Belief Propagation Stereo Matching Compared to iSGM on Binocular or Trinocular Video Data

Waqar Khan[1], Verónica Suaste[2], Diego Caudillo[2], and Reinhard Klette[1]
[1]Tamaki Innovation Campus, The University of Auckland, New Zealand
[2] CIMAT, Guanajuato, Mexico
wkha011@aucklanduni.ac.nz

***Abstract*— The report about the ECCV 2012 Robust Vision Challenge summarized strengths and weaknesses of the winning stereo matcher (Iterative Semi-Global Matching = iSGM). In this paper we discuss whether two variants of a Belief Propagation (BP) matcher can cope better with events such as sun flare or missing temporal consistency, where iSGM showed weaknesses (similar to the original SGM). The two variants are defined by the use of a census data cost function on a 5×5 window and either a linear or a quadratic truncated smoothness function. An evaluation on data with ground truth showed better results for the linear smoothness function. The BP matcher with the linear smoothness function provided then also better matching results (compared to iSGM) on some of the test sequences (e.g. images with sun flare). The third-eye-approach was used for the performance evaluation.**

I. INTRODUCTION

The ECCV 2012 Robust Vision Challenge[1] provided real-world stereo sequences for a comparative evaluation of stereo matchers. The report [12] summarized strengths and weaknesses of the winning stereo matcher iSGM, which stands short for *Iterative Semi-Global Matching* [7]. In terms of [11], this report identified *situations* which are unsatisfactorily solved by iSGM. In particular, it mentions "streaking/noise and similar to original SGM [8], bad in sun flare, temporal consistency". Figure 1 illustrates a result for sun flare.

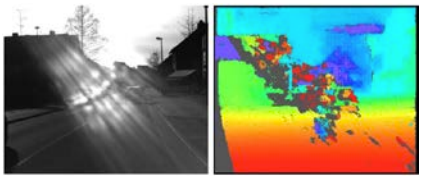

Fig. 1. Sun flare issue of iSGM as identified in [12].

In this paper we discuss whether two *Belief Propagation* (BP) matchers, characterized by either linear (linBP) or quadratic (quaBP) smoothness terms, can cope better with such situations.

The paper is structured as follows. Section II specifies briefly the used stereo matchers. Section III illustrates the performance on stereo data with ground truth (Set 2 of EISATS [2] and KITTI data [10]), and the 'appealing performance' of the census cost function on those data (see also [5], [9]). Thus, we use uniformly the census cost function in iSGM as well as in linBP and quaBP. Section IV applies then the *third-eye-approach* of [13] for evaluating the performance of those three stereo matchers on real-world data which do not come with ground truth. For this evaluation method it is required to have at least trinocular stereo data, as, for example, provided in Set 9 of EISATS [2]. Unfortunately, the ECCV 2012 Robust Vision Challenge did not provide trinocular stereo data. Thus, the paper is limited on the use of data from Set 9 of EISATS; however, this set contains one sun flare sequence and also one wiper sequence, which are of particular interest for this study. Section V introduces the novel subject of correlating video data measures with challenging events in video sequences, and Section VI concludes.

[1] hci.iwr.uni-heidelberg.de/Static/challenge2012/

II. STEREO MATCHING ALGORITHMS

We use pairs of rectified stereo images L (left or base images) and R (right or match image) of size $M \times N$ of time-synchronized video sequences. Let $\Omega = \{(x, y) : 1 \leq x \leq M \wedge 1 \leq y \leq N\}$ be the set of all pixels in an image L or R. A disparity $f_p = x_L - x_R$ is the off-set in row direction in an epipolar line y, for $p = (x_L, y)$. A dense disparity field defines a labelling f on Ω. Let d_{max} be the maximum disparity between base and match images.

A. Data Cost Functions for Stereo Matching

Cost function for stereo matching have been discussed, for example, in [5], [9]. We will refer later to common cost functions without defining them here. However, because we decided to focus on the use of the *census cost function*, we provide its definition here. This cost function is based on the *census transform* [16].

