of that disparity-based framework can be improved by fusing color into the algorithm. This strategy resolves many erroneous results of the disparity analysis at a low additional computational cost, in contrast to alternative solutions such as a high-quality camera or more advanced disparity estimation techniques. The benefit of this earlier work is shown at the bottom-left image of Fig. 1.

A key property of any ADAS is the response time, i.e. the time delay between data acquisition and the response to the result of the analysis. Since ADAS preferably need to function at high vehicle speeds, the response time of such systems should be as fast as possible. Hence, any delay that can be removed from the critical path of the analysis is beneficial to the value and applicability of the system, provided that it does not degrade the reliability of the results. Therefore, we will explore the possibility of removing the disparity analysis from the critical system path. Although there exist fast disparity estimation methods [8][11], this typically either requires relying on sub-optimal algorithms, processing at a low resolution, or acquiring customized hardware that is not commonly available. To illustrate this, even in the state-of-the-art system presented in [1], the dedicated FPGA disparity estimation takes 40 ms per frame, whereas the stixel analysis of the data takes 30 ms, when executed on a general, high-quality multi-core CPU.

Therefore, we will not rely on *strong fusion* of disparity and color in this work, even though the result presented in [2] clearly shows the qualitative benefits of that. In contrast, we propose here to process the most recent camera frame using an efficient color-only stixel segmentation. The disparity estimation and analysis, which is only required for our online color modeling, can be processed in parallel and at a lower frame rate. An example of our color-only results is shown at the bottom-right image of Fig. 1, which is very similar to the results that require disparity within the stixel segmentation.

An alternative to online color modeling is offline modeling [12], which would completely remove the need for online disparity estimation. However, we have a strong preference for an online learning approach, given the challenging nature of traffic environments, which is full of varying weather conditions, complex scenery, varying geographical settings and highly dependent on the time of the day. For instance, urban traffic scenes tend to contain predominantly gray-tones in low-light situations. We consider it more feasible to build a robust, yet discriminating color model that is tuned to that specific time and place, rather than building a generic model that holds for every environment and weather condition.

The remainder of this paper is structured as follows. First, we will provide a short description of the disparity-based Stixel World in Section II, since it serves as a basis of our work. Then, in Section III, we summarize the methods and results of our extended Stixel World with strong data fusion as presented in [2]. Section IV presents our main contributions based on [2], where we explain our new color modeling strategies. In Section V, we elaborate on our evaluation approach, including our publicly available dataset, experiments and results. Lastly, conclusions are provided in Section VI.

II. THE DISPARITY STIXEL WORLD

Let us now give a short overview of the Stixel World framework from [1], which we use as a basis of our work. The main goal of stixel segmentation is to find the optimal labeling L^* of vertically stacked, piecewise planar ground or obstacle segments for the input disparity data $\mathbb{D}$. Finding L^* can be formulated as a MAP estimation problem as specified in (1), which can be solved efficiently using Dynamic Programming. Using Bayes' theorem and assuming, among other things, independence between columns and between disparity measurements at individual pixels, the posterior probability can be written according to (2). Here, u is the column index and w the image width. The probability $P(L_u)$ models a-priori world knowledge constraining the labeling, to avoid dispensable segments and physically unlikely situations. This world model offers a way to regularize the results for image-column optimality, whereas the methods of [9] and [10] potentially lead to sub-optimal results, since they mostly analyze data locally. The details concerning $P(L)$ are presented in [1]. Finally, the specification in (3) provides the likelihood of the data given a certain labeling, where n is the segment index, N_u the number of segments in L_u, and v_n^b and v_n^t the bottom and top row-index, respectively, of segment s_n. This segment has a label $l_n \in \{g, o\}$, representing the ground and obstacle classes, respectively. The previously specified equations are given by:

$$L^* = \arg\max_{L \in \mathbb{L}} P(L|\mathbb{D}), \tag{1}$$

$$P(L|\mathbb{D}) \sim \prod_{u=0}^{w-1} P(D_u|L_u) \cdot P(L_u), \tag{2}$$

$$P(D_u|L_u) \sim \prod_{n=1}^{N_u} \prod_{v=v_n^b}^{v_n^t} P(d_v|s_n, v). \tag{3}$$

The distribution $P(d_v|s_n, v)$ as specified in (3) represents the probability of a single valid disparity measurement d_v at a certain row v, assuming that it would belong to a potential segment s_n. It consists of a mixture model containing a uniform distribution that models outliers, and a Gaussian distribution that models inliers, to assess how well the measurement fits the potential segment for each class. For ground segments, the expected disparity is a linear planar surface and for obstacle segments a fronto-parallel surface.

