Performance of Correspondence Algorithms in Vision-Based Driver Assistance using EISATS

Reinhard Klette[1], Norbert Krüger[2], Tobi Vaudrey[1], Karl Pauwels[3], Marc van Hulle[3], Sandino Morales[1], Farid I. Kandil[4], Ralf Haeusler[1], Nicolas Pugeault[5], Clemens Rabe[6], and Markus Lappe[4]

[1] The University of Auckland, New Zealand
[2] The University of Southern Denmark
[3] Katholieke Universiteit Leuven, Belgium
[4] University of Münster, Germany
[5] University of Surrey, United Kingdom
[6] Daimler Research, Germany

Abstract

The report discusses options for testing correspondence algorithms in stereo or motion analysis, designed or considered for vision-based driver assistance. It introduces a globally available database, with a main focus on testing on video sequences of real-world data. The authors suggest the classification of recorded video data into *situations*, defined by a co-occurrence of some *events* in recorded traffic scenes. About 100 to 400 stereo frames (or 4 to 16 seconds of recording) are considered to be a *basic sequence*, to be identified with one particular situation.

Future testing is expected to be on data that is reporting on hours of driving; multiple hours long video data may be segmented into basic sequences, and classified into situations. The report prepares for this expected development.

The report uses three different evaluation approaches (prediction error, synthesized sequences, labeled sequences) for demonstrating ideas, difficulties, and possible ways in this future field of extensive performance tests in vision-based driver assistance; especially also for cases where ground truth is not available. The study shows that the complexity of real-world data does not support an identification of general rankings of correspondence techniques on sets of basic sequences showing different situations. It is suggested that correspondence techniques need to be adaptively chosen in real time using some type of statistical situation classifiers.

Keywords. Vision-based driver assistance, performance evaluation, stereo analysis, optical flow, motion analysis, video data, basic sequences, situations, ground truth.

1 Introduction

Driver assistance systems are active safety measures in modern cars, also developed for comfort, fuel economy, and so forth. Vision-based driver assistance systems have been implemented in commercial vehicles since the $1990s$, first time in form of a lane-departure warning system by Mitsubishi in 1995 [34]. There is an increasing demand in evaluating such sensor-based components of modern cars, similar to crash tests being performance measures for, e.g., mechanical components of cars.

1.1 Objectives and Motivation

The main objective of this work is to report about current work in testing correspondence algorithms (i.e., motion or stereo analysis techniques), on a large variety of test data under realistic recording conditions, as typically occurring in the driver assistance domain.[1] Such 'large scale' testing is not yet supported by popular benchmarks, as used in the computer vision community; those are characterized by well controlled data sets, and, as a

[1]This certainly also generalizes to other domains such as outdoor robotics, surveillance, human pose recognition, or for mobile platforms in outdoor applications in general. However, this report is only discussing evaluation in a context of driver assistance systems.

consequence, their value is only of limited relevance for application domains dealing with outdoor data such as occurring in driver assistance.

Evaluations of correspondence methods in general have already some history in the computer vision literature (see, for example, [3, 17, 35, 44, 47]), and have contributed to the current progress in the field of those algorithms, designed for spatial or temporal matching of image data. For evaluations especially in the context of vision-based driver assistance systems, see, for example, earlier work by the authors as in [27, 36, 37, 38, 42, 48, 49]. This report is summarizing but also extending work reported in those references. In particular, the authors would like to present the various opportunities in using publicly available test data on EISATS, the *.enpeda..* image sequence analysis test site; see [12].

1.2 Basic Terms

In this report, we introduce data sets that are selected or designed for testing vision-based driver assistance systems, together with a discussion of evaluation methods (when using those data) for stereo or motion analysis. Motion is typically described in computer vision by optic flow (i.e., the visual change of image intensities from frame to frame).

The *ego-vehicle* is the car in which the driver assistance system is operating; *ego-motion* is defined by changes of the ego-vehicle in position, tilt, roll, or yaw angles. Stereo or motion analysis are low-level tasks for understanding image data, and results of this analysis are used in subsequent higher-order process for understanding traffic scenes. As an example of such a higher-order process we consider in this report the segmentation of image data into regions where the optic flow is caused either by ego-motion (of the ego-vehicle) or by *independently moving objects*, such as other vehicles, pedestrians, or bicyclists.

The EISATS database in its current form already supports evaluation of stereo, motion, or segmentation algorithms. Provided stereo image video data are sequences of some length, typically of about 100 to 400 frames (or 4 to 16 seconds of recording, assuming 25 Hz as a current standard). These sequences are considered as representing particular *situations*. One way of defining a situation is by identifying a particular co-occurrence of *events* in recorded traffic scenes. Examples of events are activities of adjacent traffic (overtaking, oncoming traffic, crossing pedestrians, and so forth), weather and lighting conditions (rain, sun strike, patterns of shadow while driving

Figure 1: Frames from sequences showing three different situations: inner city at night (left), brightness differences (middle; a changing angle towards the sun occurs here), and close objects (right).

below trees, and so forth), road geometries (flat or curved, narrow lane, entering a tunnel, driving on a bridge, a speed bump, and so forth), or particular events such as traffic signs, a wet road surface or strong light reflections at night. For example, "driving in daylight on a planar road while overtaking a truck" is an example of a situation defined by a concurrent appearance of such events. Another way of defining a situation may be by statistical properties of image data; for example, see the *visual textures* in [1]. However, in this report we stay with the first option (i.e., situations defined by events).

Situations typically change every few seconds in normal traffic, and we consider 4 to 16 seconds as a reasonable length of a recorded video sequence (also called a *basic sequence*) to be identified with one particular situation.

Examples of situations are: default driving conditions, inner city traffic at night, brightness differences between both images in the stereo pair, illumination artifacts (e.g., sun through trees along the road), or close objects in front of the ego-vehicle. See Figure 1 for images visualizing three examples of situations; those are situations as discussed at some length in Section 5.

Future testing is expected to be on data that is reporting on hours of driving; multiple hours long video data may be segmented into basic sequences, and classified into situations. The report aims at preparing for this expected development.

We are interested in identifying the impact of particular events (i.e., of real-world outdoor issues) on the performance of correspondence algorithms in the context of a given situation. This allows us to recognize a particular challenge, often a particular task for research in correspondence algorithms,

and possibly also to propose ways to overcome the underlying problem, defined by this particular event.

We name a few examples of such challenges: (i) large disparities or motions across frames for image regions showing objects that are close to the ego-vehicle (e.g., caused by passing-by cars) but also due to large rotational motions of the ego-vehicle (i.e., also of the recording cameras), (ii) variations in illumination across frames[2] (e.g., slowly due to movements of clouds or a change in viewing angle relatively to the sun, or rapidly due to driving through a forested area, having the shade of leaves on cameras for fractions of a second), or (iii) motion blur in recorded frames. These are just three examples of real-world outdoor challenges. Such challenges often severely effect the complexity of the correspondence problem, to be solved for stereo or motion analysis.

1.3 Used Acronyms for Correspondence Algorithms

For stereo analysis, we discuss in this report (BP) loopy belief propagation stereo [14], (DP) dynamic programming stereo [39] with simple spatial (s) or/and temporal (t) propagation strategies [30], encoded by acronyms DPs, DPt, or DPst (note: those propagations could also be applied for the other matching algorithms), semi-global matching (SGM) stereo [21] with cost functions defined either by mutual information (SGM-MI) or (SGM-BT) the cost function introduced by Birchfield and Tomasi in [7], and (GC) graph-cut stereo matching [28].

Other correspondence algorithms could be considered as well, but this set provides a good selection of currently favored stereo matching approaches. Note that they represent three dominant design methodologies of stereo matching: optimization along linear paths of pixels (DP, SGM), graph cut optimization (GC), or optimization by belief propagation (BP).

Each stereo matching technique is defined by selected parameters such as weights in cost functions, or the size of used neighborhoods when defining cost values. It is not the intention of this report to evaluate one particular method in detail (e.g., by aiming at optimizing such parameters for a particular situation). The report likes to point out that matching methods behave very

[2] [22] studies cost functions in stereo algorithms under the particular aspect of brightness differences between stereo frames; see also [20] for a general study of cost functions in stereo algorithms.

Figure 2: Example of stereo input data (left and middle) and a disparity map obtained with SGM-BT (right). This disparity map is fairly useless in this particular case; this is for illustrating a 'bad performance'. Note that the cameras were also recording some reflections on the windscreen in this example of a situation that might be called 'brightness differences between left and right images, and an approaching truck'; the first event (brightness differences) causes the BT cost function to fail, the second event (approaching truck) should be solvable for the SGM matching strategy in principle when using a 'better' cost function for matching than BT.

differently for different situations recorded in a traffic context, and we define and illustrate ways how to obtain such evaluations.

