Census-Based Cost on Gradients for Matching under Illumination Differences

Christos Stentoumis, Aggelos Amditis
Inst. of Communication and Computer Systems
9, Iroon Polytechniou Str., 15773
Athens, Greece
cstent *at* mail.ntua.gr

George Karras
National Technical University of Athens
9, Iroon Polytechniou Str., 15773
Athens, Greece
gkarras *at* central.ntua.gr

Abstract

Stereo-matching is an indispensable process of dense 3D information extraction for a wide range of applications. Relevant methods rely on cost functions and optimization algorithms for estimating accurate disparities. This work analyses a novel cost for stereo matching under radiometric differences in the stereo-pair, which is based on a modification of the widely used census transformation. It is proposed to define the census on image x and y gradients. The modified census (MC) on gradients is evaluated as an independent matching cost in the presence of severe radiometric differences. For this, the original and the modified census transformation (CT) are implemented in three different aggregation schemes, namely fixed rectangular windows, adaptive cross-based support regions and semi-global matching. It is shown that the MC can provide better results in the cases of local radiometric differences, such as different illumination conditions. Thus, this approach can extend the inherent capability of the original CT to address global monotonic radiometric differences.

1. Introduction

Dense 3D information represents an indispensable component in several computer vision and photogrammetric tasks (e.g. 3D reconstruction, DSM production, novel view synthesis, autonomous navigation, object detection). In terms of accuracy, cost and flexibility, image-based (passive) approaches for acquiring 3D data appear today as competitive to active ones (laser or optical scanners). Among the former, image matching, i.e. automatic determination of pixel correspondences among images, remains fundamental in both its main variations: sparse matching for camera calibration/orientation/rough surface reconstruction; and dense matching for full 3D surface reconstruction. Stereo-matching algorithms, as that presented here, usually rely on the epipolar constraint, thus typically operating on rectified images to produce a dense disparity map.

Existing and emerging application fields have guided the research interest towards a significant number of proposed stereo-matching approaches. Their effectiveness has been extensively discussed in several surveys [1]–[4]. Scharstein and Szeliski [3] have categorized algorithms according to four fundamental components: matching cost computation, support aggregation, disparity optimization and disparity refinement. In [5], [6] the issue of support region formation has been addressed, while [7], [8] provide surveys focusing on criteria for hardware implementation and real-time performance. Two most commonly used on-line platforms for benchmarking stereo-matching algorithms are that of Middlebury College [9] and the Karlsruhe Institute of Technology (KITTI) with its dataset for autonomous driving [10].

In matching function computation a similarity/dissimilarity measure is computed at each pixel for all values of potential disparity. A wide spectrum of such matching metrics have been proposed, some of which have been evaluated in [11] under various optimization algorithms. Most common matching measures are the absolute or the squared intensity difference, the normalized cross correlation, or those based on filtered images, e.g. by median, mean or bilateral filtering [12]. The Birchfield-Tomasi dissimilarity measure [12] copes with differences in image sampling; more recently, the mutual information cost [13] has been proposed for effectively handling radiometric differences. Pixel-wise descriptor measures like DAISY [14] and SIFT variations [15], on the other hand, have yielded promising results in global formulations for wide-based stereo.

During the last decade, the census transformation (CT) [16] has become increasingly popular as the core in dense matching functions; modifications have also been proposed for optimizing its performance. The original matching cost has been evaluated on different colour spaces for a typical rectangular-window-based cost aggregation [17]; no considerable improvement of disparity maps for colour spaces other than RGB was found. The modified census transformation (MCT) approach [18] compares each pixel intensity against the average intensity in a neighbourhood rather than the intensity of the central pixel; following this, [19] have suggested, in the context of optical flow, the application of MCT on bidirectional gradient images using Sobel filtering. The two latter works also propose the computation of CT on a sparse neighbourhood for speeding up the results with small loss in quality. Another variation is cross comparison census (CCC) [20], which takes into consideration not the binary relations among a pixel and its neighbours but the

relations among each pixel in the defined neighbourhood and its four adjacent pixels in a clockwise direction.

Matching cost on census transformation (as well as some of the costs referred to above) has been combined with several other costs. In [21] census cost (CCI) is combined with absolute difference on colour (CADC) to produce top-scoring results on the Middlebury benchmark. In [22] the original census transformation was extended through its application on image gradients (CG); the total cost was found by combining costs CCG and CADC with the absolute difference on gradients CADG and proved to be efficient for sub-pixel accuracy on the Middlebury benchmark and a variety of different scenarios. More recently [23], the results of [22] were improved in the vicinity of edges by the addition of a cost component based on Weber's law. Finally, a combination of CCI, or CCCC, with the mean sum of relative differences of intensities inside a window has been proposed for handling radiometric differences of the KITTI stereo-dataset [20].

