

The Gradient - A Powerful and Robust Cost Function for Stereo Matching

Simon Hermann and Tobi Vaudrey

.enpeda.. group, Department of Computer Science, University of Auckland, New Zealand

Abstract—Using gradient information for a pixel-based cost function for stereo matching has lacked adequate attention in the literature. This paper provides experimental evidence to show that the gradient as a data descriptor outperforms other pixel-based functions such as absolute differences and the Birchfield and Tomasi cost functions. The cost functions are tested against stereo image datasets where ground truth data is available. Furthermore, analysing the effect of the cost functions when exposure and illumination settings are different between the left and right camera is analysed. Not only has the performance of the cost functions been analysed, but also analysis into “why” one cost function is better than another. The analysis tests the global and spacial optimality of the cost function, showing that the gradient information returns stronger minima than the other two. These results are aimed at future research towards the design of a new smoothness prior that also depends on the characteristics of the employed cost function. This paper shows that the gradient is a simple, yet powerful, data descriptor that shows robustness to illumination and exposure differences, but is often overlooked by the stereo community.

I. INTRODUCTION AND RELATED LITERATURE

Stereo matching can be defined as an energy minimization problem. The energy to be minimized is usually a combination of a *data* and a *smoothness term*. The data term incorporates a local matching cost into the energy, that usually relies entirely on a similarity measure based on pixel information within a certain neighbourhood of tested pixels. A smoothness term adds an additional cost in order to ensure piecewise smooth disparities to solve standard stereo matching problems that occur in homogeneous areas and at depth discontinuities or occlusions. In those areas the data term alone is not sufficient, due to its inherent locality.

However, there are other problems in stereo matching that affect, primarily, the data term, such as illumination differences (between stereo images) and noise (obtained during image acquisition). Those effects can have major influence on the image data and therefore on the quality of the matching cost itself. Especially when it comes to real world image sequences [4]. Recent studies [16], [17], [1] suggest decomposing the input images into a structure and a texture component. The texture component tends to be robust against illumination changes. Additionally, bilateral filtering [15] can be used to reduce the noise in an image, since it applies a Gaussian smoothing along with a spatial constraint that preserves image discontinuities.

This study was financially supported by the German Academic Exchange Service, DAAD.

Pixel-based cost functions are used as similarity measures by state-of-the-art stereo algorithms such as *belief propagation* (BP) [5], *semi-global-matching* (SGM) [6], or *graph cuts* (GC) [3]. Those algorithms define the energy that needs to be minimized over a disparity map D as follows:

$$E(D) = E_{\text{data}}(D) + E_{\text{smooth}}(D) \quad (1)$$

The smoothness term is driven by a prior model and decides whether to keep the disparity map consistent within a local neighbourhood, or to allow a disparity jump, when a depth discontinuity is likely. Those prior models are often based on the input image data (e.g., first order derivatives).

The motivation of this paper is to generate a prior model that is not only driven by the data, but also by the characteristics of the chosen cost function. Of course this is only possible if different cost functions do actually have a distinctive characteristic. In order to avoid any bias, it is crucial to analyse the cost functions separately from any optimization technique (e.g., BP, SGM, or GC). Evaluation of *good pixel percentages* on a *winner takes all* basis is clearly important to evaluate quality of a cost function, but it is insufficient to derive any characteristics.

In this paper, four pixel-based cost functions are evaluated, namely the *absolute difference* (AD) cost function, the sampling insensitive cost function [2] by Birchfield and Tomasi (BT), the gradient absolute difference (GRAD), and an extended gradient difference (GRAD-X). While the performance under illumination changes of AD and BT have been recently studied [7], [8], the gradient was unfortunately not part of those evaluations. In [11] however, the gradient was employed in a similarity measure that was a weighted sum of AD and the gradient (using forward differences) accumulated over a 3×3 window. Another study introduces the gradient when creating a similarity measure using a multiplicative contribution of SAD, NCC and the gradient [10]. The contribution of the gradient was shown to provide a more reliable cost function. However, none of those studies were using or analysing the gradient concept isolated from other cost functions to determine the performance contribution of the gradient.