The calculation of depth or disparity values (i.e., stereo analysis) is a basic step in the understanding of the surrounding environment of an ego-vehicle. A given stereo algorithm may report very confusing depth values if the input data has not been recorded under ideal conditions. For example, Figure 2, left and middle, presents a stereo pair in which there is a large difference in brightness between both images. The right image in this figure shows the output (i.e., depth map) of SGM-BT. Note that it is difficult to recognize any 3D structure of the scene. But this only shows that SGM-BT is not well suited for image data of this particular situation; possibly some preprocessing of the input data might actually change the performance to be better, or SGM-BT might perform 'much better' if there would be no brightness differences between left and right image? Actually, our studies have shown that SGM-BT is 'not a good choice' in general, due to the small neighborhood of contributing pixels even if there are no brightness differences, but in general also due to the inherent assumption of brightness constancy in this cost function.

We also introduce a few acronyms for the discussed motion analysis algo-

rithms. For computing optical flow, we will report about (PyrHS) Pyramid Horn-Schunck [23] (by extending the basic version of the Horn-Schunck algorithm in OpenCV with a pyramidal control structure), (BBPW) the algorithm described by four authors (identified by capitals B,B, P, and W) in [8], and (TVL1) a total-variation technique using an approximation of the L_1-metric [53]. Again, this selection is obviously not covering the whole diversity of currently discussed motion analysis algorithms, but the authors aimed at having a methodologically reasonable selection of different techniques for this report.

Where available, stereo analysis or optic flow sources have been downloaded either from author's websites and adapted by us, or fully implemented by us (partially also in contact with those authors); the source of basic HS was downloaded from OpenCV.

1.4 Stability and Robustness

Correspondence methods may be ranked for given situations based on their performance. *Stability* of one method is defined with respect to a particular situation, and a constantly good performance of this method for different basic sequences all showing this situation. *Robustness* of one method is defined by good performance on various situations.

The ranking of correspondence algorithms with respect to stability can be rather different for considered situations. A 'top performer' for one situation may not be robust within a given set of situations. For example, BP (which assumes intensity constancy in the data term of its cost function) is often a very good choice if there are no brightness differences in the stereo image data, but its results quickly degrade if there are lighting differences.

See Figure 3 for complete diagrams of performance values of considered stereo algorithms for some selected situations, for stereo image (sub-)sequences of a length between 80 to 140 frames. The used NCC measure will be specified later. However, with reference to those diagrams we already note that rankings may differ within one sequence from frame to frame, and the overall mean performance from sequence (i.e., situation) to sequence.

For example, the results indicate that SGM-MI may be called 'stable' on the *Brightness Difference* sequence, but it ranks low in general on the other four sequences. DPt ranks twice fairly high, on the *Ordinary Conditions* sequence and on the *Close Objects* sequence, and by analyzing the corresponding image data we noticed that this correlates to scenes where a high

Figure 3: Performance of stereo algorithms on five different situations, using a *normalized cross-correlation* (NCC) measure (a larger value is 'better') on trinocular sequences. Top row: Left: *Ordinary Conditions* sequence. Right: *Illumination Artifacts* sequence. Middle row: Left: *Inner City at Night* sequence. Right: *Brightness Difference* sequence. Bottom: *Close Objects* sequence.

percentage of pixels shows the surface of a planar road (and this corresponds on a theoretical level exactly to the underlying model of the used temporal propagation). Sometimes there is a strong correlation in the ups and downs of all the stereo algorithms (e.g., at frame 55 in the *Close Objects* sequence), and sometimes just particular algorithms fail (e.g., those depending on the intensity constancy assumption about at frame 110 in the *Brightness Difference* sequence).

Similar variations in performance may be noticed for motion analysis algorithms. Here, TVL1 proved to be superior in general; the other two considered motion analysis techniques were good for some situations only. TVL1 may be called robust, at least with respect to currently known motion analysis methods.

There was no 'clear winner' for stereo analysis algorithms for all the considered situations. Some studied algorithms may be called 'stable' for some situations, but we do not call any of the considered stereo matching algorithms 'robust' for some kind of extensive class of different situations. However, we only have studied a few different situations so far (what was already quite time-consuming), and it might be possible to identify such classes in future, for example by sharing more evaluation results for long stereo sequences as available on the EISATS database.

1.5 Contributions of this Report

As a technical contribution we introduce the EISATS data base [12] which provides appropriate data as well as evaluation tools for benchmarking and analysis of correspondence algorithms within a driver-assistance context, on a much larger scale and variation than currently used benchmarks which are still defined by small sets of test images. Since the visual information in the driver assistance domain is very different than the commonly used rather controlled stereo and optic flow benchmarks, their value is limited for practical applications in the driver assistance domain.

We develop tools and perform large scale benchmarking on long sequences with high variations. This allows us to characterize stability or robustness; something what was not discussed earlier for small sets of benchmark data. Moreover, there are many situations (e.g., overtaking a truck, or driving into a tunnel) in which established correspondence algorithms actually fail, meaning that their performance drops such that there is no way to ensure a qualitatively correct 3D scene analysis. This points to still existing funda-

mental problems in stereo processing that are not visible in currently used data bases.

As a further scientific contribution, our analysis shows that the optimal stereo algorithm in general depends on the actual situation, and hence some categorization of situations is needed, to assign whole classes of situations to particular correspondence algorithms. This defines a new direction of research.

The report is structured as following: First, the next section introduces the EISATS database as available by December 2010, which is structured into Sets 1 to 7. Next we prepare for stereo and motion performance evaluation; at first we define error measures, also proposing new summarizing methods for error measures on image sequences, and in the next section prediction error analysis for stereo methods based on calibrated trinocular image sequences.

After all those preparations we demonstrate comparative stereo analysis on EISATS data, with the intention to highlight the study of situations. We also use synthesized EISATS data for evaluating motion analysis algorithms. We show that different situations lead to different rankings in stereo or motion algorithms.

Finally we briefly discuss the use of Set 3 for discussing methods for the detection of independently moving objects in traffic scenes. The report ends with conclusions and comments about future work.

2 The EISATS Data

Testing on extensive and varying data sets helps to avoid a bias which occurs when using only selective (e.g., 'small') sets of data. The report introduces several sets of 'long' test sequences that are made available on the net as well as used evaluation measures for comparing different algorithms.[3]

Data relevant for driver assistance applications have basically an unlimited range of variations ('expect the unexpected'), due to the potential range of events, and thus of their combinations into situations. Selective ('small') sets of test data, say with a focus on rendered or engineered (good lighting, indoor) scenes, are insufficient for serious testing. The *.enpeda..* (environ-

[3]Actually, 'long' in this report still translates into durations of sequences in multiple seconds only, rather than of minutes, hours, or days; however, sequences of 150 or 400 stereo frames already allow us to illustrate the potentials of such 'long' sequences for testing and improving correspondence algorithms.

Table 1: Data sets offered on the EISATS website in December 2010.

Set	Comments
1	Night vision stereo sequences (Daimler AG) These seven stereo night vision sequences (12 bit, between 220 and 300 pairs of frames each) have been provided by Daimler AG, Germany, in June 2007 (group of Dr. Uwe Franke). These sequences come with ego-motion data and time stamps for each frame.
2	Synthesized stereo sequences (*.enpeda..* & Daimler AG) These synthesized stereo sequences (with ground truth) have been provided by Tobi Vaudrey (.enpeda..) and Clemens Rabe (Daimler AG).
3	IMOs in color stereo sequences (Drivsco) These three day-time, color stereo sequences have been provided by the European Drivsco project. Independent moving objects ground truth and gaze data is now available.
4	"Normal camera" binocular stereo sequences (Hella Aglaia Mobile Vision & *.enpeda..*) A few of those day- or night-time, grayscale stereo sequences have been provided by Hella Aglaia Mobile Vision GmbH, Germany; most of them have been recorded by students in the *.enpeda..* project.
5	"Normal camera" trinocular stereo sequences (*.enpeda..*) Three-camera stereo sequences (rectified by pairs) captured with HAKA1.
6	Grayscale stereo sequences with range scans (HU Berlin, *.enpeda..* & Daimler AG) So far three stereo vision sequences where the test vehicle drives through a car park; ground truth from a laser scanner; SGM, block and cross matcher disparity maps are also included.
7	Grayscale stereo sequences for scene labeling (Daimler AG) Stereo sequences with ground truth for scene labeling analysis (segmentation).

ment perception and driver assistance) Image Sequence Analysis Test Site (EISATS), see [12], is not (!) focused on one particular set of data or one particular evaluation strategy, but open to researchers in vision-based driver assistance systems for applying those data in their evaluations, as well as for contributing more (best: verified) data. The website contains recently seven different sets of test sequences, provided by different research groups in vision-based driver assistance, and of relevance for particular evaluation strategies. We illustrate the use of some of those data sets in this report. We are not discussing data in Sets 6 and 7 of EISATS in this report; more sequences with accompanying range scans or segmentation ground truth are in preparation, and this will be a subject of more specialized papers.

