Stereo Fusion from Multiple Viewpoints

Christian Unger[1,2], Eric Wahl[2], Peter Sturm[3], and Slobodan Ilic[1]

[1] Technische Universität München, Germany
Slobodan.Ilic@in.tum.de
[2] BMW Group, München, Germany
{Christian.Unger,Eric.Wahl}@bmw.de
[3] INRIA Rhône-Alpes and Laboratoire Jean Kuntzmann, France
Peter.Sturm@inrialpes.fr

Abstract. Advanced driver assistance using cameras is a first important step towards autonomous driving tasks. However, the computational power in automobiles is highly limited and hardware platforms with enormous processing resources such as GPUs are not available in serial production vehicles. In our paper we address the need for a highly efficient fusion method that is well suited for standard CPUs.

We assume that a number of pairwise disparity maps are available, which we project to a reference view pair and fuse them efficiently to improve the accuracy of the reference disparity map. We estimate a probability density function of disparities in the reference image using projection uncertainties. In the end the most probable disparity map is selected from the probability distribution.

We carried out extensive quantitative evaluations on challenging stereo data sets and real world images. These results clearly show that our method is able to recover very accurate disparity maps in real-time.

1 Introduction

Dense real-time multi-view stereo allows for a wide spectrum of useful applications including automotive driver assistance or robotics. Although a large amount of research has been devoted to the stereo problem using image pairs [1,3,5,7,6,8,14,17] and using multiple cameras [2,10,12,13,15,16], obtaining dense high-quality disparity maps in real-time is still a challenging problem. Traditional real-time stereo methods [6,17] still lack accuracy compared to methods which do not impose time constraints. A few multi-view stereo methods [11,18] may achieve real-time performance, but only by using the enormous processing power of graphics cards. However, such hardware is usually not available on mobile platforms and therefore it is absolutely necessary that all calculations can be performed in real-time on a standard mobile CPU at video frame rate.

Another important problem based on traditional pairwise stereo methods is *motion-stereo*. In Fig. 1 we show an example of our automotive driver assistance application where the camera is mounted laterally on a vehicle. The disparity maps can be computed from consecutive image frames over time while the vehicle is moving. From these disparity maps, we build a model of the environment, in

A. Pinz et al. (Eds.): DAGM/OAGM 2012, LNCS 7476, pp. 468–477, 2012.
© Springer-Verlag Berlin Heidelberg 2012

(a) (b) (c) (d)

Fig. 1. Real-time motion-stereo for automotive driver assistance. When the vehicle moves, depth is inferred via motion-stereo. (a) A camera on the side of the vehicle observes the lateral space. (b) One frame captured by the side camera. (c) Disparity map obtained by pairwise real-time stereo matching. (d) Result of our proposal.

order to mitigate collisions or to find lateral parking spaces. Even at higher velocities, the disparity maps obtained over time exhibit a large overlap and thus depth information is highly redundant. At the same time, due to the real-time stereo method used, disparities are very error prone. The question is how to fuse all those disparity maps to improve the accuracy of the disparity map defined by a reference image pair, for example, the last two images in case of motion-stereo.

In our paper we assume that a set of disparity maps is available and that they were computed using any available short baseline stereo technique. Then, given any other reference view pair, we propose a novel *stereo fusion* method to produce an accurate disparity map of the given reference view pair by fusing all available disparity maps. In our approach, we first project all disparity maps to the reference view pair. After maintaining visibility constraints, we estimate a probability density function over all valid disparities in the reference view using uncertainties of these reprojections. Finally, this allows us to select the most probable disparity map from this distribution.

We tested our method on the challenging datasets of Middlebury [14] and compared it to the fusion methods of [11] and [19]. The experiments show that our technique is very robust and that the quality is significantly improved, especially at object boundaries. We also show results on real-world sequences acquired from a camera attached to a vehicle. A very important fact is that our method allows real-time operation on CPU without dedicated hardware.