In the methodology presented below, the four cost functions are evaluated under differing illumination and exposure settings. Other than a combination, the pure gradient is used as the cost function. The performance comparison is done without any optimization strategy. We recall, that the goal is (at some stage) to identify characteristics of cost functions to employ them in an optimization prior model. Looking at how

many true disparities are identified when employing a winner takes all approach is not sufficient for deriving characteristics. We therefore define two cost function descriptors that should give an intuition about the property of the identified minimum costs, e.g., is the identified minimum strong or weak?

The analysis is expanded by evaluating those cost functions when accumulated over a small 3×3 window, as is commonly employed for disparity estimation. This usually provides a performance gain in terms of quality of the match, while taking a performance degradation in terms of computation time. However, as shown in the experimental evidence, the gradient clearly outperforms the other cost functions in all measures for all illumination and exposure settings, which supports our agenda to promote the gradient as a strong candidate to serve as a matching cost in disparity estimation.

The paper is structured as follows. The following section introduces the pixel-based matching costs. This is followed by the methodology, data sets, and testing measures used for evaluation. This leads onto a discussion of the experimental results, which is then finalised by a conclusions section.

II. PIXEL-BASED COST FUNCTIONS

In a rectified stereo image pair we consider a *base* and a *match* image. The base image is assumed to be the left image L . The match image R is usually the right image. The images are of same size with a dimension of $n \times m$ (width times height) pixels within the image domain Ω . We only consider intensity images (ignoring colour) in this paper with values between 0 and I_{\max} . Any cost function Γ defines a global mapping $\Gamma(L, R) = C$ that takes rectified stereo images L and R as input, and outputs a 3D cost matrix C with elements $C(i, j, d)$, representing the cost when matching a pixel at (i, j) in L with a pixel at $(i - d, j)$ in R , for any relevant disparity d in the range $[1, d_{\max}] \subset \mathbb{N}$ (zero is used for an “invalid” disparity, such as for occlusion). The ranges for i and j are $[0, n] \subset \mathbb{N}$ and $[0, m] \subset \mathbb{N}$, respectively. We simplify notation as we are working with rectified images (epipolar lines are aligned to the x -axis), and we consider a fixed image row j in both the base and match image. Let p_i denote a pixel location in L at column i . Let L_i be the value at this location in the base image; q_{i-d} denotes the pixel location $(i - d, j)$ in the match image R with intensity R_{i-d} . The cost can be abbreviated to omit the row $C(i, d)$.

A *pixel-based* cost function determines the matching cost for a disparity on the basis of a descriptor that is defined for one single pixel. Pixel-based cost functions can easily be extended to window-based matching costs by integrating pixel-based costs within a certain (usually square) neighbourhood.

The following four cost functions are categorized in the literature as pixel-based cost functions. They are presented briefly with an additional computational cost estimation. For simplicity, a multiplication operation is considered equal to an addition or subtraction and also to a sign switch or a min/max evaluation.

A. Absolute Difference

The *absolute difference* (AD) of base and match pixel is the simplest and cheapest (in terms of computational cost) measure:

$$C_{\text{AD}}(i, d) = |L_i - R_{i-d}| \quad (2)$$

The calculation cost at each pixel for every valid disparity is one subtraction and one sign switch, which gives us the total computational cost of:

$$2d_{\max}nm \quad (3)$$

There is no preprocessing possible in order to speed up calculation.

B. Birchfield and Tomasi Metric

Another commonly used pixel-based cost function (BT) was presented in [2]. The intention of this cost is to be more insensitive to image sampling by using absolute differences between identified extrema of interpolated intensities. In a first step, intensities in L and R are interpolated using either a previous or a subsequent pixel along the epipolar line. For example, $L_{i-1/2} = \frac{1}{2} (L_i + L_{i-1})$ is an interpolated value at p_i , with respect to the previous pixel. $\mathcal{L}_i = \{L_{i-1/2}, L_i, L_{i+1/2}\}$ is a set containing the intensity at p_i in L as well as the interpolated intensities with previous and subsequent pixels. Analogously, \mathcal{R}_{i-d} is a set containing the intensity at q_{i-d} in R as well as the interpolated intensities with previous and subsequent pixels. The BT cost function is then as follows:

$$C_{\text{BT}}(i, d) = \min\{a, b\} \quad (4)$$

where

$$a = \max\{L_i - \max(\mathcal{R}_{i-d}), \min(\mathcal{R}_{i-d}) - L_i, 0\} \quad (5)$$

$$b = \max\{R_{i-d} - \max(\mathcal{L}_i), \min(\mathcal{L}_i) - R_{i-d}, 0\} \quad (6)$$