For the case of stereo and optic flow, we demonstrate that performance of established algorithms on small sets of test data, such as [35, 44, 47], is

not necessarily describing their performance on data as used in experiments in vision-based driver assistance. This reflects a common problem that (e.g.) engineers working in certain application areas find only little indications on the actual relevance of an algorithm in their specific scenario on a very selective set of test data.

We see the EISATS data base as a dynamic forum for relevant data and benchmarks for vision-based driver assistance. See Tab. 1 for a brief characterization of available sets of image sequences.

Each of the EISATS image sequences [12], as used in this report, represents a few seconds of driving (i.e., typically showing one situation). Sets 1-5 provide already some diversity of situations, and the goal is to extend those in a more systematic way. These sequences are still only representing a very small segment of possible situations in vision-based driver assistance (e.g., our discussion does not yet cover sequences recorded in the rain with moving wipers, against the sun, and so forth).

However, the given sets already allow us to go to a new quality of performance evaluation for correspondence algorithms compared to basically "no-sequence" data sets on sites such as [6, 10, 32, 33], which do not support studies about the influence of illumination artifacts, temporal filtering, or having low-contrast images with rapid changes due to events happening in real-world driving situations.

3 Error Measures

Since we are dealing with long sequences, we can analyze results over time, where frames are indexed by t. There are hundreds of error measures available (e.g., see [11]), and we aim at using general error measures that apply to many types of evaluation data.

If ground truth is available, we may calculate at a pixel position p the Euclidean distance

$$E_t(p) = \|A_t(p) - B_t(p)\|_2$$

of a generated result $A_t(p)$ from the ground truth $B_t(p)$. Doing so for all available (e.g., non-occluded) pixel positions p, we obtain an error image E_t.

In general, A_t and B_t are both n-valued functions (e.g., for stereo we have one disparity value and thus $n = 1$, optical flow is a field of 2D vectors with $n = 2$, or scene flow combines disparity with optical flow, and we have $n = 3$ in this case).

From such an error image E_t we can derive various measures, such as the mean μ, standard deviation σ, zero-mean standard deviation σ_0 (also referred to as *root mean squared error*), or simply maxima max.

We can explain this better by translating to some common error metrics already used in the community. In the case of stereo matching, a common metric is the root mean square error. For example, in [44] this is simplified to

$$\sigma_0(E_t) \text{ with } n = 1$$

For optical flow, the common measure is the mean end point error [3]

$$\mu(E_t) \text{ with } n = 2$$

To be even more specific, consider the case of stereo algorithms first. Assume that we have to compare two images A_t and B_t (e.g., calculated depth map with ground truth depth map) at time t, at all pixel locations p in a set Ω_t (e.g., all non-occluded pixels). The applied evaluation measures are point-wise root mean square

$$R_p(t) = \sqrt{\frac{1}{|\Omega_t|} \sum_{p \in \Omega_t} [A_t(p) - B_t(p)]^2} = \sigma_0(E_t)$$

spatial root mean square, where we compare Gaussian means of local neighborhoods (around the reference pixels),

$$R_s(t) = \sqrt{\frac{1}{|\Omega_t|} \sum_{p \in \Omega_t} [\mu(G_\sigma(p) * E_t(p))]^2} = \sigma_0(G_\sigma * E_t)$$

where $G_\sigma(p)*$ is a Gaussian convolution. Another measure is normalized cross correlation (NCC)

$$N(t) = \frac{1}{|\Omega_t|} \sum_{p \in \Omega_t} \frac{[A_t(p) - \mu_t^{(A)}][B_t(p) - \mu_t^{(B)}]}{\sigma_t^{(A)} \sigma_t^{(B)}}$$

Those errors are calculated along the given sequences, frame by frame, and conclusions are drawn based on mean errors and error variances along sequences, or due to particular error patterns at particular sub-sequences (e.g., for the occurrence of large occlusions, or of brightness alterations). In the case of motion vector fields, we calculate either average angular errors

or mean end-point errors $\mu(E_t)$ between vectors at corresponding pixel positions.

The NCC mean m_N and standard deviation σ_N of stereo matching techniques,

$$m_N = \frac{1}{T}\sum_{t=1}^{T} N(t) \quad \text{and} \quad \sigma_N^2 = \frac{1}{T}\sum_{t=1}^{T}[N(t) - m_N]^2$$

on individual sequences (say, with $T = 100$ or more stereo frames) allows one to identify a *winner* (as always, defined by the order of the means) for the recorded situation, and its *steadiness* (standard deviation).

However, the winner algorithm might not be the best in every frame within a given sequence. To measure this, we compare each two algorithms using *sums of direct comparison*. Given two algorithms, say C and D, and the corresponding NCC values $C(t)$ and $D(t)$ for every frame t in a given sequence of length T, the sum of direct comparisons between C and D is given by

$$S_{DC}(C, D) = \sum_{t}^{T} \Delta(t) \quad \text{where} \quad \Delta(t) = \begin{cases} 1, & C(t) > D(t) \\ 0, & C(t) = D(t) \\ -1, & C(t) < D(t) \end{cases}$$

Note that the absolute values of $S_{DC}(C, D)$ and $S_{DC}(D, C)$ are equal. Thus, we are able to define a ranking for each sequence based on sums of direct comparisons. Let

$$s_N(S_j) = \sum_{i=1}^{6} S_{DC}(S_i, S_j), \quad S_j \neq S_i$$

where the $S_i s$ represent the six selected stereo algorithms SGM-BT, DP, DPt, DPs, BP, GC, and S_j is the particular stereo algorithm for comparison.

Taking multiple sequences for the same situation, or even all sequences for all situations, the NCC mean and standard deviation defines the *robustness* of methods on those data. We illustrate this for five different situations.

4 Virtual Views for Stereo Evaluation

Set 5 of EISATS offers five trinocular image sequences, where the third camera may be used for prediction error analysis on stereo image sequences [37], similar to the prediction error analysis in [45] for motion analysis.

The prediction error strategy is a valuable tool to objectively evaluate the performance of stereo algorithms when ground truth is unavailable or basically impossible to acquire at full range and sufficient accuracy. The prediction error strategy requires only that input data is captured with (at least) three cameras: Two of them are used as input of the algorithms; the remaining one (the *third* camera) is used for evaluation purposes. Calculated depth data are used to map one of the stereo images (say, that of the 'left' camera) into the pose of the third camera, thus defining the *virtual* view. The similarity between virtual and third view characterizes the quality of the used stereo algorithm. Because of possible brightness differences between left and third view, normalized cross correlation (rather than root mean squared error) is used to quantify this similarity. The set Ω is defined by all pixel positions in the virtual view which receive a mapped image value of the left image.

See Figure 4 for an example of a recorded third view, a calculated virtual view, and the disparity map (a result of applying BP stereo matching) which was used for calculating this virtual view. The use of several trinocular sequences for prediction error analysis has been demonstrated in [27], and more of such sequences are now available on EISATS. The geometric approach of the prediction error methodology was specified in [36], only using Sequence 1 of Set 5 of [12]) as a long real-world sequence at that time. All contributing cameras are calibrated [18], thus making mapping of data into defined poses possible.

Figure 4: An application for a sequence in Set 5. Left: third view. Middle: virtual view for the disparity map shown on the right. The applied matching algorithm was belief propagation (BP) stereo analysis. The specularity (see image on the left), apparent both in the recorded left and right image, causes a 'defect' in the calculated disparity data.

Algorithm	Mean	St. Dv.
DPt	0.81	0.03
DP	0.78	0.03
BP	0.75	0.02
SGM-MI	0.72	0.02
GC	0.67	0.03
SGM-BT	0.58	0.02

	BP	BT	DP	DPt	GC	MI	s_N	Rank
BP	0	-	-	-	-	-	146	3
SGM-BT	-110	0	-	-	-	-	-550	6
DP	76	110	0	-	-	-	334	2
DPt	106	110	70	0	-	-	506	**1**
GC	-110	110	-110	-110	0	-	-328	5
SGM-MI	-108	110	-108	-110	108	0	-108	4

Table 2: Overall results for the Ordinary Conditions sequence of Set 5 (see Figure 3 for the complete diagram of values per frame). Left: Mean and standard deviation. Right: Sums of direct comparisons for (SGM-BT - BP), (DP - BP), (DP - SGM-BT), (DPt - BP), and so forth.

