

# A Methodology for Evaluating Illumination Artifact Removal for Corresponding Images

Tobi Vaudrey<sup>1</sup>, Andreas Wedel<sup>2</sup>, and Reinhard Klette<sup>1</sup>

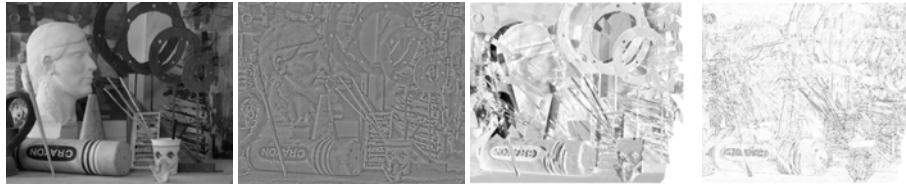
<sup>1</sup> The *.enpeda..* Project, The University of Auckland, Auckland, New Zealand

<sup>2</sup> Daimler Research, Daimler AG, Stuttgart, Germany

**Abstract.** Robust stereo and optical flow disparity matching is essential for computer vision applications with varying illumination conditions. Most robust disparity matching algorithms rely on computationally expensive normalized variants of the brightness constancy assumption to compute the matching criterion. In this paper, we reinvestigate the removal of global and large area illumination artifacts, such as vignetting, camera gain, and shading reflections, by directly modifying the input images. We show that this significantly reduces violations of the brightness constancy assumption, while maintaining the information content in the images. In particular, we define metrics and perform a methodical evaluation to identify the loss of information in the images. Next we determine the reduction of brightness constancy violations. Finally, we experimentally validate that modifying the input images yields robustness against illumination artifacts for optical flow disparity matching.

## 1 Introduction

Previous studies have shown that when using correspondence algorithms (i.e., stereo and optical flow) to provide reliable information, the results on synthetically generated data (e.g., [10]) do not compare well with results on realistic images [16]. Further studies have shown that illumination artifacts (such as shadows, reflections, and vignetting) and differing exposures have the worst effect on the matching [11]. This effect is especially highlighted in driver assistance systems (DAS), where illumination can change drastically in a short amount of time (e.g., going through a tunnel, or the “dancing light” from sunlight through trees).



**Fig. 1.** Example for removing illumination artifacts due to different camera exposure in the *Art* image (left) by using its residual component (2nd from left). The brightness difference between the plain intensity images (3rd from left) shows laminar errors. The brightness difference of the residual images (right) contains spatially distributed noise but no large area illumination artifacts.

For dealing with illumination artifacts, there are three basic approaches: simultaneously estimate the disparity matching and model brightness change within the disparity estimation [5], try to map both images into a uniform illumination model, or map the intensity images into images which carry the illumination-independent information (e.g., using colour images [9, 18]).

Using the first option, only reflection artifacts can be modelled without major computational expense. From experiments with various unifying mappings, the second option is near impossible. The third approach has more merit for research; we restrain our study to using the more common grey value images.

An example of mapping intensity images into illumination-independent images is the structure-texture image decomposition [1, 12] (an example can be seen in Figure 1). More formally, this is the concept of *residuals* [7], which is the difference between an intensity image and a smoothed version of itself. A subset of residual operators has been recently evaluated together with different matching costs in the context of stereo disparity matching in [6]. In this paper we systematically evaluate and compare residual operators as basic approach for preprocessing corresponding images, to reduce the effect of illumination variances.

The main contribution of this work is we provide a valid methodology for analysing information loss compared to illumination removal effects for an arbitrary filter. The methodology is based on first showing that information is not lost by applying the filter, using co-occurrence matrix [4] based measures. The second contribution is high-lighting these effects using correspondence images as validation. This is done by using ground truth correspondence data, comparing the differences in illumination, and summarising the information with an error metric. We go on to show that using residual images removes illumination artifacts, by using a mixture of synthetic and real-life images [3, 10]. The illumination effects are highlighted more drastically when the illumination and exposure conditions of the corresponding images are not the same. The chosen filters are the TV-L<sup>2</sup> [12], median, mean, sigma [8], bilateral [14], and trilateral filter [2]. All are effectively “edge preserving” filters, except the mean filter.

