Proceedings Paper:
Saleem, NH, Chien, H-J, Rezaei, M orcid.org/0000-0003-3892-421X et al. (1 more author) (2017) Improved Stixel Estimation Based on Transitivity Analysis in Disparity Space. In: Computer Analysis of Images and Patterns (Lecture Notes in Computer Science series). International Conference on Computer Analysis of Images and Patterns, 22-24 Aug 2017, Ystad, Sweden. Springer , pp. 28-40. ISBN 9783319646886

https://doi.org/10.1007/978-3-319-64689-3_3

Improved Stixel Estimation based on Transitivity Analysis in Disparity Space

Noor Haitham Saleem[1], Hsiang-Jen Chien[1], Mahdi Rezaei[2], and Reinhard Klette[1]

[1] School of Engineering, Computer, and Mathematical Sciences, EEE Dep.
Auckland University of Technology, New Zealand
[2] Faculty of Computer and Information Technology Engineering, CE Dep.
Qazvin Islamic Azad University, Iran

Abstract. We present a novel method for stixel construction using a calibrated collinear trinocular vision system. Our method takes three conjugate stereo images at the same time to measure the consistency of disparity values by means of the transitivity error in disparity space. Unlike previous stixel estimation methods that are built based on a single disparity map, our proposed method introduces a multi-map fusion technique to obtain more robust stixel calculations. We apply a polynomial curve fitting approach to detect an accurate road manifold, using the v-disparity space which is built based on a confidence map. Comparing the depth information from the extracted stixels (using stixel maps) with depth measurements obtained from a highly accurate LiDAR range sensor, we evaluate the accuracy of the proposed method. Experimental results indicate a significant improvement of 13.6% in the accuracy of stixel detection compared to conventional binocular vision.

1 Introduction

Vision-based driver assistance systems (VB-DAS) contribute to the current transition process towards autonomous vehicles. They are already widely used in current modern cars [1]. Cameras are one type of sensors that are commonly installed in modern cars. In particular, stereo vision contributes to systems that aim at distance measurements, surface modelling, or object detection [2]. This is important, for example, for scene analysis [3], feature descriptors [4], optimising learning time [5], or for reducing processing efforts in general [6].

In 2009 a novel "super-pixel representation" has been proposed for urban road scenes. The method is known as *stixel* (from "stick elements"). It groups vertically space cubes which belong to an on-road object [7]. The representation yields a highly efficient modelling of scene objects in urban traffic environments [9]. Recently, joint stixel representations, combining semantic data and depth, are proposed to integrate both categories in terms of a joint optimized scene model [10].

To construct a "stixel world" (see Fig. 1), multiple independent techniques may have to be cascaded.[3] These may include mapping disparities into occu-

[3] We adopt a *semi-global matching* (SGM) algorithm [12] for disparity calculation.

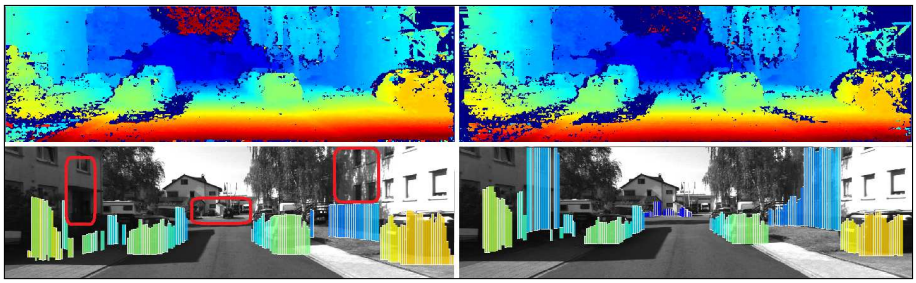

Fig. 1. A stixel world for a scene in KITTI's *residential* dataset [11]. *Top-left*: Disparity map using an SGM-variant visualized by applying a color key. *Top-right*: Improved disparity map. *Bottom-left*: Stixels on a ground plane using binocular vision (red rectangles indicate missing stixels). *Bottom-right*: Proposed stixel estimation.

pancy grids, ground manifold estimation, object height detection, and finally stixel extraction.

