ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3/W3, 2013
CMRT13 - City Models, Roads and Traffic 2013, 12 – 13 November 2013, Antalya, Turkey

AN INCREMENTAL MAP BUILDING APPROACH VIA STATIC STIXEL INTEGRATION

M. Muffert[a], S. Anzt[b], U. Franke[a]

[a] Daimler Research, Sindelfingen, Germany - [maximilian.muffert, uwe.franke]@daimler.com
[b] Hochschule für Technik (HFT), Stuttgart, Germany - silke.anzt@hft-stuttgart.de

KEY WORDS: Map Building, Stereo Vision, Stixel Representation, Evidential Theory

ABSTRACT:

This paper presents a stereo-vision based incremental mapping approach for urban regions. As input, we use the 3D representation called multi-layered Stixel World which is computed from dense disparity images. More and more, researchers of Driver Assistance Systems rely on efficient and compact 3D representations like the Stixel World. The developed mapping approach takes into account the motion state of obstacles, as well as free space information obtained from the Stixel World. The presented work is based on the well known occupancy grid mapping technique and is formulated with evidential theory. A detailed sensor model is described which is used to determine the information whether a grid cell is *occupied*, *free* or has an *unknown* state. The map update is solved in a time recursive manner by using the Dempster's Rule of Combination. 3D results of complex inner city regions are shown and are compared with Google Earth images.

1 INTRODUCTION

Precise ego vehicle localization in urban regions is one of the major challenges in autonomous driving applications (Kammel et al., 2008), (Thrun, 2010). For this task, it is not sufficient to only use GPS information, because well known limitations like multi-path effects or signal shadowing occur in urban regions. Thus, precise self-localization approaches are often based on prior map information. Next to the (re)detection of 3D landmarks (Lategahn and Stiller, 2012), another possibility is the use of geometrical information of the static environment, like building facades or traffic infrastructures (Larnaout et al., 2012). With the help of this information, the accuracy of self localization increases. Furthermore, modern driver assistance systems like lane change systems or collision avoidance systems (Muffert et al., 2013) would also benefit from detailed prior map information to support trajectory planning for the ego vehicle.

In this paper, the focus is on the incremental map building process in urban regions given the absolute pose and orientation of the ego vehicle. The goal is to update the map sequentially to achieve real time capability in future. For the mapping approach, the results should contain static environment information (e.g. building facades) since only this information is useful for further tasks such as ego vehicle localization. In contrast to the use of ultrasonic sensors (Pagac et al., 1998) or laser scanners (Thrun, 2010), a stereo camera system is applied which went into series production in the Mercedes-Benz S- and E-class in summer 2013. It is assumed that the approximated absolute pose of the experimental vehicle is known, e.g. by using an inertial measurements unit in combination with a GPS sensor.

For the first time we use the 3D scene representation called Stixel World (Pfeiffer and Franke, 2011) as input data for the mapping approach. It is computed from dense disparity images (Fig. 2(a)) at each acquisition time step. A Stixel is defined by its 2D position and 2D velocity, its width, and its height referring to the coordinate system of the ego vehicle (Fig. 2(b)). Hereby, the Stixel World efficiently describes dynamic and static objects, as well as free space information. In contrast to the raw input data (up to 500.000 disparity values), the 3D scene is represented by a few hundred Stixels only. As a result, this step reduces the computational burden for the following steps significantly. Due to named properties research groups of Driver Assistance Systems increasingly rely on 3D representations like the Stixel World.

Figure 1: The result of the mapping approach for a 3.000 image sequence. For the visualization and the data structure the *OctoMap* Representation (Wurm et al., 2010) is used. The purple line represents the driven path of the ego vehicle. The height of the static environment is color encoded.

Subsequently, the multi-layered Stixel World is segmented into static and dynamic object classes using the approach by Erbs *et al.* (Erbs et al., 2012), as seen in Fig. 2(c). This allows to exclude dynamic objects like driving vehicles or bicyclists, so that static environment information is only considered for the mapping approach. In Fig. 2(d), the final set of Stixels used for the mapping approach is shown.

Because of the fact that the Stixels include height information, a height layer of the complete 2D occupancy grid map is also computed. The combination of the 2D map with the height layer allows to estimate a 3D environment model which is visualized using the *OctoMap* framework (Wurm et al., 2010)(Fig 1).

The remainder of this paper is organized as follows: Section 2 gives a brief overview about related work. Then, Section 3 points out the mapping process which is based on the idea of Cartesian occupancy grid maps (Moravec and Elfes, 1985) in combination with evidential theory (Shafer, 1976). Finally, results are shown in Section 4 and Section 5 summarizes the paper.