In the remainder of the paper we will first review related work, then present our method and finally show an exhaustive experimental evaluation.

1.1 Related Work

In recent years, traditional stereo and multi-view stereo methods have been extensively studied and tested using the available Middlebury datasets [15]. While resulting in a large amount of excellent results, little attention has been spent on computational performance. However, when that was the case and real-time stereo methods were proposed [6,17], the reconstruction quality was significantly decreasing. While multi-view stereo approaches introduce assumptions on shape

priors and use robust photo-consistency measures, there are others which aim to produce consistent disparity maps [4,10,11,16,19,20]. In many cases disparity maps are produced locally using a number of overlapping views and are later fused into either a global disparity video [19] or a full 3D model [11,18]. Again, the vast majority of works aim at high quality reconstructions of single objects and only very few try to minimize the computational overhead.

Since the main motivation of our work comes from motion-stereo we tend to fuse locally overlapping disparity maps and do not aim to produce full 3D models. Works of Merrell et al. [11] and Zhang et al. [19], which explicitly deal with fusion of the disparity maps are thus directly related to our approach.

Merrell et al. [11] compute depth maps between neighboring views and fuse this information based on the *stability* of every depth. In order to keep track of occlusions, the stability is determined for every depth hypothesis and is defined by counting occlusions in the reference and other views. A valid depth is defined as the first depth hypothesis which is stable. However outliers affect the stability and such hard decisions may produce incorrect depth estimates. Further, the computational complexity grows quadratically with the number of disparity maps and in practice real-time operation is only possible with GPU hardware. In our paper, we overcome these problems. Our probabilistic approach employs reprojection uncertainties, handles outliers robustly and depth-accuracy gets improved compared to this approach.

Zhang et al. [19] impressively generalized the fusion problem by formulating it as an energy minimization problem. In their *bundle optimization* framework all disparity maps are optimized iteratively using belief propagation. In contrast to Merrell et al. [11] they do not model occlusions or visibility constraints explicitly. In their work these constraints are handled by the simultaneous use of *geometric coherence* and *color-similarity* as well as the regularization of belief propagation. The minimization of the energy functional is in practice very time consuming and thus, this method is not an option for mobile real-time applications.

Koch et al. [9] introduced the efficient *correspondence linking algorithm*: by *chaining* correspondences across many views outliers are rejected and accuracy is improved. However, no solution was provided for multiple disparity maps per view and disparities in occluded regions or outliers near the beginning of the chain are problematic. Zach [18] fuses multiple depth maps to obtain a full volumetric 3D reconstruction. It was formulated as a relatively efficient method using the GPU and produces very good results. However, the hardware requirements are too high and the volumetric representation is problematic for our application.

Compared to other fusion methods, our work focuses on both real-time performance and high quality depth maps. In our exhaustive experimentation we obtained better depth maps, especially in occluded and discontinuity areas.

2 Method

The major problem in motion and multi-view stereo are occlusions and discontinuities. Here we consider a reference view pair (RVP) in which we want

to improve disparities, especially in occluded and discontinuity areas by bringing the information from other view pairs to the RVP. For this, we propose to compute a probability density function (pdf) estimating the probabilities of the disparities in the RVP. It is done by the reprojection of all available disparity maps to this RVP. This allows us to construct a global pdf which is sampled from a relatively large number of measurements coming from all disparity maps reprojected to the RVP.