The *free space* is a region in the ground manifold "without any obstacle" [13], i.e. regions ahead of the ego-vehicle where this vehicle may potentially drive in, for example, in the next few seconds. Free-space and stixel calculations are closely related to each other; the existence of a stixel excludes free space at this place; stixels are "sitting" on the ground manifold, and the free space is a subset of the ground manifold. The detection of free-space is important for intelligent transportation control [14]. It is also crucial for collision avoidance for the *ego-vehicle* (i.e. the vehicle in which the system is operating in) and assisting a blind pedestrian [32].

Having VB-DAS as a core component over other active sensors, many *advanced driver assistance systems* (ADAS) demonstrate prominent developments in this area (e.g. [15]). An ADAS provides a better understanding of the environment in order to improve traffic safety and efficiency [16].

Accuracy of stixels requires a disparity signal of "good" quality; this quality often decreases in cases of occlusions or textureless image patches [17]. Unfortunately, these issues are common in traffic scenes, thus more efforts are needed to improve disparity signals, also aiming at more reliable free-space estimation and stixel calculations.

A binocular vision system depends on calculated disparity values which are calculated by implementing stereo matching algorithms [12, 18] on images obtained by a left and right camera.

Since noisy 3D points have a considerable impact on free-space detection, it is very important to identify unreliable disparity values before they are transformed into 3D space and used for stixel estimation. Therefore, we consider the use of confidence maps (see [19] for different options for such maps) with the aim of improving stixel segmentation.

The remainder of this paper is structured as follows. Section 2 addresses work closely related to stixel estimation. In Section 3, the proposed approach is

described in detail. In Section 4, experimental results are given and discussed. Section 5 concludes.

2 Related Work

We briefly discuss work on stixel extraction. Stixels are a compact representation towards semantic segmentation of traffic scenes; space elements above neighbouring pixels at the same depth are vertically grouped [20], according to an estimated object height at those pixels. Apparently, stixels are like rectangular thin columns on the ground manifold defined on a regular grid. A stixel starts at the top at a detected object surface and ends at the bottom on the level of the ground manifold. Free space (for the ego-vehicle) is a subset of the ground manifold not covered by stixels.

Rapid stixels describe techniques which enhance stixel extraction by reducing computational costs. In [23], a direct stixel computation is proposed by changing the parametrization from disparity space into pixel-wise cost volumes for speed improvement. In [21], the authors use deep convolution neural networks for free-space detection using monocular vision, while obstacle detection and stixel estimation are done using stereo vision. Fast stixel computation without depth maps is proposed in [22]; it allows high-speed pedestrian detection up to 200 fps.

Color fusion models compute stixels using stereo images, and also involve a combination of color appearance and depth cues for free-space and obstacle detection. Such methods have been presented in the stixel segmentation literature [17, 24, 25]. Their implementation can be done by using low-level fusion of depth or semantic information in the stixel generation process. Scharwächter et al. employed pixel classification by random decision forests [24], while in [25] semantic information via object detectors is used for a suitable set of classes. Yet another method to improve stixels is by using low-level appearance models in an on-line self-supervised framework; see [17].

Stereo confidence-based methods, on the other hand, use confidence estimation within the stereo-matching process to replace spurious disparity matches by interpolating surrounding disparity values at these locations; see [26–28] for examples. In [26], the authors incorporate three confidence measures, namely the naïve peak-ratio (PKRN), the maximum-likelihood measure (MLM), and local curve (LC) information into stixel representations. The stereo confidence measures use stereo confidence cues based on an extended Bayesian approach. In [28], an ensemble learning classifier is adopted to increase accuracy in stereo-error detection. In [27], histogram-sensor models are explored to model on a real-world application using a global formulation of 3D reconstruction through an occupancy grid.

Rapid-stixel methods may have some drawbacks; they may suffer from low-depth accuracy which affects stixel extraction negatively. In order to perform stixel segmentation, an adopted colour fusion model might not be suitable due

to shortages highlighted in [10]. With promising results achieved by adopting confidence information, this paper proposes the following:

1. altogether a low-cost architecture for reducing false-positives in stixel estimation,
2. in particular the use of a confidence measure derived from trinocular stereo matching, and
3. a method for performance evaluation of stixel estimation assuming the availability of LiDAR data.

3 Stixels in Trinocular Stereo Vision

We consider a trinocular calibrated video recording system which allows us to perform stereo matching on one of the three possible camera pairs. Thus we may have up to three different left-right disparity maps; they may be fused and warped to a selected reference camera (one of the three). Based on the fused (and thus enhanced) disparity map, the *ground manifold* (i.e. a generalisation from a plane) is estimated using a *v-disparity* technique. This is followed by detections of base- and top-points of stixels applying means of membership voting and a cost image. In a final step, base- and top-points are used for extracting the stixels.