III. Color Extensions

The key idea of our previous work [2] is to increase the robustness of the system against adverse conditions such as low light, bad weather, or a low-quality sensing system. These conditions are common in real-world scenarios, but typically degrade the disparity signal. In the original framework, a degraded disparity signal leads to the detection of many false obstacles, rendering the results useless for a practical system under adverse conditions. To address this issue, we have included a color component $P(c_v|s_n,v)$ in the original likelihood function of [1], as specified by (3), which results in the following likelihood:

$$P(D_u,C_u|L_u) \sim \prod_{n=1}^{N_u} \prod_{v=v_n^b}^{v_n^t} P(d_v|s_n,v) \cdot P(c_v|s_n,v), \quad (4)$$

thereby treating disparity and color as independent signal modalities. The rightmost term $P(c_v|s_n,v)$ captures the probability of a certain color measurement given a certain segment. In [2], this term is simplified to $P(c_v|s_n,v) = P(c_v|l_n)$, under the assumption that the probability only depends on the label of the segment under evaluation, and not on its position v or other colors in the segment. Therefore, $P(c_v|l_n)$ can be directly realized by computing and normalizing a color histogram for each class. By calculating that histogram in an online fashion, the color modeling is highly adaptive to the currently relevant traffic situation.

Based on our experiments in [2], we have concluded that the optimal tradeoff of design choices for the online color modeling is considering to (a) quantize the RGB color space with 64 values, obtained with *minimum variance quantization* [13], (b) rely on the 10 most recent frames as the online learning window, and (c) selectively adopt online-training samples for free-space data from the whole segmentation mask, but select samples for obstacles data only below the horizon.

IV. Color-Only Analysis

The main contribution of this work explores the possibility of solely relying on color for the analysis of the most recent camera frame. The goal of this strategy is to reduce the delay between data acquisition and data analysis. Our approach facilitates this in two ways. Firstly, the disparity estimation can be removed from the critical path, as it is not required for the analysis of the most recent frame. Secondly, the analysis of one signal modality is less complex than analyzing two signals simultaneously. The basis of our approach is the online self-supervised learning method as described in Section III. This method processes preceding stereo frames and generates a free-space vs. obstacle labeling based on disparity. Consecutively, this labeling is exploited as self-supervised training masks for the color representation for these two classes. Our research addresses two sub-problems: (a) finding a color representation that is informative enough to separate free-space from obstacles and yet sufficiently suited for online processing, and (b) defining an efficient and robust color-based cost term for within the stixel framework. These problems are addressed in the next subsections.

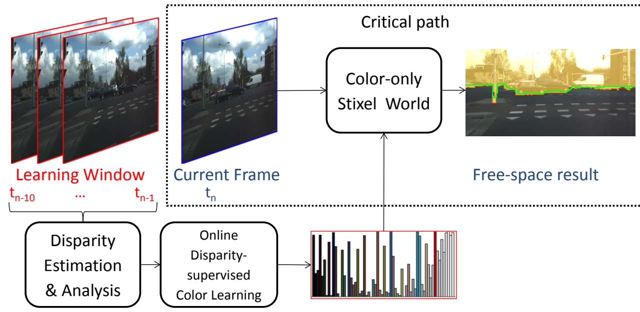

Fig. 2. Our proposed system diagram. Note that the disparity-supervised color modeling in the lower part of the scheme can be lagging or run at a lower frame rate than the critical path. The range of the Learning Window can be varied.

A. Relative Color Representation

In line with the recommendations made in [2], we employ indexed RGB colors that are adapted online to the frames in the learning window for an efficient and yet descriptive representation. In the Stixel World model, each stixel has a fixed width w_{stix}, meaning that a stixel spans several image columns. This increases the robustness and decreases the computational load at the cost of horizontal resolution in the labeling. There are several strategies to condense the image columns into a single stixel data vector. We propose and evaluate three methods below for doing this.