2 RELATED WORK

Elfes and Movarec (Elfes, 1987), (Moravec and Elfes, 1985) introduced the occupancy grids at first using wide angle sonar sensors. Thrun *et al*. (Thrun et al., 2005) give a detailed overview about probabilistic methods for 2D occupancy grid mapping techniques given the robot's pose. At first, Thrun *et al*. describe an incremental *ad hoc* approach which is based on a binary Bayes filter. A drawback is the assumption that all grid cells are independent and mostly initialized with a probability of 0.5. Furthermore, the described sensor model is considered unsuitable for stereo vision sensors in the authors view: It returns fixed probability values for *occupied* (e.g. 0.8) or *free* (e.g. 0.2) which does not capture the measurement characteristics of a stereo camera system. In addition, Thrun *et al*. formulate a maximum a posteriori approach with a descriptive sensor model which considers the dependence of all grid cells. This is a golden standard model, but the disadvantage is the non incremental map update which is a requirement for the developed mapping approach. A detailed discussion of both methods can also be found in (Merali and Barfoot, 2012).

To reconstruct the complete 3D environment of urban scenes Gallup *et al*. (Gallup et al., 2010) and Zheng *et al*. (Zheng et al., 2012) present probabilistic methods using street-level videos or photo collections. The required depth maps are estimated using either structure from motion or dense stereo techniques.

Gallup *et al*. (Gallup et al., 2010) use a n-layer height map which allows the representation of overhanging structures like balconies or bridges. Zheng *et al*. (Zheng et al., 2012) take up this idea and present an efficient incremental depth map fusion framework using wavelet based compression techniques.

The developed mapping process is formulated with evidential theory (Shafer, 1976). We take up the idea from (Moras et al., 2011) and (Yang and Aitken, 2006) which use ultrasonic and range sensors in contrast to a stereo camera system.

Yang *et al*. (Yang and Aitken, 2006) give a detailed overview of evidential map building techniques: Each grid cell state is described by its power set which is defined by the subsets *free*, *occupied* and *unknown*. Furthermore, the conflict in a cell is described. For each subset a probability assignment function is defined which formulates a detailed characterization of the sensor model. The Dempster's Rule of Combination (Shafer, 1976) makes an incremental cell update possible.

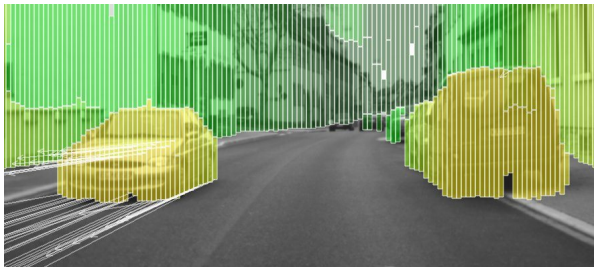
3 MAPPING PROCESS

Assuming a planar surface of the environment, a 2D Cartesian reference grid map $\mathcal{M} = \{\boldsymbol{m}_j\}, j \in \{1, .., \mathrm{J}\}$ with the cells $\boldsymbol{m}_j$ is given. For each cell hypotheses of *free* (F) and *occupied* (O) are made. As described in (Moras et al., 2011) and (Yang and Aitken, 2006), in evidential theory the power set is defined by

$$2^{\Omega} = \{\emptyset, \{O\}, \{F\}, \{O, F\}\}. \quad (1)$$

The subset $\Omega = \{O, F\}$ describes the ignorance of a cell and is denoted as the *unknown* state U. The subset $\emptyset$ is the empty set. Following the definition of Eq. (1), for each element A of the power set a *belief mass function* $m_j^*(A)$ is specified with the property $m_j^*(\emptyset) + m_j^*(O) + m_j^*(F) + m_j^*(U) = 1$. The complete state of a cell is given by the Cartesian information of the center point (X_j and Z_j) and the *belief mass function*: $\boldsymbol{m}_j = [X_j \; Z_j \; m_j^*(A)]^T$.

(a) The dense disparity image, estimated via SGM (Gehrig et al., 2009). The color represents the distance. Red stands for near, green for far away objects.



(b) The Dynamic Stixel representation (Pfeiffer and Franke, 2011). It contains free space information and object boundaries. The arrows predict the Stixel position within the next 0.5 sec.

