SEMI-GLOBAL MATCHING WITH SELF-ADJUSTING PENALTIES

E. Karkalou*, C. Stentoumis, G. Karras
(ellikarkalou *at* gmail.com, cstent *at* mail.ntua.gr, gkarras *at* central.ntua.gr)

Laboratory of Photogrammetry, National Technical University of Athens, GR-15780 Athens, Greece

Commission II

KEY WORDS: Disparity, Stereo, Matching, Semi-global, Penalties, Adaptive, Aggregation, Optimization

ABSTRACT:

The demand for 3D models of various scales and precisions is strong for a wide range of applications, among which cultural heritage recording is particularly important and challenging. In this context, dense image matching is a fundamental task for processes which involve image-based reconstruction of 3D models. Despite the existence of commercial software, the need for complete and accurate results under different conditions, as well as for computational efficiency under a variety of hardware, has kept image-matching algorithms as one of the most active research topics. Semi-global matching (SGM) is among the most popular optimization algorithms due to its accuracy, computational efficiency, and simplicity. A challenging aspect in SGM implementation is the determination of smoothness constraints, i.e. penalties P_1, P_2 for disparity changes and discontinuities. In fact, penalty adjustment is needed for every particular stereo-pair and cost computation. In this work, a novel formulation of *self-adjusting penalties* is proposed: *SGM penalties can be estimated solely from the statistical properties of the initial disparity space image*. The proposed method of self-adjusting penalties (SGM-SAP) is evaluated using typical cost functions on stereo-pairs from the recent Middlebury dataset of interior scenes, as well as from the EPFL Herz-Jesu architectural scenes. Results are competitive against the original SGM estimates. The significant aspects of self-adjusting penalties are: (i) the time-consuming tuning process is avoided; (ii) SGM can be used in image collections with limited number of stereo-pairs; and (iii) no heuristic user intervention is needed.

1. INTRODUCTION

The extraction of dense 3D information and the accurate visual recording from a set of images is a core part in various Cultural Heritage applications. Typically, accurate visual and geometric recording supports documentation, restoration and preservation activities ranging from large scale monuments to small artifacts. Lately, cultural heritage has benefited from new emerging technologies, based on 3D information, and the impressive increase in available smart mobile devices. Gamification of guided tours and story-telling approaches for the public presentation of cultural heritage are based on augmented and virtual reality tools, which all share the extraction of 3D information as a key enabling technology. As an active research topic, extraction of dense 3D information is bundled with many intermediate products and application fields in the areas of photogrammetry, computer vision and image processing. 3D reconstruction, ortho-projection, pose estimation, simultaneous localization and mapping, image stitching, recognition and novel view synthesis are but a few of the topics of interest. In this context, dense image matching is a fundamental task for every application undertaking automated 3D reconstruction from images. 3D model generation concerns an ever-growing list of diverse applications, which includes cultural heritage recording, precision agriculture and farming, automation in construction, large-scale city modeling, 3D GIS, automotive industry, industrial robotics, infrastructure inspection, and security.

While several related software has been commercially introduced, the varying application conditions, the demand for complete and accurate measuring products, as well as for computational efficiency in a variety of hardware, keep image-matching algorithms as one of the most active research topics. Stereo-matching, or multiple view stereo-matching, is indeed a challenging task when compared to multi-view matching, as it addresses the question with limited number of observations. This said, it represents an indispensable tool for case scenarios where multiple views are limited, such as in the cases of historical images, aerial images, robot vision, autonomous vehicles, mobile devices.

Scene reconstruction usually falls under two distinct processes: *sparse matching* for retrieving correspondences among images for camera extrinsic and intrinsic calibration, and *dense matching* for full 3D surface reconstruction. Dense stereo-matching, i.e. the estimation of a homology in the matching (right) image for each pixel in the base (left) image, is typically performed on rectified (epipolar) stereo pairs, and it is an essential element in both multi-view stereo or stereo-view reconstruction processes. A well-established approach in analysing and classifying stereo-matching algorithms is to typically decompose them in four basic components: matching *cost computation*, support *aggregation*, disparity *optimization* and disparity *refinement* (Szeliski, 2011). An evaluation of stereo-matching methods based on their actual results and usefulness in real life applications is quite difficult and depends on several diverging criteria. This is particularly true if one considers the variety of both applications and arising issues, e.g. depth variability, lighting conditions, reflecting surfaces, scene occlusions, image acquisition geometry, and illumination changes, just to name a few.

In the matching *cost computation* step a dissimilarity measure is given to each pixel for every value in the disparity range. The

* Corresponding author.

