ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/228646852

# Robust and Precise 3D-Modeling of Traffic Scenes based on Dense Stereo Vision

**Article**

CITATIONS
4

READS
202

**4 authors**, including:

David Pfeiffer
Daimler
**29** PUBLICATIONS **2,186** CITATIONS
SEE PROFILE

Alexander Barth
Daimler
**20** PUBLICATIONS **486** CITATIONS
SEE PROFILE

Uwe Franke
Daimler
**177** PUBLICATIONS **22,326** CITATIONS
SEE PROFILE

All content following this page was uploaded by Alexander Barth on 27 May 2014.

The user has requested enhancement of the downloaded file.

# Robust and Precise 3D-Modeling of Traffic Scenes based on Dense Stereo Vision

David Pfeiffer, Alexander Barth and Uwe Franke
Daimler AG, Sindelfingen, Germany
e-mail: [david.pfeiffer, alexander.barth, uwe.franke]@daimler.com

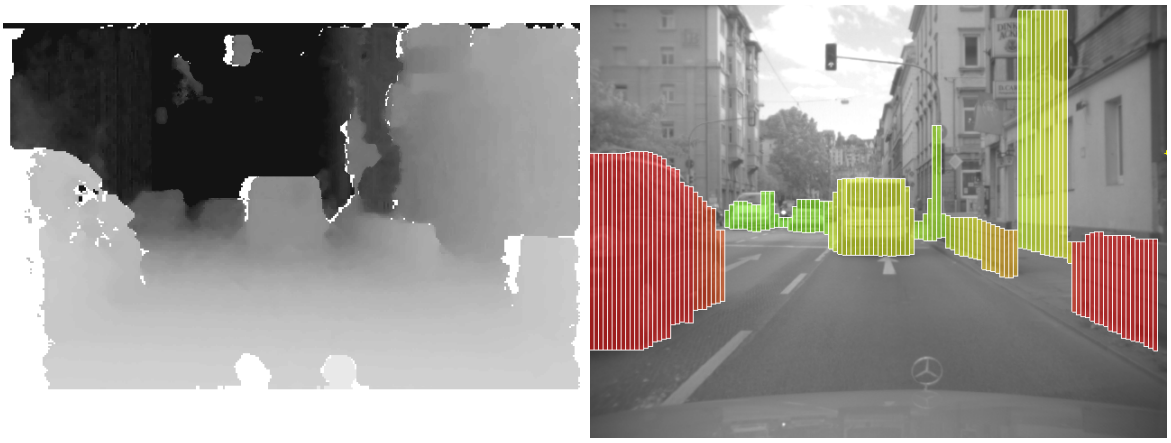
**Abstract:** Dense stereo vision is a key technology for natural scene understanding. Recent progress in real-time *dense stereo* provides high quality depth information for (almost) every pixel of an image. Based on this information the traffic environment can be modeled precisely and competely. This includes a vertical modeling of the driving corridor, the determination of the free space in front of the car and the representation of all vertical obstacles delimiting that freespace. We propose so called "stixels" to represent the 3D scene efficiently. Distance and height of each stixel are determined by the parts of the obstacle it aproximates. The stixel representation is designed to act as the common basis for the scene understanding tasks of driver assistance and autonomous systems. We show that the inherent spatial integration delivers depth information with an unprecedented accuracy.

**Keywords:** Computer Vision, Driver Assistance, Scene Modeling, Dense Stereo Vision, Vehicle Tracking, Vehicle Pose and Maneuver Estimation, Collision Detection and Prevention

## 1 Introduction

Future driver assistance systems for usage in complex urban scenarios demand a complete awareness of the situation, including all moving and stationary objects that determine the drivable free space. We are convinced that stereo vision will play an essential role for an extensive scene understanding. It incorporates information about position, size, and shape of arbitrary objects and thus allows for their detection and recognition independently of their specific appearance. Tracking these objects or even parts of them allows for estimating their motion, helping at the same time to distinguish between stationary and moving obstacles.