Firstly, we take the mode over indexed RGB values for each image row in a $[w_{stix} \times w_{stix}]$ window, located at the central image column of each stixel. Since we work with color indices, taking the mode is the most straightforward method of getting a robust, representative value.

Secondly, we also add color variation to the modeling, instead of only considering absolute color. The reason is that relative information may also be descriptive in our case, since free-space areas tend to be more homogeneous than obstacles such as cars, pedestrians and houses. Therefore, we combine the first color mode with local edge strength, to assess both absolute and relative color information. We measure local edge strength using Sobel filter responses, averaged over a $[w_{stix} \times w_{stix}]$ window.

Thirdly, we extend the relative color representation by specifically modeling color pairs instead of measuring color variation. To this end, we calculate both the first and the second mode in a $[w_{stix} \times w_{stix}]$ window (in homogeneous areas, the first and second modes are equal). This makes the relative color modeling more informative and more discriminative, as it considers both local color homogeneity and specific color pairs. The latter aspect is not accounted for when using local edge strengths.

B. Distance-aware Color Analysis

The stixel optimization process relies on probability distributions, as described in Section II. Therefore, we need to define the color-only likelihood function, specified by

$$P(C_u|L_u) \sim \prod_{n=1}^{N_u} \prod_{v=v_n^b}^{v_n^t} P(c_v|s_n, v), \tag{5}$$

which is analogous to the specifications of (3) and (4).

The term $P(c_v|s_n, v)$ should capture the probability of a certain color measurement given a certain potential free-space or obstacle segment. In [2], this term is simplified to $P(c_v|s_n, v) = P(c_v|l_n)$, under the assumption that the probability only depends on the label of the segment under evaluation, and not on its position v. This leads to a mixture model with a uniform distribution with probability p_{out} to model outliers and a normalized histogram per class over all color bins with probability $P_h(c|l)$.

The key contribution of this paper is to add distance-awareness to the histogram-based color modeling. This is motivated by the basic phenomenon that camera images naturally suffer from geometric, perspective distortion. Effectively, pixels representing areas close to the camera contain a smaller real-world surface than pixels representing areas at a large distance. Therefore, surfaces that are close to the camera dominate in regular histograms, which contain only basic pixel counts. This imbalance can result in inaccurate color modeling of far-away obstacles.

The first step in adding distance-awareness to the color modeling is to weight each pixel with its corresponding real-world surface during the process of calculating a weighted color histogram $P_{wh}(c|l)$. This results in a histogram that is more balanced towards obstacles at a large distance. To avoid that this leads to false obstacle detections close to the camera, we use both the regular and the distance-weighted histogram for defining the distance-aware posterior distribution P_{DA}:

$$P_{DA}(l_n|c_v, v) = (1 - \alpha_w(v)) \cdot P_h(l_n|c_v) + \alpha_w(v) \cdot P_{wh}(l_n|c_v). \tag{6}$$

Here, $\alpha_w(v)$ is a factor in the interval $[0, 1]$ to leverage the regular and the distance-weighted color posteriors, which are calculated from the corresponding histograms using Bayes' rule. We define $\alpha_w(v)$ empirically as the mean of a linear function and the real-world pixel length at row v. In our stereo-camera framework, the pixel size and its corresponding length can be determined from the disparity signal and the camera parameters. Since we have removed the disparity estimation from the critical path, we have to rely on a disparity signal from at least one frame earlier. Fortunately, the differences between consecutive frames are small and, on top of that, they are smoothed by the probabilistic nature of our processing. We rely on a fixed linear ground-plane model to fill any holes in the disparity map prior to determining pixel sizes. Fig. 3 illustrates these steps and their effect on the posterior distribution.

Note that our strategy of making the processing distance-aware cannot be achieved by simply computing histograms using a Bird's Eye View (BEV) representation of the image. A BEV representation can work for the ground plane, but it heavily distorts the area of obstacles, since it projects all image pixels onto the same flat plane. In contrast, our approach models each pixel surface individually, leading to a more accurate representation of obstacles in the histograms.