2.1 Reprojection

Our goal is to compute an improved disparity map $\hat{\mathcal{D}} = \hat{\mathcal{D}}_{R_1,R_2}$ for a specific RVP $(\mathcal{I}_{R_1}, \mathcal{I}_{R_2})$. To do this we transfer the disparity maps from all input view pairs (e.g. $\mathcal{D}_{0,1}$, $\mathcal{D}_{2,3}$) to the RVP $(\mathcal{I}_{R_1}, \mathcal{I}_{R_2})$. A simple triangulation and projection is sufficient [19] to perform this transfer. Independent from the transfer method used, we refer to it using the transfer function $\Theta_k^{A,B} : (\mathbf{x}_A, d_{A,B}) \mapsto \mathbf{x}_k$, which transfers the point $\mathbf{x}_A$ using input disparity $d_{A,B} = \mathcal{D}_{A,B}(\mathbf{x}_A)$ into view $\mathcal{I}_k$. So, we use functions $\Theta_{R_1}^{A,B}$ and $\Theta_{R_2}^{A,B}$ to compute a *reprojected* disparity map $\tilde{\mathcal{D}}_{A,B}$ by applying the transfer to every disparity in $\mathcal{D}_{A,B}$: $\tilde{\mathcal{D}}_{A,B}(\mathbf{x}_{R_1}) = \Theta_{R_1}^{A,B}(x_A, \mathcal{D}_{A,B}(x_A)) - \Theta_{R_2}^{A,B}(x_A, \mathcal{D}_{A,B}(x_A)) = \mathbf{x}_{R_1} - \mathbf{x}_{R_2}$. In practice, all available disparity maps are transferred to the RVP and they are used to compute the pdf of the disparities in the RVP.

2.2 Visibility Model

There are in general zero, one or even multiple disparity estimates for every pixel of a reprojected disparity map depending on the occlusions and discontinuities in $\mathcal{D}_{A,B}$. In an ideal world, the case with only one disparity occurs when cameras of the reference and input views observe only non-occluded scene points. Multiple disparities occur due to depth discontinuities where several input disparities of different scene surfaces reproject to the same location in the reference view with different disparities. In these cases we pick the closest depth estimate (i.e. the occluding surface) to maintain correct visibility. Zero disparities occur mainly due to occlusions and thus, no disparity information is available.

The Reliable Area: We must ensure that every reprojected disparity comes from a surface observable in both, reference and input view pair. If that is not the case, it means that the point corresponding to this disparity is potentially occluded or not visible in the RVP. To check this we verify if a point on the surface defined by the maximum disparity is outside the frustum of $\mathcal{I}_A$ and $\mathcal{I}_B$. In practice, for every point $\mathbf{x}_{R_1} \in \mathcal{I}_{R_1}$ we compute $\mathbf{x}_k = \Theta_k^{R_1,R_2}(\mathbf{x}_{R_1}, d_{\max})$ and check if $\mathbf{x}_k \in \mathcal{I}_k$ for $k \in \{A, B\}$. If $\mathbf{x}_k \notin \mathcal{I}_k$, then the disparity at $\mathbf{x}_{R_1}$ is invalidated, meaning that either it is occluded in the RVP or it is not visible in the input views. Here, $d_{\max}$ is the maximum disparity of view pair $\mathcal{I}_{R_1}$ and $\mathcal{I}_{R_2}$.

2.3 Probability Density Function of Disparity

We reproject all input disparities to the RVP and use them as measurements to compute a probability density function of the disparities in the reference image.

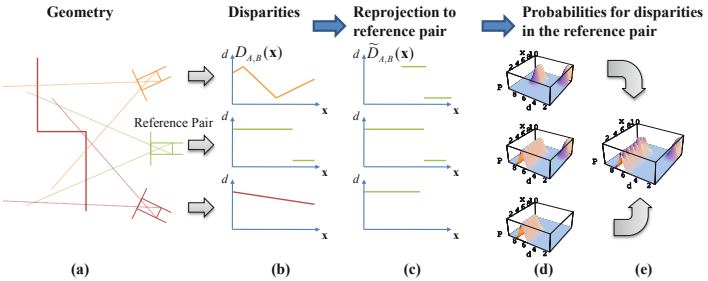

Fig. 2. The pdf estimation: (a) The 2D-geometry is observed from three different stereo cameras. (b) Disparity maps from input stereo pairs are determined. (c) The reprojected disparity maps to the reference view pair lead to (d) three pdfs for each reprojected disparity map which are finally (e) combined in one pdf.