The census transform uses intensity differences within a window around a pixel (x, y). For simplicity, assume a window of size $(2k+1)^2$ in an image I. The census transform maps pixel intensities in the window into a bit string $\xi(x, y)$ as follows:

$$\xi(x,y) = \bigsqcup_{i=-k}^{k} \bigsqcup_{j=-k}^{k} \varrho\left(I(x,y), I(x+i, y+j)\right)$$

where $\bigsqcup$ denotes the bit concatenation, and

$$\varrho\left(u, v\right) = \begin{cases} 1 & \text{if u} > \text{v} \\ 0 & \text{if u} \leq \text{v} \end{cases}$$

The cost function C_{census} is then the Hamming distance between the bit string $\xi_L(x,y)$ in the base image and the bit string $\xi_R(x+\Delta,y)$ in the match image, defined by off-set Δ.

B. Iterative Semi-Global Matching

Semi-global matching (SGM) was proposed in [8]. In the sequel, various modifications have been published with respect to search strategy, used cost functions, or build-in smoothness constraint. For example, the program iSGM [7] introduced iterations into the SGM scheme.

We briefly recall the basic SGM strategy. The cost for pixel correspondences between base and match images is first derived for all possible disparities $0 \leq \Delta \leq d_{max}$. Using dynamic programming, the correspondence cost is computed along a set (usually of two to eight) scanlines of different directions. This leads to the computation of final disparities based on a the-winner-takes-all evaluation.

Costs are accumulated along an oriented scanline, thus defining the dense labelling function f for all disparities. For a *partial energy function* $E_s(f)$ at stage s, disparities need to be assigned at the previous stages first. This is defined recursively as follows:

$$E_s(f) = D_s(\Delta) + M_s - \min_{0 \leq \Delta \leq d_{max}} E_{s-1}(f) \qquad (1)$$

where $D_s(\Delta)$ is the matching data cost between pixels in L and R defined by off-set Δ. For the data cost we decide for using the census cost function on a 3×9 window as specified for iSGM in [7]. Term M_s is defined as follows:

$$M_s = \min \begin{cases} E_{s-1}(f) \\ E_{s-1}(f-1) + P_a \\ E_{s-1}(f-1) + P_a \\ \min\limits_{0 \leq \Delta \leq d_{max}} E_{s-1}(f) + P_b \end{cases} \qquad (2)$$

where P_a and P_b are regularization penalties for enforcing disparity consistency along a scan-line, basically only defining a 'two-step' Potts model (i.e. penalty P_a if disparity difference 1 at an adjacent pixel along the path, or penalty P_b if the difference exceeds 1). Input parameters are penalties P_a and P'_b, and P_b is derived as follows:

$$P_b(x,y) = \max\left\{\frac{P'_b}{|L(x-1,y)-L(x,y)|}, P_a + \kappa\right\} \qquad (3)$$

where $\kappa > 0$. As P_b is derived from image data differences, and assuming that depth discontinuities 'usually' occur at intensity discontinuities, P_b aims at improving the matching performance at such places. For further details, in particular for a description of the used program iSGM in our experiments, see [7].

C. Belief Propagation

The use of belief propagation (BP) for stereo matching was popularized by [3]. Disparities between 0 and $d_{\max}$ are *labels* in this context, and a labelling function f is updated in the array of size $M \times N$. Each pixel has assigned belief values for any of the possible disparities, and influences the decisions of the adjacent pixels (in the next iteration) by its belief values. Figure 2 illustrates the propagation of belief values assuming 4-adjacency for the underlying belief propagation network.