This work evaluates a novel matching cost for the purposes of stereo matching, which is a modification of the widely used census transformation (CT). It is proposed to define the modified census (MC) on image x and y gradients, as in [22], but, going beyond this, MC is evaluated as an independent matching cost in the presence of severe radiometric differences. In the past, image gradients have been used for matching, although in very different contexts ([24], [25]). The Middlebury 2006 dataset was selected for the evaluation, since it is composed of 21 stereo pairs with three different illumination and three different exposure settings. The MC is compared against CT using three different optimization schemes, namely fixed rectangular windows, adaptive cross-based support regions and semi-global matching. All matching schemes, i.e. the combinations of cost functions and optimization methods, were carefully tuned through a grid search, and results obtained using the optimum parameters are reported. It is shown that the MC can provide significantly better results in the cases of local radiometric differences, such as different illumination conditions. Thus, this approach can extend the inherent capability of the original CT to address global monotonic radiometric differences, e.g. exposure changes.

In Section 2 the modified census is described; the optimization methods are presented in Section 3; in Section 4 the dataset is described; Section 5 presents are evaluates the results, followed by the conclusions in Section 6.

2. Census transformation on image gradients

In this contribution, the *modified census transformation* (MC) is defined; the gradient of each pixel **p** in i direction is compared to the gradient of its neighbour **q** in a rectangular neighbourhood $N_{\mathbf{p}}$:

$$c_i(\mathbf{p},\mathbf{q}) = \begin{cases} 0, \partial I/\partial i(\mathbf{p}) \le \partial I/\partial i(\mathbf{q}) \\ 1, \partial I/\partial i(\mathbf{p}) > \partial I/\partial i(\mathbf{q}) \end{cases} \quad (1)$$

A binary vector is then formed from the concatenation of the above results, which maps the neighbours with weaker i-gradients. A final vector T_{MC} is formed from the concatenation of the two vectors which correspond to the x and y directions of the gradient:

$$T_{MC}(\mathbf{p}) = \bigotimes_{i \in \{x,y\}} \bigotimes_{\mathbf{q} \in N_{\mathbf{p}}} c_i(\mathbf{p},\mathbf{q}) \quad (2)$$

Symbol $\otimes$ denotes the act of concatenation, following the initial definition of the census non-parametric transformation [16]. The typical census transformation T_{C} is defined on the image intensity function I, whereas here its definition separately on the two gradients in x and y directions is proposed. Hence, the modified census transformation T_{MC} provides a binary vector of doubled length.

The matching cost C between a pixel $\mathbf{p}(x,y)$ of the reference image (*ref*) and its corresponding pixel $\mathbf{p}(x\text{-}d,y)$ in the matching image (*mat*) is the Hamming distance, which is the number of unequal elements in the two binary vectors:

$$C = T_{MC}^{ref}(\mathbf{p}(x,y)) \oplus T_{MC}^{mat}(\mathbf{p}(x-d,y)) \quad (3)$$

The relation of pixel **p** to its neighbour **q**, which belongs to neighbourhood $N_{\mathbf{p}}$, is arbitrarily altered if a local change, like illumination, is applied on the image. In Figure 1 such an instance is presented. The first row displays a detail of the *Bowling1* stereo-pair (see Section 4). The left image refers to default illumination and exposure settings, whereas the right image is the result of a different illumination setting (i.e. mode 3). The blue cross pinpoints the correct pixel correspondence, while the red cross indicates the wrong match obtained by typical census T_{C}. In the second row, the intensities in the neighbourhood of **p** on the left image, of its corresponding pixel **p′** on the right image and of the wrong match are illustrated. The intensity values of each pixel are noted on these image patches. The third row presents the corresponding T_{C} patches. It is seen that local illumination changes have disturbed the relation between the central pixel **p** (seen in grey) and its neighbours. As a consequence, the binary vector of the census transformation is significantly different, and the matching cost is higher at the true disparity position d_{true} than at the wrongly estimated position d_{est}. The respective x-gradients are illustrated on the fourth row, the T_{MC} results are shown on the fifth. Although the relation of the actual pixel **p** intensity to its neighbours' has changed, the relation of the gradient at **p** to its neighbouring gradients has been kept almost unaltered.