(c) The results of the Stixel segmentation (Erbs et al., 2012). The vehicle on the left side is classified as an oncoming object. The corresponding Stixels are red. The static environment is colored in dark grey.

(d) The input data for the mapping approach: The static Stixels are colored in yellow to highlight the input data. Here, Stixels with a distance of 25 m to the ego vehicle coordinate system are used.

Figure 2: Pipeline of the generation of the input data.

3.1 Input data

With the help of the rectified image sequences of the stereo camera, dense disparity images are estimated using semi-global matching (SGM) (Gehrig et al., 2009) as seen in Fig. 2(a). Then, the multi-layered Stixel World (Pfeiffer and Franke, 2011) is computed. The Stixels represent the relevant information of the current 3D scene such as free space and dynamic and static obstacles (see Fig. 2(b)). A single Stixel is a vertically oriented rectangle with a fixed width in the image (*e.g.* 5 px) and a variable height.

Referred to the 3D Cartesian space, each Stixel $\boldsymbol{S}_i = [\boldsymbol{s}, \dot{\boldsymbol{s}}, H, W]_i^T$ with $i \in \{1, .., \mathrm{I}\}$ is parametrized by its position vector $\boldsymbol{s} = [X, Z]^T$ with the lateral and longitudinal components X and Z, its height H and its width W. Consequently, obstacles are described by planar, vertical oriented surfaces which is a common assumption in urban regions. As a result, facades behind parking vehicles can be mapped, as shown in Fig. 5.

A well known challenge in mapping approaches is that dynamic objects have to be detected to be not taken into account in the fusion step. To overcome this problem, the Stixels are tracked over time which is achieved by the 6D-Vision principle (Franke et al., 2005). This scheme uses Kalman filter to estimate both the position and velocity of each Stixel. Performing this step makes it possible to obtain a motion state $\dot{\boldsymbol{s}} = [\dot{X}, \dot{Z}]^T$.

The separation of the scene into moving or stationary obstacles is achieved by a multi-class traffic scene segmentation (Erbs et al., 2012) which is based on a conditional random field framework. With the help of this segmentation, from now, $\boldsymbol{S}_i$ describes only the static Stixels which are used as input data. An example is shown in Fig. 2(d).

With the help of the pinhole camera projection equation (projection from the disparity(d)-column(u)-space to the 3D space with given variances σ_d^2 and σ_u^2) and using Gaussian error propagation it could be shown that the covariance matrix $\Sigma_i^{(s)}$ of the position $\boldsymbol{s}_i$ of a Stixel is estimated by

$$\begin{aligned}\Sigma_i^{(s)} &= \begin{pmatrix} \sigma_i^{2(X)} & \sigma_i^{(XZ)} \\ \sigma_i^{(XZ)} & \sigma_i^{2(Z)} \end{pmatrix} \\ &= \begin{pmatrix} \left(\frac{Z}{f}\right)^2\sigma_u^2 + \left(\frac{XZ}{fb}\right)^2\sigma_d^2 & \frac{XZ^3}{f^2b^2}\sigma_d^2 \\ \frac{XZ^3}{f^2b^2}\sigma_d^2 & \left(\frac{Z^2}{fb}\right)^2\sigma_d^2 \end{pmatrix},\end{aligned} \tag{2}$$

including the baseline b and the focal length f of the stereo rig. Due to the assumption that the ego vehicle's position and orientation is known the Stixels are transformed into the reference coordinate system of the map $\mathcal{M}$ at each acquisition time step.

3.2 Sensor Model

The sensor model describes in which way the Stixel position $\boldsymbol{s}_i$ influences a set of cells $\mathcal{M}_{s_i} \in \mathcal{M}$. Due to the fact that each Stixel obtains information about free space and obstacles (see Section 3.1), the cells $\mathcal{M}_{s_i}$ are partitioned into cells which support a set of *occupied* cells $\mathcal{M}_{s_i}^{(O)}$ and a set of *free* cells $\mathcal{M}_{\widehat{s}_{s_i i}}^{(F)}$, as seen in Fig. 3. Additionally, cells which are not influenced by a Stixel (i.e. $\mathcal{M} \backslash \mathcal{M}_{s_i}$) are defined with an *unknown* state U.