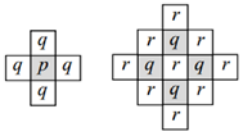

Fig. 2. Belief propagation: p influences pixels q in one step, and thus (indirectly) also pixels r in the next step

Labels are assigned by energy (cost) minimization for the generated labelling function f. The cost is defined by data cost $E_d(p)$ at a pixel p (which was called D_s above for SGM) in the base image and smoothness (or discontinuity) cost E_s aiming at spatial consistency of the generated labels. Assume $m^t_{p \to q}$ to be the message from pixel p to its adjacent pixel q at iteration t (see Figure 2, left):

$$m^t_{p \to q}(f_q) = \min_{0 \leq l_p \leq d_{max}} \{(E_s(f_p, f_q) + H_{p,\,q}(f_p))\} \qquad (4)$$

where

$$H_{p,q}(f_p) = E_d(f_p) + \sum_{r \in A(p) \setminus \{q\}} m^{t-1}_{r \to p}(f_p) \qquad (5)$$

and $A(p) \setminus \{q\}$ denotes adjacent nodes of p excluding q (see Figure 2, right).

An optimum labelling f would minimize

$$\sum_{p \in \Omega} E_d(f_p) + \sum_{(p,q) \in A} E_s(f_p, f_q) \qquad (6)$$

where A is the chosen adjacency relation (such as 4-adjacency). If BP would run such that every pixel communicates with every pixel, the global minimum could be achieved. However, there are time limits, and practically we can only use a limited number of iterations, possibly combined with a hierarchical or alternating strategy for being more efficient.

Parameter	Value
Iterations t	24
Pyramid levels	1
Data cost function	Census
Cost function window (width $\times$ height)	5×5
Scaling factor b of discontinuity function	1.0
Truncation value c	$d_{max}/8$

TABLE I

CONFIGURATION PARAMETERS FOR BELIEF PROPAGATION

Our BP implementation follows [1]. However, for E_d we decide for a 5×5 census data cost function. For the smoothness term E_s we use two options, a truncated linear function (defining program linBP) or a truncated quadratic function (defining program quaBP). For truncation, we use $d_{max}/8$ in both cases. Table I summarizes the used BP parameters for programs linBP and quaBP.

III. VIDEO SEQUENCES WITH GROUND TRUTH

Matching accuracy can be determined by the absolute difference between computed and ground truth disparity, assuming that a 'reasonable' ground truth is available. An accuracy measure (i.e. an error) is defined by the percentage of pixels in Ω where ground truth assigns a positive disparity, but the difference to the assigned disparity exceeds a threshold τ, such as 1 or 2.

A. Set 2 of EISATS

Set 2 of the EISATS benchmark website (by Daimler A.G. and the *.enpeda..* group Auckland, see [15]) provides synthetic stereo sequences with accurate ground truth. From this set we discuss experiments with Sequence 2 in this paper. By being a synthetic sequence, the capturing environment is ideal, without any lighting variations between images in a stereo pair. The images are of resolution $M \times N = 640 \times 480$, with $d_{max} = 58$ pixels.

1) linBP vs quaBP: We use the first 90 images of this sequence for a comparative evaluation of linBP versus quaBP, with $\tau = 2$. See Figure 3. Clearly, linBP outperforms quaBP for the selected data. Thus, we only discuss linBP in the sequel in this brief paper.

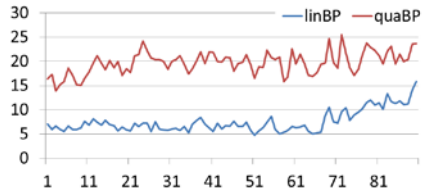

Fig. 3. linBP vs. quaBP: % of pixels with error > 2 on Sequence 2

2) Data cost functions for linBP: We use linBP to evaluate various data cost functions on a 5×5 window including the absolute difference (ABS), Birchfield-Tomasi (BT), and census. Figure 4 shows (as a typical example) the percentage of error pixels for $\tau = 2$ pixels on the first 20 images of the considered sequence. This is just a brief re-confirmation of the discussion in [5], [9] for this synthetic sequence.