Stereo disparity estimation methods commonly rely on a correlation scheme. ASIC as well as FPGA stereo solutions have been developed for automotive applications. Recently, the dense stereo algorithm "Semi-Global Matching" (SGM) has been proposed, which offers accurate object boundaries and smooth surfaces under the constraint of real-time capability [1]. Due to the computational complexity, in particular the required memory bandwidth, the SGM algorithm is still too demanding for a general purpose CPU. However, FPGA implementations of the SGM algorithm exist that allow for realtime applicability [2]. Fig. 1(a) shows the result of the SGM method applied to a standard urban traffic situation.



(a) Dense disparity image (SGM result) (b) Stixel representation

Figure 1: (a) Dense stereo results overlaid on the image of an urban traffic situation. The colors encode the distance, red means close, green represents far. Note that SGM delivers measurements even for most pixels on the road.
(b) Stixel representation for this situation. The freespace (not explicitly shown) in front of the car is limited by the stixels, the colors encode the distance.

Recent improvements in disparity estimation yield a sub-pixel accuracy in the range of 0.2 pixel. This information is the basis for our precise modeling of the scene. First, we estimate the vertical shape of the road in front. Secondly, we determine the freespace, i.e. the area in front of the car containing no obstacles. Thirdly, we model the 3D-situation by a set of rectangular sticks named "stixels" as shown in Fig. 1(b). Each stixel is defined by its 3D position relative to the camera and stands vertically on the ground, having a certain height and a fixed width. The totality of all stixels limits the freespace and approximates the objects boundaries. If the width of the stixels is set to 5 pixels, a scene shown in a VGA image can be represented by 640/5=128 stixels only.

Section 2 sketches the used dense stereo vision algorithm. Section 3 describes the analysis of the freespace incorporating a vertical road estimation. The freespace is utilized to build the stixel-world as described in Section 4. We present a direct application of our proposed representation in Section 5 where stixels are utilized in an object tracking process. Section 6 concludes the paper.

## 2 Modern Dense Stereo Vision

Most real-time stereo algorithms based on local optimization techniques (e.g. correlation) yield sparse disparity data. In contrast, SGM aims to find a dense disparity image close to the global optimum by minimizing a two-dimensional energy in a dynamic-programming fashion on multiple (8 or 16) 1D paths across each pixel. The energy consists of three parts: a data term enforcing photo-consistency, a term to penalize minor depth changes and a larger penalty to prevent undesired depth discontinuities.

The implementation of this stereo algorithm on an FPGA allows for running this method in real-time in our demonstrator vehicle for daily use. Fig. 1(a) shows that SGM is able to model object boundaries precisely. In addition, the smoothness constraint used