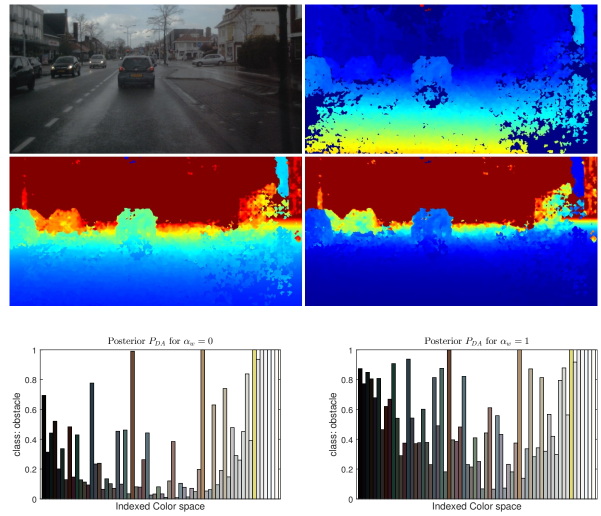

Fig. 3. Illustration of the proposed depth-aware processing (DA). Top row: the brightened camera image (left) with the corresponding disparity signal (right). Middle left: the distance image, where holes are filled in using a static, linear, ground-plane assumption. The distance saturates at 35 meters. Middle right: the corresponding real-world surface of each pixel, where the quadratic course of pixel surfaces is visible in the colorization. Bottom row: the posterior for the obstacle class without (left) and with DA (right).

V. EVALUATION

This section evaluates the performance of the proposed color-based free-space scene segmentation. All our results will be compared both to the disparity-only baseline approach of [1] and to the color-extended method of [2].

A. Test Setup

We employ two datasets for the validation. The first one is the EHV-road-ITSC14 dataset from in [2]. The second one, EHV-road-ITSC15, is newly introduced in this paper.

Both datasets are acquired in an urban environment, using a BumbleBee2 camera, mounted behind the windshield of a car, just below the rear-view mirror. The camera has a baseline of 12 cm, a resolution of 1024×768 pixels and a frame rate of 20 Hz. Both datasets are publicly available[1].

Whereas EHV-road-ITSC14 contained both frames with bright and dim lighting conditions, EHV-road-ITSC15 is solely focused on dark, clouded, low-light and rainy frames. It consists of 114 frames that have a road and a free-space annotation (road including pavement, grass, etc.). For each annotated frame, the 10 preceding frames are also available to facilitate the online color modeling. The sequences are selected in such a way that they do not contain windshield wipers or obstacles in close proximity that would hamper appropriate disparity estimation. The two datasets combined contain a large variety of relevant traffic situations, such as crowded streets, road repair sites, large crossings and

[1] http://www.willemsanberg.net/research

highways. They contain asphalt as well as paved roads of several colors (black, gray, red).

To obtain disparity measurements, we employ a multi-threaded version of the OpenCV implementation of the Semi Global Block Matching algorithm of [14] with similar settings as [2]. We have set the bounds of the disparity estimator to $d_{min} = 1$ and $d_{max} = 48$. Additionally, we employ a *winner margin* of 20 to force the algorithm to provide only measurements with a high confidence at the cost of a reduced density in the disparity signal. This is beneficiary for the baseline Stixel World method, since it can handle missing values better than erroneous ones. This can be seen as a simplification of the work from [15], in which disparity estimates are accompanied by a confidence measure to adaptively set an outlier probability. In our approach, this confidence is binary with a threshold on the *winner margin*.

As described, our camera has lower resolution and a smaller stereo-baseline than, for example, the camera used for the original Stixel World [1], resulting in lower quality disparity estimates. To compensate for this deficiency and to obtain more favorable results for the baseline method, we have made improvements to the baseline framework, as presented in [2]. For example, we learn the ground-plane model online instead of using a single fixed model and have tuned the label-based transition probabilities defined in $P(L)$ to boost the performance of the baseline method even further.

In our experiments, we have adopted a stixel width of 11 columns and subsample the disparity and color signals vertically with a factor of 3, prior to segmentation. Note that we exploit the full-image data to compute look-up tables and color models, which is comparable to the approach in [1]. At present, the version of the proposed color-only Stixel World method is a MATLAB-based implementation. For a real implementation in C/C++, we have estimated that the complexity is comparable to that of the disparity-only baseline, so that real-time executions can be obtained [1].