3.1 Transitivity Error in Disparity Space

Given a collinear m-camera configuration, we have $m(m-1)/2$ left-right disparity maps. It has been shown that the accumulative transitivity error among these maps can be effectively used as a confidence indicator on a stereo matcher [29]. Let $(u, v) \in \mathbb{R}^2$ denote a pixel location in left-image coordinates. A disparity map $\delta : \mathbb{R}^2 \to \mathbb{R}_0^+$ finds its corresponding pixel in right-image coordinates $(u - \delta(u, v), v)$. A disparity map can therefore be used to define the warping of a function $\mathcal{M} : \mathbb{R}^2 \to \mathbb{R}$ as follows:

$$\phi(\mathcal{M}, \delta)(u, v) = \mathcal{M}(u - \delta(u, v), v) \tag{1}$$

The warping function ϕ is used to construct the concatenation of two disparity maps

$$\tau(\delta_{01}, \delta_{12})(u, v) = \delta_{01}(u, v) + \phi(\delta_{12}, \delta_{01})(u, v) \tag{2}$$

where δ_{01} and δ_{12} are the disparity maps with respect to camera pairs $(0, 1)$ and $(1, 2)$, respectively, in a trinocular configuration.

Let $\bar{\delta}_{02} = \tau(\delta_{01}, \delta_{12})$ be a *combined* disparity map, and δ_{02} the explicitly computed one for camera pair $(0, 2)$. We define our new, say, *trinocular confidence measure* by

$$\Gamma(u, v) = \frac{1}{\|\delta_{02}(u, v) - \bar{\delta}_{02}(u, v)\| + 1} \tag{3}$$

where the absolute difference $\|\delta_{02}(u, v) - \bar{\delta}_{02}(u, v)\|$ is the *transitivity error in disparity space* (TED). See Fig. 2 for an example of this confidence indicator; $\bar{\delta}_{02}$ is also called the *TED-based disparity*.

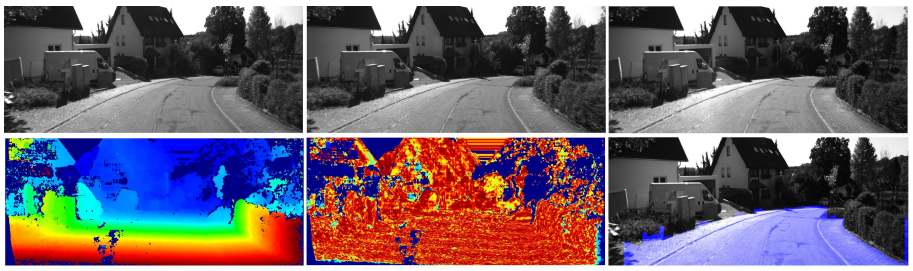

Fig. 2. Trinocular confidence and free space. *Top row*: Trinocular stereo pair from the KITTI *road* dataset. *Bottom left*: TED-based disparity. *Bottom middle*: Red and blue pixels indicate high and low confidence values, respectively. *Bottom right*: Calculated free-space (using v-disparity, confidence map, and polynomial curve fitting).

3.2 Detection of Base-Points of Stixels

Our stixel calculation works on TED-based disparities. We propose a new polynomial curve-fitting technique to identify the lower envelop in the common v-disparity space. This identification supports the base-point calculation of stixels. We consider base-points $b_1, b_2, ..., b_{N_{\text{col}}}$ of obstacles in row v.

The v-disparity map is computed by accumulating pixels with the same disparity value in one row v, $1 \leq v \leq N_{\text{row}}$, of the disparity map:

$$V(v,d) = \text{card}\{u : 1 \leq u \leq N_{\text{col}} \wedge \text{int}(\delta(u,v)) = d\} \tag{4}$$

where $0 \leq d \leq d_{\max}$ defines the quantized disparity range for δ in the $N_{\text{row}} \times N_{\text{col}}$ disparity map; int is the nearest integer.

In Eq. (4), each element in the disparity map is considered equally. In this work we propose to use a weighted sum of our trinocular confidence values:

$$V(v,d) = \sum_{1 \leq u \leq N_{\text{col}} \ \wedge \ \text{int}(\delta(u,v))=d} \Gamma(u,v) \tag{5}$$

Here, elements with higher TED-based confidence become more influential.