The cost here is somewhat more expensive. At each pixel the interpolation is needed with the successor and the predecessor along the epipolar line. This can however, be pre-calculated to speed up time. Each interpolation required two multiplications and one addition. There are about¹ $n \times m$ interpolations necessary. At each pixel we have four subtractions along with seven min/max decisions. The total computational cost is therefore

$$3nm + 11d_{\max}nm = (11d_{\max} + 3)nm \quad (7)$$

This is however a lower estimate than found in practice, as min/max operations have a much higher computational cost when compared to multiplication, sign checks, and addition.

¹There are only $(n-1) \times m$ interpolations required, but we keep it simple.

C. The Gradient

This cost function [11] employs the spatial distance of the end points of the gradient vectors as the similarity measure. It therefore falls into a different category than the previous two functions, because it is based on first order approximations rather than on intensity data. It is defined as:

$$C_{\text{GRAD}}(i, d) = |\nabla L_i - \nabla R_{i-d}|_1 \quad (8)$$

where ∇ is estimated using central differences² and $|\cdot|_1$ is the L^1 norm.

The computational costs are as follows. For the gradient calculation we need to calculate the gradient at each pixel using two subtractions (central differences) in a preprocessing step. Calculating the costs for each disparity needs two sign checks, one addition, and two subtractions. This results in the total computational cost of:

$$2nm + d_{\text{max}}nm = (5d_{\text{max}} + 2)nm \quad (9)$$

D. A Gradient Extension

The gradient operator in its standard form takes a scalar field (like an image) and converts it into a 2D vector field, where the vectors consist of the derivatives in horizontal x and vertical y directions, i.e., $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)^T$. We use central differences in x and y approximating the derivatives. However, this incorporates only the 4-neighbourhood of a pixel. We may as well define central differences in diagonal direction to include the whole 8-neighbourhood. Thus we extend the gradient formulation to a 4D operator

$$\nabla_4 = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial(x + \frac{\pi}{4})}, \frac{\partial}{\partial(y + \frac{\pi}{4})} \right)^T \quad (10)$$

where $\partial(x + \frac{\pi}{4})$ denotes the approximate changes of the scalar field when changing the reference direction by 45 degrees (i.e., $\frac{\pi}{4}$). We define

$$C_{\text{GRAD-X}}(i, d) = |\nabla_4 L_i - \nabla_4 R_{i-d}|_1 \quad (11)$$

The computational cost would be twice that of the standard gradient plus one extra addition. So the total cost in our reference system is

$$(11d_{\text{max}} + 4)nm \quad (12)$$

We included the computational cost evaluation to keep a certain level of fairness, when we compare the performance of the cost functions. Clearly, using central differences to obtain the gradient information incorporates a one pixel neighbourhood in both the horizontal and vertical direction. Therefore, the information from four pixels is used to describe the matching cost at each pixel, or 8 pixel when looking at the extended version. The same happens with BT. Because interpolation is performed on intensities forward and backward along i , three pixels are incorporated for the BT descriptor. It is not

²For the experiments in this paper, central differences are used, however, other gradient operators can be used, and may provide better, depending on the image data used.

surprising that extending the area of influence for a descriptor may have a positive impact on the matching quality, but this also has a negative impact on computational costs. These two issues need to be considered when comparing results in the discussion.