## 2 Methodology

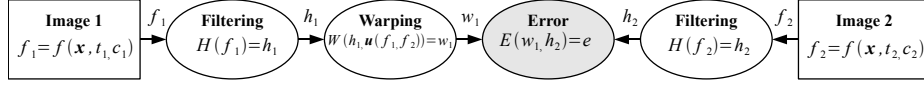
Here we define the methodology of our process. It is defined by two parts; firstly, identifying if the images loose information, and secondly, determining reduction of the effect of illumination artifacts.

**Co-occurrence Matrix and Metrics.** The co-occurrence matrix has been defined for analysing different metrics about the texture of an image [4]:

$$C(i, j) = \sum_{\mathbf{x} \in \Omega} \sum_{\mathbf{a} \in \mathcal{N} \setminus \{(0,0)\}} \begin{cases} 1, & \text{if } h(\mathbf{x}) = i \text{ and } h(\mathbf{x} + \mathbf{a}) = j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\mathcal{N} + \mathbf{x}$  is the neighbourhood of pixel  $\mathbf{x}$ ,  $\mathbf{a} \neq (0, 0)$  is one of the offsets in  $\mathcal{N}$ , and  $0 \leq i, j \leq I_{\max}$ , for maximum intensity  $I_{\max}$ .  $h$  represents any 2D image (e.g.,  $f$ ). All images are scaled  $\min \leftrightarrow \max$  for utilizing the full  $0 \leftrightarrow I_{\max}$  scale.

In our experiments we chose  $\mathcal{N}$  to be the 4-neighbourhood, and we have  $I_{\max} = 255$ . The loss in information is identified by the following metrics: *homogeneity*



**Fig. 2.** Outline of the methodology used to obtain an error image. For our study  $H = R$ .

$T_{homo}(h) = \sum_{i,j} \frac{C(i,j)}{1+|i-j|}$ , *uniformity*  $T_{uni}(h) = \sum_{i,j} C(i,j)^2$ , and *entropy*  $T_{ent}(h) = -\sum_{i,j} C(i,j) \ln C(i,j)$ . An increase in homogeneity represents the image having more homogeneous areas, an increase in uniformity represents more uniform areas, and a decrease in entropy shows that there is less information contained in the image. To get a better representation of the effect of filters, we scale the result by the original image's metric, i.e.,  $T_*(h)/|T_*(f)|$ , where  $h$  is the processed image (obviously,  $h = f$  gives a value of 1). Previous studies [15] have shown that homogeneity and entropy can define the information loss in an image. In this study, we only use the homogeneity (see results section).

**Testing Illumination Artifact Reduction.** Correspondence algorithms usually rely on the brightness consistency assumption, i.e., that the appearance of an object (according to illumination) does not change between the corresponding images. However, this does not hold true when using real-world images, this is due to, for example, shadows, reflections, differing exposures and sensor noise. It is well known, that illumination artifacts propose the biggest problem for correspondence algorithms; a recent study has shown that illumination artifacts may, in fact, be the worst type of error [11]. Figure 2 shows our proposed approach for evaluating the effectiveness of a filter. In this paper, we chose the filtering operator  $H$  to be the residual image ( $H = R$ ).

**Image Warping.** One way to highlight this (i.e., that the errors from residual images are lower than the errors obtained using the original images) is to warp one image to the perspective of the other (using ground truth) and compare the differences. The forward warping function  $W$  is defined by the following:

$$W\left(h_1(\mathbf{x}), \mathbf{u}^*(\mathbf{x}, h_1, h_2)\right) = w\left(\mathbf{x} + \mathbf{u}^*(\mathbf{x}, h_1, h_2)\right),$$

where  $h(\mathbf{x})$  is the image value at  $\mathbf{x} \in \Omega$ , and  $\mathbf{u}^*$  is the 2D ground truth warping (remapping) vector from  $h_1$  to the perspective of  $h_2$ . In practice, the warping is performed using a lookup table with interpolation (e.g., bilinear or cubic). In the stereo case,  $\mathbf{u}^*$  is the ground truth disparity map from left to right (all vertical translations would be zero). Another common example is optical flow, where  $\mathbf{u}^*$  is the ground truth flow field from the previous to the current frame.