To formulate the mapping process with evidential theory, at first, the *belief mass function* of the subset $m_j^*(O)$ has to be defined. Next to the cell which is directly hit by a Stixel a region of cell neighbors is considered in $\mathcal{M}_{s_i}^{(O)}$. The size of the region depends on the estimated uncertainties of $\boldsymbol{s}_i$. With the help of $\Sigma_i^{(s)}$, the *belief mass function* for the subset O is defined by

$$m_j^*(O) = \begin{cases} f_O(\boldsymbol{m}_j \mid \mathcal{N}_2(\boldsymbol{s}_i, \Sigma_i^{(s)})) & , \text{ if } \boldsymbol{m}_j \in \mathcal{M}_{s_i}^{(O)} \\ 0 & , \text{ else .} \end{cases} \tag{3}$$

The function $f_O(\boldsymbol{m}_j \mid \mathcal{N}_2(\boldsymbol{s}_i, \Sigma_i^{(s)}))$ returns an occupancy value of evidence given the 2D normal distribution $\mathcal{N}_2(\boldsymbol{s}_i, \Sigma_i^{(s)})$ as long as the current cell $\boldsymbol{m}_j$ is an element of $\mathcal{M}_{s_i}^{(O)}$.

As shown in Fig. 3, the space between the ego vehicle and the Stixels $\widehat{\boldsymbol{S}}_i \in \boldsymbol{S}_i$ which represents the first obstacles in the image rows is free space. A set of K rays $\mathcal{R}_{\widehat{s}_i} = \{\boldsymbol{r}^k\}_{\widehat{s}_i}$ with $k \in \{1, .., K\}$ from the ego vehicle to the Stixel are defined where K depends on the width of the Stixel W_i. All cells intersecting with those rays are defined as $\mathcal{M}_{\widehat{s}_i}^{(F)}$. Subsequently, the rays are partitioned into ray segments as illustrated in Fig. 3.

In a more mathematical way, the intersection $\boldsymbol{m}_j \cap \boldsymbol{r}^k$ of an arbitrary cell $\boldsymbol{m}_j$ with a ray $\boldsymbol{r}^k$ returns the adequate ray segment with its length $\mid \boldsymbol{m}_j \cap \boldsymbol{r}^k \mid$. The assumption is made that the more a cell includes rays segments, the greater is the evidence of *free*.

With the summarized length $L = \sum_K \mid \boldsymbol{r}_{\widehat{s}_i}^k \mid$ of all rays the *belief mass function* $m_j^*(F)$ for the subset F is defined by

$$m_j^*(F) = \begin{cases} L^{-1} \sum_K \mid \boldsymbol{m}_j \cap \boldsymbol{r}_{\widehat{s}_i}^k \mid & , \text{ if } \boldsymbol{m}_j \in \mathcal{M}_{\widehat{s}_i}^{(F)} \\ 0 & , \text{ else .} \end{cases} \tag{4}$$

Due to the properties of the *belief mass function*, $m_j^*(U)$ and $m_j^*(\emptyset)$ are defined as

$$p_{\boldsymbol{m}_j}^*(U) = \begin{cases} 1 - f_O(\boldsymbol{m}_j \mid \mathcal{N}_2(\boldsymbol{s}_i, \Sigma_i^{(s)})) & , \text{ if } \boldsymbol{m}_j \in \mathcal{M}_{s_i}^{(O)} \\ 1 - L^{-1} \sum_K \mid \boldsymbol{m}_j \cap \boldsymbol{r}_{\widehat{s}_i}^k \mid & , \text{ if } \boldsymbol{m}_j \in \mathcal{M}_{\widehat{s}_i}^{(F)} \\ 1 & , \text{ else} \end{cases} \tag{5}$$

and $p_{\boldsymbol{m}_j}^*(\emptyset) = 0$.

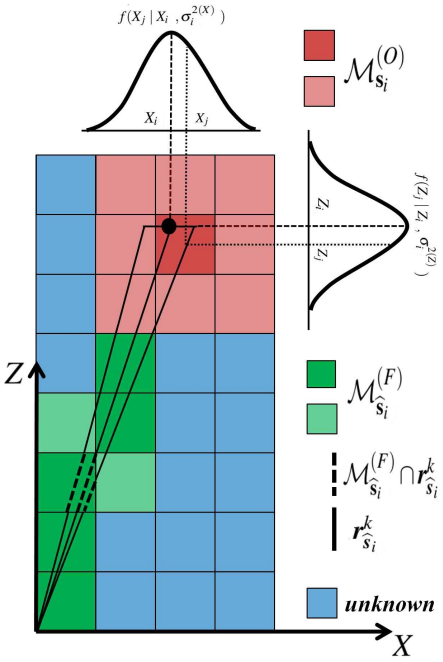


Figure 3: The sensor model of the mapping approach. The cell which is directly hit by a Stixel is dark red, the immediate cell neighbors are bright red. For these cells an evidential value of *occupied* is estimated which is based on a two dimensional Gaussian distribution. The distribution is defined by the position of a Stixel $\boldsymbol{s}_i = [X, Z]_i$ and the covariance matrix $\Sigma_i^{(s)}$ which includes the variances $\sigma_i^{2(X)}$ and $\sigma_i^{2(Z)}$. The cells which support *free* regions are green. The more rays pass through a cell the greater is the evidence of *free*. Cells which are not influenced have an *unknown* state (blue).