E. Winner Takes All Matching

To calculate the optimal disparity map, a winner takes all approach is used. This approach finds the minimum cost for C at every pixel (i, j) . Formally, this is calculated as follows:

$$D(i, j) = \min_{d \in [1, d_{\text{max}}]} \{C(i, j, d)\} \quad (13)$$

If there is no unique minimum (i.e., if there are two or more costs that have the same minimum value) then the optimal disparity $D(i, j)$ is set to the disparity value of the first minimum cost that is found. It would also be possible to handle these ambiguities (non-unique situations). We choose not to, since ambiguities are rarely considered in optimization techniques so far. The high ambiguity however, could be a distinctive characteristic for the AD and BT cost functions. Considering this in the prior model of a subsequent optimization technique may improve results (e.g., ignoring the data term when ambiguities are detected).

III. METHODOLOGY AND DATASETS

Illumination issues have been proven to cause major issues when it comes to stereo matching and may, in fact, be the worst type of noise for stereo matching [14]. The methodology here tests the cost functions under normal lighting conditions, as well as with different exposures and illuminations between the left and right camera. Furthermore, the tests are performed using single pixel data (1×1 window) and a 3×3 window. Incorporating more data is likely to provide better results overall.

The dataset and methodology are presented below. Not only is the best performer analysed (comparing estimates to ground truth), but also the properties of the cost function that aims to answer to the question “why is this cost function better?”

A. Dataset

Stereo images where ground truth is available is used to evaluate the cost functions. Furthermore, the stereo images are recorded under different lighting and exposure settings, to provide test data where illumination/exposure could cause issues. Such data is available, for example, from the Middlebury Stereo Vision webpage [8]. Figure 1 shows an example (in this case, the *Art* images) of the dataset used in this paper.

The cost functions are tested against the following images from the dataset: *Art*, *Books*, *Dolls*, *Laundry*, *Moebius*, and *Reindeer*. For each image pair used, the base image is using the exposure setting of 1 and illumination setting of 2. The left image is kept at this setting, but both illumination and exposure are varied in the right hand image. For each measure (outlined below) three tests are performed using different right hand images:

- 1) Identical lighting conditions (exp. 1, illum. 2)



Fig. 1. Top row: shows the left reference (base) and right (match) image, respectively, of the *Art* input pair under identical lighting conditions. Bottom row: shows the right image with illumination and exposure change, respectively.

- 2) Illumination difference (exp. 1, illum. 1)
- 3) Exposure difference (exp. 0, illum. 2)

In the experiments we see how those cost functions based purely on intensity fail terribly, which is obvious because of the high reliance on the intensity consistency between images.

B. Good Pixel Percentage

Let T be the ground truth image of the corresponding data set where T_i encodes the *true disparity* at pixel p_i . The *good pixel percentage* is defined below:

$$GPP = 100\% \times \frac{1}{|\Omega|} \sum_{(i,j) \in \Omega} \begin{cases} 1, & \text{if } |D(i,j) - T_i| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where Ω is the set of all pixels where $T_i \neq 0$, as 0 is used to identify occlusions, and $|\Omega|$ is the cardinality.

In other words, if the minimum cost disparity is within one disparity distance of the ground truth, it is a good pixel. This measure was used as the base test for answering the question: “which cost function provides the best results?”

C. Local Data Term Descriptor

A local data term descriptor indicates the strength of the minimum cost, in a local context. This approach is closely related to confidence measures [9]. The chosen measure used here is the angle α between neighbouring pixels. This measure helps to answer the question of “why does the function perform better?”

$$\alpha = \arccos \left(\frac{\mathbf{v}^1 \cdot \mathbf{v}^2}{|\mathbf{v}^1|_2 |\mathbf{v}^2|_2} \right) \quad (15)$$

