YORK
UNIVERSITÉ
UNIVERSITY

redefine THE POSSIBLE.

Stereoscopic Datasets and Algorithm Evaluation for Driving Scenarios

Mikhail Sizintsev and Richard P. Wildes

Technical Report CSE-2013-06

June 10 2013

Department of Computer Science and Engineering
4700 Keele Street, Toronto, Ontario M3J 1P3 Canada

Stereoscopic Datasets and Algorithm Evaluation for Driving Scenarios

Mikhail Sizintsev
York University
Toronto, ON, Canada
sizints@cse.yorku.ca

Richard P. Wildes
York University
Toronto, ON, Canada
wildes@cse.yorku.ca

May 26, 2013

Abstract

This report presents novel binocular stereo video datasets that capture automotive driving relevant scenarios as well as evaluation of five disparity estimation algorithms on the acquired imagery. Binocular stereo has great potential as a component technology for driving assistance, as it provides an approach to recovering 3D distance relative to a vehicle and thereby provide critical information for a variety of driving tasks. For incorporation into an overall driving system, candidate algorithms must have their performance specified precisely. The acquired imagery and comparative algorithm evaluation respond to this need by providing detailed qualitative and quantitative evaluation of alternative disparity estimation approaches on driving relevant data.

Chapter 1

Introduction

Analogous to the role of vision in human perception, computer vision-based sensing in automobiles is anticipated to play a key role in the quest to build human-like perception around vehicles. Along these lines, estimates of 3D scene structure are of particular importance to a variety of driving tasks, including collision avoidance and more general navigation in the presence of other objects. For principled selection between competing technologies for a particular driving task and incorporation into a larger automotive system, it is critical that components have their individual performance capabilities characterized precisely. Toward such ends, this paper presents comparative empirical evaluation of binocular stereo algorithms on driving relevant datasets.

While technologies such as the Global Positioning System (GPS) are likely to form the cornerstone of the sensing systems of next generation vehicles [3], vision-based sensing is identified as a key complementary contributor [22, 14]. Active sensing technologies (e.g., radar, lidar) would also play a role in vehicle guidance. However, they have a number of limitations that do not apply to vision-based sensing, including reliance on special purpose hardware, dependency on the working media and external lighting as well as susceptibility to interference when multiple sensors are in a crowded area (congested traffic). Furthermore, vision-based sensing naturally extends beyond distance measurements to include lane following, object recognition and sign reading. Conceptually, a single video camera might be enough for navigation; however, having multiple cameras has potential for providing increased overall performance. Along these lines, binocular stereo, providing the minimal multiview setup, is a strong candidate for driving applications.

Considerable previous research has addressed the comparative empirical eval-

uation of stereo vision algorithms without consideration for specialized applications [17, 13, 4, 11, 24]. Closer to the focus of the present research, various efforts have considered stereo vision evaluation specifically for driving, including the evaluation of various match metrics under both local and global matchers [22], the comparison of various non-local matchers with a single match metric [14] and the comparison of stereo *vs.* motion-based algorithms [21]. Significantly, a good deal of field data encompassing a variety of automotive driving scenarios has appeared, *eg.* [8]. These datasets capture a great variety of road situations, weather conditions and camera modes with one significant drawback – complete absence of ground truth, or very limited sparse depth measurements for still frames using lidar or similar technology. On the other hand, significant effort has been devoted to generating complex synthetic (i.e., computer graphics generated) datasets with full ground truth (depth and motion); while these data sets are modeled on driving scenarios, they fail to capture realistic scene materials, lighting and sensor characteristics, which are crucial considerations in practice. Not surprisingly, the conclusions of algorithm evaluations on synthetic scenes may turn out to be significantly different from the field data experimental results [26].

In the light of previous research, the current effort makes the following contributions: (i) To fill the gap between evaluation of stereo vision algorithms for automotive applications on largely ungroundtruthed field data and completely artificial computer graphics generated data, a novel driving relevant dataset of laboratory acquired stereo videos with dense disparity data groundtruth for every frame is presented; (ii) A complementary field dataset acquired from a prototype stereo video camera equipped car is presented; (iii) Results of evaluating five disparity estimation algorithms on both the laboratory and field datasets are presented. In distinction from previous efforts, the current paper covers a wider range of stereo vision algorithms on driver relevant datasets.

Chapter 2

Datasets for driving scenarios

2.1 Field driving data

The field data was provided by an automotive manufacturer. In essence, it comprises a long video (7240 frames) acquired from a car equipped with a prototype calibrated binocular video camera with 640×480 resolution at 30 fps. A followed car performs a variety of basic maneuvers, including accelerating, decelerating, lane changing and turning with respect to the camera bearing car. No groundtruth 3D distance currently accompanies this dataset. Example frames are presented in Fig. 4.6. Essentially, this video sequence serves as a prototype for the laboratory dataset described below; therefore, experimental results from the laboratory data can be more readily extrapolated to the real world.

2.2 Laboratory driving emulation data

Acquisition of disparity groundtruth in field conditions is notoriously difficult. Although datasets with laser range scanned groundtruth have appeared (e.g., [8]), their spatial and temporal density are severely compromised. In response to this state of affairs, the remainder of this section documents the construction of a scale model driving relevant dataset with dense disparity groundtruth.

The developed 3D scale model is depicted in Fig. 2.1. The horizontal surface of the model consists of a T-junction roadway and grass as well as various roadside structures, including houses, trees and stop sign. Stereo cameras are mounted on a computer controlled motion platform at one end of the model to simulate video capture from a car moving through the scene along the roadway. A model car is