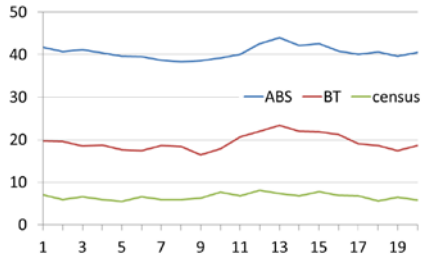

Fig. 4. Data cost functions: % of pixels with error > 2 on Sequence 2

B. The KITTI Dataset

These real-world sequences come with ground truth provided by a laser range-finder [10], [4]. Shown scenes have high complexity and there are other challenges like lighting variations among stereo pairs. The training dataset consists of

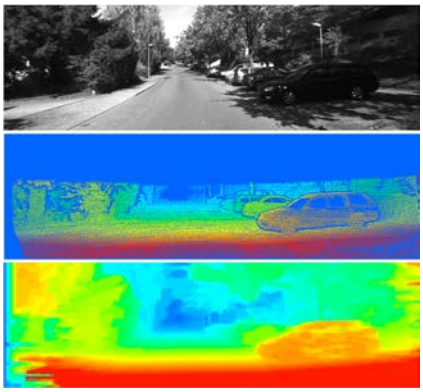

Fig. 5. KITTI training data example. Top: reference image, middle: ground truth, bottom: linBP disparity map

194 images, where ground truth ($d_{max} = 255$) is provided with or without occlusion. For an example of the KITTI dataset, see Figure 5. It shows colour-coded disparities for linBP and the ground truth disparity map.

1) Data cost functions for linBP: We repeat the experiment for linBP as specified before for the synthetic sequence. We show results for the first 20 KITTI training stereo frames, compared with no-occlusion ground truth data and $\tau = 2$; see Figure 6. This again shows that census has the best matching accuracy compared to other two cost functions, especially its robustness with respect to brightness differences.

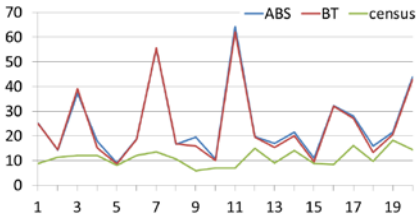

Fig. 6. Data cost functions: % of pixels with error > 2, KITTI training data

2) linBP vs. iSGM: Using the census cost function for both, we compare linBP with iSGM using the complete training dataset with non-occluded ground truth disparities. Figure

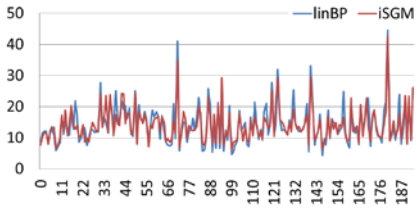

Fig. 7. linBP vs iSGM: % of pixels with error > 2 on KITTI training data

7 shows this comparison for $\tau = 2$. For this data set and $\tau = 2$, iSGM outperforms linBP with respect to the mean error. See Table II for mean error and standard deviation. However, this order changes for $\tau = 3$. Errors are computed using the KITTI stereo development kit.

To further analyse the matchers behaviour over the whole sequence, we compute the winning count: a matcher having

	Mean (%)		Standard deviation (%)	
	iSGM	linBP	iSGM	linBP
2 pixel	12.22	12.45	5.23	6.02
3 pixel	9.40	9.23	4.41	5.14

TABLE II

LINBP VS ISGM: COMPLETE KITTI TRAINING DATASET

lesser error % wins for that image. For $\tau = 2$, surprisingly, the winning count is 99 for linBP, and 95 for iSGM. These numbers show that linBP is winning more often, even though its mean error % is higher than that of iSGM. Because linBP also has a higher standard deviation than iSGM, this suggests that the iSGM % error is more often closer to its mean error. Whereas, the linBP error is more often below the mean, but to nullify that, for a few images the error is much higher than the mean. This explanation makes sense to us, as occasionally, BP message passing fails to propagate in certain regions (e.g. across 'strong' intensity edges) leading to more errors in disparity matching.

The winning count for $\tau_t = 3$ further develops to 124 for linBP and 70 for iSGM.