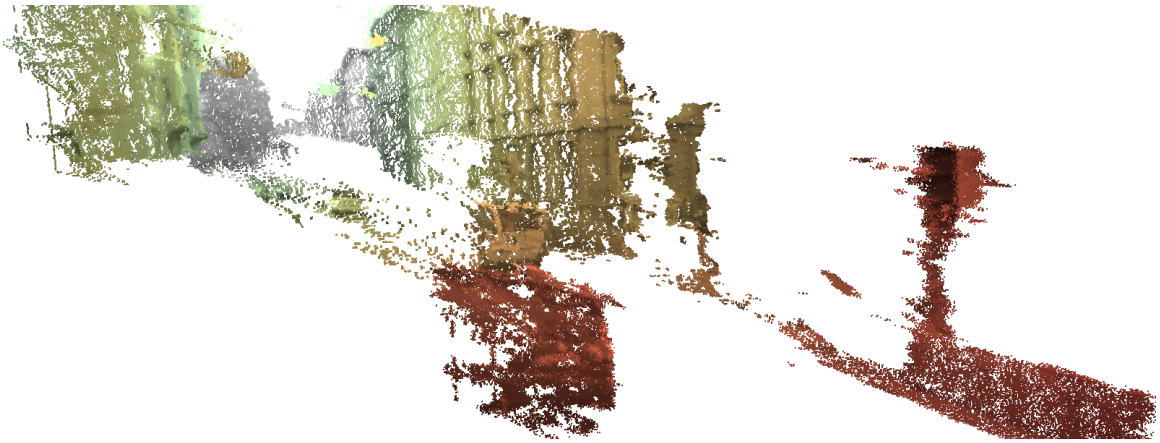
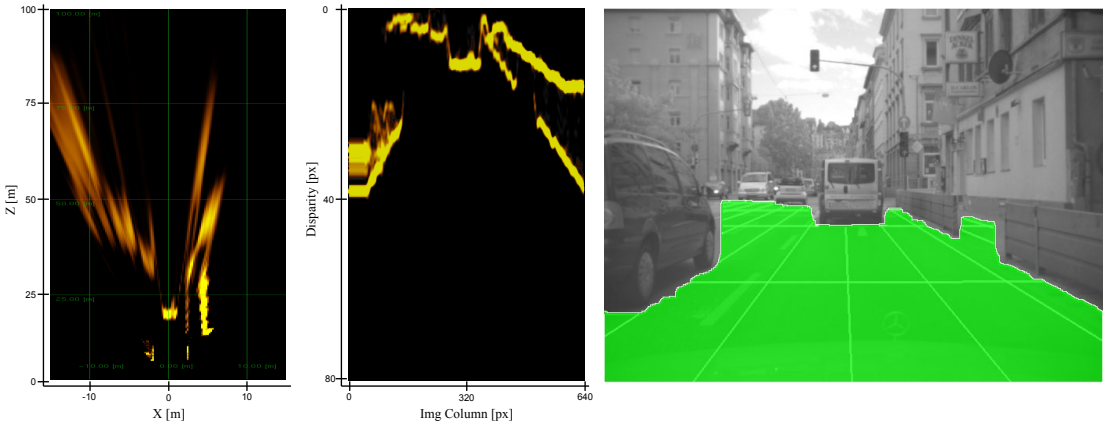


Figure 2: Shows a 3D point cloud representing derived stereo disparities using SGM and a 640×480 image as input. All 3D measurements on the road surface have been suppressed in order to highlight the quality of the derived stereo information. Obviously no measurements can be obtained for occluded areas.

in the algorithm leads to smooth estimations in low contrast regions, exemplarily seen on the street and the untextured parts of the vehicles and buildings. A 3D visualization is given in Fig. 2.

# 3 Free Space Computation

The stereo disparities are used to build a stochastic occupancy grid. An occupancy grid is a two-dimensional array which models occupancy evidence of the environment and thus approximates the real world. Occupancy grids are computed in real-time using the method presented in [3] which allows to propagate the uncertainty of the stereo disparities onto




(a) cartesian occ. grid (b) column/disparity (c) freespace in image

Figure 3: Occupancy grids: Fig. (a) and (b) show the occupancy grid obtained from the disparity image shown in Fig. 1(a) in cartesian and polar representation respectively (brightness encode the likelihood of occupancy). Fig. (c) shows the resulting freespace within the image.

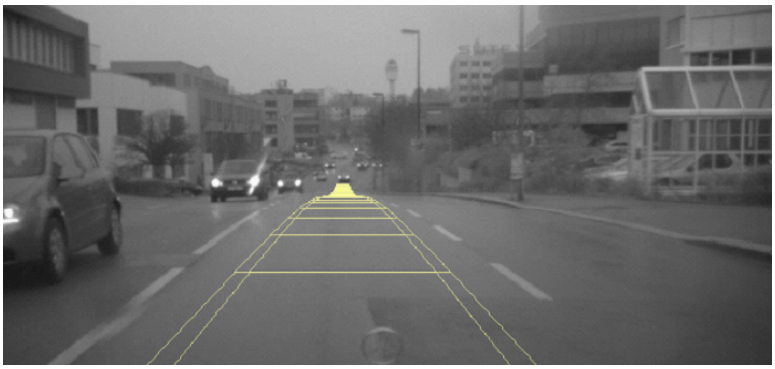


Figure 4: A vertical road estimate computed from a Kalman filtered B-Spline model.

the grid. Only those 3D measurements lying above the road are registered as potential obstacles in the occupancy grid. Fig. 3(a) shows an example of a cartesian occupancy grid obtained from the disparity image shown in Fig. 1(a).