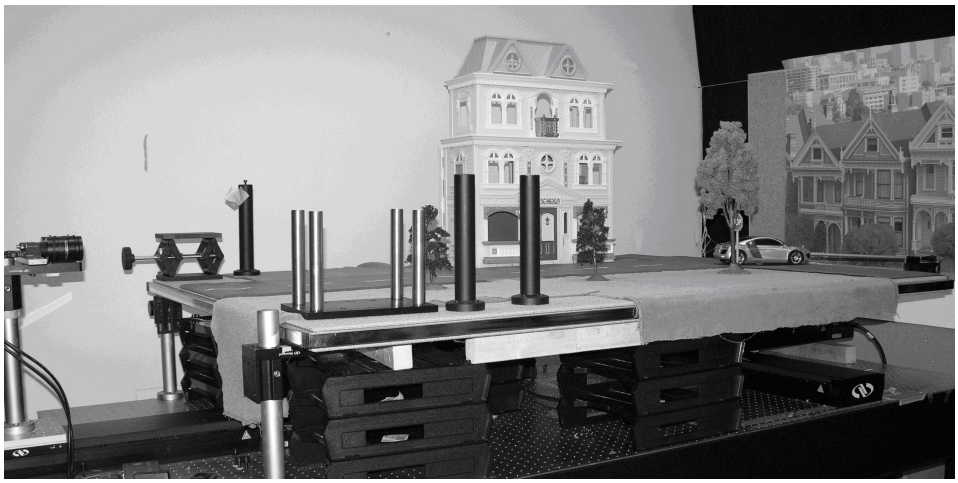

Figure 2.1: Lab setup for the scale driving model.

connected to a second computer controlled motion platform to simulate motion of a second car through the scene that can be captured from the first. The scale of the entire mode is approximately $1:15$. The stereo baseline is chosen to be 8 cm, which is well within the limits of the model car width, 13 cm. Illumination is provided by ambient overhead lighting in the laboratory to approximate overhead, outdoor illumination.

Nine experimental conditions were captured using the model that differ according to the relative motions and positions of the camera and car. Table 2.1 groups the scenarios and explains them in detail. These conditions were chosen to comprise a representative set of abstractions from actual driving conditions captured in the field data described in Sec. 2.1. For all conditions, 101 frames were captured. All motions were generated by having the camera/car mounted on the computerized motion control platforms configured to span the desired space of motions.

For image capture, CCD video cameras with 8-bit monochrome at 1024×768 spatial resolution were used. The actual depth estimation analysis in Sec. 4 was done on half resolution at 512×384. Example frames from three of the scenarios described in Tab. 2.1 are shown in the top row of Fig. 2.2. While the acquired imagery does not appear completely like natural imagery, it it does capture an interesting range of variables, including, appropriately scaled scene components; car with metallic paint, specular windows and realistic detailing; textured road surface; diffuse, sky-type lighting; also a range of driving maneuvers are encom-

Set	Description
000	No motion: Both camera and car are at rest on the road
	Same lane motion: The camera and car move along the same lane
001a	Camera and car move at same speeds
001b	Camera moves faster than car
001c	Camera moves slower than car
	Lane change: The camera moves along one lane of the road and the car changes lanes
002a	Car starts in same lane as camera and switches to opposite lane
002b	Car starts in opposite lane and switches to same lane as camera
	Turn: The camera moves down one lane of the road; the car starts in the same lane, subsequently turns into the orthogonal road at the T-junction and then continues to move along that road
003a	Car turns left
003b	Car turns right
004	Intersection: The camera moves along the road before the T-junction intersection and the car moves along the orthogonal road

Table 2.1: Evaluation datasets and brief distinctive description for each scenario.

Name	Colour	Description
car	red	the car followed by the cameras
road	green	the ground plane consisting of the paved road and the grassland to the right
background	blue	far houses and the sky on a frontoparallel plane
building	cyan	building to the left of the road
trees	purple	trees, foliage and stop sign
all		full scene

Table 2.2: Scene segmentation annotation

passed. Moreover, it will be shown in Sec. 4 that depth estimation results on the scale model imagery are in good accord with those recovered from driving field data (unlike previous experiments with computer graphics generated imagery [26]), which further suggests that the developed model captures key driving relevant factors.

To obtain pixelwise groundtruth disparity, a structured light approach [18] was applied to each binocular frame across the entire videos, as employed elsewhere in comparable previous grounthruthing [16, 13]. To provide a fine grained analysis of algorithm performance as a function of scene components, images are segmented according to meaningful object categories, as documented in Tab. 2.2 and visualized in Fig. 2.2, second row. The segmentation has been performed in a semi-automatic fashion. The car is manually segmented in the first frame and tracked throughout the remainder of the sequence using robust affine template matching based on the Lucas-Kanade algorithm [2]. If the car changes appearance significantly, *eg.* while turning as in 003a and 003b, it is hand segmented in additional frames within the sequences. After the **car** is localized, the **background** is automatically detected by robustly fitting a plane to the disparity groundtruth in the region above the car, while the **road** is found in a similar fashion by robustly fitting a plane using the points other than **car** and **background**. Finally, **building** and **trees** are trivially localized once the **car**, **road** and **background** are excluded, since they manifest as the remaining portions of the left and right half scenes, resp.

While it is common in general purpose stereo evaluation to report results for different regions using more "generic" labeling (e.g., near *vs.* away from 3D boundaries) [17], in the present evaluation emphasis is instead given to segmentation by more directly driving relevant object categories, e.g., which will allow

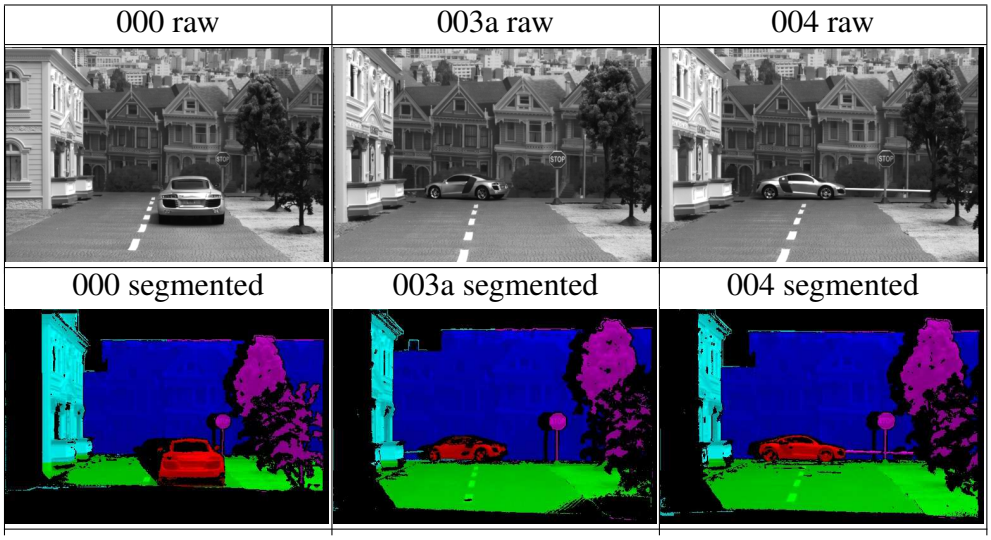

Figure 2.2: Sample raw frames and segmentation results

for comparison of ability to recover cars *vs.* road surfaces *vs.* various road side structures.