5 Evaluations for Situations

In this section we illustrate the use of sequences, as provided on EISATS, for evaluating the performance of stereo algorithms for particular situations. The classification of sequences into situations was done manually, just by subjective evaluation of contributing events.

5.1 Ordinary Driving Conditions

Ordinary conditions are those in which the traffic is relatively light, the brightness differences between the stereo pair is minimum, the sun still high in the sky and there are no objects in the borders of the road which may create *illumination artifacts* (see Section 5.2). Shadows and specularities are minimum. Note that this sort of conditions can also be present in a cloudy environment. Sequences in Sets 1 and 4 of EISATS are mostly in this category.

We show results for the Ordinary Driving Conditions sequence in Set 5. The algorithm with better performance (with respect to the NCC mean) was DPt, followed by DP and BP. All the algorithms presented their worst performance in that sequence when the followed and incoming vehicles are closer to the ego-vehicle. DPt, the winning algorithm, did not show the best performance in every single frame, for around ten frames it performed worse than DP. See Table 2 and Figure 3 top row, left.

5.2 Illumination Artifacts

Illumination artifacts (e.g., while driving below trees) are present in most of the sequences of Sets 1, 4 and 5. We show results for the Illumination Artifacts sequence in Set 5. This sequence was recorded over a road surrounded by trees. This situation, in general, did not modify drastically the brightness between the stereo pairs, but introduced a considerable number of different dark and bright patches (caused by the foliage) in the left and right images.

Algorithm	Mean	St. Dv.
GC	0.88	0.08
DP	0.82	0.03
DPt	0.81	0.02
BP	0.78	0.10
SGM-MI	0.77	0.03
SGM-BT	0.64	0.08

	BP	BT	DP	DPt	GC	MI	s_N	Rank
BP	0	-	-	-	-	-	-6	4
SGM-BT	150	0	-	-	-	-	-446	6
DP	4	150	0	-	-	-	270	2
DPt	4	150	-112	0	-	-	48	3
GC	144	146	138	142	0	-	712	**1**
SGM-MI	4	150	-142	-148	-142	0	-278	5

Table 3: Results for the 150 frames of the Illumination Artifacts sequence (see Figure 3 for the complete diagram of values per frame). Left: Mean and standard deviation. Right: Sums of direct comparisons for (SGM-BT - BP), (DP - BP), (DP - SGM-BT), (DPt - BP), and so forth.

It also introduced a fast change in the lighting conditions between subsequent frames.

For this particular sequence we noticed that, when the trees are closer to the right side of the road (due to the fact that this sequence was recorded in the late afternoon, when the sun was in a low position in the left side of the ego-vehicle), the difference in brightness between the stereo pair is reduced, improving the performance for most of the algorithms. The top performing algorithm was GC, followed by the two dynamic programming ones, see Figure 3 top row, right. An interesting point to note with this sequence is that BP had a slightly better performance than SGM-MI with respect to the mean, however, the latter algorithm had a better performance in a larger number of frames than the former one, according to the Table 3.

Algorithm	Mean	St. Dv.
GC	0.79	0.06
BP	0.76	0.06
DP	0.74	0.07
DPt	0.73	0.07
SGM-MI	0.66	0.04
SGM-BT	0.64	0.05

	BP	BT	DP	DPt	GC	MI	s_N	Rank
BP	0	-	-	-	-	-	332	2
SGM-BT	-150	0	-	-	-	-	-696	6
DP	-88	150	0	-	-	-	60	3
DPt	-92	150	-26	0	-	-	0	4
GC	138	150	146	150	0	-	734	**1**
SGM-MI	-140	96	-118	-118	-150	0	-430	5

Table 4: Results for the 150 frames of the *Inner City at Night* sequence (see Figure 3 for the complete diagram of values per frame). Left: Mean and standard deviation. Right: Sums of direct comparisons for (SGM-BT - BP), (DP - BP), (DP - SGM-BT), (DPt - BP), and so forth.

5.3 Inner City at Night

A sequence of an "Inner City at Night" situation (Set 5) is recorded after sunset with regular to dense traffic on the road; the scene is illuminated by the lights of the other vehicles, and lights and specularities cause large white regions of missing dynamics in intensity values. Using NCC as a quality metric, GC was the algorithm with the best performance on the selected original sequence, while SGM-BT showed the worst performance; see Table 4, left. In the first 30 frames, the performance of all the algorithms is poor as a consequence of a close object present in the scene (see Figure 3, middle row, left).

For this sequence, the ranking with the NCC mean and the one obtained with sums of direct comparisons was the same. However, according to the Table 4 right, DP had a better performance than BP in almost the half of the frames.

5.4 Brightness Differences

Brightness differences between the input stereo pair is a common issue in driver assistance. For example, by changing the viewing angle with respect to the sun, one camera may record a brighter sequence of frames than the other. Of course, inter-camera-communication may somehow relax this issue in the future.

The output of correspondence algorithms is severely affected in such a situation of brightness differences. In the Brightness Differences sequence of Set 5, there are brightness differences in every frame, and they increase in the last 40 frames. The algorithm with the best performance on this sequence was SGM-MI, followed by DP and DPt (see Table 5, left). This ranking is in accordance with the results obtained in [37] in the case for the brightness altered sequence, supporting the idea that the prediction error is a good technique to evaluate stereo algorithms in the absence of ground truth. BP showed a good performance (second) until the difference in brightness becomes extreme, in which its performance is the second worst. This situation was reflected in the ranking defined by the sums of direct comparisons, in which BP was the second best, due to the fact that its performance was degraded until the last third of the sequence. See 3, middle row, right.

Algorithm	Mean	St. Dv.
SGM-MI	0.86	0.01
DPt	0.77	0.02
DP	0.77	0.02
BP	0.75	0.08
GC	0.63	0.03
SGM-BT	0.60	0.05

	BP	BT	DP	DPt	GC	MI	s_N	Rank
BP	0	-	-	-	-	-	174	2
SGM-BT	-150	0	-	-	-	-	-584	6
DP	-58	150	0	-	-	-	28	4
DPt	-42	150	64	0	-	-	172	3
GC	-74	-16	-150	-150	0	-	-540	5
SGM-MI	150	150	150	150	150	0	750	**1**

Table 5: Overall results for the 150 frames of the *Brightness Differences* sequence. Left: NCC mean and standard deviation. Right: Sums of direct comparisons for (SGM-BT - BP), (DP - BP), (DP - SGM-BT), (DPt - BP), and so forth.

5.5 Close Objects

Sequences with close objects (people, other vehicles, static structures, etc.) are very important to be investigated as this situation is the main characteristic during a traffic jam, a potential conflict, and so forth. In this situation it is likely that the implemented driver assistance systems should contribute to adaptation and optimization of driving. In the analyzed Close Objects sequence of Set 5, two pedestrians appear in front of the vehicle (without any actual danger because the ego-vehicle stopped earlier).

The ranking of the algorithms differs from the two previous ones: DP is the best overall, as seen in the Table 6 left and Figure 3 lower row. However, four different algorithms have the best performance throughout the sequence for particular intervals of time. The performance of all algorithms is below standards around the middle of the sequence (except for SGM-MI), when the two pedestrians are very close to each other. Even the top performing algorithm, DP, had a worst performance than the worst overall algorithm

(SGM-BT) for about 15 frames, according to the right side of Table 6.

Algorithm	Mean	St. Dv.
DP	0.65	0.07
BP	0.64	0.06
DPt	0.63	0.08
SGM-MI	0.63	0.03
GC	0.59	0.05
SGM-BT	0.53	0.03

	BP	BT	DP	DPt	GC	MI	s_N	Rank
BP	0	-	-	-	-	-	123	2
SGM-BT	-71	0	-	-	-	-	-331	6
DP	35	65	0	-	-	-	241	**1**
DPt	-5	61	-43	0	-	-	81	3
GC	-77	55	-67	-57	0	-	-201	5
SGM-MI	-5	79	-31	-11	55	0	87	4

Table 6: Overall results for 79 frames of the *Close Objects* sequence. Left: Mean and standard deviation. Right: Sums of direct comparisons for (SGM-BT - BP), (DP - BP), (DP - SGM-BT), (DPt - BP), and so forth.

5.6 Summary

We summarize briefly the detected robustness of the algorithms across the five situations presented in here (see Table 7). Using the mean of NCC over all the situations, DPt outperforms (by a small difference) all the other algorithms, despite the fact that it only performs the best in two of the presented situations. However, the performance of DPt heavily depends on the percentage of pixels showing a planar road surface. On the other hand, GC finalized as the second worst algorithm, even that it was the best in two of the sequences.