**Image Scaling.** For the purposes of this paper,  $h$  is discrete in the functional inputs ( $\mathbf{x}$ ), but continuous for the value of  $h$  itself. For a typical grey-scale image, the information is discrete ( $0 \leq h \leq 2^n - 1 \in \mathbb{N}^2$ , where  $n$  is usually 8 or 16). However, we find it easier to represent image data continuously by  $-1 \leq h \leq 1 \in \mathbb{Q}^2$ , which takes away the ambiguity for the bits per pixel (as any  $n$ -bits per pixel image can be scaled to this domain). We scale all images to this domain using  $h(\mathbf{x}) = h(\mathbf{x}) / \max_{\mathbf{x} \in \Omega} |h(\mathbf{x})|$ .

**Error Images and Metrics.** An error image  $e$  is the magnitude of difference between two images,  $E(h, h^*) = e(\mathbf{x}) = \|\mathbf{h}(\mathbf{x}) - \mathbf{h}^*(\mathbf{x})\|$ , where, usually,  $\mathbf{h}$  is the result of a process and  $\mathbf{h}^*$  is the ground truth. For this paper, the error image is between  $\mathbf{h}^* = h_2$  and the warped image  $\mathbf{h} = W(h_1)$ .

A common error metric is the *Root Mean Squared (RMS) Error*. The problem with this metric is that it gives an even weighting to all pixels, no matter the proximity to other errors. In practice, if errors are happening in the same proximity, this is much worse than if the errors are randomly placed over an image. Most algorithms can handle (by denoising or such approaches) small amounts of error, but if the error is all in the same area, this is seen as signal. We define the *Spatial Root Mean Squared Error* (Spatial-RMS) to take the spatial properties of the error into account:

$$RMS_S(e) = \sqrt{\frac{1}{M} \sum_{\mathbf{x} \in \Omega} (G(e(\mathbf{x})))^2} \quad (2)$$

$M$  is the number of pixels in the (discrete) non-occluded (when occlusion maps are available) image domain  $\Omega$ , and  $G$  is a function that propagates the errors in a local neighbourhood  $\mathcal{N}$ . For our experiments, we chose a Gaussian error propagation using a standard deviation  $\sigma = 1$ .

**Smoothing Operators and Residuals.** Let  $f$  be any frame of a given image sequence (or stereo camera setup), defined on a rectangular open set  $\Omega$  and sampled at regular grid points within  $\Omega$ .

$f$  can be defined to have an additive decomposition  $f(\mathbf{x}) = s(\mathbf{x}) + r(\mathbf{x})$ , for all pixel positions  $\mathbf{x} = (x, y)$ , where  $s = S(f)$  denotes the *smooth component* (of an image) and  $r = R(f) = f - S(f)$  the *residual* (Figure 1 shows an example of the decomposition). We use the straightforward iteration scheme:

$$s^{(0)} = f, \quad s^{(n+1)} = S(s^{(n)}), \quad r^{(n+1)} = f - s^{(n+1)}, \quad \text{for } n \geq 0.$$

The concept of residual images was already introduced in [7] by using a  $3 \times 3$  mean for implementing  $S$ . We use the mean operator and also an  $m \times m$  median operator in this study. The other operators for  $S$  are defined below.