B. Scoring Metric

As a key quantitative analysis method, we propose to evaluate the detected free-distance per stixel. For each stixel, we calculate the true free-space from the hand-made drivable-surface annotation. For this process, we rely on a static linear ground-plane assumption, effectively translating an image row-index to a distance. For the detection results, we calculate the deviation as a percentage of the true free-space. For robustness reasons, free-space detections are counted as correct when they are within the range of 30% too short or 15% too long. This asymmetrical range reflects the fact that missing an obstacle is more dangerous than detecting one too close. For the same reason, we distinguish the incorrect stixels into obstacle misses (free space too long) and false obstacle detections (free space too short). Note that, although a deviation of 30% may seem a lot, it corresponds to only a couple of pixels after several meters (and only some centimeters before that).

In essence, this metric is comparable to our drivable-distance metric in [2], but at a finer horizontal resolution, since it uses the width of a stixel instead of the width of a car. The latter is rather coarse for in-depth analysis. A single false obstacle nearby heavily degrades the achieved recall, and detecting a big obstacle only in a single stixel already results in a high precision of the drivable distance. Our new stixel-resolution metric provides a more detailed insight about the best settings for obtaining reliable results.

C. Experiments and Results

This section presents the experiments assessing the color representation and the distance-aware processing of our Color-only Stixel World framework, together with their quantitative results. Fig. 4 shows qualitative results of our best method in comparison with the results of the disparity baseline. It can be seen that our color-only stixel algorithm provides similar or even better results with respect to the disparity method. In the bottom-right case, the artifacts in the disparity segmentation are consistent throughout the full learning window. This causes the image areas with light reflections to be modeled as obstacles in the color model, leading to false obstacle detections in our color-only analysis.

The quantitative results of our most relevant experiments are shown in Fig. 5. Note that we show the results for the two different datasets separately (middle and right graph) and combined (left graph). First of all, since our new metric is more strict and considers a larger range, our previous work, the Extended Stixel World (run *b*), obtains a lower score. The earlier algorithm was tuned to reduce the number of false detections. This is confirmed by the graph in Fig. 5, which shows that the percentage of stixels with false obstacle detections is reduced. However, the number of stixels with missed obstacles increases, resulting in a lower number of correct stixels. With our new method, which considers color pairs and is distance-aware (run *k*), the total number of correct stixels is increased significantly compared to the fused method (run *b*) and even, although to a smaller extent, compared to the disparity baseline (run *a*).

Fig. 6 provides a detailed comparison between the results of the disparity-color fusion method and our new color-only method without and with distance-aware processing. The figure clearly shows that the fusion method tends to miss parts of obstacles, specifically at large distances, where the uncertainty in the disparity signal is high and the color contrast is typically low. Our new approach with a more informative color modeling reduces these errors to a large extent. On top of this, the distance-aware processing gives a further improvement and makes the results more consistent. The added value of DA is also visible quantitatively in Fig. 5, by comparing runs *i*, *j* and *k* to *f*, *g* and *h*, respectively.

The best free-space segmentation results are achieved by using a combination of histogram equalization, color pairs, and distance-aware color processing (run *k*). For these settings, we have evaluated several learning window parameters. Two exemplary results are provided in Fig. 5, using only frames $t-10$ to $t-3$ (disregarding the two most recent frames, run *l*), or using the full range ($t-10$ to $t-1$), but at a lower frame rate by skipping two of every three frames