This report only discusses five sequences (situations). Summarizing our more general experience, also taking our experimental results into account which are not reported in this report, we may conclude that

Algorithm	Mean	St. Dv.
DP	0.65	0.07
BP	0.64	0.06
DPt	0.63	0.08
SGM-MI	0.63	0.03
GC	0.59	0.05
SGM-BT	0.53	0.03

Table 7: Overall NCC results over the five different situations considered in Section 5.

- cost functions should not depend on brightness constancy (as an alternative, some kind of preprocessing methods may map the given stereo sequences into data where the impact of brightness differences has been reduced, e.g., by using redials with respect to smoothing),
- the well-known streaking-effect of DP also limits the use of this simple matching approach in the given application context,
- SGM is potentially able to deal with scenes of high depth complexities,
- BP may be preferred in scenes with larger homogenous regions, and
- GC has a tendency to create convex regions of nearly constant depth.

These findings are also accompanied by progress in real-time implementations of DP, SGM, and BP, but the lack of fast implementations of GC variants.

6 Synthesized Video Sequences

Synthetic data with ground truth do have already a history in computer vision. Long synthetic sequences, as in Set 2 of EISATS, are very useful for studying defined variations in image data, for analyzing their impact on the performance of a selected stereo correspondence algorithm.

In [37], the performance of several stereo algorithms was tested over different adverse conditions (blurred images, stereo pairs with differences in brightness, and images corrupted with Gaussian noise) by modifying the rendered sequence No. 1 of Set 2 of [12]. An objective evaluation (using

the root mean squared error and the percentage of miscalculated bad pixels), based on the available ground truth for such a rendered sequence, showed indeed that the ranking of the studied algorithms varied depending on modifications applied to the sequence. However, a ranking of methods on such rendered sequences is not well correlated to a ranking on real-world sequences for particular situations. This is due to at least two facts: those synthesized sequences are not yet perfectly photo-realistic and physics-based, thus different from the real-world sequences recorded with specific cameras, and they are also very limited with respect to the covered situations or (unpredictable) events in the real world.

Synthesized data are especially important for motion analysis algorithms;

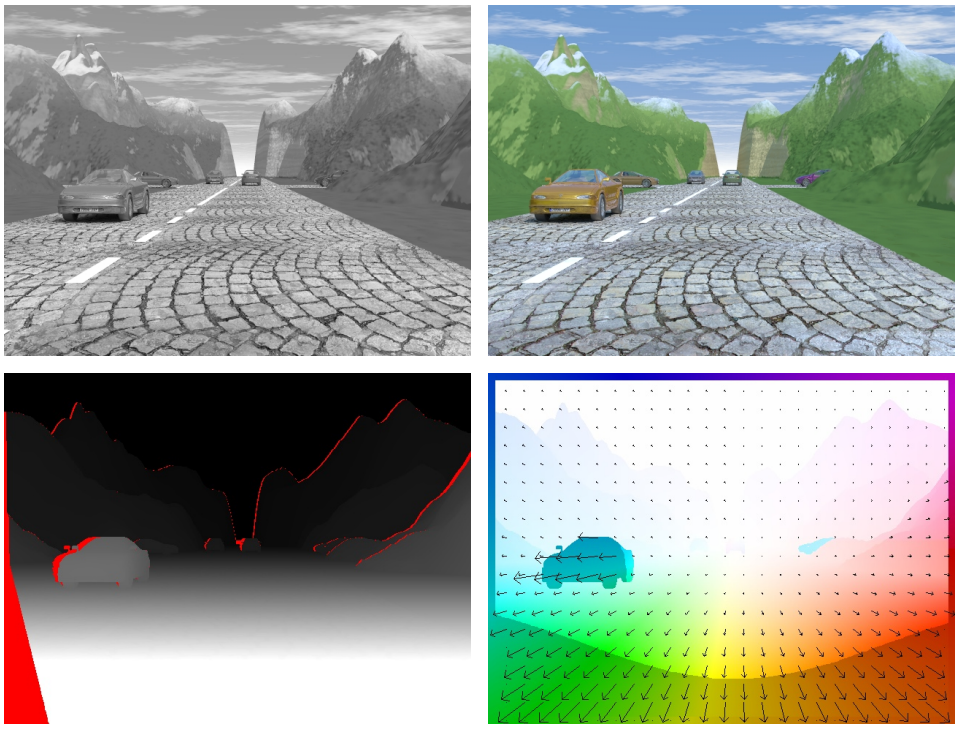

Figure 5: Top: example images (frame 42) from the synthetic sequence No. 1; grayscale (left) and color (right). Bottom: disparity map (left) with light to dark as near to far, and red as occlusion; flow map (right) with the color as direction (see border of image) and saturation as length, a vector map is overlayed for additional information.

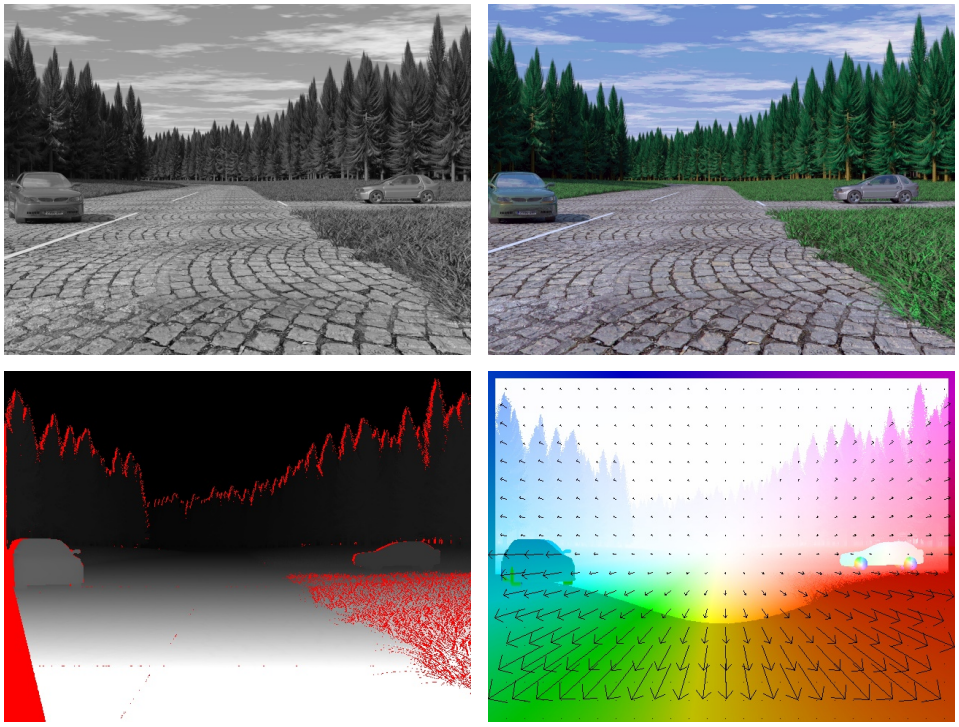

Figure 6: Top: example images (frame 219) from the synthetic sequence No. 2; grayscale (left) and color (right). Bottom: disparity map (left) and flow map (right), color encoding as in Figure 5.

relatively slow recording of video sequences (e.g., at 25 Hz) does not allow to use prediction error analysis [45] for evaluation, and there are still not many studies on performance of motion analysis on real world sequences (such as, e.g., in [52]).

6.1 Sequences 1 and 2

The synthetic sequence No. 1 (a synthetic POV-ray sequence of 100 stereo frames) was introduced in [48] and made publicly available (with stereo and motion ground truth) in Set 2 on [12]. This was the first long stereo sequence, with ground truth data for optical flow (both x and y directions), disparity, and disparity rate (change in disparity between frames - for scene flow ground truth). This data is generated with ray-tracing and texture mapping, generating a very clean looking image. Figure 5 shows an example of

the original images and ground truth interpretations. Furthermore, it is one of the first stereo databases containing $>$ 8-bit dynamic range; the sequence contains 12-bit grayscale and 3×12-bit color depth. This is comparable to top of the line commercially available machine vision cameras (e.g., [4, 41]). This scene has been used to compare stereo [37], optical flow [38], and scene flow algorithms [51] in various papers.

There is an advanced sequence available, synthetic sequence No. 2 in Set 2 [12]. This sequence contains a more realistic driving situation, also including trees and grass. This sequence aims to be more challenging for the optical flow and stereo algorithms. Example frames can be seen in Figure 6. It has all the same qualities as sequence No. 1 (i.e., high dynamic range input, with ground truth available). Furthermore, the ground truth ego-motion (i.e., fundamental matrix) is available for every sequential pair of images. This gives the ground truth movement of the cameras from frame $t-1$ to t. This allows one to use this information to create "biased" algorithms, and also test ego-motion/fundamental matrix algorithms. Ego-motion estimation is a very important aspect for driver assistance, as vibrations and variations in the road cause very large rotational ego-motion between frames.