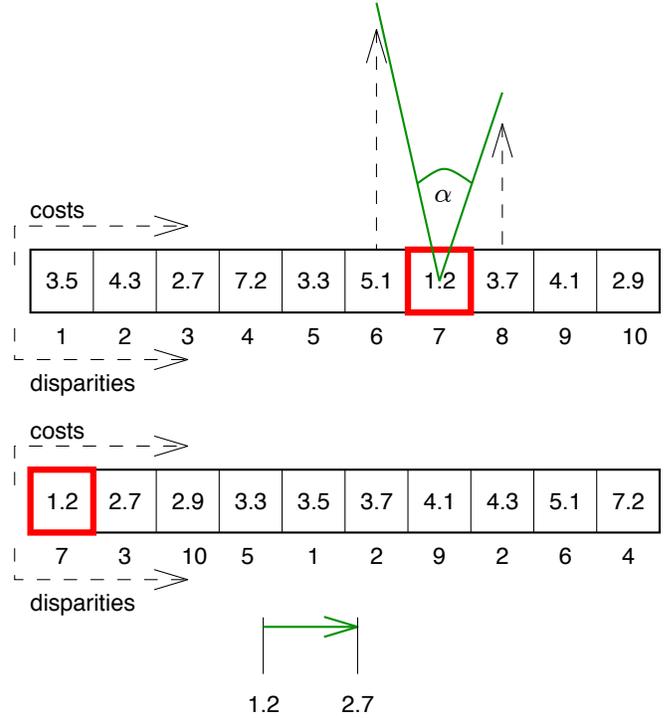


Fig. 2. Visual interpretation of the data descriptors. Top: local data descriptor, which takes into account the neighbouring left and right pixels, and calculates the angle. Bottom: global data descriptor, which reorders the cost values globally, and calculates the relative difference between the bottom two costs vs. the cost range.

where \cdot denotes the dot product of two vectors and

$$\mathbf{v}^1 = (-1, C(i, d_{w-1}) - C(i, d_w))^T \quad (16)$$

$$\mathbf{v}^2 = (1, C(i, d_{w+1}) - C(i, d_w))^T \quad (17)$$

where d_w is the index (position) of the optimal disparity. To summarise the data over an image, the angle α is averaged over Ω .

We calculate the angle in degrees to have a better visual understanding of the results (compared to radians values).

The top half of Figure 2 shows an example of this data descriptor. Here, a disparity range of $1, \dots, d_{max}$ where $d_{max} = 10$ is assumed. In this case, the minimum disparity index d_w is 7. The angle α is calculated between $6 \rightarrow 7 \rightarrow 8$. The larger the angle, the less descriptive the cost function in a local context.

Examples of the pixel-wise local descriptor are shown in Figure 3. Here, the *Laundry* stereo pair is used for the input images.

D. Global Data Term Descriptor

This measure is used to analyse how definitive the minimum cost is in a global context (at each pixel). In order to test this, we sort the costs at each pixel, then measure the absolute distance between the lowest two costs.

This measure further helps answer the question of “why does the function perform better?” It shows how descriptive

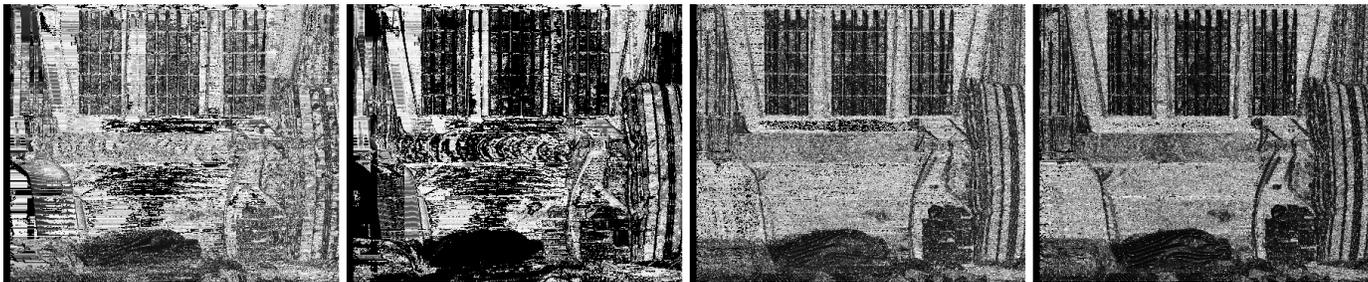


Fig. 3. Angle based local descriptor at each pixel for the *Laundry* stereo pair. The images are intensity coded with low to high representing the size of the angle (results are scaled to 8-bit for visualisation). The lower the angle, the more descriptive the cost at that point. Left to right shows results for AD, BT, GRAD and GRAD-X, respectively.

the cost function is, when comparing it globally within the disparity range. This is defined formally below.