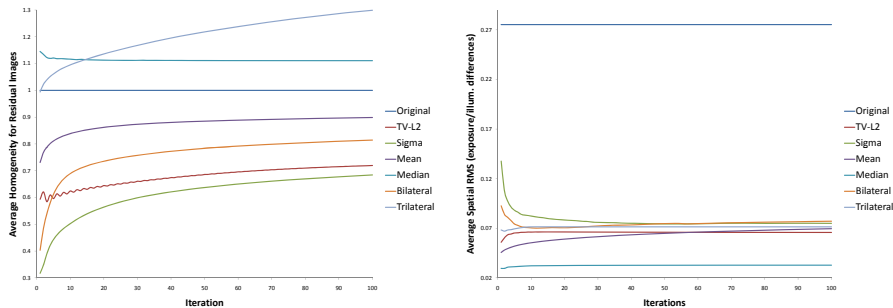
*TV-L<sup>2</sup> filter.* [12] used the definition of  $f = s + r$  (as above), where  $s$  is assumed to be in  $L^1(\Omega)$  with bounded TV (in brief:  $s \in \text{BV}$ ), and  $r$  is in  $L^2(\Omega)$ . We use the residual image from this idea as implemented and exploited in [17].

*Sigma filter.* This operator [8] is effectively a trimmed mean filter; it uses an  $m \times m$  window, but only calculates the mean for all pixels with values in  $[a - \sigma_f, a + \sigma_f]$ , where  $a$  is the central pixel value and  $\sigma_f$  is a threshold. We chose  $\sigma_f$  to be the standard deviation of  $f$  (to reduce parameters for the filter).

*Bilateral filter.* This edge-preserving Gaussian filter [14] is used in the spatial domain (using  $\sigma_2$  as spatial  $\sigma$ ), also considering changes in the colour domain (e.g., object boundaries). It therefore only takes into consideration values within a Gaussian kernel within the colour domain ( $\sigma_1$  as colour  $\sigma$ ).

*Trilateral filter.* This gradient-preserving smoothing operator [2] (i.e., it uses the local gradient plane to smooth the image) only requires the specification of one parameter  $\sigma_1$ , which is equivalent to the spatial kernel size. The rest of the parameters are self tuning.

All filters have been implemented in OpenCV, where possible the native function was used. For the TV-L<sup>2</sup>, we use an implementation (with identical parameters) as in [17]. All other filters used are virtually parameterless (except a window size) and we



**Fig. 3.** Left: average scaled homogeneity of the residual images, averaged over dataset [10]. Right: Spatial-RMS graph for the same data. Notice the huge benefit from using residual images.

use a window size of  $m = 3$  ( $\sigma_1 = 3$  for trilateral filter<sup>3</sup>). For the bilateral filter, we use color standard deviation  $\sigma_1 = I_r/10$ , where  $I_r$  is the range of the intensity values (i.e.,  $\sigma_1 = 0.2$  for the scaled images).

**Datasets.** We illustrate our arguments with the Middlebury dataset [10] and the EISATS [3] synthetic data (Set 2).

This highlights the major importance of removing illumination artifacts. For the Middlebury dataset we include both the 2005 and 2006 datasets (provided by [6, 13]). This data has 3 different exposures and 3 different illuminations (for both the left and right images). This enables us to test the brightness consistency assumption under extreme conditions. Again, we only use images with ground truth available. For the 2005 set, that includes: *Art, Books, Dolls, Laundry, Moebius, and Reindeer*. For the 2006 set: *Aloe, Baby1-3, Bowling1-2, Cloth1-4, Flowerpots, Lampshade1-2, Midd1-2, Monopoly, Plastic, Rocks1-2, and Wood1-2*. We are not interested in “good quality” situations. Therefore, we only use images with differing exposure and illumination. To do this, for each image pair, we keep the left image with illumination = 1 and exposure = 0 (as defined by [10]). But for the right image, we make use of all the differing illumination (1, 2, 3) and exposure (0, 1, 2) settings (excluding the exact same illumination = 1 and exposure = 0). This is a total of 8 different illumination/exposure combinations, for each image pair. That brings the total dataset to 216 ( $27 \times 8$ ).