The next step is to extract a ground manifold from the generated v-disparity map. The ground manifold is identified with an approximated lower envelope in the v-disparity space.

Assuming a ground plane, a Hough transform is used in [30, 31] to detect a lower envelop function in form of a straight line in the v-disparity space. In order to construct this envelop function, the method starts at first with a lower and upper envelop. The envelop estimation is based on calculating the intensity sum of all pixels along a considered curve in the v-disparity image, and then selecting the envelop for which this sum becomes a minimum.

Considering that a road surface is not a perfect plane, and possibly also more irregular in slope changes than a continuous curve, we consider polynomial curve fitting for extracting best fits to the v-disparities, defining a polynomial as being our envelop function; see Fig. 3 for an example.

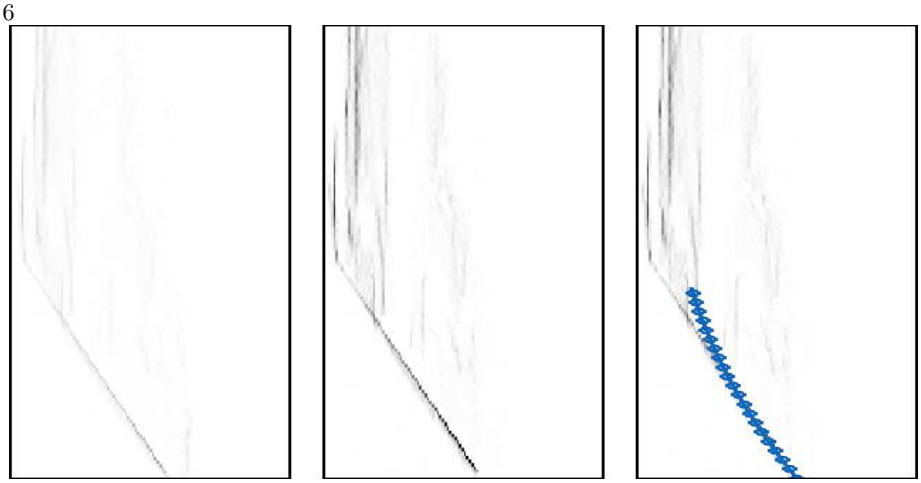

Fig. 3. Demonstration of v-disparity. *Left*: Common cardinality-based v-disparity map. *Middle*: Novel TED-based v-disparity map. *Right*: Detected curve using polynomial fitting.

In general, there is always room for improvements in curve fitting. (A ground manifold also remains to be approximated to some degree only when assuming identical height across one image row v.) We apply a polynomial curve fitting technique to find the coefficients of a polynomial $P(x)$ of degree n that best fits the lower envelop in the v-disparity image:

$$y = P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \tag{6}$$

where $a_0, a_1, \ldots, a_n$ are the coefficients, and the degree n is selected according to accuracy requirements for the algorithm. In order to generate the coefficients of the polynomial according to the degree specified, we need to compute a least-square polynomial for a given set of data. Following the least-square principle, we obtain the parameters $a_0, a_1, \ldots, a_n$, which minimize the total square error:

$$E(a_0, a_1, \ldots, a_n) = \sum_{i=1}^{m} \left[y_i - (a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0) \right]^2 \tag{7}$$

where $m \geq n$ is the number of samples. The optimal coefficients can be solved linearly.

The computed curve then defines a value $d_R(v)$ for row v, and function d_R altogether estimates for "on-road disparities". The profile is used to find the base points b_u of obstacles in column u, following [31].

3.3 Detection of Top-Points of Stixels

The height of obstacles (which "stand" on the ground manifold) is obtained by seeking an ideal segmentation between *foreground* and *background* disparities. The height-of-obstacle calculation begins with selecting membership votes. Next we estimate a cost image to approximate $t_1, t_2, ..., t_{N_{\text{col}}}$, the upper boundary of

obstacles based on the method proposed by [20]. Briefly, the membership values rely on the selection of every disparity of each column from the disparity for its member to the foreground obstacle.