Let $C(i, j, d)_{0 < d \leq d_{max}}$ be the finite sequence of all costs at a pixel location (i, j) for the chosen disparity range. We can order this sequence from the lowest to the highest cost. We chose the absolute distance from the lowest to the second lowest cost as a global descriptor for the cost functions at a specific pixel.

The bottom half of Figure 2 provides an example of this data descriptor. Here, the disparity range of $1, \dots, d_{max}$ is sorted (ascending) by cost. Then the difference between the two lowest costs is taken. This gives a absolute descriptor of the cost distance between first and second lowest cost.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Summarising the paper so far, the experiments were run for each pixel-based cost functions (with two different window sizes), against the Middlebury data set with three different settings: no lighting changes, illumination difference, and exposure difference. For each image pair, the good pixel percentage was calculated, along with the local and global data descriptors. The results are summarised in Figure 4. For GPP and the global descriptor a higher value is better than a lower value, and for the local angular descriptor a lower value is better than a higher one.

When looking at the 1×1 window good pixel percentage results for the cost functions (top row of Figure 4) the first thing we notice is that both GRAD and GRAD-X outperform BT and AD, for all situations. This makes sense if we think about the information captured from each cost function. GRAD-X is using eight pixels of information, GRAD is using four, BT is using three, and AD is using one. Surprisingly, AD is higher than BT, thus AD would be an obvious choice between the two as it also has a lower computational cost. However, the concept of BT may become of relevance when applied within a subsequent optimization scheme; as the cost function was introduced in combination with one [2] and only then used with other optimization techniques. We notice that as soon as there are illumination issues (two right graphs) AD and BT become almost unusable; this is expected because the constant intensity assumption is violated. Illumination issues degrade the quality of GRAD and GRAD-X, but not by too

much. From this we can declare that gradient based operators provide better results than intensity based measures, and are also robust to illumination conditions.

Analysing the local and global descriptor for the 1×1 cost functions provides some interesting results. The local descriptor ordering is the same as for GPP; GRAD-X has the lowest angle (thus strongest descriptor), followed by GRAD, AD, then BT. There is no way to compare the global descriptors against each other, but the consistent range of values for each cost function will guide decisions for choosing smoothing priors. Illumination changes do not have a consistent effect on the descriptors. The aim of analysing these descriptors is to help drive the penalty choice globally (using the global descriptor) along with a spatially adapting penaliser (using the local descriptor).

Continuing with the local descriptors, Figure 3 shows the pixel-wise data descriptor. Visual inspection shows that angles in the gradient cost functions seem to be more consistent, providing clear object boundaries and low information on homogeneous areas. This is favourable for an optimisation scheme if spatially varying penalisers are to be used. However, this has to be further analysed in order to exploit this information for incorporating this into optimization techniques.

Results get more interesting when we look at the results for good pixel percentages on the 3×3 window cost functions. For the three different illumination conditions we notice that AD and BT are overlapping, and so are GRAD-X and GRAD. This makes sense if we remember that AD and BT are both intensity based, so summarising the data over a small window negates the interpolation of BT. Similarly, GRAD-X has two extra angles of gradient when compared to GRAD, but this extra information is almost negligible on a window as only four extra pixels of information are being included. Results are improved for all cost functions, with the intensity-based cost functions showing some robustness to lighting change, but not exposure change. GRAD-X only slightly out performs GRAD; if you take into consideration the computational difference (GRAD-X is twice that of GRAD) then GRAD would probably be the best choice here.

Looking at the local and global descriptor for the 3×3 window does not provide any extra information. The results are similar when compared to a 1×1 window.

Summarising the results, there is one major characteristic difference between the two groups of cost functions: gradient-based (GRAD, GRAD-X) and intensity-based (AD,BT). We can say from these experiments that gradient-based cost functions do not only provide better results, but have stronger local minima and show high robustness to illumination changes. The GRAD-X operator provides better results for single pixel operators, but this advantage is reduced dramatically for higher window sizes (where GRAD would be favourable due to the much lower computational cost). Unfortunately, ordering based cost functions like census [18] were not included, but are planned for future extensions with analysis of optimization techniques.