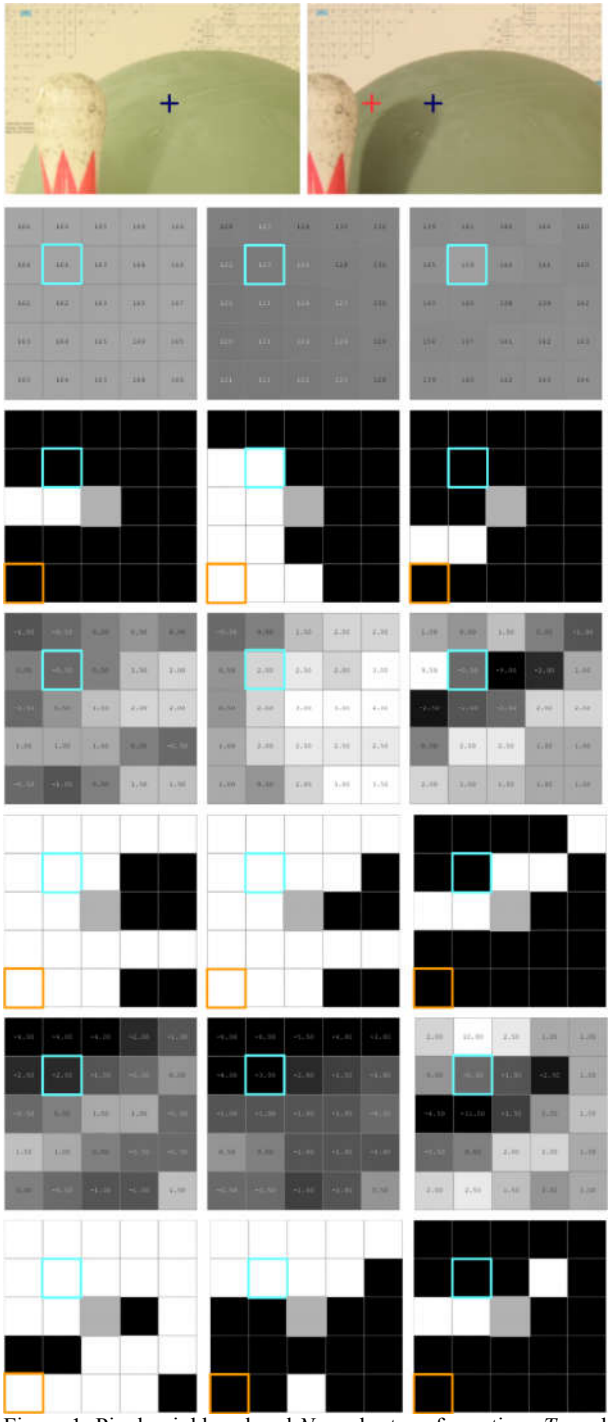

Figure 1: Pixel neighbourhood $N_{\mathbf{p}}$ under transformations T_C and T_{MC}. First row: a detail of the *Bowling1* stereo-pair; left: image with default radiometry; right: image of illumination mode 3. The blue cross is the correct correspondence, while the red cross indicates the wrong match obtained by typical census T_C. Second row: the intensities within neighbourhoods of **p**, **p′** and of the wrong match. Third row: corresponding T_C patches. Fourth row: x-gradients as 2nd row. Fifth row: T_{MC} results. Rows 6 and 7 refer to the y-gradient.

Furthermore, x-gradients around the false estimation d_{est} are very different, thus matching cost is high. Similar results are illustrated in the last two rows which refer to the y-gradients. The cyan boxes mark a pair of pixels **p** and **q** and the values of census in each case. Orange boxes indicate a pixel whose transformation leads to erroneous results both on intensity field and y-gradient, but not on x-gradient; thus, the final T_{MC} absorbs the error in the y-gradient, and the estimated match is correct.

In more general terms, the modified census is computed on gradients, which are more stable against changes in illumination. The original census depends on the relation of the **p** intensity to its neighbour's **q** intensity; in the modified census this relation is "filtered" by the gradient, i.e. the relation of **p** and **q** to their *own immediate* neighbours. This relation is more resistant to change than the relation of the **p** and **q** intensities themselves. Hence, a change in position of a lighting source modifies intensities, but normally it is not expected to simultaneously produce a substantial change of gradients in both the x and y directions. The original census transformation depends on how a pixel relates to its surroundings within the image patch; thus, it can handle radiometric distortions that do not disturb the ordering of intensity values. The proposed modified census, on the other hand, is robust not only in the presence of linear (or monotonic) changes, but also against illumination changes which are non-modelled local changes.

3. Aggregation/optimization algorithms

Matching cost computed pixel-wise appears as a rather weak option for allowing accurate estimation of a disparity map. Hence, in *local* methods the cost of assigning a disparity value to each pixel is supported by a suitably selected region around it during the *cost aggregation* step. Three main approaches are available for addressing this issue: use of support weights; arbitrary window shapes; variations of rectangular windows. A review of aggregation methods for the purposes of real-time stereo matching is found in [5]. A more thorough evaluation of both performance and accuracy for the most important aggregation methods has been conducted in [6]. A global approach, on the other hand, performs *disparity optimization* on an energy function defined over all image pixels by simultaneously posing a smoothness constraint; the most important techniques are based on partial differential equations [26] and graph-cuts [27]. Between local and global methods a class of algorithms have been derived for *semi-global* matching [13]. In this contribution, the raw (pixel-wise) cost and three representative aggregation (or optimization) methods are used, in order to illustrate the efficiency of the modified census-on-gradients matching cost against the typical census transformation on the intensity field.

3.1. Rectangular windows

The first aggregation method tested in conjunction with census transformation variations uses a typical fixed-sized rectangular window. The aggregation of pixel-wise costs C was implemented in typical rectangular windows of constant size L_N in both directions:

$$C(\mathbf{p},d)=\sum_{\mathbf{q}\in N_\mathbf{p}} C(\mathbf{q},d) \tag{4}$$

Although this is a 'naïve' matching algorithm, it can still be useful in some applications thanks to its computational efficiency.

3.2. Hierarchical cross-based regions

Cross-based aggregation [28] is a highly adaptive method which intuitively implements the assumption that pixels **q** in a support region $N_\mathbf{p}$ ought to have similar colours, and is based on decreasing spatial coherence by expanding the arms of a cross around each pixel **p**. Hierarchical cross-based aggregation, as proposed in [22], is a modification of the initial method and was proved efficient in accuracy and computation loads for large scale images.