Later, we draw from this pdf the most probable disparity at every pixel location of the reference view as illustrated in Fig. 2.

First we build the set $\mathcal{S}$ of reprojected disparity maps by reprojecting all N input disparity maps to the RVP. Now we use these disparity maps as measurements to sample the pdf of disparity d at every given pixel location $\mathbf{x}$ in the reference image. The unknown pdf p can be modeled as:

$$p(\mathbf{x}, d) = \sum_{\tilde{\mathbf{x}} \in \mathcal{I}_{R_1}} \sum_{\tilde{d} \in \mathcal{S}(\tilde{\mathbf{x}})} p(\mathbf{x}, d \,|\, \tilde{\mathbf{x}}, \tilde{d})\, p(\tilde{\mathbf{x}})\, p(\tilde{d}) \tag{1}$$

where $p(\mathbf{x}, d|\tilde{\mathbf{x}}, \tilde{d})$ is the joint probability of disparity d at pixel location $\mathbf{x}$ given a measurement $\tilde{d} \in \mathcal{S}(\mathbf{x})$ at measured location $\tilde{\mathbf{x}}$ of a reprojected disparity map in $\mathcal{S}$. We assume that all the measurements of locations and disparities are equally probable. Therefore we consider them constant and write after marginalization:

$$p(\mathbf{x}, d) \sim \sum_{\tilde{\mathbf{x}} \in \mathcal{I}_{R_1}} \sum_{\tilde{d} \in \mathcal{S}(\tilde{\mathbf{x}})} p(\mathbf{x}, d \,|\, \tilde{\mathbf{x}}, \tilde{d}) \tag{2}$$

The probability $p(d, \mathbf{x} \,|\, \tilde{\mathbf{x}}, \tilde{d})$ depends on the reprojection uncertainty defined by the probability $p_L(\mathbf{x}, d \,|\, \tilde{\mathbf{x}}, \tilde{d})$ that the scene point $\mathbf{X}(\tilde{\mathbf{x}}, \tilde{d})$ (computed from the uncertain correspondence $\mathbf{x}_A \leftrightarrow \mathbf{x}_B$ before reprojection) projects to the location $\mathbf{x}$ in the image $\mathcal{I}_{R_1}$. It further depends also on the probability $p_D(\mathbf{x}, d \,|\, \tilde{\mathbf{x}}, \tilde{d})$ that the disparity of $\mathbf{X}$ is d in the RVP. So we write it as:

$$p(\mathbf{x}, d \,|\, \tilde{\mathbf{x}}, \tilde{d}) = p_L(\mathbf{x}, d \,|\, \tilde{\mathbf{x}}, \tilde{d}) \cdot p_D(\mathbf{x}, d \,|\, \tilde{\mathbf{x}}, \tilde{d}) \tag{3}$$

These uncertainties are naturally coming from the input image pairs and can be directly estimated there. In the following, we use the transfer function Θ to relate the uncertainties to the RVP. The location uncertainty at pixel position $\mathbf{x}$ is measured by the discrepancy between the true location $\mathbf{x}_A = \Theta_A^{R_1,R_2}(\mathbf{x}, d)$ and the measured location $\tilde{\mathbf{x}}_A = \Theta_A^{R_1,R_2}(\tilde{\mathbf{x}}, \tilde{d})$ in the input image obtained by back-projections of the true $\mathbf{x}$ and measured $\tilde{\mathbf{x}}$ locations from the reference

image. Thus, p_L has its maximum value when the true $\mathbf{x}_A$ and measured $\tilde{\mathbf{x}}_A$ back-projections coincide and it decreases with increasing distance. So we use:

$$p_L(\mathbf{x}, d|\tilde{\mathbf{x}}, \tilde{d}) \sim \exp\left(-\frac{1}{2\sigma_x^2}\left\|\Theta_A^{R_1,R_2}(\mathbf{x}, d) - \Theta_A^{R_1,R_2}(\tilde{\mathbf{x}}, \tilde{d})\right\|_2^2\right) \tag{4}$$