3.3 Incremental Map-Updating via Dempster's Rule of Combination

Up to this point, the sensor model and the definition of the *belief mass function* are described. Now, an incremental map update between consecutive time steps $t-1$ and t is realized to recursively merge new Stixel measurements into the existing map. The Dempster's Rule of Combination is used which combines two independent states $m^*_{j,t}(A)$ and $m^*_{j,t-1}(A)$ at consecutive time steps. The general formulation is defined by

$$m^*_{j,t}(A) \oplus m^*_{j,t-1}(A) = \begin{cases} \frac{1}{1-K} \sum\limits_{B \bigcap C = A} m^*_{j,t}(B)\, m^*_{j,t-1}(C) \text{ if } A \neq \emptyset \\ 0, \quad \text{else} \end{cases} \tag{6}$$

with

$$K = \sum_{B \bigcap C = \emptyset} m^*_{j,t}(B)\, m^*_{j,t-1}(C)\,.$$

In this contribution, the fusion step is formulated as

$$\begin{aligned} m^*_j(O) &= \frac{m^*_{j,t}(O)m^*_{j,t-1}(O)+m^*_{j,t}(O)m^*_{j,t-1}(U)+m^*_{j,t-1}(O)m^*_{j,t}(U)}{\left(1-K\right)}, \\ m^*_j(F) &= \frac{m^*_{j,t}(F)m^*_{j,t-1}(F)+m^*_{j,t}(F)m^*_{j,t-1}(U)+m^*_{j,t-1}(F)m^*_{j,t}(U)}{\left(1-K\right)}, \\ m^*_j(U) &= \frac{m^*_{j,t}(U)m^*_{j,t-1}(U)}{\left(1-K\right)} \end{aligned} \tag{7}$$

and

$$K = m^*_{j,t}(O)m^*_{j,t-1}(F) + m^*_{j,t-1}(O)m^*_{j,t}(F)\,.$$

For a better understanding of the update step from Equation 7, results of a simulation are shown in Fig. 4:

At the beginning ($t = 0$), all cells are initialized with an *unknown* state which is visualized by a dark blue map. Because of this, the power set of a cell is defined as $m^*_{j,0}(O) = m^*_{j,0}(F) = m^*_{j,0}(\emptyset) = 0$ and $m^*_{j,0}(U) = 1$.

At the next time step ($t = 1$), the figure points out the integration of one Stixel measurement into the map. The evidence for *occupied* grows in these cells which are directly hit by the Stixel. In the cells between the ego position (black cross) and the Stixel the subset of *free* grows slowly, visualized by the color green.

The influence of an additional measurement and the recursive filter characterization can be seen at the following time step $t = 2$: In overlapping areas the evidence of *free* and *occupied* grows which is particularly shown by the free space area.

During the update step K is not equal to zero, as long as a conflict occurs in the cells.As an example, this case happens if a current Stixel measurement overlaps free space area from the last time step. Specifically, the conflict is significant if Stixel outliers occur. This effects a robustness of our algorithm.

As seen in Eq. 7 the subsets O, F and U are normalized by the term $(1-K)$ at each time step to satisfy the properties of the *belief mass function*.

To obtain the mentioned height layer, the Stixel height $H_{i,t}$ is taken into account. For each cell a height $H_{j,t}$ is estimated over time by

$$H_{j,t} = \frac{H_{i,t}m^*_{j,t}(O) \;+\; H_{j,t-1}}{1 + m^*_{j,t}(O)}\,. \tag{8}$$

Note that the height $H_{j,0}$ is initialized by zero. For the visualization of the 3D environment model, the open source framework *OctoMap* (Wurm et al., 2010) is used which is based on a 3D Octree representation (see Fig. 1).

