Stixel-Based Target Existence Estimation under Adverse Conditions

Timo Scharwächter

Daimler AG, Dept. Environment Perception

Abstract. Vision-based environment perception is particularly challenging in bad weather. Under such conditions, even most powerful stereo algorithms suffer from highly correlated, "blob"-like noise, that is hard to model. In this paper[1] we focus on extending an existing stereo-based scene representation – the Stixel World – to allow its application even under problematic conditions. To this end, we estimate the probability of existence for each detected obstacle. Results show that the amount of false detections can be reduced significantly by demanding temporal consistency of the representation and by analyzing cues that represent the geometry of typical obstacles.

1 Introduction

The importance of stereo image processing in the context of modern driver assistance increases steadily. Depth information is used to detect obstacles in the driving corridor [5] or to support pedestrian and vehicle classifiers by guiding the attention to relevant areas in the image [4].

There is no doubt, that for safety critical applications high accuracy and specificity of depth measurements is of utmost importance, as false detections can lead to unnecessary emergency maneuvers. At the same time, the range of stable system operation must be as wide as possible and should cover both well-conditioned input data and adverse conditions such as darkness, rain, reflections etc. Some of such typical scenarios a vision system is confronted with in real world applications are shown in Figure 1.

(a) Mist (b) Darkness (c) Reflections (d) Snow

Fig. 1. Input images under adverse conditions in the automotive context, that only show a few examples of the high variety of environmental effects

[1] Recommended for submission to YRF2012 by Dr. Uwe Franke, Daimler AG

To maintain a stable mode of operation as often as possible, it is necessary to explicitly address those situations. Therefore, this contribution focuses on the typical challenges of an automotive vision system under adverse conditions. To this end, the Stixel[2] representation proposed in [8] is extended to cope with adverse weather and input conditions by estimating the probability of existence for each Stixel. The estimation process is performed by analyzing the temporal and spatial behavior of the extracted Stixel representation.

2 System Overview

In this work, we rely on the semi-global matching (SGM) stereo algorithm introduced in [3]. SGM combines local pixel matching with approximated global smoothness. Steingrube et. al. [9] show that SGM outperforms other recent stereo algorithms in terms of detection performance (particularly in bad weather).

From the resulting disparity map, we extract the Stixel representation (see Figure 2), which provides a compact (∼500 Stixels compared to ∼400.000 pixels) description of free space and obstacles by means of vertically stacked surfaces. For our purpose, the central objective of using this representation is that it performs a spatial regularization of the input data, which already reduces the impact of strong depth outliers. As a result, we obtain a compact and spatially smoothed representation of potential obstacles in the depicted scene.

However, at the time of this writing *temporal* coherence and *horizontal* (spatial) dependencies are not considered inherently in the Stixel representation. Under adverse conditions, this leads to so-called *phantom* Stixels that represent obstacles, although in fact no real obstacle is present. An analysis conducted prior to this work by Pfeiffer [6] shows that – under adverse conditions – such false Stixels pop up spontaneously from time to time (i.e. are temporally uncorrelated) and are spatially small. Some of those phantoms can be seen in the two example images in the middle of Figure 2, which depicts the general proceeding of the developed system.

To exploit these findings, we introduce a tracking scheme that allows to remove temporally uncorrelated Stixels and perform existence estimation based on cues that contain knowledge about the geometry of obstacles. In the following sections, Stixel tracking is discussed briefly and the existence estimation component is reviewed in more detail. Section 5 contains an evaluation of the developed framework and Section 6 ends with a conclusion.

3 Stixel Tracking

In dynamic environments temporal integration requires to track objects of interest to allow correct data association. The tracking component developed in the context of this work adopts the 6D-Vision principle presented in [2]. This allows to estimate the velocity and simultaneously improve the 3D world position of

[2] Derived from the two words *stick* and *pixel*

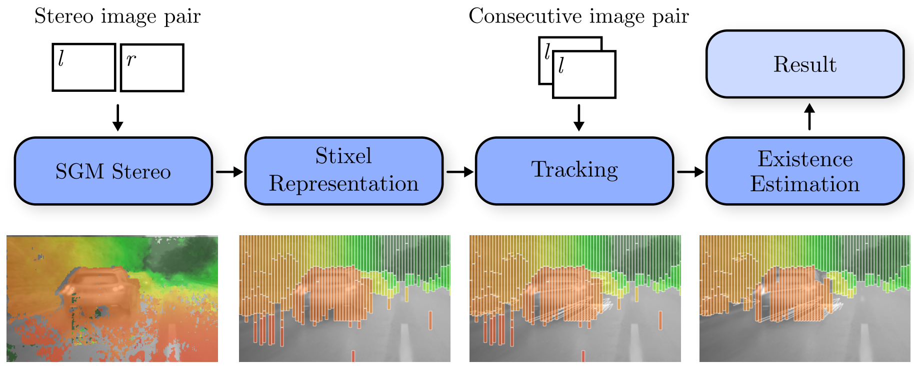


Fig. 2. General proceeding with result images for each respective component, applied to the first image of Figure 1 (best viewed in color)

Stixels (w.r.t. the observing camera) by means of Kalman filters. For our purpose, the filtered information provides a more robust basis for the subsequent cue computation.