The four cross segments of a cross-based support region are expanded for length l around each pixel **p**, thus defining two pixel sets H(**p**) and V(**p**) in the horizontal and vertical directions, respectively, restrained by the threshold τ [22]:

$$\tau(l)=-\frac{\tau_{\max}}{L_{\max}}\times l+\tau_{\max} \tag{5}$$

This threshold is controlled by the two parameters $\tau_{\max}$ and $L_{\max}$ defining the maximum colour dissimilarity and spatial distance, respectively. Neighbours of pixel **p** are all pixels **q** belonging to V(**p**) plus those which belong to each corresponding set H(**q**). The combined support region of the two cross windows ($N(\mathbf{p}(x,y))\cap N(\mathbf{p}(x\text{-}d,y))$) formed on the two images of the stereo-pair is used, in order to accommodate the local projective distortions and radiometric differences. Thus, support region $N(\mathbf{p},d)$ is differently shaped for each possible disparity value. The final cost C_t is normalized by the size of $N(\mathbf{p},d)$:

$$C_\mathrm{t}(\mathbf{p},d)=\sum_{\mathbf{q}\in N(\mathbf{p},d)} C(\mathbf{q},d)/\left\|N(\mathbf{p},d)\right\| \tag{6}$$

The support neighbourhood of each pixel per each disparity is formed in every scale s of the hierarchical scheme. Simultaneously, it defines the range of potential disparities for **p**, since it is accepted that each **q** member of $N(\mathbf{p},d)$ can adequately define the disparity range of **p** via its approximate disparity $d_\mathbf{q}^{(s-1)}$ computed in the coarser layer:

$$d_\mathbf{p}^{s}\in\left\{min\ d_\mathbf{q}^{(s-1)},...max\ d_\mathbf{q}^{(s-1)}\right\},\forall\mathbf{q}\in N(\mathbf{p},d) \tag{7}$$

3.3. Semi-global matching

The semi-global matching (SGM) algorithm has been introduced [13] in order to compensate for the high computational load of global optimization methods and provide a robust optimization scheme. This is achieved by splitting the global 2D energy function E of the disparity map D into several 1D approximate equivalents in the directions of L paths throughout the image. Typically, the energy function in (8) consists of a pixel-wise data term $C(\mathbf{p},D(\mathbf{p}))$ and a smoothness term regarding the neighbourhood $N_\mathbf{p}$ of a pixel $\mathbf{p}(x,y)$:

$$E(D)=\sum_\mathbf{p}\left(\begin{array}{l}C(\mathbf{p},D(\mathbf{p}))+\sum_{\mathbf{q}\in N_\mathbf{p}} \mathrm{P}_1 T\left[\left|D(\mathbf{p})-D(\mathbf{q})\right|=1\right]\\ +\sum_{\mathbf{q}\in N_\mathbf{p}} \mathrm{P}_2 T\left[\left|D(\mathbf{p})-D(\mathbf{q})\right|>1\right]\end{array}\right) \tag{8}$$

The smoothness term adds cost penalties P_1 or P_2 when the disparity of a neighbour **q** differs from the disparity of **p**. Penalty P_1 is enforced if the disparity change is low, i.e. up to 1 pixel, whereas larger change (depth discontinuities) is penalized via the P_2 cost. The 1D energy function is computed by following 1D paths L in several directions **r** along the image:

$$\begin{array}{l}L_\mathbf{r}(\mathbf{p},d)=\\ C(\mathbf{p},d)+\\ \min\left(\begin{array}{l}L_\mathbf{r}(\mathbf{p}-\mathbf{r},d),L_\mathbf{r}(\mathbf{p}-\mathbf{r},d-1)+\mathrm{P}_1,\\ L_\mathbf{r}(\mathbf{p}-\mathbf{r},d+1)+P_1,\min_i\left(L_\mathbf{r}(\mathbf{p}-\mathbf{r},i)\right)+\mathrm{P}_2\end{array}\right)-\\ \min_k\left(L_\mathbf{r}(\mathbf{p}-\mathbf{r},k)\right)\end{array} \tag{9}$$

In detail, the partial cost $L_\mathbf{r}(\mathbf{p},d)$ for a pixel **p** and x-disparity d is calculated as the combination of three terms. The first is the pixel-wise matching cost $C(\mathbf{p},d)$. Next is the minimum cost from the path to the preceding pixel (**p-r**) for the same d value, the previous one (d-1), the next one (d+1), or over the range i of disparities, including penalties P_1, P_2. Finally, the minimum path cost of the previous pixel is subtracted, in order to ensure low summation values. The partial costs from all paths $L_\mathbf{r}$ are accumulated to each pixel over all possible disparities, forming the final cost C_t in the representation of the disparity space image:

$$C_\mathrm{t}(\mathbf{p},d)=\sum_\mathbf{r} L_\mathbf{r}(\mathbf{p},d) \tag{10}$$

The disparity map D is calculated from the optimal pixel disparity, selected by a simple winner-takes-all decision on C_t:

$$D(\mathbf{p}) = \underset{d}{argmin}\left(C_t(\mathbf{p},d)\right) \tag{11}$$

A class of SGM based algorithms has been developed. Here, a variation of the algorithm has been implemented, which uses a function for adjusting the P_2 cost penalty, based on the image I intensity differences of pixel **p** with its preceding pixel **q** in direction **r** [29]:

$$P_2 = max\left(\frac{P_{2i}}{1+\left|I(\mathbf{p})-I(\mathbf{p}-\mathbf{r})\right|/w}, P_1\right) \tag{12}$$

with w being a weighting parameter and P_{2i} the initial penalty value. Finally, the 1D optimization is performed for 8 paths, i.e. the four principal and the four diagonal directions.