Table III shows the comparison of linBP with iSGM on the KITTI test dataset. Out-Noc is % of erroneous pixels in non-occluded areas, and Out-All is % of erroneous pixels in total.[2] Avg-Noc is the ratio "average disparity / end-point error" in non-occluded areas (1.2 for iSGM and 1.7 for linBP), and Avg-All is the ratio "average disparity / end-point error" in total (2.1 for iSGM and 2.7 for linBP.

	Out-Noc (%)		Out-All (%)	
	iSGM	linBP	iSGM	linBP
2 pixel	8.04	11.78	10.09	13.87
3 pixel	5.16	8.66	7.19	10.81
4 pixel	3.87	7.06	5.84	9.22
5 pixel	3.15	6.02	5.03	8.18

TABLE III

EVALUATION ON KITTI TEST IMAGES FOR LINBP AND ISGM

From those summarizing numbers it appears that iSGM outperforms linBP on those real-world data. Because the ground truth of KITTI test dataset is unavailable to the end-users, it is not possible to compute the winning count on this dataset.

IV. VIDEO SEQUENCES WITHOUT GROUND TRUTH

The KITTI data do not come with challenges such as sun flare, rain, wipers, or low light. We use the *third-eye-approach*, as defined in [13], for evaluating the quality of stereo matchers on real-world video data for the common case that disparity ground truth is actually not available. This requires trinocular recording (at least). The ECCV 2012 Robust Vision Challenge did not provide trinocular stereo data. However, there are trinocular test sequences available, for example in Set 9 on EISATS [2].

[2] With these results, linBP would have ranked 13th at [10] on 5^{th} Jan, 2012, while iSGM on second place.

A. The Third-eye-approach

Recording is with three calibrated and time-synchronized cameras. Two cameras provide the base (or left) and match (or right) images L and R for stereo analysis, respectively, and the third camera provides image T which is used for evaluation (and could also be used for trinocular stereo analysis).

Calculated disparities and calibration data allow us to map the base image L into a virtual image V at the pose of the third camera. Due to the geometric transform and due to occlusions, some pixel values in the virtual image V remain undefined. Because there might be brightness issues between images L and T, we are using the *normalised cross correlation* (NCC) for comparing defined image values in V with those at the same pixel location in T, at any time t (only specified in the following formula for the set Ω_t of pixel locations (x, y) with defined values in V) of the recorded sequence:

$$S_{NCC}(t) = \frac{1}{|\Omega_t|} \sum_{(x,y)\in\Omega_t} \frac{[T(x,y) - \mu_T][V(x,y) - \mu_V]}{\sigma_T \sigma_V}$$

Symbols μ and σ denote mean and standard deviation of the corresponding images, and $|\Omega_t|$ is the cardinality of this set. This *similarity measure* S_{NCC} equals 1 in case of absolute identity, and decreases in magnitude with the ratio of differences between V and T at positions in Ω_t. For reducing the effect of mismatches within homogeneously textured regions in L (or T), paper [6] also suggested to use a modified measure, where Ω_t only contains pixel locations which are in distance 10 or less to an edge pixel in L. This defines similarity measure $S_{NCCmask}$.

B. Set 9 of EISATS

We are now testing iSGM and linBP on Set 9 of EISATS [2]. This set contains eight trinocular sequences, each of 400 stereo frames, grey-level images of size $M \times N = 640 \times 480$, recorded at 10 bit per pixel [6], [14]. Sequence are 'Barriers' (crossing a bridge with road blocks on one side, very close to the ego-vehicle), 'Dusk' (a sun flare sequence), 'Harbor bridge' (same bridge but without road blocks), 'Midday' (hilly suburban road), 'Night' (with dense traffic), 'People' (pedestrians crossing in front of the ego-vehicle), 'Queen street' (driving along a CBD street), and 'Wiper' (same scene as in 'Midday', but now with running wipers).