### 3 Experimental Results

A previous study, has already pointed out that the results for slight illumination artifacts are improved using residual images [15]. We now show that these results get even better when illumination is a major issue (not just a minor one).

**Co-occurrence Metrics.** This subsection demonstrates that the important information for correspondence algorithms is contained in the residual image  $r$ . The residual image is, in fact, an approximation of the high frequencies of the image, and the smoothed image  $s$  is an approximation of a low-pass filter. Obviously, by iteratively running a smoothing filter, you will get a more and more smoothed image (i.e., you will be getting lower and lower frequencies, thus reducing the higher frequencies). In [15] the metrics were shown to represent this effect accurately.

<sup>3</sup> The authors thank Prasun Choudhury (Adobe Systems, Inc.) and Jack Tumblin (EECS, Northwestern University), for their implementation of the trilateral filter.

# Its.	1		40		Time / iteration		Rank	
	Ave.	S.D.	Ave.	S.D.	$470 \times 370$	$752 \times 480$	$T_{homo}$	RMS
Original	0.282	0.136	0.282	0.136	-	-	5	7
TV-L <sup>2</sup>	0.068	0.028	0.080	0.030	30 ms	60 ms	2	3
Sigma	0.168	0.036	0.090	0.023	30 ms	100 ms	1	6
Mean	0.055	0.023	0.072	0.025	1 ms	2 ms	4	2
Median	0.039	0.020	0.041	0.022	7 ms	15 ms	6	1
Bilateral	0.113	0.021	0.086	0.026	160 ms	340 ms	3	5
Trilateral	0.085	0.017	0.082	0.016	5,000 ms	11,000 ms	7	4

**Table 1.** Average (Ave.) and Standard Deviation (S.D.) for the methodology performed on dataset [10]. Average running times per iteration are also included (right) for two image resolutions. The rank of the filters is also given for both evaluations.

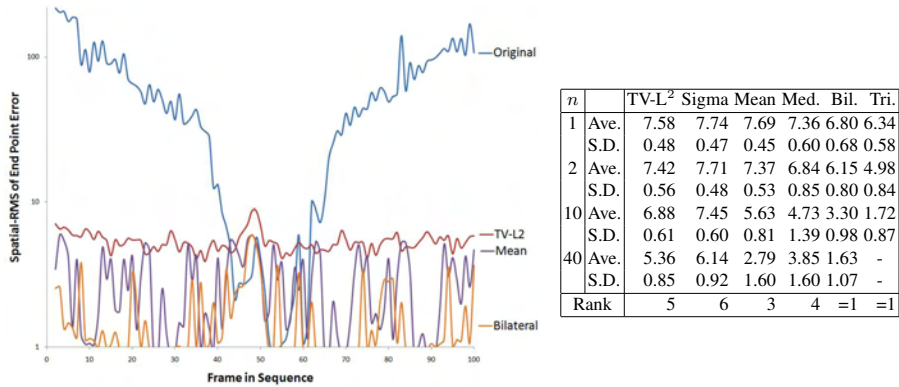
The residual of an image is an approximation of the high frequencies of the image, so the information should not be reduced. We average the co-occurrence metric results over dataset [10] to highlight this (Figure 3). This graph shows that the residual images do not lose information, in fact the homogeneity is slightly reduced (except for the trilateral and median filter). The increase in information could be seen as noise from the filter, or an increase in emphasis of the high-frequencies.

**Illumination Differences.** This subsection uses again the dataset [10]. A qualitative example of error images  $e$  can be seen in Figure 1. This specific error image is generated using the *Art* right image, with illumination and exposure both equal to 1 (left image is 1 and 0, respectively). The image is from [6], and has ground truth available (warping from left to right). The original error image (left) clearly shows how increasing the exposure (250 to 1000 ms) has very big consequences on the illumination differences between the left and right image. The error image using the TV-L<sup>2</sup> residual (right) reduces the error dramatically. Furthermore the magnitudes of the maximum errors are less; the original image is 1.83 and the TV-L<sup>2</sup> residual image is 1.25.