Similarly, p_D is maximal at $\tilde{d}$ and decreases for differing depths:

$$p_D(\mathbf{x}, d|\tilde{\mathbf{x}}, \tilde{d}) \sim \exp\left(-\frac{1}{2\sigma_d^2}\left\|(\Theta_B(\tilde{\mathbf{x}}, \tilde{d}) - \Theta_A(\tilde{\mathbf{x}}, \tilde{d})) - (\Theta_B(\mathbf{x}, d) - \Theta_A(\mathbf{x}, d))\right\|_2^2\right) \tag{5}$$

where $\Theta_A = \Theta_A^{R_1,R_2}$, $\Theta_B = \Theta_B^{R_1,R_2}$ and σ_x is the location uncertainty defined by pixelwise sampling and σ_d is the accuracy of the disparity estimation. Note that $\tilde{\mathbf{x}}$ and $\tilde{d}$ are taken from the set of reprojected disparity maps. If the disparities d and $\tilde{d}$ are the same, the point defined by $(\mathbf{x}, d)$ will project to exactly the same input locations $\tilde{\mathbf{x}}_A$ and $\tilde{\mathbf{x}}_B$ and define the same disparity in the input view, which will result in the maximum value. Otherwise, points with different disparities or locations will back-project to locations away from the measurement $(\tilde{\mathbf{x}}, \tilde{d})$ and get lower values.

2.4 Disparity Estimation

Finally we estimate the most probable disparity map from the estimated pdf. From $p(\mathbf{x}, d) = p(d|\mathbf{x})p(\mathbf{x})$ and assuming that image positions $\mathbf{x}$ are equiprobable we get: $\hat{d} = \text{argmax}_d\, p(d|\mathbf{x}) = \text{argmax}_d\, p(\mathbf{x}, d)$.

3 Results

We evaluated our method using classical stereo datasets with ground truth [14] and real world data. In our experiments we used $\sigma_d = 1$ and $\sigma_x = 1$. The standard two-frame stereo datasets from Middlebury [14] contain up to 9 images from which we computed 72 (*Venus*, *Teddy*, *Cones*) or 42 (*Art*, *Moebius*, *Aloe*) disparity maps from all possible image combinations. After that, we fused these disparity maps to the standard reference view pair (e.g. $(2, 6)$ for *Teddy*) and computed the percentage of erroneous pixels (disparities that differ by more than 1). For stereo processing we used Geodesic Support Weights [8] (GSW). We used constant parameters for stereo among all baselines and datasets.

3.1 Comparison to Other Fusion Methods

We compare our method to other fusion algorithms, in particular the stability-based algorithm of Merrell et al. [11] using our own implementation running on CPU and the bundle optimization of Zhang et al. [19] using their implementation (without their stereo-matching and without final bundle adjustment). We used

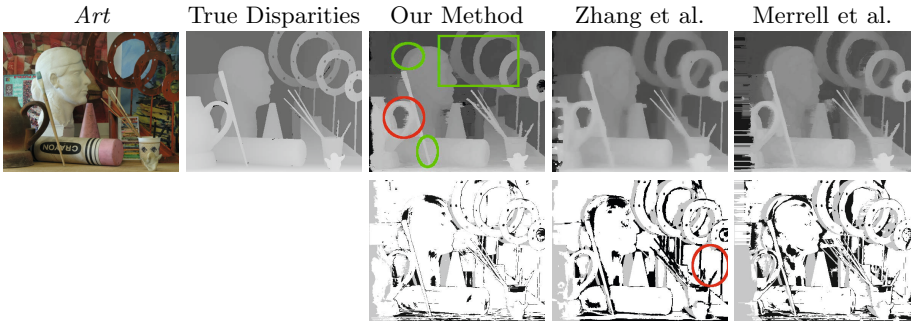

Fig. 3. The disparities and bad pixels of different fusion methods for the dataset *Art*.