C. Similarity Measures

We analysed the information provided by similarity measures S_{NCC} and $S_{NCCmask}$, and also a simple sum-of-squared-differences (SSD) comparison between virtual and third images. The SSD measure appears to be of not much significance for understanding the quality of stereo matchers when using this third-eye-approach. When comparing S_{NCC} with $S_{NCCmask}$, it appears in general that obtained confidence results with both measures are 'well correlated' with visually evaluated accuracy of stereo matching. Measure $S_{NCCmask}$ provides a more detailed analysis, with more 'valleys' in its curve indicating difficulties (where designers of a stereo

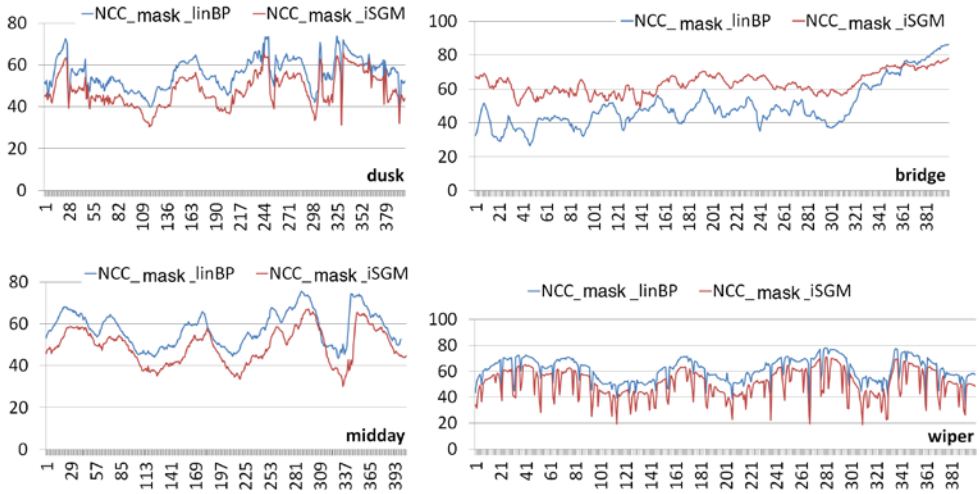

Fig. 9. Comparison of iSGM with linBP on four of the sequences of Set 9. iSGM performs better on Bridge, but linBP wins on Dusk, Midday, and Wiper

matcher need to check for reasons and possible solutions). See Fig. 8 for both functions for the sequence Midday when using iSGM, and for sequence Dusk when using linBP. In conclusion, we decide to use measure $S_{NCCmask}$ only in the sequel.

D. NCC_mask Comparison of iSGM with linBP

Figure 9 shows four graphs for discussing the performance of iSGM and linBP on four sequences of Set 9 of EISATS, showing functions $S_{NCCmask}$ only for both stereo matchers. For many of the stereo frames in those eight sequences of Set 9, we visually evaluated that iSGM is the better algorithm compared to linBP due to 'appealing accuracy' at occlusion edges. However, when summarizing by using the $S_{NCCmask}$ measure, surprisingly the opposite appears to be the dominating event (i.e. linBP appears to win more frequently).

V. DATA MEASURES

The third-eye-approach might be possibly avoided by correlating the quality of stereo matchers directly with data measures on stereo videos. For example, let sigma_left be the standard deviation of the base image L, NCC_leftright the NCC between L and match image R, and sigma_Sobel the standard deviation of the Sobel edge values of L. How do such measures relate to the performance of stereo matchers on challenging real-world data?

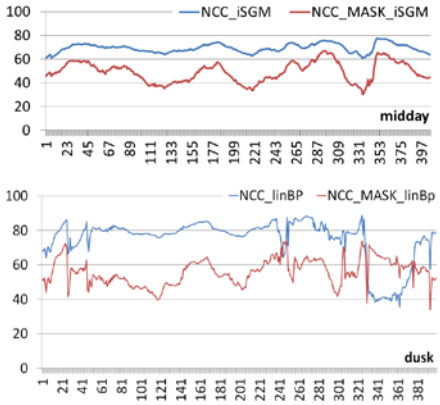

Fig. 8. Functions S_{NCC} and $S_{NCCmask}$ for iSGM on sequence Midday, and for linBP on sequence Dusk

For comparing functions we unify their means and variances, taking the mean and variance of function $S_{NCCmask}$ as uniform goal for all the other functions. Figure 10 shows plots of functions $S_{NCCmask}$ for iSGM on the sequences in Set 9 of EISATS together with normalized plots of those three data measures listed above.