6.2 Results on Provided Data

The following results are for synthetic sequence No. 1. For computing optical flow, we decided for testing PyrHS, BBPW and TVL1 (see Section 1.3). This subset of algorithms highlights the basic to state-of-the-art algorithms for optical flow. The results for mean end point errors (as introduced in Section 3) can be seen on the left of Figure 7.

Further to this, we can alter the input data to contain noise that is present in real-world imagery. This was done in [38] and highlighted that illumination differences cause the major problems in both stereo and optical flow algorithms. This is obvious, because both types of correspondence algorithms rely on the *intensity consistency assumption*, i.e., that the pixel on an object will look identical between corresponding images.

Sample results can be seen on the right side of Figure 7. The shape is because the brightness differences are large ($\pm$100 intensity values) at the start of the scene, and reduce to zero at the mid-point, before increasing back up to $\pm$100. Obviously, TVL1 is more sensitive to major illumination differences, compared to the other algorithms. Furthermore, with a difference of only $\pm$10 (see around the middle of the sequence), the algorithms rankings

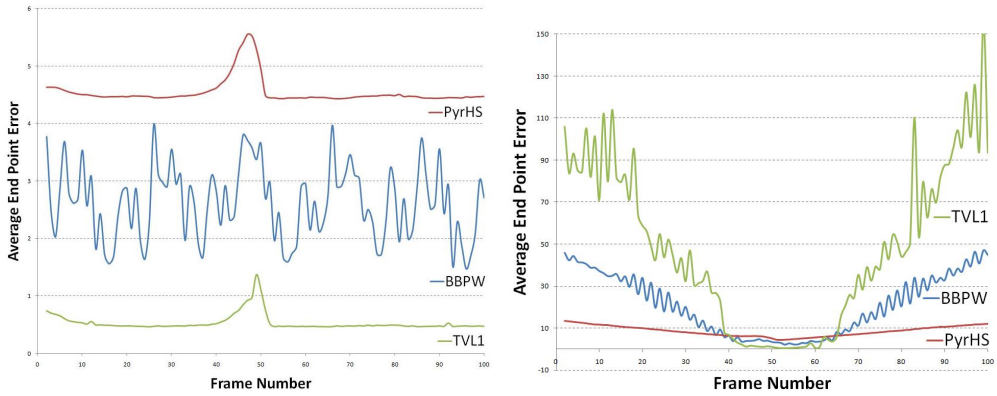

Figure 7: Mean end point error results across image sequence, compared to ground truth (left: on original data, right: on illumination altered data). The total average was 4.5 (PyrHS), 2.6 (BBPW), and 0.53 (TVL1) for the original data, and 8.6 (PyrHS), 21.0 (BBPW) and 48.3 (TVL1) for the illumination altered data.

are the same as for no illumination difference.

This evaluation was only given as an example of what can be done with the provided data with ground truth. A much more extensive test, varying parameters and noise properties could be investigated to exploit this data. One major hole in the literature is an extensive evaluation of the importance of having a high-dynamic range for machine vision. From a practical point of view, we have experienced that the stereo and optical flow results are of a much higher quality when the dynamic range is high. This is obvious because the cost functions have an easier discretization between possible matches. This effect needs to be studied in detail, and the data provided here makes that possible.

6.3 Possible Future Extension of Set 2

Most studies work on adding artificial noise to generated scenes (as done in [37, 38]). This noise needs to be more realistic. One way of doing this is by introducing the noise generation into the image generation process. We demonstrate opportunities of physics-based rendering for camera modeling, which is planned to be used in further sequences in Set 2 of EISATS.

Figure 8 shows a 3D model of an urban road intersection, and stereo

sequences (path tracing) are rendered either with a simple ray tracer or LuxRender. This involves the use of a realistic model of atmosphere and sunlight. A simulation of realistic specular highlights or blooming is of importance, as these events occur frequently in imagery of outdoor scenes and cause major problems to correspondence algorithms. A scene with specular highlights or reflections can not be seen by a cameras as 'perfect' as shown in the upper row (left, or second to the left). An image with moderate bloom and some chromatic aberration simulates some realistic distortion as appearing for common cameras. An image with severe blooming (upper row, right) simulates a defective camera (or a camera with over-exposure - which often happens in outdoor environments).

We also studied the behavior of BP stereo on such synthetic images. As expected, depth maps derived from ray-traced stereo pairs contains only minor errors in image regions showing reflections. However, depth maps from images with moderate blooming and chromatic aberration are not significantly degraded compared to results from undistorted data. Images severely degraded by blooming show impaired results. Further studies in this area hope to identify which noise effects results for stereo and optical flow the most, thus giving the community tools on where they should be trying to adapt their algorithms.

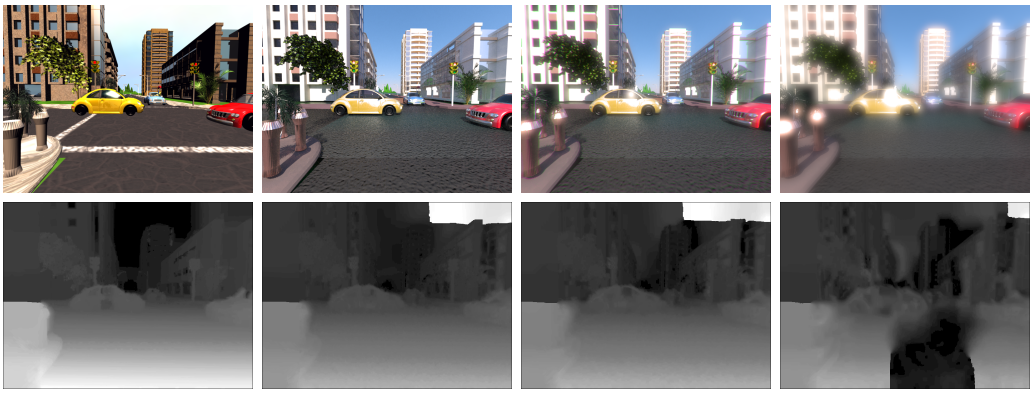

Figure 8: Upper row, left to right: 3D model rendered with simple ray tracing, tone-mapped output of LuxRender (some issues with material support of this render engine are apparent such as missing road marks), also with moderate blooming and chromatic aberration, and severe blooming. Lower row, left to right: corresponding BP depth maps.

7 IMOs in Color Stereo Sequences

The sequences in Set 3 of EISATS were especially designed for studying the detection of independently moving objects. This set provides two situations, both with (very) long image sequences. For the detection of independent moving objects (IMOs), we also provide ground truth in terms of labeled regions with associated information (such as the type of IMO, on which lane the car is driving and additional occlusion properties of the IMOs). Moreover, we define several measurements that allow for comparative evaluation of IMO detection algorithms. The two sequences show situations 'Suburban Bridge (851 frames)' and 'Suburban follow' (1182 frames) with ground truth information in terms of labeled IMOs.

We discuss the used data labeling. In the following, true IMOs are denoted $A = (x, y, w, h)$ where the position of the IMO's center is denoted by (x, y), and its width and height by (w, h). Detected IMOs are denoted by $B = (x', y', w', h')$. The *overlap value* for A and B is an estimate for the distance between a true labeled IMO and a detected one, and is calculated as follows (see Figure 9): $a = \left(\max\{x + \frac{w}{2}, x' + \frac{w'}{2}\} - \min\{x - \frac{w}{2}, x' - \frac{w'}{2}\}\right)$ and $b = \left(\max\{y + \frac{h}{2}, y' + \frac{h'}{2}\} - \min\{y - \frac{h}{2}, y' - \frac{h'}{2}\}\right)$ and

$$O(A, B) = \frac{a \cdot b}{\max\{wh,\ w'h'\}} \tag{1}$$

The defined overlap value is not a metric: it is symmetric $O(A, B) = O(B, A)$, we have $O(A, A) = 0$, but $O(A, B) = 0$ does not mean that $A = B$, and this measure does also not satisfy the triangularity constraint $O(A, C) \leq O(A, B) + O(B, C)$, for sets of pixels A, B and C. However, the overlap value

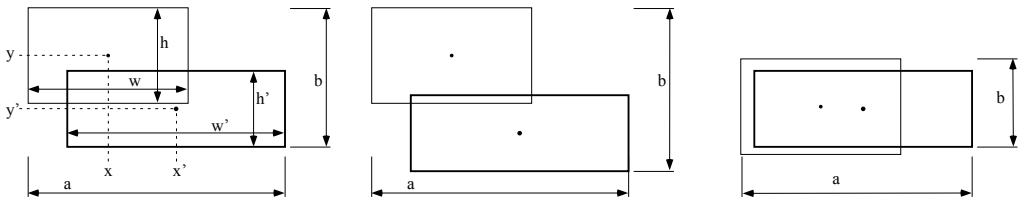

Figure 9: Left: because of $w'h' > wh$ in the shown case, the estimate is equal to $ab/w'h'$. Middle: ab increases resulting into an increase of the overlap value $> 0{\cdot}5$. Right: ab decreases resulting into an overlap $< 0{\cdot}5$.

is easy to calculate and proved in our experiments to be a reasonable estimate. On the other hand, the cardinality of the symmetric difference between A and B divided by the cardinality of the union of both sets would be a metric; for a proof see [26]; but this measure is more costly to calculate.