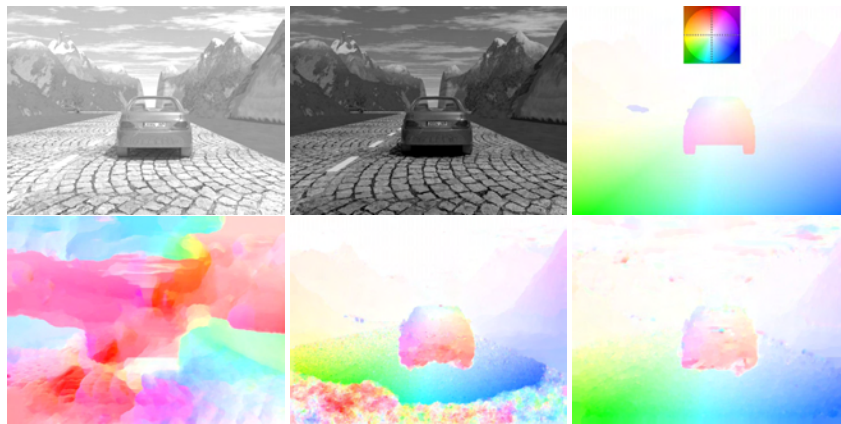
See again Figure 3. The trilateral filter was stopped at iteration 10. It is immediately obvious that the original images are far worse than residual images, around 3 times worse on average. This again highlights that with extremely different exposures and illuminations, the residual images provide the best information for matching.

Since most of the filters stabilize around iteration 40 (TV-L<sup>2</sup>, sigma, bilateral, and median), we have presented statistical results of the RMS after 1 and 40 iterations. These results are shown in Table 1. You can see from these results that all the statistics for the original images are higher than any of the filters. The mean, trilateral, and median filter seem to be the most robust; showing the lowest standard deviation. The TV-L<sup>2</sup>, mean, and median filters have the best average. The timing information provided in this table is the average time per iteration, on two sizes of images ( $470 \times 370$ ,  $752 \times 480$  pixel resolution), this is to highlight the scalability of the filters. The tests were under Windows, the CPU was an Intel Core 2 Duo 3GHz (multi-core processing not exploited), with 4GB memory.

**Optical Flow on EISATS Dataset.** For this subsection, we computed optical flow using TV-L<sup>1</sup> optical flow [19] (one of the top performing algorithms), on the EISATS dataset [3]; see [16] for Set 2. We altered the data to resemble illumination differences in time, as performed in [11]; the differences start high between frames, then go to zero at frame 50, then increase again. The flow field is computed using  $U(h_1, h_2) = \mathbf{u}$ . This



**Fig. 4.** Left: Flow end point error over entire EISATS sequence using number of filter iterations  $r^{(40)}$ , graph is logarithmically scaled ( $\log_{10}$ ) (only TV-L<sup>2</sup>, mean, and bilateral shown). Right: results using different number of filter iterations  $r^{(n)}$ , the original average (Ave.) is 61, and standard deviation (S.D.) is 53 (much higher than the rest).



**Fig. 5.** Top row: frame 1 (left) and 2 (middle) from EISATS scene. Ground truth flow with key (HSV circle for direction, saturation for vector length) is shown on the right. Bottom row: optical flow results using; original images (left) and residual images, TV-L<sup>2</sup> (middle) and trilateral (right), respectively.

is to show that a residual image  $r$  provides better data for matching, than for the original image  $f$ . Figure 5 shows an example of this effect, obviously the residual image vastly improves optical flow results. We calculated the end-point-error using the error image  $e = E(\mathbf{u}, \mathbf{u}^*)$  and Spatial-RMS.

We computed the flow using  $U(r_1^{(n)}, r_2^{(n)})$  with  $n = 1, 2, 10$ , and 40 to show how each filter behaves. The results are compared to optical flow on the original images  $U(f_1, f_2)$ . The results can be seen in Figure 4. It is immediately obvious that the original image results are much worse quality than the residuum results. Residual images are more robust to illumination differences than standard images.