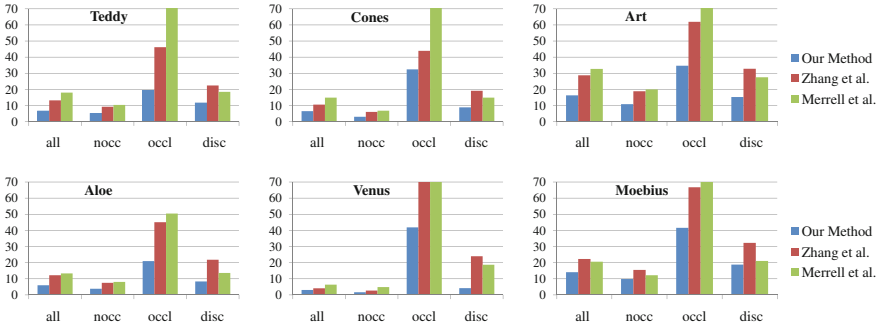

Fig. 4. The performance of different fusion methods. Disparity maps were computed using GSW [8]. Error bars show percentages of disparities that differ by more than 1 from the ground truth in the whole image (all), non-occluded (nocc) or occluded pixels (occl) and regions near discontinuities (disc). We fused up to 72 disparity maps.

the same input data (i.e. disparity maps) for all fusion methods. The method of [19] seems to be optimized for short baselines (the video sequences of [19] have much smaller baselines than the datasets of [14]). Our method works better with larger baselines, which is our target application. The error bars in Fig. 4 show that our method performs very well. It is also visible in Fig. 3 and Fig. 5 where our method preserved sharp object boundaries and thin structures.

Analysis: In Merrell, visibility-constraints are enforced using their expensive definition of stability (having a complexity of $\mathcal{O}(N^2)$ – please note that the computation of $\mathcal{S}$ is $\mathcal{O}(N)$ and that for every disparity of $\mathcal{S}$, $N-1$ projections are performed). However, visibility can be maintained more efficiently using our reprojection and the reliable area (having $\mathcal{O}(N)$, because at every entry of $\mathcal{S}$ we only update the global pdf by summation). This also has the big advantage that projection uncertainties can be used later, whereas in Merrell it is not possible. Moreover, for optimal stability calculation it is important that the number of outliers having a negative stability is equal to the number of outliers

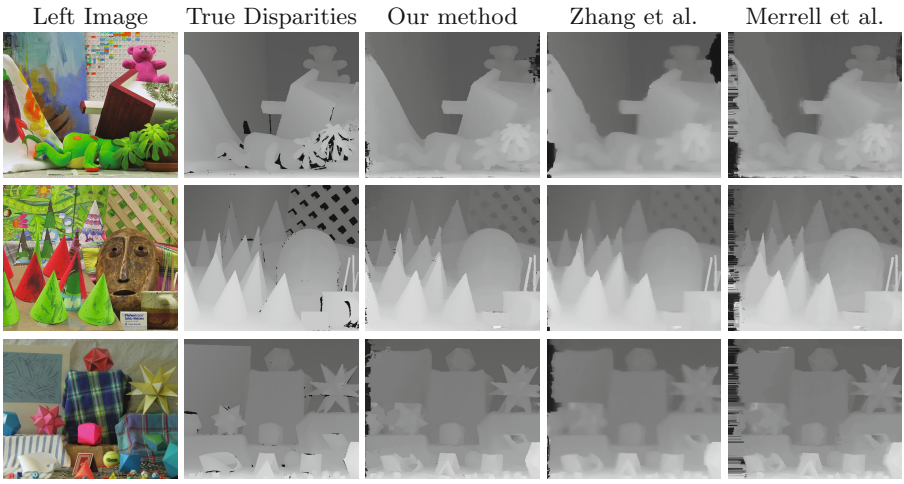

Fig. 5. The disparity maps of different fusion methods for the datasets *Teddy* (first row), *Cones* (second row) and *Meobius* (bottom row).

with positive stability. Our experiments suggest that this assumption is often violated in occluded regions, where usually many outliers are present.