V. CONCLUSIONS

This paper provided experimental evidence on the differences between gradient-based and intensity-based cost functions. The study included stereo data where illumination causes an issue for most disparity estimation algorithms. The gradient-based functions provided better results, that were also robust against the illumination changes. The top performer for single pixel cost functions was the new gradient operator introduced in this paper. However, the advantage of the operator is lost when using larger window sizes. In addition to the good pixel percentage, a local and global descriptor was introduced to guide a user on the penalisers that should be used in an optimisation scheme.

Future works include analysing ordering based cost functions, such as the census function. Furthermore, these cost functions need to be incorporated into different optimisation strategies to analyse how the local and global descriptors can help guide the optimisation.

REFERENCES

- [1] Aujol, J.-F., Gilboa, G., Chan, T. and Osher, S.: Structure-texture image decomposition – modeling, algorithms, and parameter selection, *Int. J. of Computer Vision*, 67:111-136, 2006.
- [2] Birchfield S. and Tomasi C.: Depth Discontinuities by Pixel-to-Pixel Stereo. *Int. J. of Computer Vision*, volume 35(3), pages 269–293, 1999.
- [3] Boykov Y., Veksler O., and Zabih R.: Fast Approximate Energy Minimization via Graph Cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- [4] EISATS (*enpeda.. Image Sequence Analysis Test Site*): <http://www.mi.auckland.ac.nz/EISATS>
- [5] Felzenszwalb P. F. and Huttenlocher D.: Efficient belief propagation for early vision. *Int. J. of Computer Vision*, volume 70, pages 41–54, 2006.
- [6] H. Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *IEEE Conf. Computer Vision Pattern Recognition*, volume 2, pages 807–814, 2005.
- [7] Hirschmüller H. and Scharstein D.: Evaluation of Cost Functions for Stereo Matching. In *IEEE Computer Science Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, Minneapolis, June 2007.
- [8] Hirschmüller H. and Scharstein D.: Evaluation of Stereo Matching Costs on Images with Radiometric Differences. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31(9):1582–1599, 2009.
- [9] Hu, X. and Mordohai, P.: Evaluation of stereo confidence indoors and outdoors, In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, TBA, 2010.
- [10] El-Mahassani, E.D.: New Robust Matching Cost Functions for Stereo Vision. In *Proc. 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA)*, pages 144-150, 2007.
- [11] Klaus, A., Sormann, M. and Karner, K.: Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. In *Proc. Intl. Conf. Patt. Rec. (ICPR 2006)*, volume 3, pages 15 – 18, 2006.
- [12] Scharstein D., Pal C.: Learning Conditional Random Fields for Stereo. In *IEEE Computer Science Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*, Minneapolis, June 2007.
- [13] Scharstein, D., Szelinski, R.: Middlebury Stereo Vision Page: <http://vision.middlebury.edu/stereo/>
- [14] Morales,S., Woo, Y.W., Klette, R. and Vaudrey, T.: A study on stereo and motion data accuracy for a moving platform, In *Proc. FIRA RoboWorld Congress - Advances in Robotics*, Springer, pages 292-300, 2009.
- [15] Tomasi C. and Manduchi R.: Bilateral Filtering for Gray and Color Images. In *Proc. IEEE International Conference on Computer Vision*, pages 836– 846, 1998.
- [16] Vaudrey, T. and Klette R.: Residual Images Remove Illumination Artifacts! In *Lecture Notes in Computer Science (DAGM)*, volume 5748, pages 472–481, Jena, Germany, 2009.
- [17] Vaudrey, T., A. Wedel, and Klette R.: A methodology for evaluating illumination artifact removal for corresponding images. In:*CAIP, LNCS* volume 5702, pages 1113 –1121, Springer, Berlin, 2009.
- [18] Zabih, R. and Woodfill, J.: Non-parametric Local Transform for Computing Visual Correspondence. IN *Proc. ECCV 1994*, volume 2, pages 151–158, Stockholm, Sweden, 1994.