## 4 Conclusions and Future Research

We have identified a methodology for analysing the effect of illumination reducing filters using numerical comparisons. We went on to show that the results for this test do align with the optical flow performance, on a scene with drastic illumination variation. The tests showed that generating a simple mean residual image, produces acceptable improvements, while being the fastest (and easiest) to implement. Future work should test the limits of the proposed methodology. Other smoothing algorithms and illumination invariant models need to be tested. Finally, a larger dataset can be used to further verify the illumination artifact reducing effects of residual images.

## References

1. Aujol, J. F., Gilboa, G., Chan, T., and Osher, S.: Structure-texture image decomposition - modeling, algorithms, and parameter selection. *Int. J. Computer Vision*, **67**:111-136 (2006)
2. Choudhury, P., and Tumblin, J.: The trilateral filter for high contrast images and meshes. In Proc. *Eurographics Symp. Rendering*, pages 1–11 (2003)
3. .enpeda.. dataset 2 (EISATS): <http://www.mi.auckland.ac.nz/EISATS>
4. Haralick, R. M., and Bosley, R.: Texture features for image classification. In Proc. *ERTS Symposium*, NASA SP-351, pages 1219-1228 (1973)
5. Haussecker, H. and Fleet, D. J.: Estimating optical flow with physical models of brightness variation. *IEEE Trans. Pattern Analysis Machine Intelligence*, **23**:661–673 (2001)
6. Hirschmüller, H., and Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Analysis Machine Intelligence*, to appear.
7. Kuan, D. T., Sawchuk, A. A., Strand, T. C., and Chavel, P.: Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Trans. Pattern Analysis Machine Intelligence*, **7**:165–177 (1985)
8. Lee, J.-S.: Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, **24**:255–269 (1983)
9. Mileva, Y., Bruhn, A. and Weickert, J.: Illumination-robust variational optical flow with photometric invariants. In Proc. *Pattern Recognition - DAGM*, pages 152–162 (2007)
10. Middlebury dataset: <http://vision.middlebury.edu/stereo/data/>
11. Morales, S., Woo, Y. W., Klette, R., and Vaudrey, T.: A study on stereo and motion data accuracy for a moving platform. Technical report, MI-tech-32, <http://www.mi.auckland.ac.nz/>, University of Auckland (2009)
12. Rudin, L., Osher, S., and Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D*, **60**:259–268 (1992)
13. Scharstein, D., and Pal, C.: Learning conditional random fields for stereo. In Proc. *IEEE Conf. Computer Vision and Pattern Recognition*, (2007)
14. Tomasi, C., and Manduchi, R.: Bilateral filtering for gray and color images. In Proc. *IEEE Int. Conf. Computer Vision*, pages 839–846 (1998)
15. Vaudrey, T., and Klette, R.: Residual images remove illumination artifacts. In Proc. *Pattern Recognition - DAGM*, to appear (2009)
16. Vaudrey, T., Rabe, C., Klette, R., and Milburn, J.: Differences between stereo and motion behaviour on synthetic and real-world stereo sequences. In Proc. *IEEE Image and Vision Conf. New Zealand*, Digital Object Identifier 10.1109/IVCNZ.2008.4762133 (2008)
17. Wedel, A., Pock, T., Zach, C., Bischof, H., and Cremers, D.: An improved algorithm for TV-L<sup>1</sup> optical flow. In Post Proc. *Dagstuhl Motion Workshop*, to appear (2009)
18. van de Weijer, J. and Gevers, T.: Robust optical flow from photometric invariants. In Proc. *Int. Conf. on Image Processing*, pages 1835–1838 (2004)
19. Zach, C., Pock, T., and Bischof, H.: A duality based approach for realtime TV-L<sup>1</sup> optical flow, In Proc. *Pattern Recognition - DAGM*, pages 214–223 (2007)