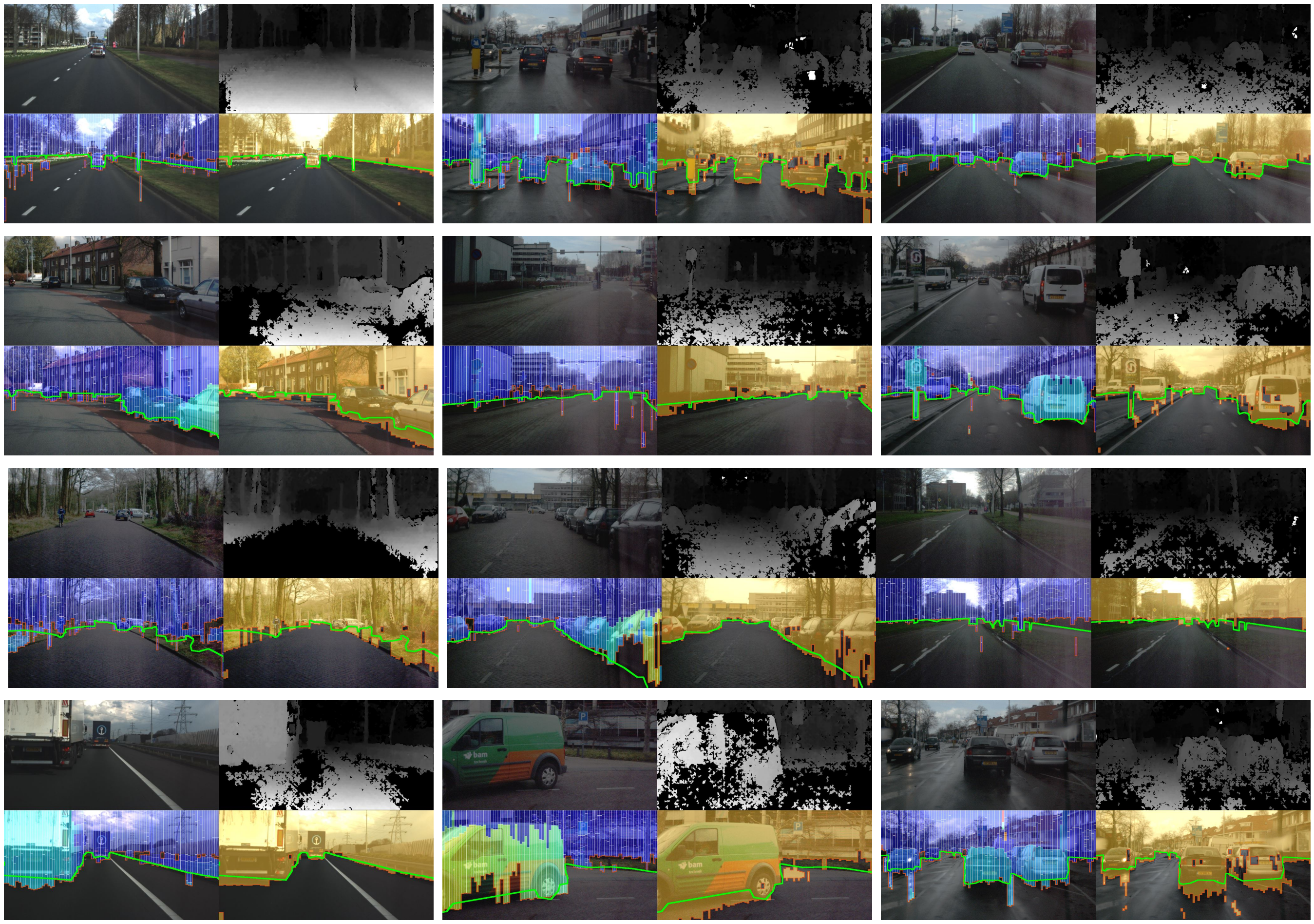

Fig. 4. Qualitative results of our proposed method (HEQ; Mode1&2; DistAw). Each cluster picture contains four (input and result) images. The top two images show the rectified and cropped left camera image and the corresponding disparity image. The bottom two images show the disparity baseline result (left) and our new improved result (right). The green line indicates the border of the ground-truth annotation of the drivable surface. In the disparity-based result, the stixels are colored by their depth (red (close) to blue (far)). In the color-only results, a homogeneous overlay of the detected obstacle region is visualized. All cluster pictures show that our color-only results provide similar or better results in various situations. The bottom-right picture illustrates a case were our color-modeling cannot resolve all artifacts in the disparity-based learning window.

(run m). Both runs l and m provide very similar results to run k. This illustrates the robustness of our method with respect to the online training strategy. Our algorithm does not require all preceding frames and does not require the directly preceding frames for the current evaluation. This is an important result, since it shows that we can remove the disparity estimation from the critical path to lower the computational requirements for real-time processing of ADAS. It is worth noticing that our method achieves the highest results on the new dataset with very low-light conditions (right graph in Fig. 5). This shows that color can be exploited reliably in situations that are difficult to handle with disparity alone.