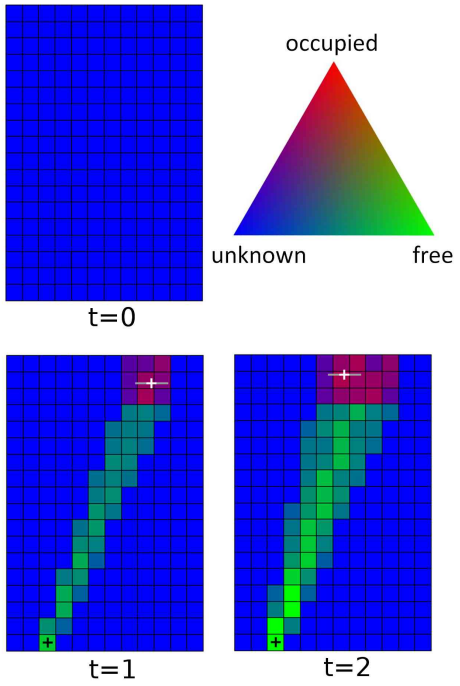


Figure 4: Example of the time recursive update step: At $t = 0$, the map is initialized as *unknown* (dark blue). At time step $t = 1$ the integration of one Stixel into the map is shown. The cell which is directly hit is red, the immediate cell neighbors are purple. The field of view between ego position (black cross) and the Stixel position is free space which is encoded with the color green. With an additional Stixel measurement at time step $t = 2$ history and current measurements are merged. The stronger the color saturation the stronger the evidence of *free*, *occupied* or *unknown*.

4 RESULTS

In this work, mapping results of a 3.000 image sequence of an urban drive (see Fig. 1) are presented. The used stereo camera system is mounted behind the windshield of the experimental vehicle. It has 1024×440 px image sensors and records with 25 Hz. With the help of an inertial measurement unit from iMar[1] in combination with a GPS sensor, a global vehicle ego motion was determined. An image Stixel width of 5 px and a grid cell size of 0.04 m^2 is chosen. We assume a disparity uncertainty of $\sigma_d = 0.5\,px$ and a column uncertainty of $\sigma_d = 0.25\,px$. To reduce the number of outliers only Stixels with a distance less than 25 m to the ego vehicle's position are used.

Fig. 5 shows the occupancy grid results of a 20 m×30 m example map which reveals *occupied* (red), *free* (green) and *unknown* (blue) grid cells. The stronger the color saturation the stronger the evidence of *free*, *occupied* or *unknown*. On the right side of the map an entrance gate, two parking vehicles and the building facades are visualized. The area between the vehicles up to the building facade is *unknown* because this area is not visible. On the left side the poles of the parking area are mapped correctly which underlines the high level of detail of the Stixel World.

[1] http://www.imar-navigation.de/index.php/en.html

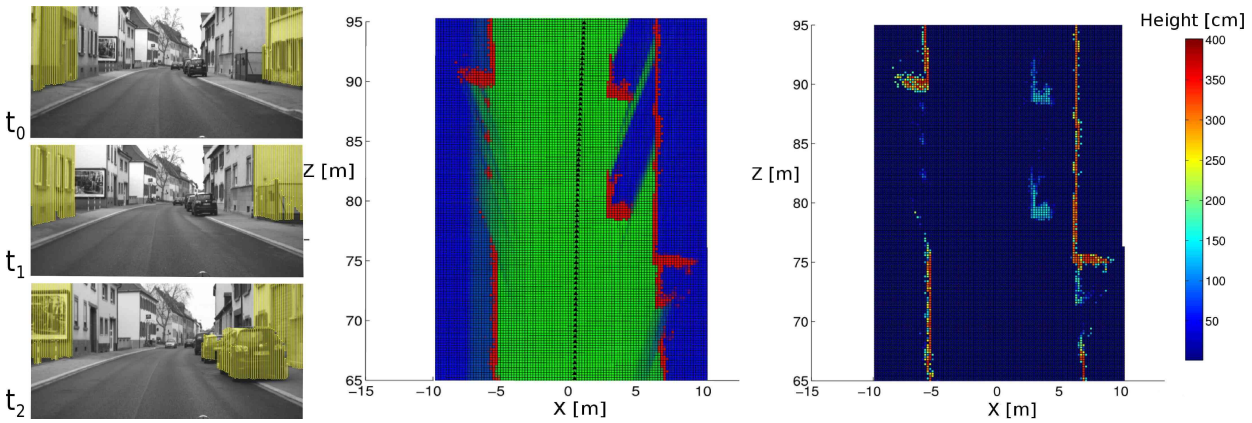


Figure 5: An exemplary mapping result is shown. On the left side the original images with the input data of three time steps are presented. The map in the center shows the results using the described approach. The color reveals *occupied* (red), *free* (green) and *unknown* (blue) areas of the environment. The stronger the color saturation the stronger the evidence of *free*, *occupied* or *unknown*. On the right side of the map an entrance gate, two parking vehicles and the building facades are mapped correctly. Furthermore, poles on the left side are presented, too. The cells "behind" the vehicle have an *unknown* state. On the right side of the figure the height layer is shown. There, the color encodes the height from zero (dark blue) to about 4 m (dark red). The facades are about 4 m high; vehicles, the entrance gate and the poles have heights between 0.5-1.5 m.