To obtain information on the road height, we estimate its pose by fitting a B-Spline surface to the 3D data as proposed in [4]. Fig. 4 shows a road with a sinusoidal shape in its vertical direction. The overlay displays the estimated course.

From the occupancy grid the free space is computed as described in [5]. Instead of using a threshold operation for every column independently, dynamic programming is used to find the optimal path cutting the polar grid from left to right minimizing an energy functional. This avoids heuristics and allows to favor temporal and spatial smoothness of the solution. For details see the mentioned paper. Fig. 3(c) displays the obtained freespace after a backprojection into the original image.

# 4 Building the Stixel-World

A polygonal approximation yields a very compact representation of the freespace. In order to obtain a representation of all objects above ground, which is robust, compact as well as complete at the same time, we suggest to build a medium level representation named the *stixel world*. It assumes that objects reside on the ground and have approximately vertical surfaces. A stixel representation of a common urban traffic situation is shown in Fig. 1(b). The height and distance are determined as follows:

Initially the distance of the stixel is taken from the freespace boundary. Subsequently the height is estimated within an optimization step that separates foreground (object) disparies from background disparities. In [5] it is shown that this task can be solved optimally by dynamic programming as well.

Given the base-point and height, the determination of the highly accurate stixel distance is straightforward. A histogram analysis and spatial integration of the disparities measured at a stixel allows for a significant gain in depth accuracy. An average stixel covers hundreds of disparity values.

The obtained stixel measurements are depicted in Fig. 5. Despite the high quality of SGM one can observe that the raw stereo information is spread over a large scale along the depth axis, especially at horizontal object boundaries. However, planar object surfaces are accurately reconstructed by neighboring stixels, although their final position

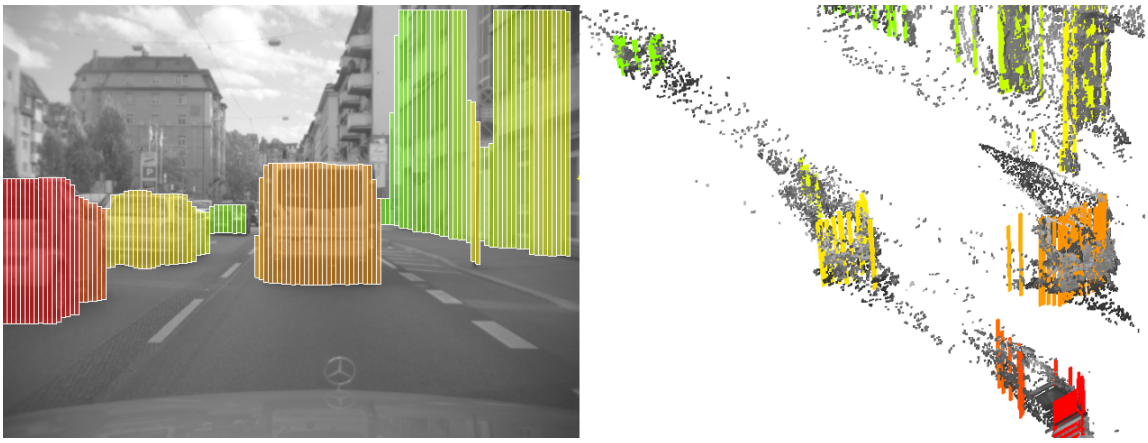
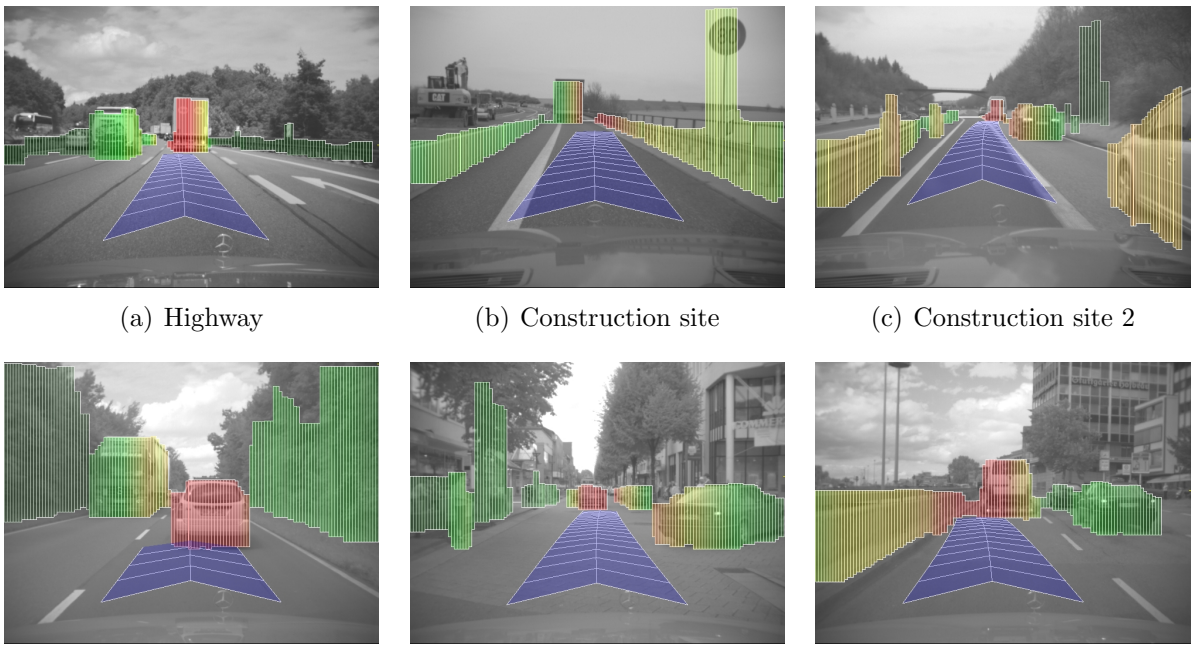