Note that prior to this work, a different Stixel tracking approach has been presented [7] which also applies the 6D-Vision principle. However, the approach developed for our purpose allows to track several Stixels per column and retains the original grid the Stixel representation is given on. These are two properties required by subsequent analysis steps which are not fulfilled by the tracking presented in [7]. An example of the tracking result can be seen in Figure 2, indicated by the arrows at the bottom of Stixels covering the leading vehicle.

To exploit the large area of a Stixel, all optical flow measurements covered by a Stixel are used to calculate a single robust flow estimate. In our case, we apply a KLT-based tracking scheme and use the median to determine the flow result for a Stixel.

4 Existence Estimation

The existence of each Stixel is expressed as the posterior probability of a binary hypothesis, where $\exists$:= "exists" and $\bar{\exists}$:= "does *not* exist" and $P(\bar{\exists}_t \mid \mathbb{Z}^t) = 1 - P(\exists_t \mid \mathbb{Z}^t)$. The term $\mathbb{Z}^t = \{z_0, z_1, \ldots, z_{t-1}, z_t\}$ denotes the set of all frame-wise measurements (cues) z up to the current time step t. To separate real obstacles from phantoms, two cues are introduced later in Section 4.1.

To further take into account the temporal context obtained from the tracking component, the estimation problem is formulated time recursively according to a Bayes filter. This leads to the typical iteration between *prediction* (Equation 1) and *update* (Equation 2), where $P(\exists_t \mid \mathbb{Z}^t)$ is the sought quantity. The prediction is performed via a simple transition model, depicted in Figure 3.

$$P(\exists_t \mid \mathbb{Z}^{t-1}) = P(\exists_t \mid \exists_{t-1})\, P(\exists_{t-1} \mid \mathbb{Z}^{t-1}) + \\ P(\exists_t \mid \bar{\exists}_{t-1})\, \left[1 - P(\exists_{t-1} \mid \mathbb{Z}^{t-1})\right] \tag{1}$$

$$P(\exists_t \mid \mathbb{Z}^t) = \frac{P(z_t \mid \exists_t)\, P(\exists_k \mid \mathbb{Z}^{t-1})}{P(z_t \mid \exists_t)\, P(\exists_t \mid \mathbb{Z}^{t-1}) + P(z_t \mid \bar{\exists}_t)\, [1 - P(\exists_t \mid \mathbb{Z}^{t-1})]} \tag{2}$$

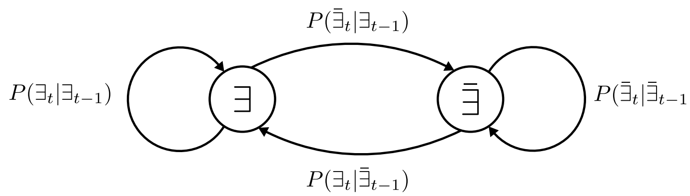


Fig. 3. Transition model to predict the current probability of existence before it is updated with new measurements

To avoid frequent label changes, the probability $P(\exists_t \mid \exists_{t-1})$ to remain in state $\exists$ is chosen relatively large while the probability $P(\exists_t \mid \bar{\exists}_{t-1})$ that a potential phantom changes to a real obstacle is rather small. However, it is still more likely than the change from a real obstacle to a phantom, which is a conservative and safe assumption. For our experiments we set $P(\exists_t \mid \exists_{t-1}) = 0.99$ and $P(\exists_t \mid \bar{\exists}_{t-1}) = 0.2$. A similar proceeding can also be found in [1], where a confidence measure for vehicle tracking is proposed.

4.1 Stixel-Wise Cues about Existence

In the following, two cues are introduced. From various cues tested, these two showed best performance for our purpose. The combination of cues is performed according to the naive Bayes approach, i.e. all N cues c_i are assumed independent, such that $P(z_t \mid x_t) = \prod_{i=1}^{N} P(c_{i,t} \mid x_t)$ with $x \in \{\exists, \bar{\exists}\}$. The two cues are:

Stixel Cluster Size Cue: The Stixel algorithm performs an optimal segmentation of the disparity data along the vertical image axis. As a consequence, horizontal smoothness is lost in favor of computational efficiency. To regain this lost information, we perform horizontal region growing on Stixel-level to cluster Stixels with similar depth and height[3]. The resulting cue value c_{CS} for each Stixel is the size of it's neighborhood region, such that $c_{\mathrm{CS}} = w_{\mathrm{cluster}} \times h_{\mathrm{cluster}}\,[m^2]$. As the cluster size of phantom Stixels turns out to be rather small, the cue value is clipped to 1 m^2 if it exceeds this value.