VI. CONCLUSIONS

We have presented a framework for color-based free-space vs. obstacle segmentation. Our system learns color appearance models for free-space and obstacle classes in an online and self-supervised fashion. To this end, it applies a disparity-based segmentation, which can run in the background of the critical system path, either with a time delay of several frames or at a frame rate that is only a third of that of the color-based algorithm. In parallel, the most recent video frame is analyzed solely with these learned color appearance models, without an actual disparity estimate and the corresponding latency. This translates into a reduced response time from data acquisition to data analysis, which is a critical property for high-speed ADAS. Our algorithm is based on two key contributions: (i) an informative color-pair representation using the first and second mode of an online-adapted indexed color space, and (ii) distance-aware color-histogram processing, based on real-world pixel surfaces. We have evaluated our contributions on two publicly available datasets, one of which is newly introduced in this paper. This evaluation shows that the color-only analysis can achieve similar or even better results in difficult imaging conditions, compared to the disparity-only method or the disparity-color fusion method. In conclusion, our system improves the quality of the free-space analysis while simultaneously lowering the latency and the computational load.

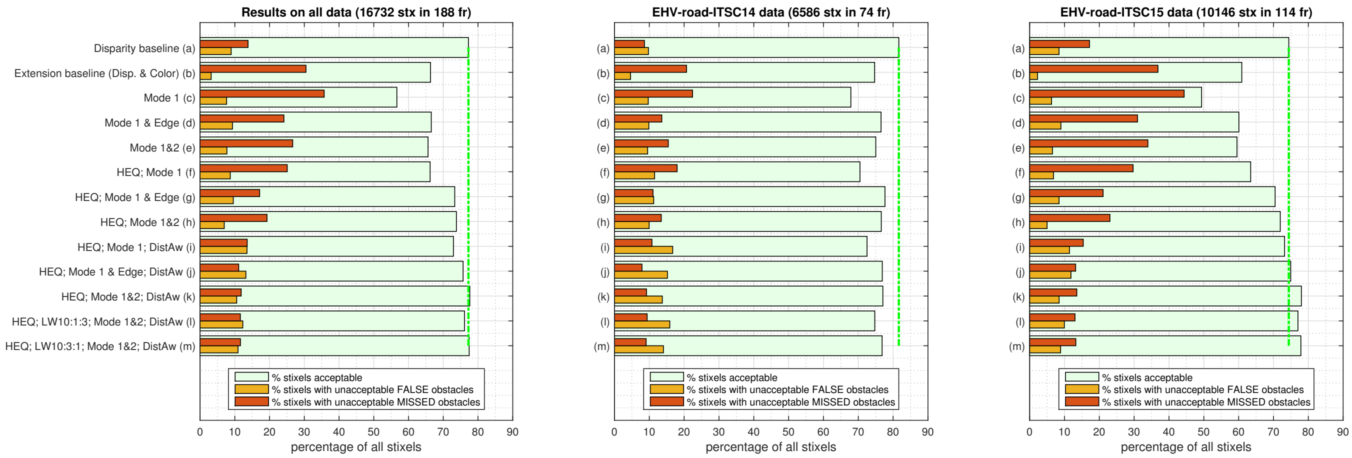

Fig. 5. Quantitative results of free-space segmentation for the baselines and several of our runs. The color experiments all rely on the indexed RGB color space and most of them apply histogram equalization before processing (HEQ). The learning window parameters are indicated with LW start:step:end. If not indicated otherwise, the learning window is LW10:1:1, meaning that it it uses all 10 preceding frames. Run 'l' (LW10:1:3) disregards the two previous frames, run 'm' (LW10:3:1) skips two frames out of three. Runs 'i' to 'm' exploit our distance-aware histogram processing (DistAw).