Figure 5: Visualization of the derived stixels in the image as well as within the 3D clouds derived from the dense stereo algorithm. For the sake of clarity road points have not been visualized. Please note the accuracy and concision by which the stixel measurements reside in the wide spreading stereo clouds.

is determined independently. This is noteworthy and a result of spatial integration.

Yet another advantage of the stixels is that arbitrarily shaped objects can be approximated with any desired accuracy by simply varying the width of the stixels. In our experiments we use a fixed width of 5 pixels as a good compromise between compactness and precision.



(a) Highway (b) Construction site (c) Construction site 2

(d) Rural road (e) Urban traffic (f) Urban traffic 2

Figure 6: Evaluation of stixels in different real world road scenarios showing a highway, a construction site, a rural road and an urban environment. The color encodes the lateral distance to the estimated driving corridor.

# 5 Applications

The stixel representation proves itself usefull for different traffic scenarios, e.g. urban areas, rural roads as well as highways. A set of typical examples for crowded traffic situations is depicted in Fig. 6.

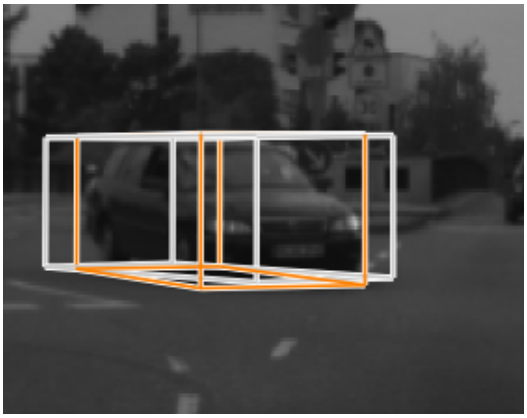
Intersections are of particular interest since they turn out to be a hotspot for traffic accidents. In order to be able to recognize potentially dangerous situations one has to detect and to track other traffic participants reliably. This implies the estimation of their shape and pose as well as the determination of their complete motion state.