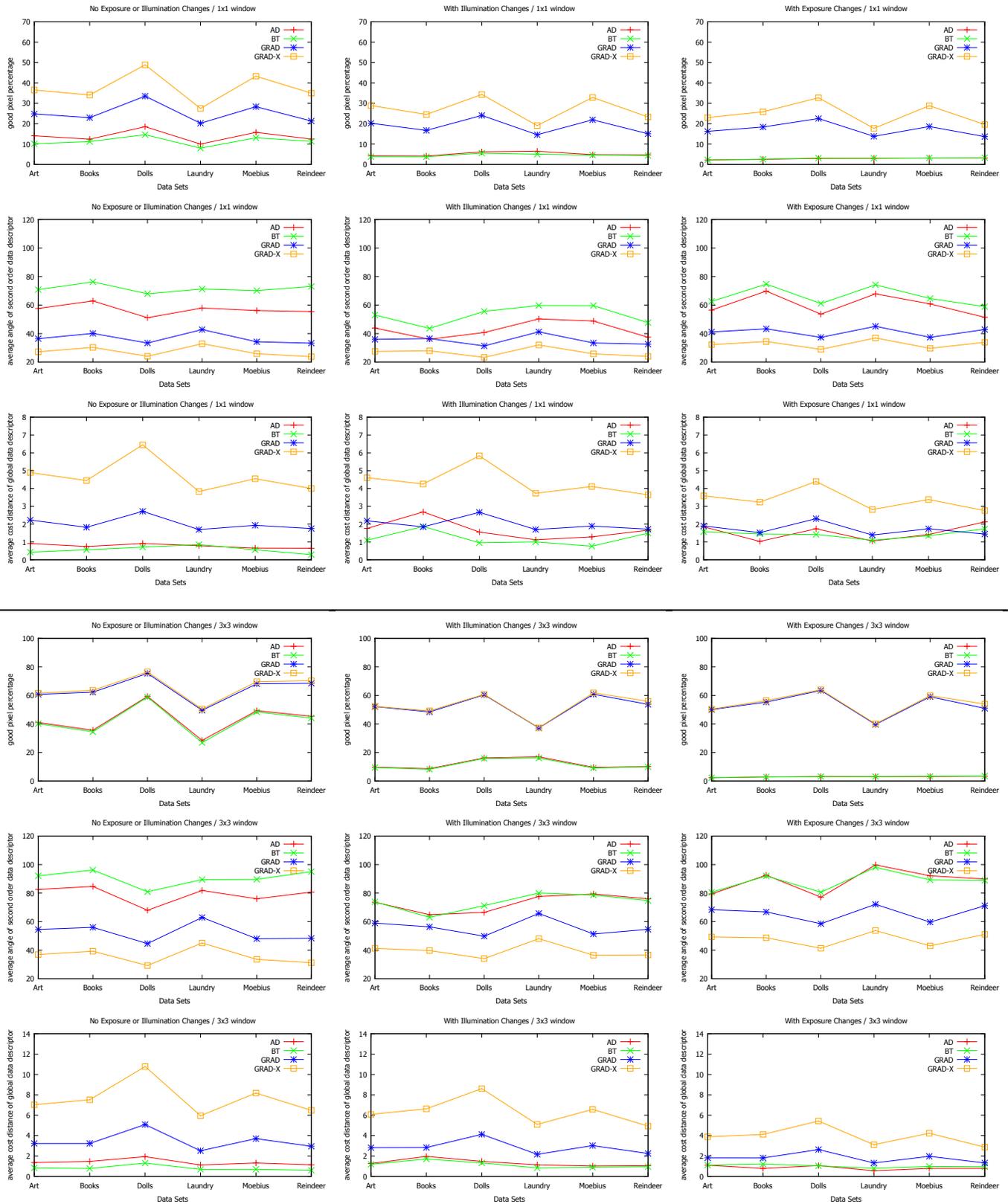


Fig. 4. Experimental results on the stereo dataset. The upper half (first 9 graphs) represent the 1×1 window cost functions, and the lower half represent 3×3 window cost functions. Within each section, the top row is the evaluation of the good pixel percentage, the second row displays the local descriptor, and the third row the global descriptor. The columns (left to right) show the analysis using identical illumination condition, with illumination differences, and then with exposure changes.