Chapter 3

Stereo algorithms for driving

Decades of intensive research in computational stereo have resulted in a wide variety of algorithms. Standard taxonomies of binocular disparity estimation divide algorithms into local and global methods emphasizing their difference in complexity and, consequently, in speed and accuracy [17]. Correspondingly, a classic local block-matching algorithm [6] is considered and contrasted with a standard (arguably the best [23]) global matcher, graph cuts [5]. Still, this bimodal taxonomy does not reasonably capture other useful algorithmic instantiations. In particular, three additional considerations that have played a significant role in the design of stereo matchers that should be captured include the combination of local and global matching, the use of multiresolution, coarse-to-fine processing and the use of high confidence matches to constrain operations in less well defined regions. Correspondingly, three more exemplars are included. First, the semiglobal stereo matcher is considered [12]. As its name suggests, this matcher can be seen as a blend of local and global approaches. Second, a coarse-to-fine matcher is considered [20]. As with all coarse-to-fine matchers, this algorithm makes use of initial coarse spatial resolution matching to guide subsequent finer resolution refinement; additionally, it makes use of adaptive windowing to ameliorate poor resolution of 3D boundaries, a standard shortcoming of multiresolution matching. Third, a region growing matcher is considered [7]. This approach makes use of initial sparse, but high confidence matches to constrain subsequent matches at the expense of match density. In summary, five algorithms have been considered, as summarized in Tab. 3.1.

The current study does not investigate the performance of different pointwise and area-based match metrics, as considerable previous investigations have concluded that real data requires normalization or rank-based measurements to get

Alg.	Description
NCC	dense block matching [17]
CTF	coarse-to-fine adaptive block matching [19]
SGM	semiglobal matching [12]
GC	graph cuts stereo [5]
RG	region-growing stereo [7]

Table 3.1: Algorithms considered in evaluation.

reliable results [13, 4]. Thus, all 5 algorithms rely on normalized cross-correlation as their match measure computed over a 5×5 window, except *NCC* which relies on 9×9 windows to obtain adequate match aggregation.

Chapter 4

Experimental results

The *NCC* algorithm was implemented by the current authors, while the original implementations of *CTF*, *SGM*, *GC* and *RG* were provided by [19], [12], [5] and [7], resp. Parameters were set individually for each algorithm so as to achieve best overall accuracy across the entire test dataset, while keeping average overall density of estimates between algorithms reasonably similar so that error statistics will not be biased by some algorithms providing notably sparser results than others. Here, density refers to the percentage of points where an algorithm returns a valid match. Once selected, parameters were kept fixed between datasets. All methods employ the same disparity search range and use Left-Right consistency check failure [6] to detect occlusions as well as to discard more generally invalid matches.

4.1 Laboratory results

4.1.1 Qualitative results

Example recovered disparity maps for middle frames from representative sequences 001a and 004 together with ground truth and error maps are shown in Fig. 4.1. The disparity maps are shown with occlusions and other regions where no result is returned by a particular algorithm (e.g., due to left/right checking failures) overlayed in dark blue and error maps are coloured according to the scene labeling (colour coded as explained in Tab. 2.2) with intensity proportional to the disparity error absolute difference. The supplementary video shows the actual sequences for all datasets [1].