We also calculate L_1-distances $d_1(f, g)$ between normalized functions f and g by taking $\sum |f(i) - g(i)|$ for all frames i in the sequence and dividing by the number of frames.

Questions are now, for example, whether a low or high value of $d_1(S_{NCCmask}, g)$, where g is one of the three data measures, does have a particular meaning for analysing stereo matchers. A low value indicates that the measure might replace the third-eye-approach on this particular sequence.

Table IV shows the distance value d_1 between the three data measure functions and the NCC-mask function for those five sequences where values were 'reasonably' small. Indeed, the plots for sequences Dusk, Midday, and Queen in Fig. 10 also indicate a 'good' correlation between $S_{NCCmask}$ and data measures. We consider this as a starting point for doing further in-depth analysis for reasons of degrading behaviour of stereo matchers.

	Dusk	Midday	Queen	Wiper	People
sigma_left	3.69	1.52	3.78	5.16	6.90
NCC_leftright	4.18	1.26	3.33	4.59	5.21
sigma_Sobel	5.79	2.30	3.29	5.42	5.80

TABLE IV

DISTANCE VALUES OF NORMALISED DATA MEASURES TO $S_{NCCmask}$

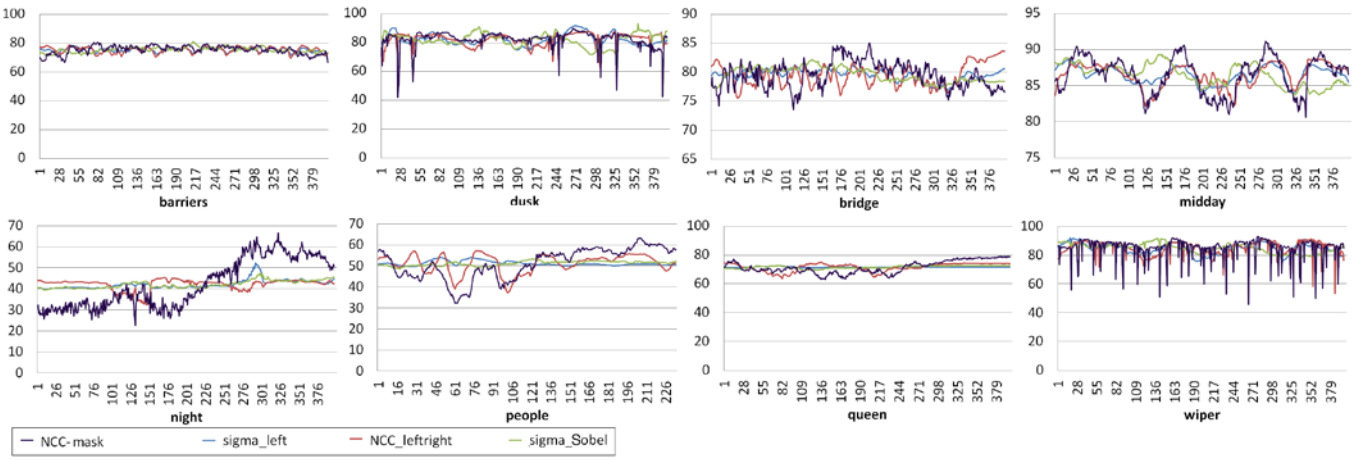

Fig. 10. Comparison of NCC_mask of iSGM on Set 9 of EISATS with NCC_mask-normalized functions sigma_left, NCC_leftright, and sigma_Sobel

VI. CONCLUSIONS

The results came partially as a surprise to us. However, there is also a good logic behind. BP is propagating belief uniformly, without taking particular attention to dominating horizontal propagations or accuracy along vertical disparity discontinuities, as iSGM does. But the accuracy within object regions, away from disparity discontinuities, appears to be better with linBP. For example, in the sequence Bridge we do have many structural details and not such a high ratio of homogeneous regions - accordingly, iSGM is winning here. In the Wiper sequence, the NCC_leftright data measure indicates correctly the actual input data situation, corresponds to the performance of the stereo matchers, and BP is able to cope better with those data (see also already [14]).