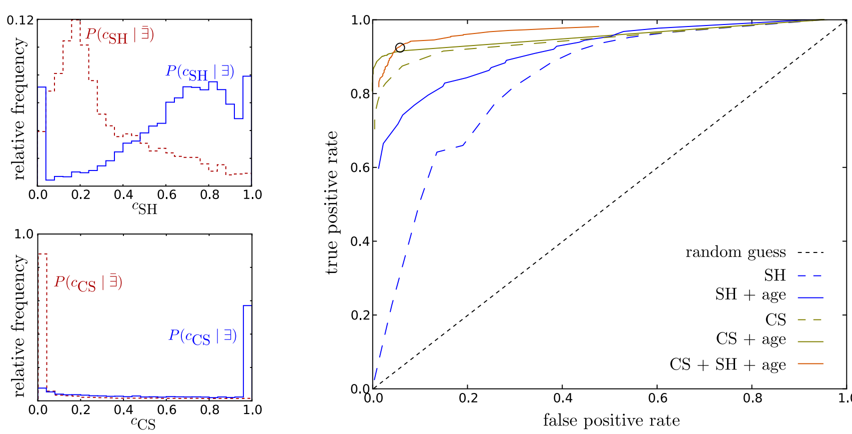
[3] The stop criteria for region growing correspond to the geometry of typical obstacles

Stixel Hypothesis Cue: The Stixel algorithm differentiates ground surface and obstacles by analyzing the progression of the disparity value along the vertical image axis. For the obstacle hypothesis, a constant disparity is assumed across the whole Stixel. The Stixel hypothesis cue evaluates the two likelihood functions (outcome of the Stixel optimization algorithm), that encode *how well* the disparity data matches both hypotheses. The final cue c_{SH} is chosen as the posterior probability of the object hypothesis, i.e. normalized to the range $[0, 1]$ by applying the law of total probability.

For both cue values it applies that a value close to one represents strong tendency towards a real obstacle while a cue value close to zero is more likely to belong to a phantom Stixel.

5 Evaluation

The detection performance of the proposed approach is presented by means of a receiver operator characteristic. To generate ground truth, we rely on an automated labeling strategy that we apply to 16 video sequences with 250 frames each, recorded under extremely bad conditions. Stixels falling into the unoccupied driving corridor are marked as negative examples (phantoms) and Stixels covering radar confirmed objects are marked as positive examples (real obstacles). This method of ground truth generation is also applied in [9]. In total we obtain roughly 12.000 negative and 55.000 positive examples which are split into a training and a test set.



(a) Cue histograms

(b) Detection performance (ROC)

Fig. 4. Distribution of the three cue values for real obstacles and phantoms (a) and the resulting detection performance shown as ROC curve (b). The term "age" denotes temporal filtering here

The distributions of the introduced cues (see Figure 4 (a)) are learned from the training set and are then used on the test set to evaluate the existence probability and generate the ROC curve (see Figure 4 (b)).

The results show, that both cues separate real obstacles from phantoms quite well. This also shows up in the final ROC result. It can be seen that temporally filtered cues (solid lines) always outperform the single-frame unfiltered cues (dashed lines) in terms of overall detection performance, which shows the effectiveness of the tracking component. Best performance with 93 % correctly classified Stixels is achieved with a combination of cluster size cue, hypothesis cue and temporal filtering (marked with a circle in Figure 4 (b)).

6 Conclusion

In this paper we showed a possibility to improve the bad weather performance of the Stixel representation by introducing existence estimation and temporal filtering for each Stixel. The evaluation shows a significant reduction of phantom Stixels while maintaining the majority of real obstacles. Furthermore, the introduced tracking scheme provides absolute velocity information for each Stixel that can be used for safety-relevant analysis, such as time-to-collision estimation.

References

1. Altendorfer, R., Matzka, S.: A confidence measure for vehicle tracking based on a generalization of Bayes estimation. In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV). San Diego, CA (2010)
2. Franke, U., Rabe, C., Badino, H., Gehrig, S.: 6D-Vision: Fusion of Stereo and Motion for Robust Environment Perception. In: German Association for Pattern Recognition (DAGM). pp. 216–223 (2005)
3. Hirschmüller, H.: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 807–814 (2005)
4. Keller, C., Enzweiler, M., Rohrbach, M., Llorca, D., Schnörr, C., Gavrila, D.: The benefits of dense stereo for pedestrian detection. Intelligent Transportation Systems, IEEE Transactions on (99), 1–11 (2011)
5. Perrone, D., Iocchi, L., Antonello, P., Fiat, C.: Real-time Stereo Vision Obstacle Detection for Automotive Safety Application. In: Intelligent Autonomous Vehicles. vol. 7, pp. 240–245 (2010)
6. Pfeiffer, D.: The Stixel World: A Compact Medium-level Representation for Efficiently Modeling Dynamic Three-dimensional Environments. Ph.D. thesis, Humboldt-Universität zu Berlin (2012)
7. Pfeiffer, D., Franke, U.: Efficient Representation of Traffic Scenes by Means of Dynamic Stixels. In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV). pp. 217–224. San Diego, CA (2010)
8. Pfeiffer, D., Franke, U.: Towards a Global Optimal Multi-Layer Stixel Representation of Dense 3D Data. In: British Machine Vision Conference (BMVC). Dundee, Scotland (2011)
9. Steingrube, P., Gehrig, S., Franke, U.: Performance evaluation of stereo algorithms for automotive applications. Computer Vision Systems pp. 285–294 (2009)