4. Dataset

Experiments have been carried out on the Middlebury College stereo datasets. The standard dataset comprises the four well-known stereo pairs used for online evaluation, but the 2005 and 2006 datasets are more challenging since they include more real-life scenarios; besides, matching algorithms have not been 'over-adapted' to them. Both datasets have been created under a variety of controlled radiometric differences; here the 2006 datasets are addressed since they include a relatively large set stereo pairs, i.e. of 21 scenes (http://vision.middlebury.edu/stereo/data/scenes2006/).

Resolution is ~1.5 megapixels, yet the 1/3 sized versions are used here. The 2006 stereo dataset of Middlebury College is in fact unexploited, probably due to its complexity and lack of on-line evaluation platform. It comprises indoor scenes that involve a range of predefined radiometric differences. Each scene has been recorded from 7 viewpoints with three different illuminations and three different exposures. The stereo-pairs are formed here using *views* 1 and 5, since true disparity maps are provided only for these views; this results in a total of 18 stereo-pairs for each scene, i.e. 9 with constant illumination and combinations of exposures, 9 with constant exposure and combinations of illuminations. In Figure 2 the radiometric differences for the *Bowling1* scene are illustrated. Changes in illumination can be seen as local changes, since they are caused by changes in position of the lighting source, whereas exposure changes may be seen as global changes, which approximate a linear change in intensities. It is noticed that when the exposure is set to 0 several under-exposed (zero-valued) pixels exist in the images.

To our knowledge, this dataset has not been exploited, as a whole, in evaluations considering illumination and exposure changes. Although it is also limited to indoor scenes, i.e. it is not adequately representative of outdoor real-world scenes, it is far more challenging (complex scene geometries, limited texture) than the widely used on-line Middlebury datasets. Hence, results are expected to be valid and more helpful for designing multi-purpose algorithms for real-life applications.

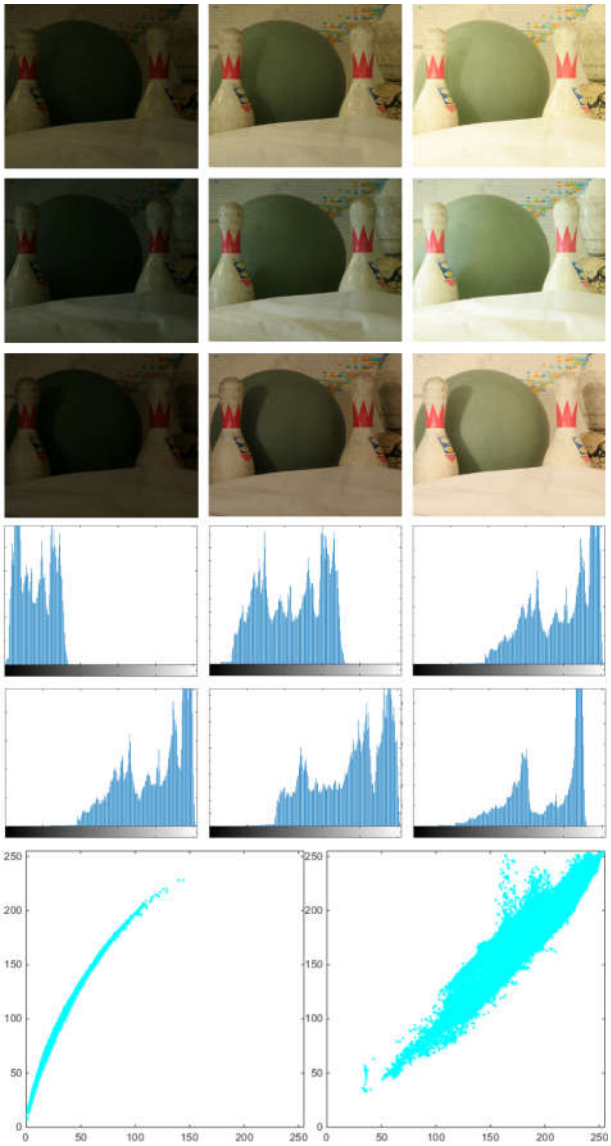

Figure 2: Radiometric differences illustrated for the *Bowling1* stereo pair. *Top*: the three rows show different variations of the left image; the different exposures are from left to right; the illumination changes are from top to bottom. The next two rows are the histograms of the different exposures under illumination 1 and of the illumination changes under exposure +2. *Bottom*: the relation of intensities between the left and right image for different exposures {0,+1} (left figure) and different illumination modes {1,2} (right figure).