Finally, the counts of hit, miss or false alarms provide an evaluation of the quality and reliability of the detection. A detected IMO is considered as being true if the overlap value with a true IMO is less or equal to 0·5, where equal to zero means that both circumscribing rectangles coincide.

We have developed a two-stage vision system for extracting driving-relevant information from stereo cameras mounted in a moving car (for details, see [5, 40, 42]). For the gaze target identification, we provide the position of the gaze point within the frame, and identity the gaze target as being the ground truth. The position is given in normalized (x, y) coordinates, ranging from $(0, 0)$ in the lower left, to $(1, 1)$ in the upper right corner of a recorded frame. Gaze target identities were classified by a human observer into one of 15 active target classes (e.g., lane markings on the right, and tangent point on lane markings at the center of the road).

Independently moving objects were hand-labelled frame-by-frame and tracked across successive frames. For each IMO, the following parameters are given: its identification number, type (car, truck, motorcycle, bike, pedestrian), a flag indicating whether it is partially occluded (1) or not (0), and finally the lane it is traveling on (1-same as the test car, 2-opposite, 3-side road left, 4-side road right). In respect to the image frame, the IMO's center (x, y) and extension (width and height) are given. This data range on both axes from 0 to 1, with the origin $(0, 0)$ being in the left lower corner. Figure 10 shows an example frame taken from the Suburban Bridge sequence. On frame 875, IMOs 3 and 4 are to be seen and the former partly occludes the latter.

Gaze point targets were classified by hand on a frame-to-frame base into one of 15 active and three error classes: Lane markings on the left side, center or right side of the road, boundary posts to either side of the road, tangent points on these lane markings, where applicable, the road surface ahead (i.e., on the first 20-30 m in front of the car), or farther away, street signs and traffic lights, IMOs ahead (on the same lane), upcoming (on the opposite lane), or moving on cross roads. Gazes to any other point (including the dashboard) were classified into the residual class, whereas errors were either associated with the start or the end of the recording or were classified as a general error. The predictive value of eye movements on car-directed actions by the

driver has recently been demonstrated [24]. Next to the IMO information mentioned above, Figure 10 also shows the position of the gaze point (number plate of the first car).

Human drivers are facing a dual task: on the one hand, they need to steer the car through straight and curved sections of the roads; for this part they usually direct their gaze to the tangential point or the road surface (i.e., two points that allow to infer the required steering angle by identifying simple geometric means); see, for example, [29, 24, 25]. On the other hand, they need to attend quickly to upcoming possible obstacles such as IMOs or points of interest such as crossings. While there are a number of algorithms built to segment the scene and identify *salient* points of interest in a bottom-up manner (see above), the combination of saliency and relevance (top-down processes) into a priority map [13] is a subject of current research, and there are no databases available for benchmarking so far.

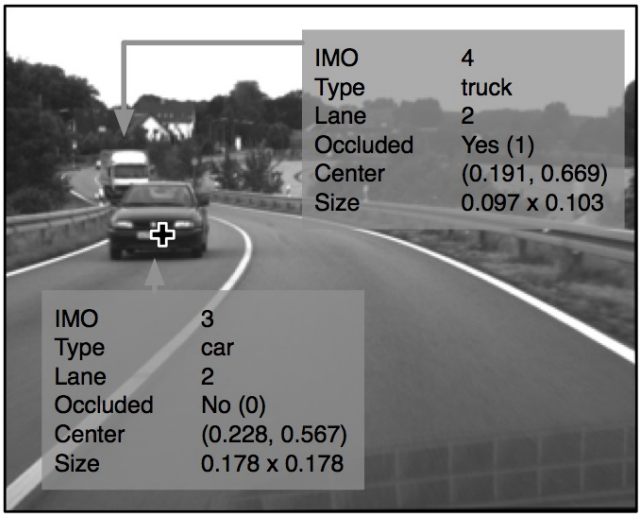

Figure 10: Frame 875 of the Suburban Bridge sequence (left camera) shows two IMOs, for which hand-labelled data are provided. Furthermore, gaze positions are available, both in coordinates and classified targets. In the ego-vehicle we observe the number plate of IMO 3 (black and white cross) and further data, also on IMO 4.

8 Conclusions

Within the *.enpeda..* project, hundreds of sequences have been recorded so far, and many of them have been analyzed using prediction error analysis and the NCC measure (in case of stereo analysis), or using approximate scene geometry (in case of motion analysis [52]).

Rankings of stereo or motion correspondence methods change often along one sequence, and benchmarking on only a few frames of such a sequence would be meaningless. Various events in outdoor driving define many challenges for stereo or motion matching, and often all the methods experience difficulties with the same event, just at different scales. The importance of testing on such sequences is actually to identify such events, and to aim at improving matching for those particular events. However, the authors also imagine that an adaptive strategy may finally be best, selecting stereo or motion matching methods out of a given *toolbox* in dependence on automatically detected situations. For example, white balancing in cameras is already such an adaptation, which needs to be refined and expanded to further layers of data processing.

We presented methodologies for performance evaluation of correspondence techniques and discussed experiments with test data relevant for outdoor vision in the driver assistance domain. Based on those investigations, we draw the following conclusions:

- There are reasonable solutions for stereo analysis in outdoor environments, on sequences taken from moving cameras, but none of the techniques was superior in all the tested situations; an adaptive selection might be the way to go.

- A careful evaluation of stereo or motion algorithms (comparable to efforts when performing car crash tests for physical performance) requires a testing on very large and representative data sets. Testing on data representing very different situations goes beyond common test behavior in today's computer vision community. The EISATS data base offers evaluation tools as well as relevant data to do such large scale analysis.

- There is a significant variation of performance of different correspondence algorithms across and even within individual sequences (situations), and adaptive selection of techniques (from a toolbox where avail-

able) would require time-efficient higher level mechanisms that identify situations.

- An investigation of the influence of particular events in images, such as typical for certain situations, on the performance of different stereo and optic flow algorithms is a meaningful task which allows for the improvement of techniques, either directly or by means of data preprocessing. The classification of such events is likely to require high-level scene analysis. In this context, the EISATS data base already contains some benchmark data relevant for IMO detection as well as human gaze data.

- Rendered data may be manipulated to simulate particular events. We demonstrated that different stereo algorithms degenerate to a different degree in case of brightness differences or lighting artifacts, depending on parameters when generating those events. The EISATS data base provided rendered sequences in which these particular effects are simulated.

The EISATS website [12] provides data for the analysis and benchmarking of stereo and optic flow as well as for high-level scene interpretation in the driver assistance domain. The authors invite other researchers to contribute with relevant benchmark data in driver assistance. Obviously, this is and will happen at various places, such as at the Daimler pedestrian benchmark data set at [15], and for many other traffic-related application areas (e.g. driver fatigue analysis [43]) which also have their particular needs for test data.

We suggest that different stereo or optic flow algorithms (identified as being stable top performers for at least one situation) are taken into a *tool box*, and that an *adaptive strategy* should be applied. A classification of incoming stereo video data into situations could guide the selection of algorithms from the tool box. A statistical approach for identifying situations might be more appropriate for this adaptive strategy than an intuitive high-level identification scheme of situations (by events).

A future statistical categorization of situations may be possibly based on distributions of selected features in the Fourier domain of signals (as in [1]), on simple features such as mean intensity or variance in randomly selected windows, or on the density of significant scale-invariant features or locally adaptive regression kernels [46] in randomly selected image rows, and so forth. We already tested scale-invariant features for this purpose [19], and

a 'sparse feature' approach appears to be reasonable for some clustering of image sequences into different categories.

Acknowledgements. This research was supported by the EU project Drivsco (FP6-IST-FET, contract 016276-2).

References

[1] A. Briassouli and I. Kompatsiaris. Change detection for temporal texture in the Fourier domain. In Proc. *ACCV*, LNCS, to appear, 2011.