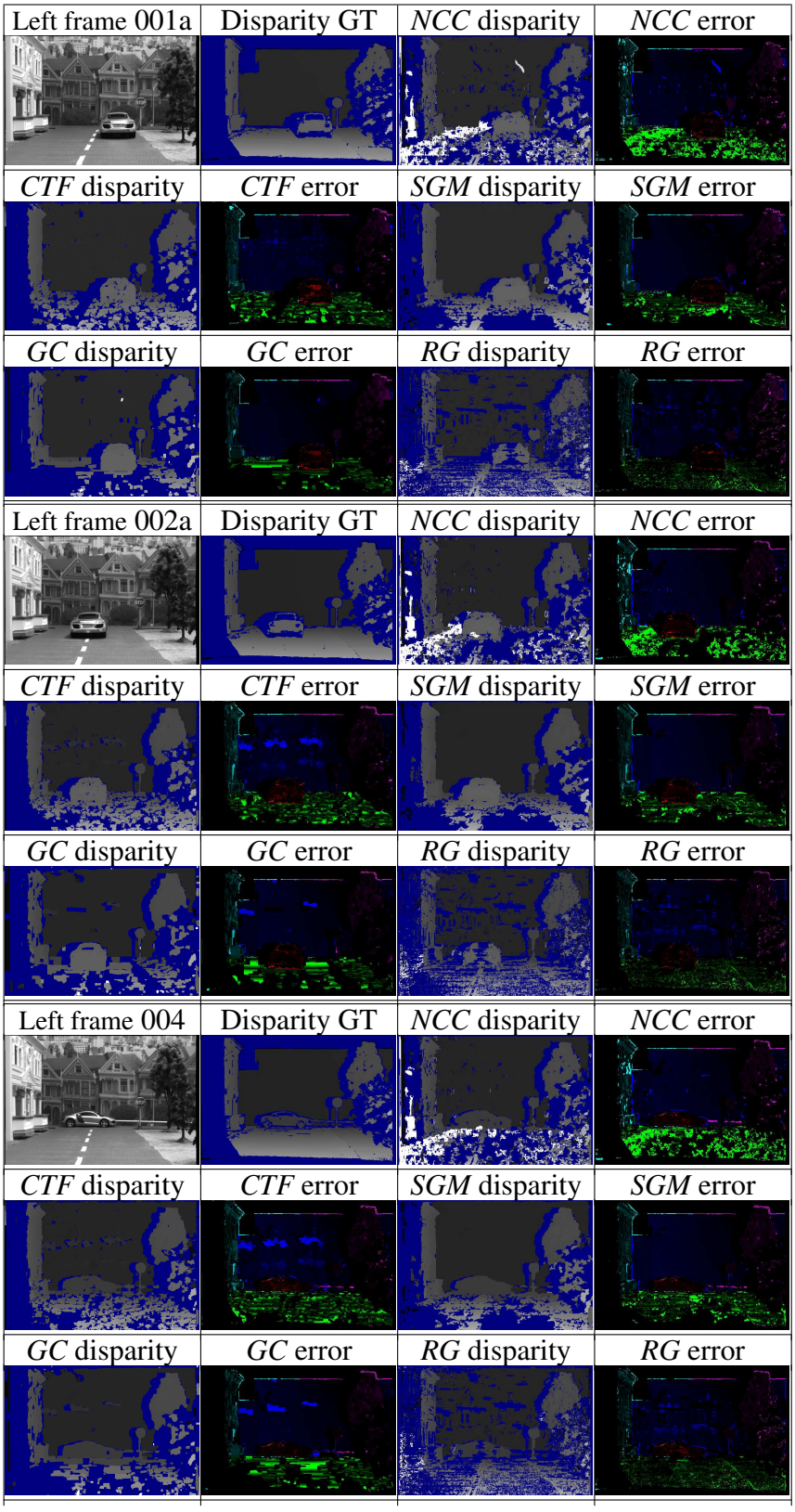

Figure 4.1: Example frame from sequences 001a, 002a, 004 for all algorithms evaluated. See text for details.

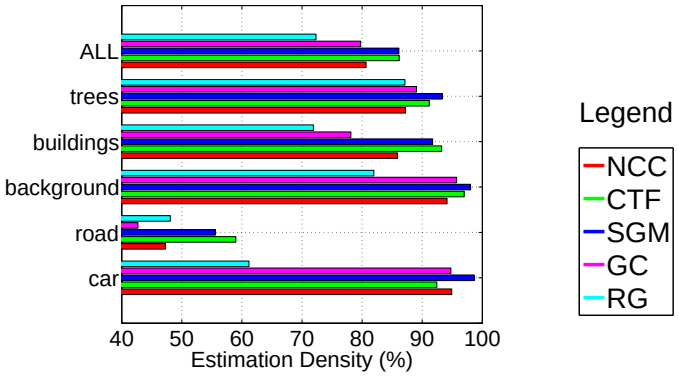

Figure 4.2: Estimation density results (percentage of groundtruthed points where algorithm returns a valid match) for all algorithms for every object category averaged across datasets.

In can be observed from the disparity and error maps that all algorithms generally recover the 3D structure of the scenes, with *SGM* providing particularly reasonable estimates and *NCC* the most gross errors. The most obvious gain of the only true global method is in the **background** regions, which are well approximated by a frontoparallel plane and thereby well suited to the *GC* match propagation approach. Interestingly, the road regions appear to be the most problematic as the majority of errors fall into the **road** object category. Finally, it appears that *CTF* and *SGM* provide the densest results; whereas, *RG* yields the sparsest.

4.1.2 Estimation density

In the following, only points where grountruth is available and where valid matches have been recovered are considered. Since different algorithms yield somewhat different densities for each object category (even given an attempt to control overall density across algorithms, as noted above), densities are reported along with accuracy and both should be considered in tandem to appreciate the results; see Fig. 4.2 for density results. Since preliminary evaluation showed that density depends on algorithm and object category (**car** *vs.* **road**, *etc.*), but not on the particular dataset (driving scenario), results are collapsed across datasets.

Regarding algorithm effects, it is seen that *CTF* and *SGM* yield the densest estimates across all categories. In contrast, the global match propagation of *GC* inconsistently over or under smooths its results in left- vs. right-based matching, which reduces its density under left-right checking match validation in compari-

son to the more local methods. Density performance of *NCC* is similar to that of *GC*. The sparsest results are returned by *RG*, which is consistent with its design principle of being guided by high confidence matches [15, 7]. While sparser estimates might be justified by relatively higher accuracy, results reported in the next section show that such a trend is not present.

Regarding object category effects, it is seen that only the **road** category yields noticeably different lower density results. This result derives from the fact that the road surface is relatively weakly textured (except near the lane markers) and non-Lambertian in reflectance and thereby provides a particularly strong challenge to all of the algorithms.

4.1.3 Cumulative error statistics

To quantify algorithm accuracy, cumulative error statistics have been calculated, which show the percentage of image points where disparity error in comparison to grountruth is within a certain number of pixels. The results are shown in Fig. 4.3 for all algorithms as executed on all datasets and broken out by object category.

It is seen that the major dependence of algorithm accuracy on driving scenario is restricted to **car** object regions (first column of Fig. 4.3). This result is to be expected: Only the camera and the followed car are in motion. Further, the camera traces the same trajectory across all scenarios (varying only in speed 001a - 001c, and not fast enough to yield motion blur); so, no effect of the static scene structures with scenario is expected. With respect to the followed car, in 000 - 002b the view of the followed car is essentially the same: The rear, with variations provided by distance from the camera and position in the scene depending on lane changes. It is seen that distance is not a problem, as the actual disparity range (varying from 45 to 82 pixels) is always within range of the matchers and the number of pixels on target (varying from approximately 3000 to 10000 pixels) prevents it from becoming too small to support locally stable estimates. In contrast, for the turning scenarios 003a and 003b and the intersection scenario 004, matters become interestingly different, as a side view of the car becomes available. The side view provides for a larger number of pixels on target (at any given distance, there are approximately twice as many pixels in a side view *vs.* a rear view). Further, the side view provides additional texture detail in comparison to the rear view (e.g., from doors and detailing). Correspondingly, it is seen that accuracy increases in the **car** regions for side viewing conditions. The effect is particularly noticeable when comparing *CTF* performance in the **car** region on 003a and 003b in comparison to 000 - 002b and for all algorithms when comparing performance in the