A membership value can be positive if it does not exceed the maximum distance of the expected obstacle disparity; otherwise it will be negative. This Boolean representation brings the challenge to identify a threshold value for the distance; if this value is too large then all disparities will be chosen from the foreground membership, and vice-versa. Therefore, the application of Boolean membership in continuous variation is a better alternative with an exponential function of the form

$$M(u,v) = 2^{1-\varepsilon(u,v)} - 1 \tag{8}$$

where

$$\varepsilon(u,v) = \left[\frac{\hat{d}_u - \delta(u,v)}{\hat{d}_u - Z^{-1}\left(Z(\hat{d}_u) + \triangle Z\right)} \right]^2 \tag{9}$$

where $\hat{d}_u = \delta(u, b_u)$ is the disparity of an obstacle's base point in column u, and Z is the disparity-to-depth conversion function; $\triangle Z$ as a defined soft constraint range in depth.

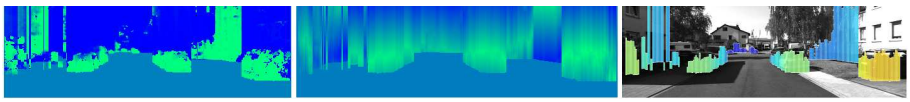

Fig. 4. Stixel world on KITTI data. *Left:* Membership votes. *Middle*: Cost image (data term). *Right*: Extracted stixels.

A visualization of membership votes is illustrated in Fig. 4. Green represents true positives (belonging to an object), pale-blue shows free-space, and blue shows true negatives (background).

From the membership values, the cost image is computed as follows:

$$C(u,v) = \sum_{j=1}^{v-1} M(u,j) - \sum_{j=v}^{N_{col}} M(u,j) \tag{10}$$

A result of the membership cost image, used for the height segmentation, is shown in Fig. 4, middle. The figure shows the height cost of foreground and background disparities. As can be seen, there are bright values which show a high likelihood for performing a foreground-background separation.

The obstacles' top-points $t_1, t_2, ...t_{N_{\text{col}}}$ are obtained from the computed cost image C following the approach proposed in [20].

3.4 Stixel Extraction

By combining base-points $b_1, b_2, ...b_{N_{\text{col}}}$ found in Section 3.2 and top-points $t_1, t_2, ...t_{N_{\text{col}}}$ found in Section 3.3, we extract the stixels.

In this paper we adopt a column grouping technique proposed in [7, 8]. Given $w \in \mathbb{N}^+$ as a predefined width of stixels, every w neighbouring columns are grouped across the whole image, resulting in $\lfloor \frac{N_{\text{col}}}{w} \rfloor$ non-overlapping stixels in one row. For the i-th stixel we have a set of w base-points $B_i = \{b_{u_i}, b_{u_i+1}, ..., b_{u_i+w-1}\}$ and a set of w top-points $T_i = \{t_{u_i}, t_{u_i+1}, ..., t_{u_i+w-1}\}$ where $u_i = (i-1)w + 1$.

The rectangle spanned from column $u = u_i$ to $u = u_i + w - 1$ and row $v = \min(T_i)$ to $v = \max(B_i)$ defines the *scope of a stixel*. Instead of using only base-points' disparities, we integrate all the disparities within the scope to yield a more robust estimate of the stixel's depth z_i, by means of a histogram-based regression technique proposed in [7] with w set to 5 pixels.

Figure 5 shows resulting stixels. The colours of the stixels encode the distance to the ego-vehicle. Red-scale colours represent objects farther away. Stixels of "minor height" have been ignored. The figure illustrates that a stixel represents the height of the first "substantial" obstacle facing the ego-vehicle along a viewing direction.