5. Evaluation

5.1. Tuning of parameters

As mentioned above, the Middlebury College 2006 stereo dataset, which has radiometric changes in illumination and exposure settings, was used for evaluation. The mean error of the estimated disparities of the non-occluded pixels over all 21 stereo-pairs served as the evaluating criterion. This measure expresses in % the number of pixels whose disparities differ by >1 pixel from the true values. The error is estimated on the disparity map of the left image with no further refinements and constraints (e.g. uniqueness constraint, cross-checking). All parameters of the matching algorithms were first tuned through a grid search. Tuning included all combinations of the MC and CT matching costs and all aggregation strategies, namely rectangular windows (RW), cross-based regions (CB) and semi-global matching (SGM). Tuning was performed only on the default exposure setting and illumination condition. The optimal parameters were used for evaluating each matching scenario under radiometric differences. Tuning was performed in the following range of values for the parameters: CB $\{L, t, k\}$ = {10…55, step 5; 5…30, step 5; 7…15, step 2}; RW $\{L, t, k\}$ = {10…55, step 5; 5…30, step 5; 7…15, step 2}; SGM $\{L, t, k\}$ = {10…55, step 5; 5…30, step 5; 7…15, step 2}. In the above, parameter k denotes the length of the census neighbourhood. The results of tuning are seen in Table 1.

Table 1: Optimum parameters for matching algorithms

	CB		RW			SMG	
	CG	CI	CG	CI		CG	CI
k	9	9	13	11	k	9	9
L	40	30	13	15	P_1	35	35
τ_{max}	10	10			P_2	350	250
					w	6	6

5.2. Results

Figure 3 shows the comparison of the proposed MC and the CT matching costs. It illustrates the disparity errors with respect to the combinations of illumination settings of the left and right image under the default +2 exposure. The top graph refers to the fixed rectangular windows; the middle graph to the hierarchical cross-based approach; the bottom graph to semi-global matching. It is established that MC yields by far more accurate results than CT in the presence of illumination differences. For the CB aggregation this difference is up to 10%. In cases where both images of the pair are under the same illumination conditions, the differences are insignificant and in cases of SGM CT is better by ~1.5%. Overall, CB algorithm using MC cost yields the best results with or without illumination differences. The comparison

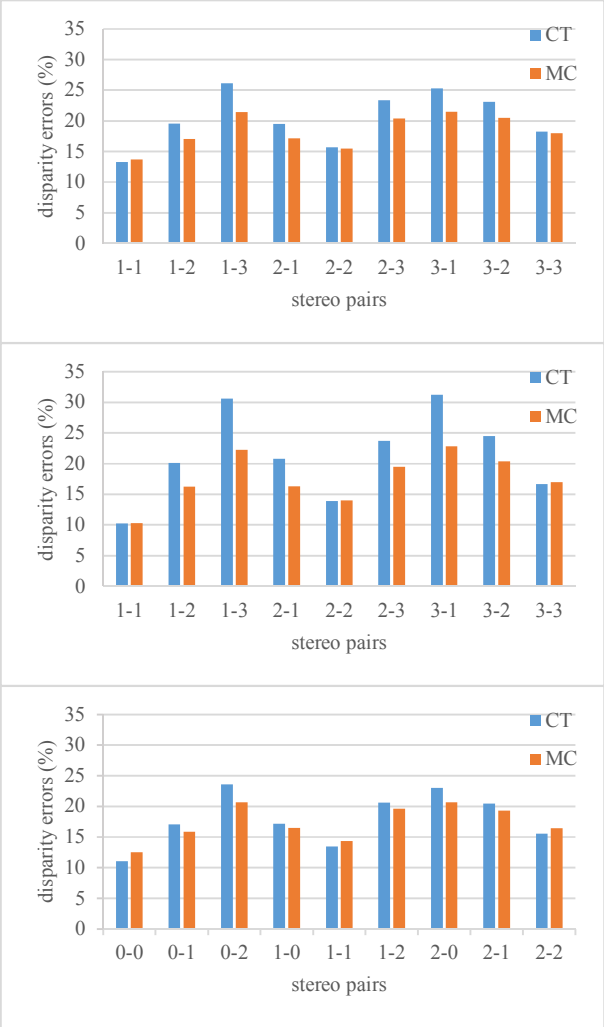

Figure 3: Comparison of the proposed MC (orange) and CT (blue) matching costs for combinations of different illumination settings. *Top*: fixed rectangular windows; *middle*: hierarchical cross-based; *bottom*: semi-global matching.

of the proposed MC and the CT matching costs for combinations of different exposures under the same illumination mode 1 is seen in Figure 4. The graph refers to the rectangular windows aggregation. It may be noticed that CT performs slightly better than MC. As mentioned, differences in exposure are global nearly-linear changes, and the original census can already address differences that do not modify the relative order of intensities (monotonic global change). Hence, MC cannot add any value to the original definition in such cases. The results regarding exposure differences are similar for all three aggregation methods. Overall, CT relies on the sign of intensity differences between a central pixel to its neighbours, whereas MC also addresses illumination differences in matching, as it relies on the sign of the differences *of the intensity differences of pixels from their immediate vicinity*. On the other hand, the