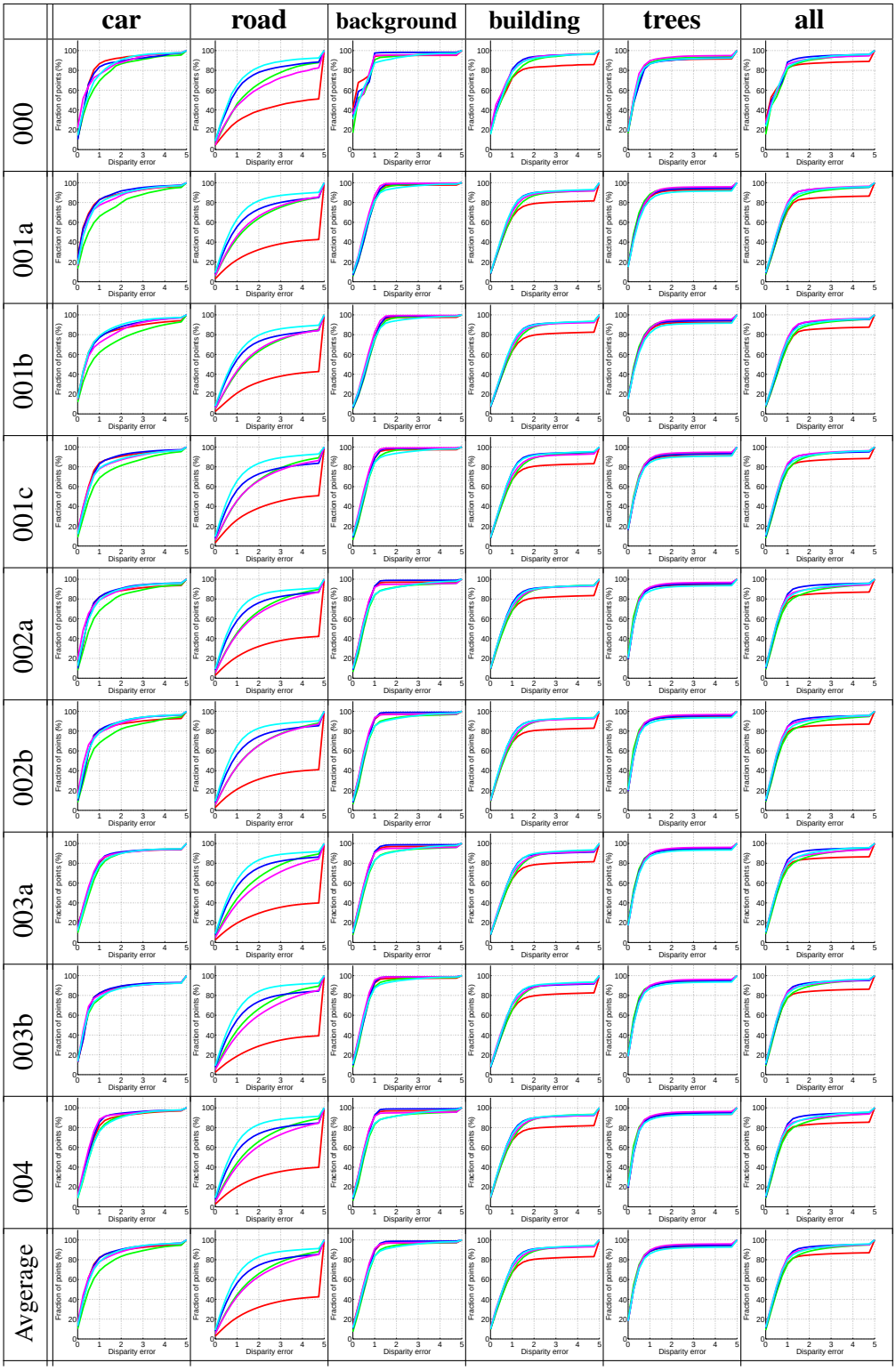

Figure 4.3: Cumulative error plots for all object categories, datasets and algorithms. Algorithm colour coding is consistent with Fig. 4.2.

car region on 004 *vs.* all other cases.

In addition to dependence on driving scenario, it is interesting to examine how accuracy depends on scene object category. To do so it is useful to consider object categories collapsed across scenario, as shown in the bottom row of Fig. 4.3. From the perspective of cumulative error, it is seen that all algorithms perform reasonably well and roughly equally in the **car** regions (with a minor exception for *CTF*), which is satisfying given that detection and positioning of other cars on the road is of great importance in driving applications. In contrast, all algorithms perform relatively poorly (indeed worst) for the highly driving relevant **road** category. Here, it is seen that *NCC* performance is weakest of all, e.g., with only 30% of its results within a reasonably useful 1-2 pixel error. The **background** regions prove to be the easiest for all algorithms as they are relatively well textured and fronto-parallel, which particular suits the smoothing favoured by (semi)global *SGM* and *GC*. In contrast, while the **building** regions remain largely planar and well textured, they are decidedly not frontoparallel and accuracy drops for all algorithms, especially *NCC*. Finally, the **trees** regions are characterized by fine texture detail and therefore also exhibit relatively strong performance across all algorithms.

4.1.4 Framewise error statistics

Since driving data is acquired across time, it is of interest to consider framewise algorithmic accuracy. Figure 4.4 shows framewise overall error plots for disparity estimates within 2 pixels of groundtruth for all algorithms as executed on each dataset. (Framewise errors as function of scene category were not found to provide additional insight with respect variation across time and are suppressed in the interest of space.) It is seen that error performance does not change dramatically with time, which indicates the consistency and stability of the algorithms. The relative ordering of the alternative algorithms is largely preserved across the duration of each sequence, with the (semi)global *SGM* and *GC* providing results that show the least variation. Finally, recall that sequence 000 is completely static and all error variation can be attributed solely to sensor noise.