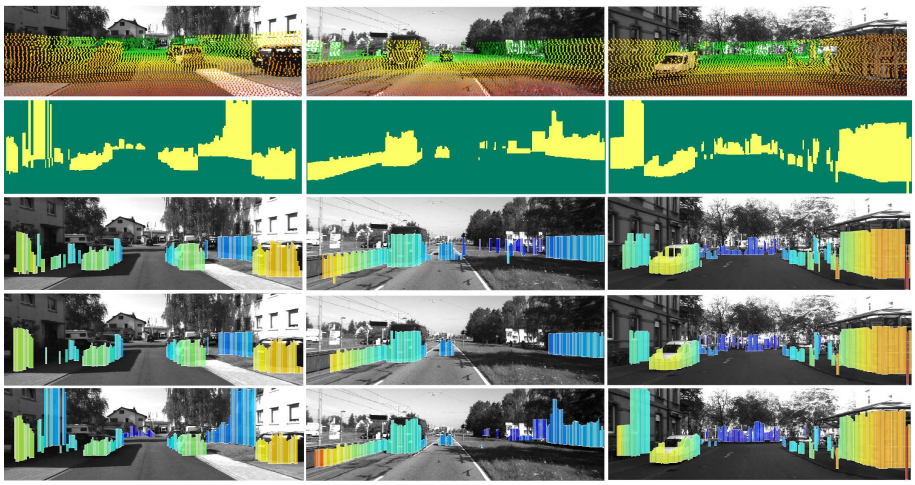

Fig. 5. Qualitative results using KITTI *residential* (first column), *road* (second column), and *city* (third column) data. *First row*: LiDAR projection. *Second row*: Stixel map. *Third row*: Stixels estimated using binocular stereo and ground plane. *Fourth row*: Stixels estimated using trinocular stereo and ground plane. *Fifth row*: Stixels estimated using TED-based disparities and polynomial ground manifold approximation.

4 Experimental Results

This section evaluates the performance of three methods for stixel detection: binocular-based occupancy grid (i.e. assuming a plane as ground manifold) stixels, trinocular-based occupancy grid stixels, and stixels using TED-based disparities on a polynomial manifold.

We selected 655 stereo images from KITTI's `city`, `residential`, and `road` datasets which include cars, pedestrians, trees, and traffic signals. Previous literature states challenges in evaluating stixels using KITTI data. The challenges are given by a lack of annotated road images, or a lack of stixel ground truth. It is also of limited relevance to evaluate the quality of the 3D reconstruction subjectively based on manually observed disparity images.

Since 3D laser scanners are accurate as reference sensors, we employ the Velodyne LiDAR data obtained by a 3D laser scanner which are publicly available [11].

We evaluate all stixels in every frame individually for understanding the efficiency of the proposed trinocular stixels in terms of distance errors. This comprises several processes:

1. Generate a stixel map which forms stixels above the ground manifold, as shown in Fig. 5 (third to fifth row) for "dominating" stixels.
2. Project LiDAR points (X_j, Y_j, Z_j) into image coordinates (u_j, v_j). Such an exemplary LiDAR point projection is also illustrated in Fig. 5 (second row). The projections are used to build a LiDAR-stixel correspondence function β_{ij}, where $\beta_{ij} = 1$ if LiDAR point j hits stixel i, otherwise $\beta_{ij} = 0$.
3. The degree of correspondence of these images verifies the accuracy of the estimated stixels. Hence, the comparison of LiDAR depths with corresponding stixel depths form the error measurement using the root-mean-square error computed by

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_{\mathrm{stx}}} \sum_{j=1}^{N_{\mathrm{pts}}} \beta_{ij} (z_i - Z_j)^2}{N_{\mathrm{hit}}}} \tag{11}$$

where N_{stx} and N_{pts} represent the number of stixels and LiDAR points, respectively, and N_{hit} is the number of non-zero elements in β.

Figures 5 and 6 demonstrate qualitative and quantitative results, respectively. `city` and `residential` data differ from `road` data by also showing pedestrians, cyclists, and sometimes cyclists having a baby stroller. As shown in Fig. 6, the error rate is constant from frames 1 to 50 while it drops for frames 50 to 80 due to unexpected interference by a cyclist. A similar pattern is noted after frame 80 till frame 150. After that, the error rate fluctuates roughly at the first road junction (frame 155) shows a different pattern of traffic while it increases again at the end of the data sequence due to a round-about.

This occurs for all the three methods. Detection using binocular stereo on a plane shows the highest error rate and the highest false-positive rate too, due to degrading disparities. Detection using trinocular stereo on a plane performs better than the binocular stereo method. Our proposed method (TED-based trinocular and polynomial ground manifold) covers more valid disparities compared to others, and appears to be insensitive to slope changes. It outperforms others regarding a smaller rate of false alarms.