In Zhang's method, the correct disparity is supported by the **simultaneous** combination of *geometric coherence* and *color similarity*. Geometric coherence alone supports also background disparities of surfaces occluded by foreground objects in the reference view, because visibility is not determined and this is problematic in cases where fore- and background objects are of similar color. The optimization using belief propagation ensures smoothness in these ambiguous situations, but seems to perform suboptimally in regions near discontinuities. Due to the results we obtained during our evaluation (our method does not use any kind of optimization), we believe that our pdf will also bring a huge advantage in comparison to the method of Zhang, especially near discontinuities and for wide-baseline sequences. We would like to stress that we explicitly compute visibility to disambiguate depth hypotheses at an early stage and model projection uncertainties.

Execution Times: For the dataset *Teddy* (72 disparity maps) our method took 8.7 s (not optimized), the method of [11] took 40.7 s and the method of [19] 175 minutes (i.e. 146 s/disparity map). These times do not include stereo matching and were measured on an Intel E8200 dual-core with 2.66 GHz (for our method and [11]) or an Intel E5405 quad-core Xeon CPU with 2.00 GHz (for [19]). For our real-time implementation, we use SIMD-instructions of the SSE2 instruction set and simplified the reprojection for motion-stereo. Using pre-computed kernels, we are able to fuse 16 disparity maps in just 30 ms on a mobile CPU (2 GHz; 320x240 pixels; 60 disparity levels).

3.2 Real World Sequences

We tested our method on real world sequences from a moving vehicle and we estimate the the transfer function from sensors attached to the vehicle. Fig. 6 shows a rectified camera frame, one input disparity map (computed using a real-time stereo method [17]) and one fused disparity map. For fusion we used a highly optimized implementation (using SIMD instructions) to fuse 16 adjacent input disparity maps. Due to the monocular camera system, it must be noted that objects which move parallel to the image plane might be determined with a wrong depth. However, such situations arise relatively seldom in our application.

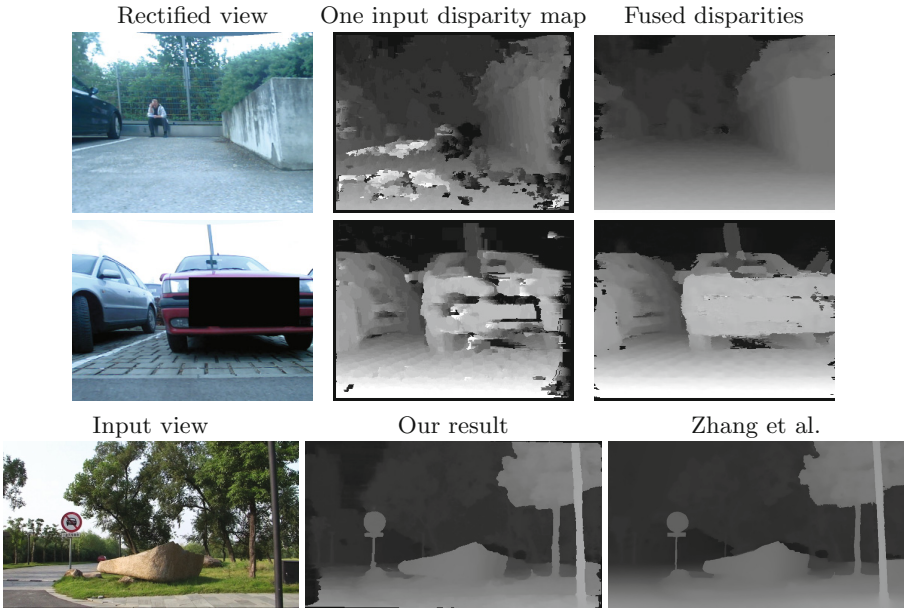

Fig. 6. First two rows: Our method applied to sequences from our vehicle using real-time stereo [17]. Last row: results for the sequence *Road* provided by [19].