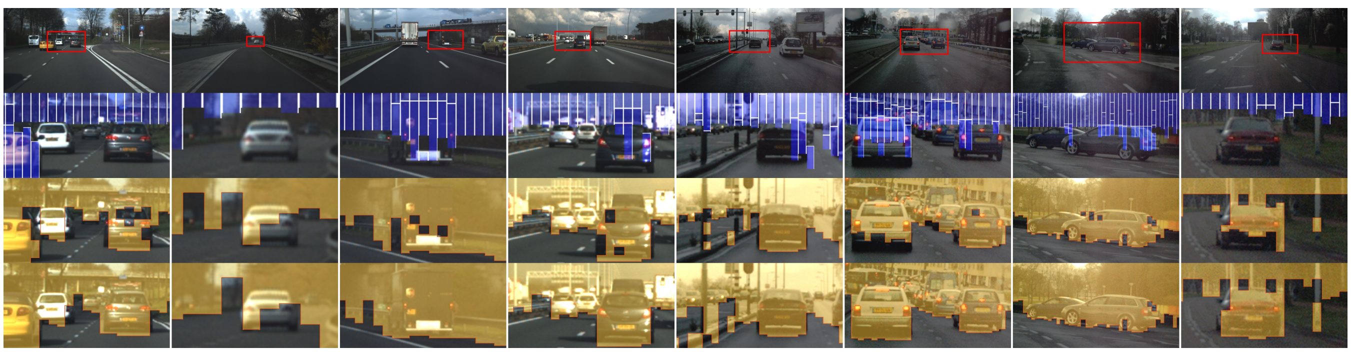

Fig. 6. Qualitative results of free-space segmentation at a large distance to compare the strong fusion method of [2] with our proposed DA color-only method. The top row shows several original images where the area of interests are indicated with a red box. The first four frames are from the EHV-road-ITSC14 dataset. The other four are from the new EHV-road-ITSC15 dataset and are brightened for visualization. The strong fusion method can lead to undetected obstacles due to high uncertainty in the disparity signal combined with low contrast in the color signal (second row). Using histogram equalization and color pairs improves those areas (third row). The bottom row shows the added value of the distance-aware processing, which increases the consistency of the analysis in these difficult imaging conditions.

References

[1] D. Pfeiffer, "The Stixel World," Ph.D. dissertation, Humboldt-Universitat Berlin, 2011.

[2] W. P. Sanberg, G. Dubbelman, and P. H. de With, "Extending the stixel world with online self-supervised color modeling for road-versus-obstacle segmentation," in *IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2014.

[3] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: survey and experiments." *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 31, no. 12, pp. 2179–95, Dec. 2009.

[4] Z. Kim, "Robust lane detection and tracking in challenging scenarios," in *IEEE Trans.on Intelligent Transportation Systems (TITS)*, vol. 9, 2008, pp. 16–26.

[5] A. De la Escalera, J. M. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," *Image and Vision Computing*, vol. 21, pp. 247–258, 2003.

[6] S. Thrun and M. Montemerlo, "Stanley: The robot that won the DARPA Grand Challenge," *J. of Field Robotics*, vol. 23, pp. 661–692, 2006.

[7] C. Urmson, C. Baker, J. Dolan, P. Rybski, B. Salesky, W. Whittaker, D. Ferguson, and M. Darms, "Autonomous driving in urban environments: Boss and the Urban Challenge," *AI magazine*, vol. 30, pp. 17–28, 2008.

[8] W. Van Der Mark and D. M. Gavrila, "Real-time dense stereo for intelligent vehicles," *IEEE Trans. on Intelligent Transportation Systems (TITS)*, vol. 7, no. 1, pp. 38–50, 2006.

[9] G. Dubbelman, W. van der Mark, J. C. van den Heuvel, and F. C. A. Groen, "Obstacle detection during day and night conditions using stereo vision," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, Oct. 2007, pp. 109–116.

[10] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation," *IEEE Intelligent Vehicle Symp. (IV)*, vol. 2, pp. 646–651, 2002.

[11] S. K. Gehrig, F. Eberli, and T. Meyer, "A Real-Time Low-Power Stereo Vision Engine Using Semi-Global Matching," in *Int. Conf. on Computer Vision Systems (ICVS)*, vol. 5815 LNCS, 2009, pp. 134–143.

[12] J. M. Alvarez, M. Salzmann, and N. Barnes, "Learning appearance models for road detection," *IEEE Intelligent Vehicles Symp. (IV)*, pp. 423–429, June 2013.

[13] P. Heckbert, "Color image quantization for frame buffer display," *Computer Graphics*, vol. 16, no. 3, pp. 297–307, 1982.

[14] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 30, no. 2, pp. 328–41, Feb. 2008.

[15] D. Pfeiffer, S. Gehrig, and N. Schneider, "Exploiting the Power of Stereo Confidences," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Portland, USA, 2013.