On the `residential` data, the error rate for all three methods was less compared to the `city` data. The used data show a car parked on the side of the road, houses, and road junctions. The accuracy of the stixels detected via the

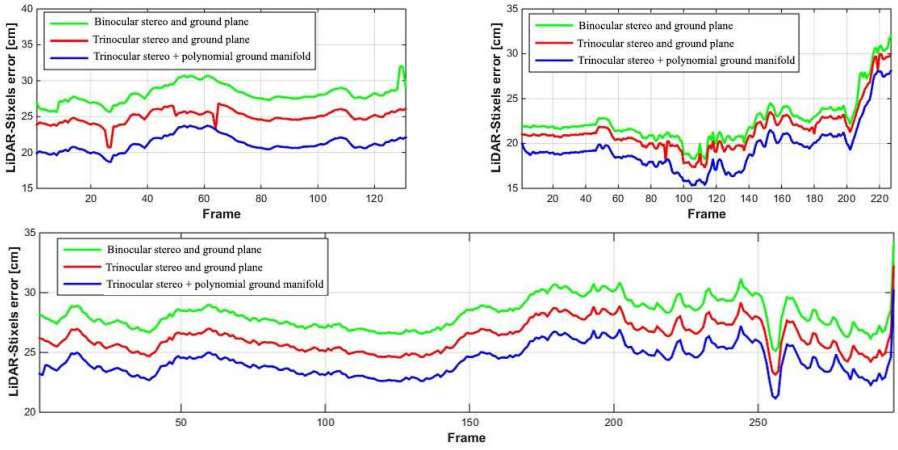

Fig. 6. The error rates represent the difference of distances between LiDAR data and stixels, shown for the three methods. *Top-left*: Error rate on `residential` data. *Top-right*: Error rate on `city` data. *Bottom*: Error rate on `road` data.

trinocular+plane method in frame 60 is very close to the one using trinocular-TED+polynomial. For the two standard methods (binocular or trinocular on plane), there are a few stixels that were not detected at the visible end of the road; see Fig. 5. They were successfully detected via the proposed trinocular-TED+polynomial. The `road` data show the effectiveness of the proposed method. It can perform better than the other two methods when dealing with open road situations. In summary, the experimental analysis illustrates an improved robustness of the proposed trinocular-TED+polynomial method across various data sets. Low texture, or changes in the slope of road surfaces are in particular cases where our method is more robust in detecting stixels.

5 Conclusions

This paper proposed a novel method for stixel construction. The stixels, built using TED-based disparities provided by trinocular vision, have been found to provide better accuracy over conventional binocular ones, especially when also using polynomial ground-manifold approximation. Our method uses a confidence map, which can vote for consistent disparity values within a trinocular stereo analysis process. Our method also includes a polynomial curve-fitting method for road geometry which is insensitive to slope changes. The main advantage of our work is to produce a low-cost architecture for reducing false-positives in stixel estimation.

In order to test our method, we used more than 600 frames including `road`, `city`, and `residential` data from KITTI. The verification has been done using LiDAR range data to verify the accuracy of the proposed method. We compared

the proposed method with two ground-plane-based standard methods (i.e. using binocular or simply unified trinocular disparities).

References

1. Klette, R.: Vision-based driver assistance. In Wiley Encyclopaedia Electrical Electronics Engineering. John Wiley & Sons, 1–15 (2015)
2. Rezaei, M., and Klette, R.: Computer Vision for Driver Assistance: Simultaneous Traffic and Driver Monitoring. Springer, (2017)
3. Kaaniche, K., Demonceaux, C., and Vasseur, P.: Analysis of low-altitude aerial sequences for road traffic diagnosis using graph partitioning and Markov hierarchical models. In Proc. Int. Multi-Conf. Systems Signals Devices, 656–661 (2016)
4. Wu, J., Cui, Z., Sheng, V. S., Zhao, P., Su, D., and Gong, S.: A comparative study of SIFT and its variants. Measurement Science Review, 13 (3): 122–131 (2013)
5. Farabet, C., Couprie, C., Najman, L., and Lecun, Y.: Learning hierarchical features for scene labeling. *IEEE Trans. Pattern Analysis Machine Intelligence*, 35 (8): 1915–1929 (2013)
6. Anders, J. , Mefenza, M., Bobda, C., Yonga, F., Aklah, Z., and Gunn, K.: A hardware/software prototyping system for driving assistance investigations. J. Real-Time Image Processing 11, 3: 559–569. (2016)
7. Badino, H., Franke, U., and Pfeiffer, D.: The stixel world - A compact medium level representation of the 3D-world. LNCS, 5748: 51–60 (2009)
8. Pfeiffer, D. and Franke, U.: Efficient representation of traffic scenes by means of dynamic stixels. In Proc. Intelligent Vehicles Symp., 217–224 (2010)
9. Scharwächter, T., Enzweiler, M., Franke, U., and Roth S.: Stixmantics: A medium-level model for real-time semantic scene understanding. LNCS, 8693 (5): 533–548 (2014)
10. Schneider, L., Cordts, M., Rehfeld, T., Pfeiffer, D., Enzweiler, M., Franke, U., Pollefeys, M., and Roth, S.: Semantic stixels: Depth is not enough. In Proc. Intelligent Vehicles Symp., 110–117 (2016)
11. Geiger, A., Lenz, P., and Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In Proc. Computer Vision Pattern Recognition, 3354–3361 (2012)
12. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Analysis Machine Intelligence, 30: 328–341 (2008)
13. Shin, B.-S., Xu, Z., and Klette, R.: Visual lane analysis and higher-order tasks: A concise review. Machine Vision Applications, 25(6): 1519–1547 (2014)
14. Onoguchi, K., N. Takeda, and M. Watanabe: Obstacle location estimation using planar projection stereopsis method. Systems Computers Japan, 32(14): 67–76 (2001)
15. Saleem, N. H. and Klette, R.: Accuracy of free-space detection for stereo versus monocular vision. In Proc. Image Vision Computing New Zealand, 48–53 (2016)
16. Seo, J., Oh, C., and Sohn, K.: Segment-based free space estimation using plane normal vector in disparity space. In Proc. Connected Vehicles Expo, 144–149 (2015)
17. Sanberg, W. P., Dubbelman, G., and deWith, P. H. N.: Color-based free-space segmentation using online disparity-supervised learning. In Proc. Intelligent Transportation Systems, 906–912 (2015)
18. Spangenberg, R., Langner, T., Adfeldt, S., and Rojas, R.: Large scale semi-global matching on the CPU. In Proc. IEEE Intelligent Vehicles Symp., 195–201 (2014)