In [6] an approach is presented where pose and motion state estimates are derived from the movement of a rigid 3D point cloud, representing the object's shape. Lateral movements are restricted to circular path motion in a Kalman filter framework based on a simplified vehicle motion model. This model allows to estimate not only the velocity and acceleration, but also the yaw rate of an object. Since there is no requirement that the point cloud completely covers the object, a precise segmentation of the object boundaries based on the sparse point cloud is not guaranteed. However, accurate knowledge of the object boundaries is essential for precise collision prediction in future driver assistance systems.

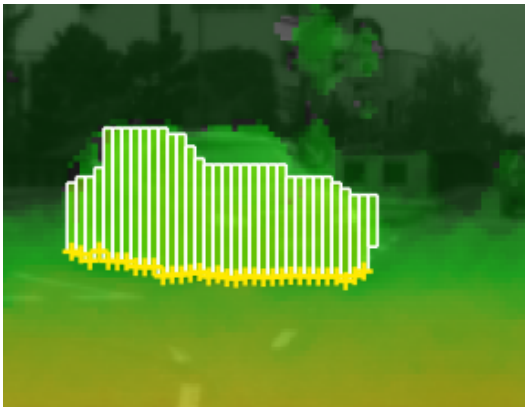
Fig. 7(a) shows a situation where the gray object box obtained by the mentioned approach does not precisely match the actual pose and dimensions of the vehicle. This problem can be overcome by using the stixel representation. As shown in Fig. 7(b), the stixels fit the silhouette of the car quite good. Obviously, this silhouette constrains the objects pose and dimension. This is demonstrated by Fig. 7(c). On the left side the red dotted box represents the estimate of the object, where the bars show the position of the stixels. The most outer visible cube corners must project onto the image columns $u_l$ and $u_r$ of the most outer stixels constraining the lateral object position. At the same time, the outer stixels define the distance of the visible outer corners (labeled with 1 and 4). The inner stixels cannot be directly assigned to particular object points. However, they provide useful information to define additional constraints on the distance to the center of visible object sides (labeled with 2 and 3).

In the Kalman filter framework this information can be used to enforce the estimation to fulfill these constraints. On the right hand Fig. 7(c) shows the improved position. The final result is also shown in Fig. 7(a) by the orange box.

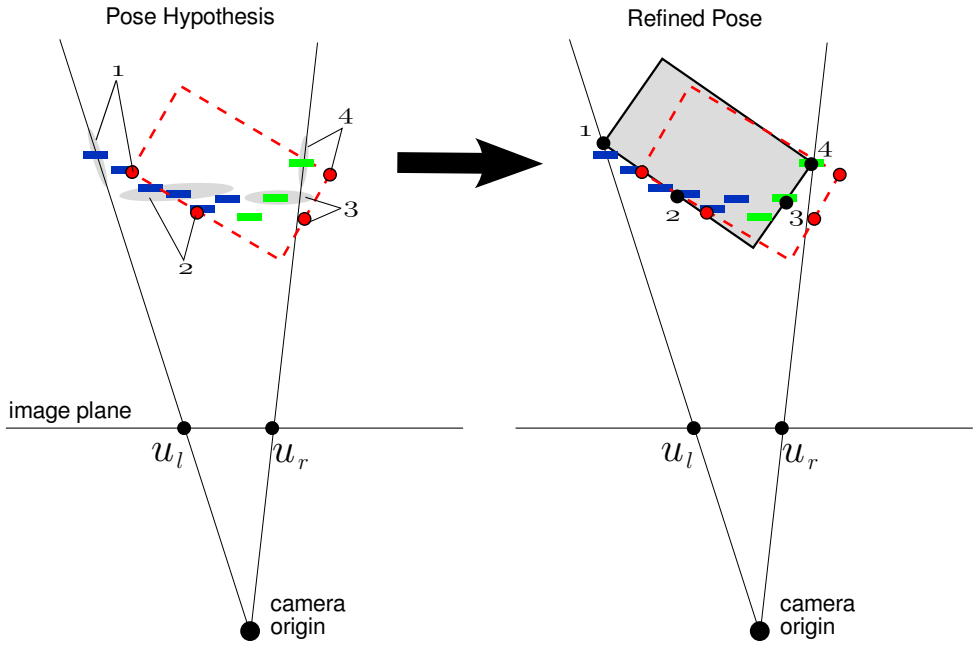
The tracking results of an exemplary sequence are shown in Fig. 8. The box indicates the current object pose and size, tracked feature points and optical flow vectors are also superimposed. The carpet on the ground represents the predicted driving path for the next second (based on the current motion estimate).