For testing, it would be of benefit to have trinocular data with ground truth for challenging situations. This would allow to correlate ground truth, similarity measures, and data measures more in detail. We assume that these are useful tools for identifying critical events for stereo matchers which need to be resolved in future.

ACKNOWLEDGMENT

The authors thank Simon Hermann for providing program iSGM for the performed experiments and discussions while writing the paper. The authors also thank Sandino Morales for providing his sources for the third-eye-approach.

REFERENCES

[1] C. C. Cheng, C. K. Liang, Y. C. Lai, H. H. Chen, and L. G. Chen. Fast belief propagation process element for high-quality stereo estimation. In Proc. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 745–748, 2009.

[2] EISATS benchmark data base. The University of Auckland, `www.mi.auckland.ac.nz/EISATS`, 2013.

[3] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *Int. J. Computer Vision*, volume 70, pages 41–54, 2006.

[4] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? The KITTI Vision Benchmark Suite. In Proc. *IEEE Int. Conf. Computer Vision Pattern Recognition*, pages 3354–3361, 2012.

[5] S. Hermann, and R. Klette. The naked truth about cost functions for stereo matching. MI-tech report 33, The University of Auckland, `www.mi.auckland.ac.nz/tech-reports/MItech-TR-33.pdf`, 2009.

[6] S. Hermann, S. Morales, and R. Klette. Half-resolution semi-global stereo matching. In Proc. *IEEE Intelligent Vehicles Symposium*, pages 201–206, 2011.

[7] S. Hermann and R. Klette. Iterative semi-global matching for robust driver assistance systems. In Proc. *Asian Conf. Computer Vision*, LNCS, 2012.

[8] H. Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In Proc. *IEEE Int. Conf. Computer Vision Pattern Recognition*, volume 2, page 807–814, 2005.

[9] H. Hirschmüller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Analysis Machine Intelligence*, volume 31, pages 1582–1599, 2009.

[10] KITTI vision benchmark suite. Karlsruhe Institute of Technology, `www.cvlibs.net/datasets/kitti/`, 2013.

[11] R. Klette, N. Krüger, T. Vaudrey, K. Pauwels, M. Hulle, S. Morales, F. Kandil, R. Haeusler, N. Pugeault, C. Rabe and M. Lappe. Performance of correspondence algorithms in vision-based driver assistance using an online image sequence database. *IEEE Trans. Vehicular Technology*, volume 60, pages 2012–2026, 2011.

[12] D. Kondermann and W. Niehsen. 2012 robust vision challenge. `hci.iwr.uni-heidelberg.de/Static/challenge2012/RVC2012_Results.pdf`, 2012.

[13] S. Morales and R. Klette. A third eye for performance evaluation in stereo sequence analysis. In Proc. *Int. Conf. Computer Analysis Images Patterns*, pages 1078–1086, 2009.

[14] K. Schauwecker, S. Morales, S. Hermann, and R. Klette. A comparative study of stereo-matching algorithms for road-modeling in the presence of windscreen wipers. In Proc. *IEEE Intelligent Vehicles Symposium*, pages 7–12, 2011.

[15] T. Vaudrey, C. Rabe, R. Klette, and J. Milburn. Differences between stereo and motion behaviour on synthetic and real-world stereo sequences. In Proc. *Int. Conf. Image Vision Computing New Zealand*, pages 1-6, 2008.

[16] R. Zabih, and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In Proc. *IEEE Int. European Conf. Computer Vision*, pages 151–158, 1994.