19. Klette, R.: Concise Computer Vision. Springer, London (2014)
20. Pfeiffer, D.: The Stixel World. Doctoral Thesis, Humboldt Universität Berlin (2011)
21. Levi, D., Garnett, N., and Fetaya, E.: StixelNet: A deep convolutional network for obstacle detection and road segmentation. In Proc. British Machine Vision Conf., 1:12 (2015)
22. Benenson, R., Timofte, R., and Van Gool, L.: Stixels estimation without depth map computation. In Proc. Int. Conf. Computer Vision Workshops, 2010–2017 (2011)
23. Benenson, R., Mathias, M., Timofte, R., and Van Gool, L.: Fast stixels estimation for fast pedestrian detection. In Proc. Europ. Conf. Computer Vision, 11–20 (2012)
24. Scharwächter, T., and Franke, U.: Low-level fusion of color, texture and depth for robust road scene understanding. In Proc. Intelligent Vehicles Symp., 599–604 (2015)
25. Cordts, M., Schneider, L., Enzweiler, M., Franke, U., and Roth, S.: Object-level priors for stixel generation. In Proc. German Conf. on Pattern Recognition, 172–183 (2014)
26. Pfeiffer, D., Gehrig, S., and Schneider, N.: Exploiting the power of stereo confidences. In Proc. Conf. Computer Vision Pattern Recognition, 297–304 (2013)
27. Brandao, M., Ferreira, R., Hashimoto, K., Takanishi, A., and Santos-Victor, J.: On stereo confidence measures for global methods, evaluation, new model and integration into occupancy grids. In Proc. Pattern Analysis Machine Intelligence, 116–128 (2016)
28. Haeusler. R., Nair, R., and Köndermann, D.: Ensemble learning for confidence measures in stereo vision. In Proc. Conf. Computer Vision Pattern Recognition, 305–312 (2013)
29. Chien, H.-J., Geng, H., and Klette, R.: Improved visual odometry based on transitivity error in disparity space: A third-eye approach. In Proc. Image Vision Computing New Zealand, 72–77 (2014)
30. Iloie, A., Giosan, I., and Nedevschi, S.: UV disparity based obstacle detection and pedestrian classification in urban traffic scenarios. In Proc. Intelligent Computer Communication Processing, 119–125 (2014)
31. Labayrade, R., Aubert, D., and Tarel, J.: Real time obstacle detection in stereovision on non flat road geometry through v-disparity representation. In Proc. IEEE Intelligent Vehicle Symp., 646–651 (2002)
32. Keller, C., Dang, T., Fritz, H., Joos, A., Rabe, C., and Gavrila, D.: Active pedestrian safety by automatic braking and evasive steering. In Proc. Intelligent Transportation Systems, 1292–1304 (2011)