4.1.5 Spread error statistics

To analyze the spread of the error distributions and thereby understand how the error ranges of the various algorithms compare, errors across object categories are shown as box plots [25] in Fig. 4.5. For the **car** regions, only *CTF* performs

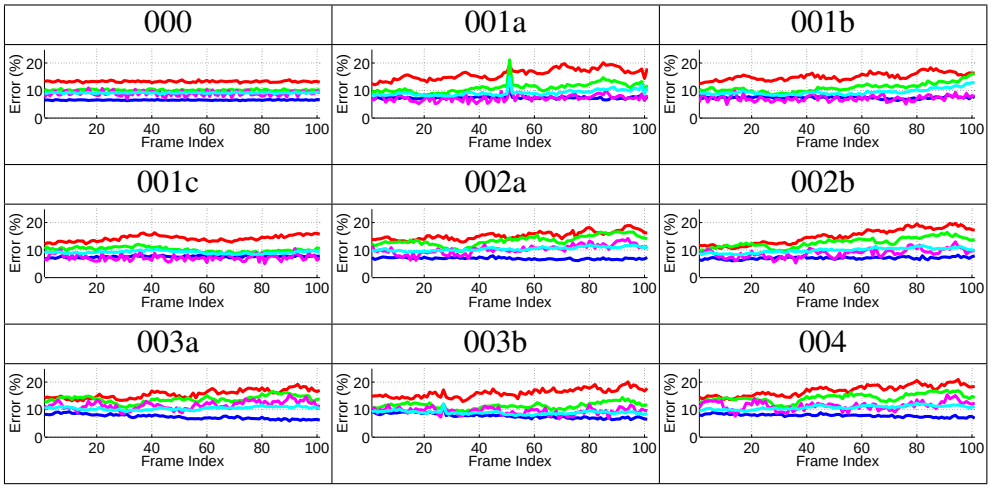

Figure 4.4: Framewise errors for threshold 2 disparity levels for each dataset averaged across all object categories. Algorithm colour coding is consistent with Fig. 4.2.

significantly worse than the others; however separate scene plots (not shown in the interest of space) show that this is only true for situations with rear-view cars in sequences 000 through 002b, which, as noted above, provide fewer pixels on target than side views. For the **road** regions, *NCC* is especially poor (all error rates exceed 50% and fall outside the plot), while *RG* is the best. For the **background** regions, it becomes apparent that *SGM* and *RG* are the strongest and weakest performers, resp. For the **building** regions, *NCC* exhibits consistently higher error rates, while all other methods are very similar in their performance. For the **trees** regions, it is fair to say that performance for all algorithms is quite similar, except that *GC* is better than *RG*. Finally, the results collapsed across all object regions underline the strong overall performance of *SGM*, e.g., with its entire interquartile range lying beneath that of all other algorithms.

4.2 Field data results

Example results from applying all algorithms to the field data are shown in Fig. 4.6 (extended video results shown in Supplemental Video [1]). Results are given as greyscale disparity maps with invalid matches superimposed as dark blue. Significantly, the field data results are quite consistent with the scale model results, with a tendency to further emphasize the weaknesses of each algorithm. The esti-

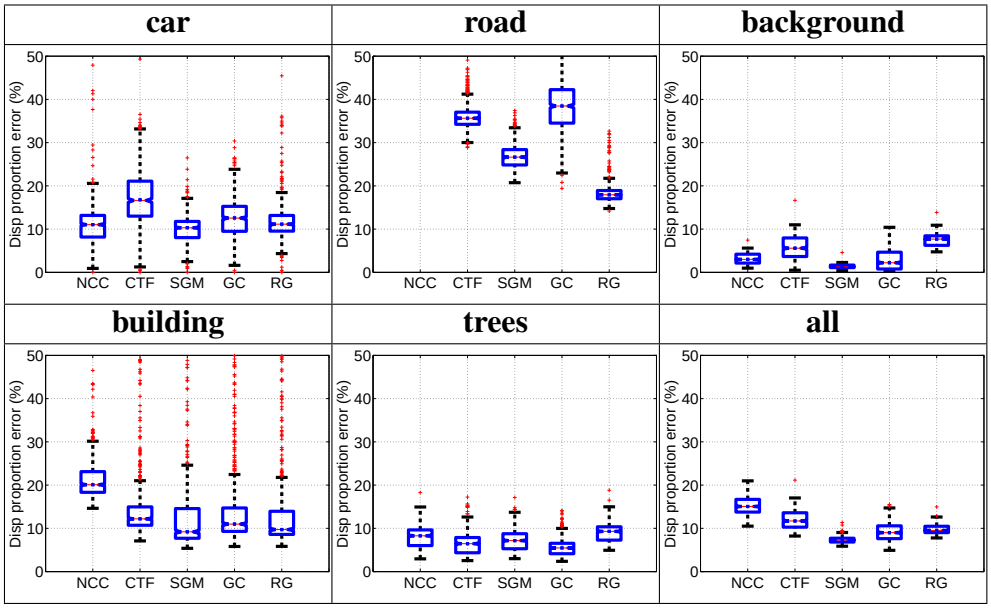

Figure 4.5: Error distributions displayed as box plots cumulative across all datasets. The vertical extent of a box covers the interquartile range; the lines below and above a box extend to the 10^{th} and 90^{th} percentiles; points beyond are shown as $+$.

mates provided by *NCC* are "noisiest", with very imprecise depth discontinuities. *CTF* provides noticeably better results, with errors concentrated in relatively textureless regions in the car interior and road surface. The results of *SGM* appear to be the best overall, with streaking line artifacts that are a well known property of dynamic-programming-based stereo matching, which forms the core of *SGM* [17]. The results of *GC* are greatly oversmoothed, which can also decrease its match density when left- and right-based smoothings are inconsistent. The results of *RG* are relatively sparse without noticeable accuracy improvements relative to the best performance of the other algorithms.