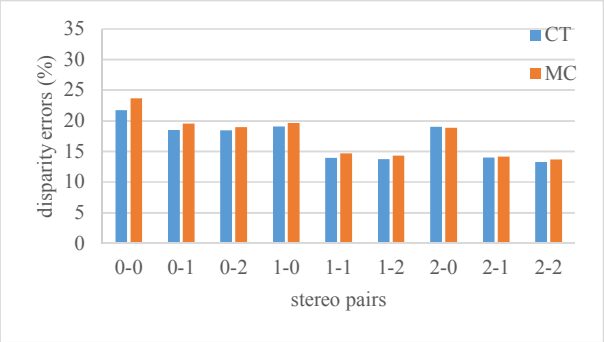

Figure 4: Comparison of the proposed MC (orange) and CT (blue) matching costs for combinations of different exposures and default illumination setting 1.

use of gradients may admittedly introduce a certain amount of noise; thus MC performs slightly worse in the cases of no illumination differences, or only exposure changes, which are inherently addressed by CT.

As an example, finally, the estimated disparity maps for three stereo pairs (*Bowling1*, *Wood2*, *Flowerpots*) are seen in Figure 5. The results of the MC and CT matching costs are shown for all implemented aggregation methods. The disparity maps refer here to the optimal parameter settings for each case and illumination mode 1 for the left image matched to illumination mode 3 of the right image. It is noticed that the improvement of matching results is clearer in areas with shadows in images (e.g. *Bowling1*). Shadows are a basic effect of changing the illumination settings of a scene. Moreover, surfaces of high relief, as that in the case of *Flowerpots*, show great differences in image texture between the images of the stereo, thus resulting in erroneous disparity estimation when CT is used; in this case, MC has showed considerable robustness.

6. Conclusion

In this contribution a modified census transformation, which is defined on image gradients, was presented and evaluated under radiometric differences in indoor scenes for the purposes of stereo matching. It was shown that, when compared to the typical census transformation, it is more efficient in coping with significant illumination differences. Hence, the typical census transformation may be further enhanced, since it is robust against monotonic changes in intensities but not when facing local changes in illumination. On the other hand, exposure changes are nearly linear; the census transformation is inherently robust to this type of changes, thus no significant differences occurred under exposure changes. As a future task, the