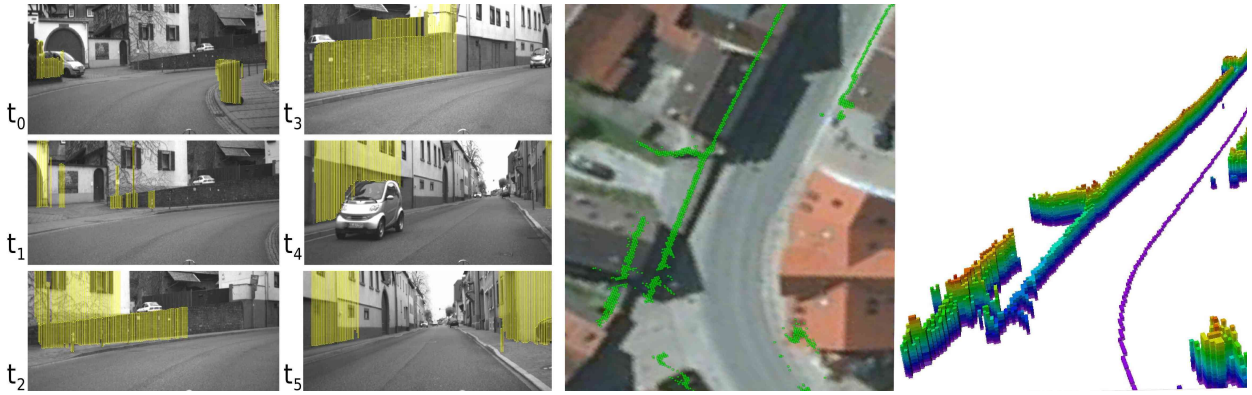


Figure 6: The first detailed example of the 3.000 image sequence is shown. The original images with the input Stixels (left), the 2D mapping results overlaid on Google Earth images (center) and the 3D representation (right) using the *OctoMap* representation are shown. Due to the segmentation of the multi-layered Stixel World into stationary and moving objects, the oncoming vehicle is not taken into account for the mapping approach. Furthermore, walls behind walls are represented. Because of the limited field of view the right building in this scene could not be mapped.

Figure 7: The second detailed example of the 3.000 image sequence: Traffic infrastructures like poles are mapped precisely into the map which is shown on the right side of the driving corridor. Because of the fact that the vehicle at the intersection is still waiting, it is classified as a stationary object. As a result, the vehicle is represented in the map. The mapped corners of buildings, the facades and the parking areas are consistent with the Google Earth images. Note that uncertainties in the ego vehicle's position have not been taken into account.

In addition, Fig. 5 shows the height layer for the map. The building facades are about 4 m high; vehicles, the entrance gate and the poles have heights between 0.5-1.5 m which are conclusive results.

In Fig. 6 and Fig. 7 detailed results from the complete image sequence are shown. For a consistency check of the pipeline of the mapping approach the results are overlaid on Google Earth images. Both figures point out that mapped corners of buildings, facades and parking areas are consistent with the Google Earth images. Note that we have not take into account uncertainties in the ego vehicle's position.

As shown in Fig. 6, the oncoming vehicle is ignored for the mapping approach thanks to the fact that the named Stixel segmentation step (Erbs et al., 2012) allows to ignore moving objects.

Fig. 7 points out that the poles next to the right driving corridor are mapped correctly. Because of the fact that the velocity of the waiting vehicle at the intersection is zero, the vehicle is classified as a static object.

In addition, the comparison with the Google Earth images points out a drawback using a stereo camera system: Due to the limited field of view the complete right building in the scene of Fig. 6 was not recorded and, as a consequence, can not be mapped. To overcome this drawback wide angle lenses should be used in future work.

5 SUMMARY

In this work, a stereo vision based incremental mapping approach for urban regions was presented. For the first time the 3D environment representation multi-layered Stixel World is used as input data. For reliable mapping results the Stixel World was segmented into static and dynamic objects. Thus, only static environment information is used for our technique.