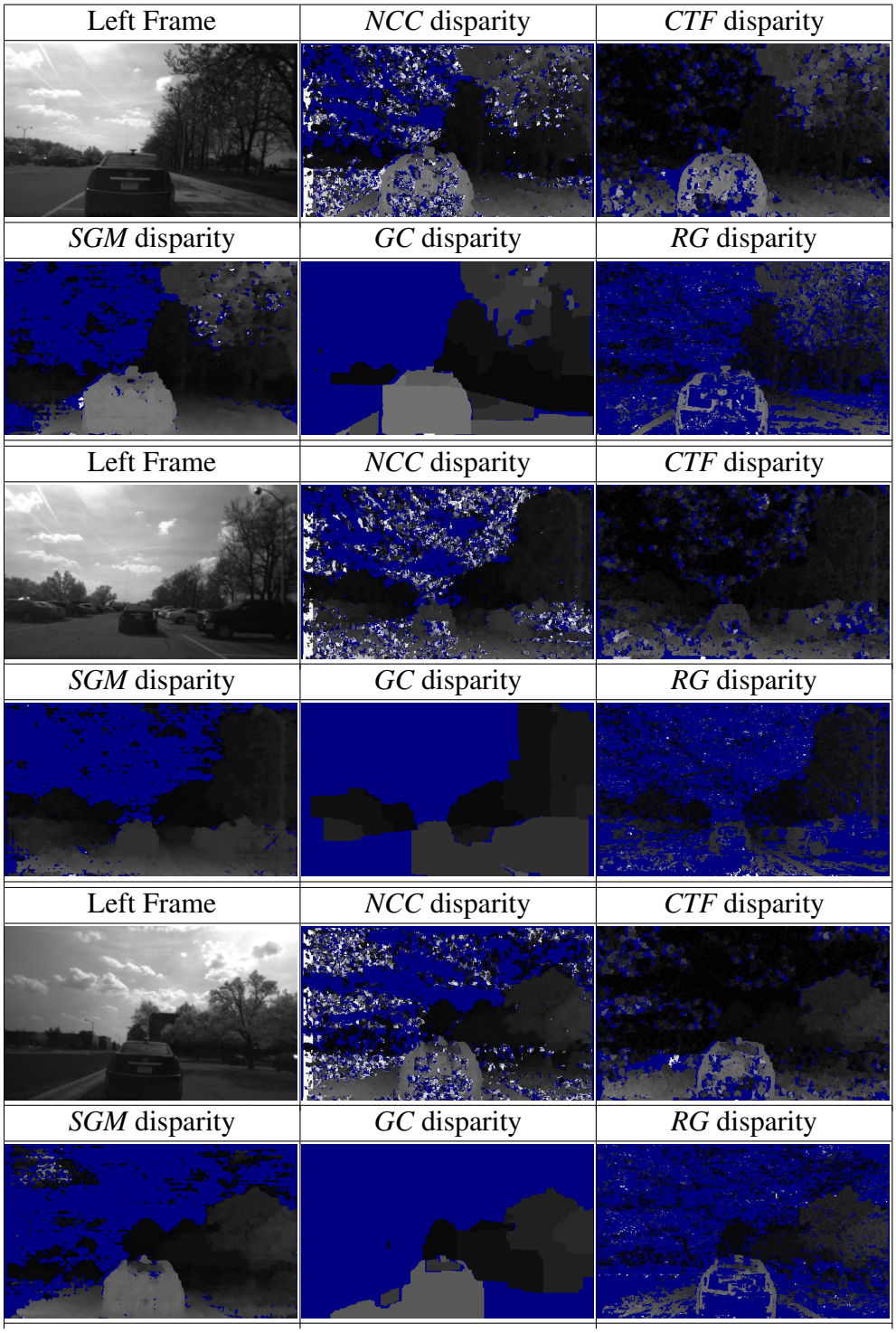

Figure 4.6: Sample results on field dataset for for all algorithms.

Chapter 5

Discussion

NCC is the most straightforward algorithm, which is easy to implement and run on most architectures. This simplicity comes at the price of providing the overall worst accuracy in the current evaluation, of particular significance given that it does not even provide the highest density. This algorithm performed especially poorly in the **road** and **building** regions and works generally poorly in resolution of 3D boundaries, which is particularly noticeable in the field dataset. Still, the algorithm showed some ability to resolve medium sized, reasonably textured objects (e.g., car, trees), which may be useful for certain tasks (e.g., obstacle detection).

The particular evaluated *CTF* algorithm has been designed to include adaptive windowing for matching in the vicinity of 3D object boundaries and has been implemented with real-time performance [20, 19]. In the current evaluation it has shown to yield accurate and high density estimates, with the exception of low texture regions as found in the sky of **background** and the interior of **car** during rear end viewing (datasets 000 - 002b). Significantly, *CTF* also showed similar, relatively strong performance on the field data.

SGM has been successfully applied in driving scenario evaluations before [26, 22, 14], and the present work verifies its promise in this application domain. It displays the highest accuracy with very little error spread both overall and in virtually all scene categories separately, except for the most challenging **road** regions. Importantly, estimation density is high and very close to the lower computational complexity [20] *CTF* alternative. Furthermore, results on field data are of very good quality in comparison to other methods. Finally, the algorithm has reasonably low complexity and real-time implementations have appeared [9, 10]. These characteristics combine to make *SGM* a very good candidate algorithm for vision-based 3D estimation in driving applications.

GC is a global stereo method with some claims to best overall general purpose performance [17, 6, 23]. In the context of the current application domain and evaluation, however, it is outperformed by *SGM*, which has shown to provide superior accuracy results with higher density. It appears that *GC*'s performance has been compromised particularly by a tendency to oversmooth its disparity estimates, particularly noticeable on the field data, even though an attempt was made to hand-optimize its parameters for the current evaluation. Moreover, *GC* has not yet yielded to real-time implementation, a significant disadvantage for driving applications.

RG is expected to yield sparser, but accurate and stable results by design. While the disparity maps returned by *RG* are indeed the sparsest of all 5 methods, estimation accuracy is never the best of all, except for the hardest **road** category, where *RG* performs especially well at the expense of density. The latter fact suggests that there may be use to some of the design principles implemented in *RG*. The estimation density tradeoff is even more dramatic for the field results depicted in Fig. 4.6, as extremely high proportion of points are reported as invalid; in fact, only disparities for the side trees are computed reliably, while road and car are virtually undetected, except when the car is very close to the camera. Further, even though *RG* is based on local matching, the approach's reliance on propagation of highly confident matches to other regions restricts its ability to be highly parallelized, which may limit real-time realizations.

Beyond comparison of algorithms, it is interesting to consider how the current results can more generally shed light on the relative difficulty of various driving relevant scene characteristics. The **car** region is of primary importance and good performance is essential. Although *SGM* is the overall winner in accuracy and density, other algorithms exhibit acceptable performance. Apparently, even though cars are made of highly reflective materials, they possess enough visible structure and silhouette to be correctly recovered in depth. The **road** regions constitute another highly driving relevant category. Here, it has been found that these regions are hardest to recover due to their relatively low texture (excepting lane markings). The fact that *RG* performed relatively well in these regions suggests that use of high confidence matches to constrain a direct ground plane fit may provide a viable approach to dealing with such regions. The **background** regions have a simple frontoparallel structure that is well suited to contemporary stereo matching techniques, especially the smoothing embodied in (semi)global matching algorithms *SGM* and *GC*. The **building** region can be perceived as a step up in challenge from the **background**, as it contains planar, but not frontoparallel structures. The challenge is immediately reflected in error rates, especially in indicat-

ing *NCC* as the weakest performer; whereas, all other approaches are comparably able to respond to the increase in difficulty. The **trees** regions are characterized by moderate to small sized objects with texture. Such characteristics are compatible with contemporary stereo algorithm capabilities and the relatively low and comparable error rates support this claim. Still, the error rates are higher than for **background** regions, due to the presence of complex 3D boundaries. Therefore, emphasis on accurate boundary recovery will aide in good recovery in such regions.