[2] S. Baker, S. Scharstein, J. P. Lewis, S. Roth, M. J. Black and R. Szelisky, "A database and evaluation, methodology for optical flow", in Proc. *IEEE Int. Conf. Computer Vision*, pages 1–8, 2007

[3] J.L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques", *Int. J. Comput. Vision*, **12**:43–77, 1995

[4] Basler Vision Technologies. [Online]. Available: `http://www.baslerweb.com/`

[5] E. Baseski, L. Baunegaard, N. Pugeault, F. Pilz, K. Pauwels, M.M. Van Hulle, F. Wörgötter, and N. Krüger, "Road interpretation for driver assistance based on an early cognitive vision system", in Proc. *VISAPP*, volume 1, pages 496–505, 2009

[6] A. Bellmann, O. Hellwich, V. Rodehorst, and U. Yilmaz, "A benchmarking dataset for performance evaluation of automatic surface reconstruction algorithms,", in Proc. *BenCOS*, pages 1–8, 2007

[7] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo", *Int. J. Computer Vision*, **35**:269–293, 1999

[8] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping", In Proc. *ECCV*, LNCS 3024, pages 25–36, 2004

[9] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods", *Int. J. Computer Vision*, **61**:211–231, 2005

[10] CMU image data base. [Online]. Available: `http://vasc.ri.cmu.edu/idb/html/stereo/`

[11] R.M. Dudley, "Probabilities and metrics : convergence of laws on metric spaces, with a view to statistical testing", Matematisk institut, Lecture notes series, report 45, Aarhus universitet, 1976

[12] *.enpeda..* image sequence analysis test site (EISATS). [Online]. Available: `http://www.mi.auckland.ac.nz/EISATS/`

[13] J.H. Fecteau and D.P. Munoz, "Salience, relevance, and firing: a priority map for target selection", *Trends Cogn. Science*, **10**:382 – 390, 2006

[14] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient belief propagation for early vision", *Int. J. Computer Vision*, **70**:261–268, 2006

[15] D. Gavrilla, "Daimler pedestrian benchmark data set", follow "Looking at people" on `http://www.gavrila.net/Research/research.html`, 2009

[16] S. Guan, R. Klette, and Y.W. Woo, "Belief propagation for stereo analysis of night-vision sequences", in Proc. *PSIVT*, LNCS 5414, pages 932–943, 2009

[17] P. Handschack and R. Klette, "Quantitative comparisons of differential methods for measuring of image velocity", in Proc. *Aspects Visual Form Processing*, Capri, World Scientific, pages 241–250, 1994

[18] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, United Kingdom, 2000

[19] R. Haeusler and R. Klette, "Benchmarking stereo data (not the matching algorithms)", In Proc. *DAGM*, to appear, 2010

[20] S. Hermann and R. Klette. The naked truth about cost functions for stereo matching. MI-tech-TR 33, University of Auckland, 2009

[21] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information", in Proc. *CVPR*, volume 2, pages 807–814, 2005

[22] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences", *IEEE Trans. Pattern Analysis Machine Intelligence*, `http://doi.ieeecomputersociety.org/10.1109/TPAMI.2008.221`, 2008

[23] B.K.P. Horn and B.G. Schunck, "Determining optical flow", *AI*, **17**:185–203, 1981

[24] F.I. Kandil, A. Rotter, and M. Lappe, "Driving is smoother and more stable when using the tangent point", *J. Vision*, **9**(1),11:1–11, 2009

[25] F.I. Kandil, A. Rotter, and M. Lappe, "Car drivers attend to different gaze targets when negotiating closed vs open curves", *J. Vision*, **10**(4),24:1–11, 2010

[26] R. Klette and A. Rosenfeld. "Digital Geometry - Geometric Algorithms for Digital Picture Analysis". Morgan Kaufmann, San Francisco, 204

[27] R. Klette, S. Sandino, T. Vaudrey, J. Morris, C. Rabe, and R. Haeusler, "Stereo and motion analysis of long stereo image sequences for vision-based driver assistance", keynote, DAGM 2009, Jena/Germany, 09 September 2009

[28] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?", *IEEE Trans. Pattern Analysis Machine Intelligence*, **26**:65–81, 2004

[29] M.F. Land and D.N. Lee, "Where we look when we steer", *Nature*, **369**:742 –744, 1994

[30] Z. Liu and R. Klette, "Dynamic programming stereo on real-world sequences", in Proc. *ICONIP*, LNCS 5506, pages 527–534, 2008

[31] Z. Liu and R. Klette, "Approximated ground truth for stereo and motion analysis on real-world sequences", in Proc. *PSIVT 2009*, LNCS 5414, pages 874–885, 2009

[32] B. McCane, K. Novins, D. Crannitch, and B. Galvin, "On benchmarking optical flow", *Computer Vision Image Understanding*, **84**:126–143, 2001

[33] Middlebury vision evaluation. [Online]. Available: `http://vision.middlebury.edu/`

[34] T. Mimuro, Y. Miichi, T. Maemura, and K. Hayafune, "Functions and devices of Mitsubishi active safety ASV". In Proc. *IEEE Intelligent Vehicles*, pages 248–253, 1996

[35] R. Mohan, G. Medioni, and R. Nevatia, "Stereo error detection, correction, and evaluation", *IEEE Trans. Pattern Analysis Machine Intelligence*, **11**:113–120, 1989

[36] S. Morales and R. Klette, "A third eye for performance evaluation in stereo sequence analysis", In Proc. *CAIP*, LNCS 5702, pages 1078-1086, 2009

[37] S. Morales, T. Vaudrey, and R. Klette, "Robustness Evaluation of Stereo Algorithms on Long Stereo Sequences", in Proc. *IEEE Intelligent Vehicles*, pages 347–352, 2009

[38] S. Morales, Y. W. Woo, R. Klette, and T. Vaudrey, "A study on stereo and motion data accuracy for a moving platform", in Proc. *FIRA RoboWorld Congress*, LNCS 5744, pages 292-300, 2009

[39] Y. Ohta and T. Kanade, "Stereo by two-level dynamic programming", in Proc. *IJCAI*, pages 1120–1126, 1985

[40] K. Pauwels, "Computational modeling of visual attention: Neuronal response modulation in the Thalamocortical complex and saliency-based detection of independent motion", PhD thesis, K.U.Leuven, 2008

[41] Point Grey. [Online]. Available: `http://www.ptgrey.com/`

[42] N. Pugeault, K. Pauwels, F. Pilz, M.M. Van Hulle, and N. Krüger, "A three-level architecture for model-free detection and tracking of independently moving objects", In Proc. *Int. Conf. Computer Vision Theory Applications*, 2010

[43] Q. Ji, Z. Zhu, and P. Lan, P., "Real-time nonintrusive monitoring and prediction of driver fatigue", *IEEE Trans. Vehicular Technology*, **53**:1052–1068, 2004

[44] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *Int. J. Computer Vision*, **47**:7–42, 2002

[45] R. Szeliski, "Prediction error as a quality metric for motion and stereo", in Proc. *ICCV*, volume 2, pages 781–788, 1999

[46] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction", *IEEE Trans. Image Processing*, **16**:349–366, 2007

[47] N.A. Thacker, A.F. Clark, J.L. Barronc, J.R. Beveridged, P. Courtneye, W.R. Crum, V. Ramesh, and C. Clark, "Performance characterization in computer vision: A guide to best practices", *Computer Vision Image Understanding*, **109**:305–334, 2008

[48] T. Vaudrey, C. Rabe, R. Klette, J. Milburn, "Differences between stereo and motion behaviour on synthetic and real-world stereo sequences", in Proc. *Int. Conf. Image Vision Computing New Zealand*, IEEE Xplore (online), 2008

[49] T. Vaudrey, A. Wedel, and R. Klette, "A methodology for evaluating illumination artifact removal for corresponding images", in Proc. *CAIP*, LNCS 5702, pages 1113-1121, 2009

[50] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for TV-L^1 optical flow", *Statistical and Geometrical Approaches to Visual Motion Analysis* (D. Cremers et al., editors), LNCS 5604, pages 23–45, 2009

[51] A. Wedel, C. Rabe, T. Vaudrey, T. Brox, U. Franke, and D. Cremers. "Efficient dense scene flow from sparse or dense stereo data", in Proc. *European Conf. Computer Vision*, 739–751, 2008

[52] X. Yang and R. Klette. "Evaluation of Motion Analysis on Synthetic and Real-World Image Sequences", in IEEE Proc. *IVCNZ*, 2010

[53] C. Zach, T. Pock, and H. Bischof, "A duality based approach for real-time TV-L^1 optical flow", in Proc. *Pattern Recognition - DAGM*, pages 214–223, 2007