The mapping approach was formulated with evidential theory which allows the explicit representation of *free*, *occupied* and *unknown* regions. A detailed sensor model describes how Stixel measurements influence the state of the cells. To fuse new Stixels into the map incrementally, the Dempster's Rule of Combination was used. In a further step, a recursive height layer for the map was estimated. Detailed mapping results of a 3.000 image sequence were shown. Furthermore, the mapping results were compared with Google Earth images for a consistency check of the pipeline of the mapping approach.

REFERENCES

Elfes, A. E., 1987. Sonar-based real-world mapping and navigation. Journal of Robotics and Automation 3(3), pp. 249–265.

Erbs, F., Schwarz, B. and Franke, U., 2012. Stixmentation - probabilistic stixel based traffic scene labeling. In: British Machine Vision Conference (BMVC), Surrey, England, pp. 71.1–71.12.

Franke, U., Rabe, C., Badino, H. and Gehrig, S., 2005. 6d-vision: Fusion of stereo and motion for robust environment perception. In: German Association for Pattern Recognition (DAGM), Vienna, Austria.

Gallup, D., Pollefeys, M. and Frahm, J.-M., 2010. 3d reconstruction using an n-layer heightmap. In: German Association for Pattern Recognition (DAGM), Darmstadt, Germany, pp. 1–10.

Gehrig, S., Eberli, F. and Meyer, T., 2009. A real-time low-power stereo vision engine using semi-global matching. In: International Conference on Computer Vision Systems (ICVS).

Kammel, S., Ziegler, J., Pitzer, B., Werling, M., Gindele, T., Jagzent, D., Schröder, J., Thuy, M., Goebl, M., Hundelshausen, F. v., Pink, O., Frese, C. and Stiller, C., 2008. Team annieway's autonomous system for the 2007 DARPA Urban Challenge. Journal of Field Robotics 25(9), pp. 615–639.

Larnaout, D., Bourgeois, S., Gay-Bellile, V. and Dhome, M., 2012. Towards bundle adjustment with GIS constraints for online geo-localization of a vehicle in urban center. In: 3DimPVT, Zürich, Switzerland.

Lategahn, H. and Stiller, C., 2012. Experimente zur hochpräzisen landmarkenbasierten Eigenlokalisierung in unsicherheitsbehafteten digitalen Karten. In: UNiDAS e.V, Walting, Germany.

Merali, R. S. and Barfoot, T. D., 2012. Patch map: A benchmark for occupancy grid algorithm evaluation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), Algarve, Portugal.

Moras, J., Cherfaoui, V. and Bonnifait, P., 2011. Moving objects detection by conflict analysis in evidential grids. In: IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, pp. 1122–1127.

Moravec, H. P. and Elfes, A. E., 1985. High resolution maps from wide angle sonar. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 116 – 121.

Muffert, M., Pfeiffer, D. and Franke, U., 2013. A stereo-vision based object tracking approach at roundabouts. In: IEEE Intelligent Transportation Systems Magazine, Volume 5, Number 2, pp. 22–23.

Pagac, D., Nebot, E. and Durrant-Whyte, H., 1998. An evidential approach to map-building for autonomous vehicles. In: IEEE Transactions on Robotics and Automation, Vol. 14number 4, pp. 623–629.

Pfeiffer, D. and Franke, U., 2011. Towards a global optimal multi layer Stixel representation of dense 3D data. In: British Machine Vision Conference BMVC, BMVA Press, Dundee, Scotland.

Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press.

Thrun, S., 2010. What we're driving at. `http://googleblog.blogspot.de/2010/10/what-were-driving-at.html`.

Thrun, S., Burgard, W. and Fox, D., 2005. Probabilistic Robotics. Intelligent Robotics and Autonomous Agents, The MIT Press.

Wurm, K. M., Hornung, A., Bennewitz, M., Stachniss, C. and Burgard, W., 2010. OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems. In: Proceedings of the ICRA 2010 Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation, Anchorage, AK, USA.

Yang, T. and Aitken, V., 2006. Evidential mapping for mobile robots with range sensors. In: -IEEE Transactions on Instrumentation and Measurement, IMTC, Ottawa, Ontario, Canada, pp. 1422–1429.

Zheng, E., Dunn, E., Raguram, R. and Frahm, J.-M., 2012. Efficient and scalable depthmap fusion. In: British Machine Vision Conference (BMVC), Surrey, England, pp. 34.1 –34.12.