Fig. 6 shows fused disparities of a sequence provided by [19], along with the camera frame and their fusion result. For stereo matching we used GSW [8] and ensured a minimal and maximal baseline of 5 and 7 frames (the baseline of adjacent frames was too small for robust matching with GSW). We fused disparity maps of only 20 adjacent frames and this explains why some disparities which are outside of the field of view are missing (black regions).

4 Conclusion

In this paper, we propose a novel probabilistic method for fusing disparity maps in classical stereo or motion-stereo setups. We achieve this by computing a

probability density function from all provided disparity maps. From this distribution, we determine the most probable disparity map for a given reference view pair.

We introduced a generic probabilistic model that uses projection uncertainties for robustness against outliers and reprojection using the reliable area for efficient and explicit visibility determination.

Acknowledgements. We want to thank Guofeng Zhang for running his method on the datasets. This work is supported by the BMW Group.

References

1. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. In: ICCV, pp. 377–384 (1999)
2. Collins, R.T.: A space-sweep approach to true multi-image matching. In: CVPR, p. 358 (1996)
3. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. IJCV 70(1), 41–54 (2006)
4. Gargallo, P., Sturm, P.: Bayesian 3D modeling from images using multiple depth maps. In: CVPR, pp. 885–891 (2005)
5. Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: CVPR, pp. 807–814 (2005)
6. Hirschmüller, H., Innocent, P.R., Garibaldi, J.: Real-time correlation-based stereo vision with reduced border errors. IJCV 47(1-3), 229–246 (2002)
7. Hirschmuüller, H.: Stereo vision in structured environments by consistent semi-global matching. In: CVPR, pp. 2386–2393 (2006)
8. Hosni, A., Bleyer, M., Gelautz, M., Rhemann, C.: Local stereo matching using geodesic support weights. In: ICIP (2009)
9. Koch, R., Pollefeys, M., Van Gool, L.: Multi Viewpoint Stereo from Uncalibrated Video Sequences. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 55–71. Springer, Heidelberg (1998)
10. Kolmogorov, V., Zabih, R.: Multi-camera Scene Reconstruction via Graph Cuts. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 82–96. Springer, Heidelberg (2002)
11. Merrell, P., Akbarzadeh, A., Wang, L., Frahm, J.M., Yang, R., Nistér, D.: Real-time visibility-based fusion of depth maps. In: ICCV, pp. 1–8 (2007)
12. Okutomi, M., Kanade, T.: A multiple-baseline stereo. PAMI 15(1), 353–363 (1993)
13. Sato, T., Kanbara, M., Yokoya, N., Takemura, H.: Dense 3-D reconstruction of an outdoor scene by hundreds-baseline stereo using a hand-held video camera. IJCV 47, 119–129 (2002)
14. Scharstein, D., Szeliski, R., Zabih, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV 47, 7–42 (2002)
15. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: CVPR (2006)
16. Szeliski, R.: A multi-view approach to motion and stereo. In: CVPR, p. 1157 (1999)
17. Unger, C., Benhimane, S., Wahl, E., Navab, N.: Efficient disparity computation without maximum disparity for real-time stereo vision. In: BMVC (2009)
18. Zach, C.: Fast and high quality fusion of depth maps. In: 3DPVT (2008)
19. Zhang, G., Jia, J., Wong, T.T., Bao, H.: Consistent depth maps recovery from a video sequence. PAMI 31(6), 974–988 (2009)
20. Zitnick, L.C., Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R.: High-quality video view interpolation using a layered representation. In: SIGGRAPH (2004)