Finally, the current scene segmentation into object categories offers suggestions not only on how to improve the estimation algorithms, but also on how to design more appropriate testing scenarios. For example, it is important to consider cars with back views and side views as performance for some algorithms can differ significantly; it is beneficial to concentrate on **building** rather than simpler **background** regions; scenes should have considerable **road** regions of different surface cover and configurations.

Chapter 6

Conclusion

The results of the evaluation indicate the relative advantages and disadvantages of the various algorithms in driving relevant scenarios, including consideration of accuracy and density of estimates. Of the algorithms evaluated, the semiglobal matching algorithm [12] appears particularly well suited to driving applications in providing accurate, dense estimates with potential for real-time performance [9, 10]. The other algorithms considered exhibited various strengths and weaknesses, e.g., *GC* and *RG* arguably showed the second best overall accuracy, albeit with compromised density (esp. *RG*) and less real-time potential; *CTF* showed high density and has real-time instantiation [19], albeit with somewhat compromised accuracy; *NCC* had arguably overall weakest performance. Finally, it is noteworthy that the qualitative results presented on a field data set parallel those found quantitatively on the scale model, laboratory dataset, which supports the usefulness of the laboratory dataset in evaluation of vision algorithms for driving applications.

Bibliography

[1] www.cse.yorku.ca/∼sizints/ivs2013-supplementary.wmv. 11, 17

[2] Simon Baker and Ian Matthews. Lucas-Kanade 20 years on: A unifying framework. *IJCV*, 56(3):221–255, 2004. 7

[3] Chaminda Basnayake. Positioning for driver assistance. *GPSWorld Magazine*, April 2009. 2

[4] M. Bleyer and S. Chambon. Does color really help in dense stereo matching? In *3DPVT*, 2010. 3, 10

[5] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *TPAMI*, 23(11):1222–1239, 2001. 9, 10, 11

[6] Myron Z. Brown, Darius Burschka, and Gregory D. Hager. Advances in computational stereo. *TPAMI*, 25(8):993–1008, 2003. 9, 11, 21

[7] Jan Cech and Radim Sara. Efficient sampling of disparity space for fast and accurate matching. In *Proc. BenCOS Workshop CVPR*, 2007. 9, 10, 11, 14

[8] EISATS. Image sequence analysis test site, http://www.mi.auckland.ac.nz/eisats, 2011. 3, 4

[9] Ines Ernst and Heiko Hirschmller. Mutual information based semi-global stereo matching on the GPU. In *ISVC*, pages 228–239, 2008. 20, 23

[10] Stefan K. Gehrig and Clemens Rabe. Real-time semi-global matching on the CPU. In *ECVW*, 2010. 20, 23

[11] Minglun Gong, Ruigang Yang, Liang Wang, and Mingwei Gong. A performance study on different cost aggregation approaches used in real-time stereo matching. *IJCV*, 75(2):283–296, 2007. 3

[12] Heiko Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *CVPR*, volume 2, pages 807–814, 2005. 9, 10, 11, 23

[13] Heiko Hirschmuller and Daniel Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *TPAMI*, 31(9):1582–1599, 2009. 3, 7, 10

[14] Reinhard Klette, Norbert Kruger, Tobi Vaudrey, Karl Pauwels, Marc van Hulle, Sandino Morales, Farid Kandil, Ralf Haeusler, Nicolas Pugeault, Clemens Rabe, and Markus Lappe. Performance of correspondence algorithms in vision-based driver assistance using an online image sequence database. *IEEE Trans. Vehicular Technonlogy*, 2011. 2, 3, 20

[15] Maxime Lhuillier and Long Quan. Match propagation for image-based modeling and rendering. *TPAMI*, 24(8):1140–1146, 2002. 14

[16] Middlebury College Stereo Vision Page. http://www.middlebury.edu/stereo/, 2008. 7

[17] D. Scharstein and R. Szeliski. Taxonomy and evaluation of dense two-frame stereo algorithms. *IJCV*, 47:7–42, 2002. 3, 7, 9, 10, 18, 21

[18] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR*, volume 1, pages 195–202, 2003. 7

[19] Mikhail Sizintsev, Sujit Kuthirummal, Supun Samarasekera, Rakesh Kumar, Harpreet Sawhney, and Ali Chaudhry. GPU accelerated realtime stereo for augmented reality. In *3DPVT*, 2010. 10, 11, 20, 23

[20] Mikhail Sizintsev and Richard P. Wildes. Coarse-to-fine stereo vision with accurate 3D boundaries. *IVC*, 28(3):352–366, 2010. 9, 20

[21] Gideon P. Stein, Y. Gdalyahu, and Amnon Shashua. Stereo-assist: Top-down stereo for driver assistance systems. In *IVS*, 2010. 3

[22] Pascal Steingrube, Stefan K. Gehrig, and Uwe Franke. Performance evaluation of stereo algorithms for automotive applications. In *ICVS*, pages 285–294, 2009. 2, 3, 20

[23] Marshall F. Tappen and William T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In *ICCV*, pages 900–907, 2003. 9, 21

[24] Federico Tombari, Stefano Mattoccia, and Luigi Di Stefano. Stereo for robots: Quantitative evaluation of efficient and low-memory dense stereo algorithms. In *ICARCV*, 2010. 3

[25] J. Tukey. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA, 1977. 16

[26] Tobi Vaudrey, Clemens Rabe, Reinhard Klette, and James Milburn. Differences between stereo and motion behavior on synthetic and real-world stereo sequences. In *IVCNZ*, pages 1–6, 2008. 3, 7, 20