(a) Refinement Result

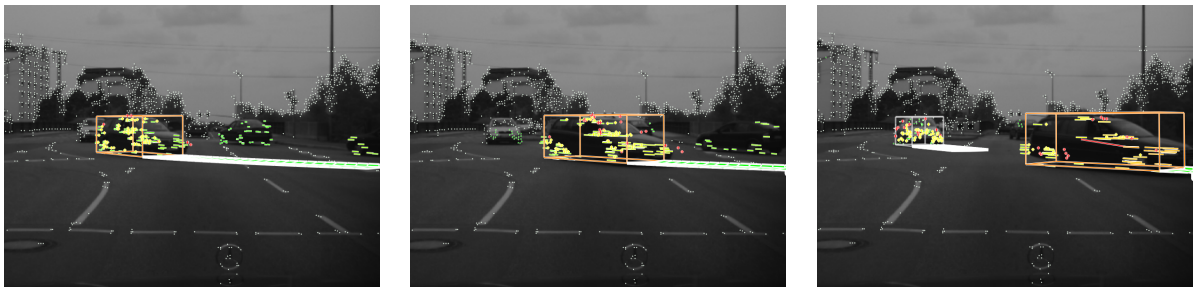


(b) Stixel Cluster




(c) Geometric Constraints

Figure 7: Several constraints on depth and lateral position are derived from a stixel cluster and assigned to four characteristic object points, that depend on visibility properties of the pose prior.



(a) (b) (c)

Figure 8: Example tracking results of an intersection scene with the estimated pose and predicted driving path superimposed.

# 6 Conclusion

A new primitive called stixel was proposed for modelling 3D traffic scenes. The resulting *stixel-world* turns out to be a robust and very compact representation of the traffic environment, modeling the freespace as well as static and moving objects.

Stochastic occupancy grids are computed from dense stereo information. The freespace is computed from a polar representation of the occupancy grid in order to obtain the base-point of the obstacles. The height of the stixels is obtained by segmenting the disparity image in foreground and background disparities using dynamic programming. Spatial integration offers a highly accurate and robust determination of the depth information.

The proposed stixel scheme serves as a well formulated medium-level representation for traffic scenes without the loss of generality. The stixel representation has been successfully applied to the task of object tracking and pose refinement, yielding accurate estimates of object boundaries. This information is essential for precise collision prediction.

# References


[1] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *CVPR*, 2005.

[2] S. Gehrig, F. Eberli, and T. Meyer, "A real-time low-power stereo vision engine using semi-global matching," in *accepted for publication at ICVS*, 2009.

[3] H. Badino, U. Franke, and R. Mester, "Free space computation using stochastic occupancy grids and dynamic programming," in *Workshop on Dynamical Vision, ICCV*, Rio de Janeiro, Brazil, October 2007.

[4] A. Wedel, U. Franke, H. Badino, and D. Cremers, "B-spline modeling of road surfaces for freespace estimation," in *Intelligent Vehicles Symposium, IEEE*, 2008.

[5] H. Badino, U. Franke, and D. Pfeiffer, "The stixel world - a compact medium level representation of the 3d-world," in *DAGM Symposium*, Jena, Germany, September 2009.

[6] A. Barth and U. Franke, "Where will the oncoming vehicle be the next second?" ¡¡¡¡¡¡¡ fas2009V1.bbl in *Intelligent Vehicles Symposium, IEEE*, 2008.

[7] A. Barth, D. Pfeiffer, and U. Franke, "Vehicle tracking at urban intersections using dense stereo," in *Submitted*, 2009. ======= in *Intelligent Vehicles Symposium, IEEE*, 2008. ¿¿¿¿¿¿¿ 1.8


View publication stats