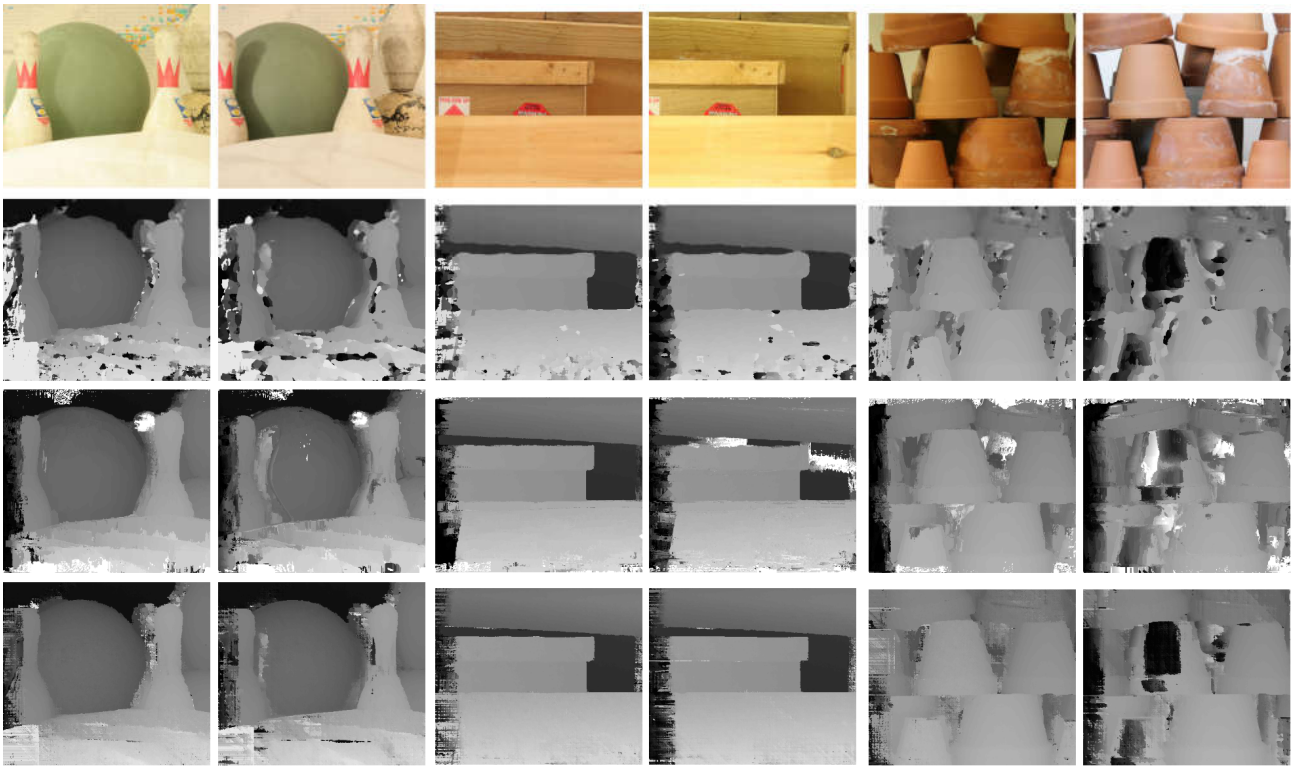

Figure 5: The left and right image (*view1*, *view5*) of the stereo pair, rectangular windows, hierarchical cross-based regions, SGM. Default exposure +2, left image illumination condition 1 and right image illumination condition 3. The first column of each dataset is census on gradients and the second the typical census on intensity. The stereo pairs are Bowling1, Wood2 and Flowerpots.

proposed modified census should be evaluated on the KITTI dataset for autonomous driving, since this presents radiometric differences within stereo pairs of real-life outdoor scenes.

Acknowledgment

"The research leading to these results has received funding from the EC FP7 project ROBINSPECT (Contract N.611145). Authors wish to thank all partners within the ROBINSPECT consortium."

References

[1] U. R. Dhond and J. K. Aggarwal, "Structure from stereo-a review," *IEEE Trans. Syst. Man Cybern.*, vol. 19, no. 6, pp. 1489–1510, 1989.

[2] J. Banks and P. Corke, "Quantitative evaluation of matching methods and validity measures for stereo vision," *Int. J. Rob. Res.*, vol. 20, no. 7, pp. 512–532, Jul. 2001.

[3] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IEEE Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 7–42, 2002.

[4] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008, 2003.

[5] M. Gong, R. Yang, L. Wang, and M. Gong, "A performance study on different cost aggregation approaches used in real-time stereo matching," *Int. J. Comput. Vis.*, vol. 75, no. 2, pp. 283–296, Feb. 2007.

[6] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda, "Classification and evaluation of cost aggregation methods for stereo correspondence," in *Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[7] L. Wang, M. Gong, M. Gong, and R. Yang, "How far can we go with local optimization in real-time stereo matching - A performance study on different cost aggregation approaches," in *International Symposium on 3D Data Processing, Visualization, and Transmission*, 2006, pp. 129–136.

[8] L. Nalpantidis, G. C. Sirakoulis, and A. Gasteratos, "Review of stereo vision algorithms: from software to hardware," *Int. J. Optomechatronics*, vol. 2, no. 4, pp. 435–462, 2008.

[9] H. Hirschmüller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[10] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI vision benchmark suite," in *Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[11] H. Hirschmüller and D. Scharstein, "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1582–1599, Sep. 2009.

[12] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 4, pp. 401–406, 1998.

[13] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[14] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[15] C. Strecha, A. Bronstein, M. Bronstein, and P. Fua, "LDAHash: Improved matching with smaller descriptors.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 66–78, May 2011.

[16] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *European Conference in Computer Vision*, 1994, vol. 801, no. May, pp. 151–158.

[17] A. Miron, A. Bensrhair, A. Rogozan, and S. Ainouz, "Cross-comparison census for colour stereo matching applied to intelligent vehicle," *Electron. Lett.*, vol. 48, no. 24, pp. 1530–1532, Nov. 2012.

[18] B. Froba and A. Ernst, "Face detection with the modified census transform," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 91–96.

[19] P. Puxbaum and K. Ambrosch, "Gradient-based modified census transform for optical flow," in *International Symposium on Visual Computing*, 2010, vol. 6453, pp. 437–448.

[20] A. Miron, S. Ainouz, A. Rogozan, and A. Bensrhair, "A robust cost function for stereo matching of road scenes," *Pattern Recognit. Lett.*, no. 38, pp. 70–77, Nov. 2014.

[21] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *ICCV Workshop on GPU in Computer Vision Applications*, 2011, pp. 467–474.

[22] C. Stentoumis, L. Grammatikopoulos, I. Kalisperakis, and G. Karras, "On accurate dense stereo-matching using a local adaptive multi-cost approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 91, pp. 29–49, May 2014.

[23] Q. Wu, X. Zhao, and X. Li, "Stereo Matching Algorithm of Color Image Based on Weber's Law and Census," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2014, pp. 199–202.

[24] D. Scharstein, "Matching images by comparing their gradient fields," in *International Conference on Pattern Recognition*, 1994, vol. 1, pp. 572–575.

[25] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *IEEE International Conference on Pattern Recognition*, 2006, p. (3) 15–18.

[26] O. Faugeras and R. Keriven, "Variational principles, surface evolution, PDE's, level set methods, and the stereo problem," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 336–344, Jan. 1998.

[27] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *IEEE International Conference on Computer Vision*, 2001, vol. 2, pp. 508–515.

[28] K. Zhang, J. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 7, pp. 1073–1079, Jul. 2009.

[29] K. Zhu, P. D'Angelo, M. Butenuth, U. Stilla, F. Rottensteiner, H. Mayer, and B. Jutzi, *A performance study on different stereo matching costs using airborne image sequences and satellite images*, vol. 6952. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.