matching measures may be simple (for instance, absolute pixel differences) but they could also involve image transformations such as the non-parametric Census transformation and its variations to produce robust results based on binary relationships of pixels with their vicinity. One of the most recent reviews evaluates an extensive collection of matching cost functions (Hu & Mordohai, 2012). Computed cost volumes need to be smoothed against noise, while usually exploiting the 'fronto-parallel' assumption, thus the pixel-wise cost is *aggregated* within a support neighbourhood. A thorough review is presented in Tombari et al. (2008). A common distinction is between local and global methods; disparity selection in *local* methods is typically carried out in the winner-takes-all (WTA) mode, while *global* methods rely on energy minimization systems to optimize disparity over all image pixels against the need to keep continuous surfaces and satisfy pixel-wise matching criteria. Between local and global methods a class of algorithms for semi-global matching (SGM) has been presented by Hirschmüller (2005). In addition, the class of non-local methods attempt to extend the kernel of the local ones onto the whole image (Huang et al., 2016; Yang, 2012). Zhang et al. (2015) combine the information from different scale spaces to efficiently exploit the image pyramid in addressing issues in texture-less regions and restricting the disparity search space. Li et al. (2016) reduce the 'fronto-parallel' effect in disparity estimation over support aggregation neighbourhoods by proposing the formation of slanted support windows which greatly improve the results for non-frontal surfaces. Following the most recent trend in computer vision research and state-of-the-art applications, deep learning approaches, i.e. convolutional neural network (CNNs) schemes, are constructed for the purposes of stereo-matching in the matching cost computation step. Some of the top-ranking algorithms in the evaluation platforms are based on such formulations. Thus, Zbontar & Le Cun (2015) train a convolutional neural network on small image patches of known disparity, and the result is used as an initial cost volume. On the other hand, Luo et al. (2016) estimate a product layer from the inner product of the two representations of the typical Siamese network in order to simplify the process and exponentially speed up the process to real-time applications. The promising idea of exploiting the strengths and avoiding the weaknesses of different matching functions is proposed in Spyropoulos & Mordohai (2015), where an ensemble classifier is trained to decide the appropriate cost functions on a certain pixel. Lately, it has been discussed that the cost aggregation process is the key process for most local methods and an important component for many global ones (Yang et al., 2009; Wang & Zheng, 2008). In Georgousis et al. (2016) such a hybrid method refining the global estimations by local support windows has been presented.

One of the most cited, publicly available databases of stereo images, which at the same time serves as an online evaluation platform, is that of Middlebury College*. The images used here have been taken from the Middlebury 2006 stereo-pairs and the newest Middlebury 3 high resolution dataset, which has separate training and testing stereo-pairs. Furthermore, stereo-pairs from the EPFL multi-view datasets of external architectural scenes† were chosen for evaluating the proposed approach. Finally, it is noted that the KITTI datasets‡ provide a series of images of urban driving scenes. On the evaluation sites new stereo-matching algorithms are being constantly reported.

In this paper, an improved approach of the Semi-Global Matching (SGM) algorithm is presented, which eliminates the need for scenario-specific tuning of the SGM penalty parameters. Thus, its main contribution is that it introduces a method for automatically estimating penalties P_1 and P_2 of SGM and methods derived from it. This is achieved after computing certain statistical properties of the Disparity Space Image (DSI), which is estimated during the matching cost computation. The presented method of self-adjusting penalties (SGM-SAP) was evaluated using internal stereo-images from the Middlebury online evaluation platform datasets, as well as images from external architectural scenes selected from the EPFL multi-view datasets.

Next, Section 2 reviews the specifics of SGM and penalty definitions; Section 3 analyses the process of self-adjusting of the penalty values; Section 4 evaluates the results of our tests; the paper is concluded with final remarks and possible future tasks.

2. SEMI-GLOBAL MATCHING AND PENALTIES

Semi-global matching (Hirschmüller, 2005, 2008) is among the top-ranking dense matching algorithms. Its main advantages are accuracy, computational efficiency and simplicity in implementation when compared to high performance global and local methods. Consequently, it is used in stereo as well in multi-view stereo scenarios from real-time to large-scale satellite applications. In this Section, the SGM algorithm is briefly reviewed for the purposes of completeness, and some variations relevant to this work are presented.

SGM is employed in the optimization step, as it defines a global 2D energy function E that depends on the disparity map D:

$$E(D)=\sum_{\mathbf{p}}\left(\begin{array}{l} C(\mathbf{p},D(\mathbf{p}))+\sum_{\mathbf{q}\in N_\mathbf{p}}\left(P_1 T\left[\left|D(\mathbf{p})-D(\mathbf{q})\right|=1\right]\right)+ \\ \sum_{\mathbf{q}\in N_\mathbf{p}}\left(P_2 T\left[\left|D(\mathbf{p})-D(\mathbf{q})\right|>1\right]\right) \end{array}\right) \quad (1)$$

The global function contains a data term $C(\mathbf{p}, D(\mathbf{p}))$ as well as a smoothness term for each pixel $\mathbf{p}$. The latter adds a penalty P_1 or P_2 to each pixel $\mathbf{q}$ in the neighbourhood $N_\mathbf{p}$ of $\mathbf{p}$, if the disparity of $\mathbf{q}$ differs by 1 or more pixels from the disparity of $\mathbf{p}$, respectively. SGM suggests approximating the global function by following 1D paths L in several directions r through the image:

$$\begin{array}{c} L_r(\mathbf{p},d)=C(\mathbf{p},d)+ \\ min\left(\begin{array}{l} L_r(\mathbf{p}-r,d),L_r(\mathbf{p}-r,d-1)+P_1, \\ L_r(\mathbf{p}-r,d+1)+P_1,\min_i\left(L_r(\mathbf{p}-r,i)\right)+P_2 \end{array}\right) \\ -\min_k\left(L_r(\mathbf{p}-r,k)\right) \end{array} \quad (2)$$

In each of the r = 8 paths, the optimized cost $L_r(\mathbf{p},d)$ for every pixel $\mathbf{p}(x,y)$ and every x-disparity d is estimated from the sum of three terms. The first two are the matching cost $C(\mathbf{p},d)$ and the minimum path cost of the preceding pixel ($\mathbf{p}$-r); the latter is computed after comparison of the path costs of the previous pixel in the same (d), the lower (d-1), the higher (d+1) or all the disparity range (i), while taking into consideration penalties P_1 and P_2. Finally, the minimum path cost of the preceding pixel is subtracted. P_1 penalizes slightly slanted surfaces, P_2 penalizes discontinuities. The costs from all paths $L_r(\mathbf{p},d)$ are summed up to each pixel for all possible disparities, resulting in the aggregated cost $S(\mathbf{p},d)$:

$$S(\mathbf{p},d)=\sum_r L_r(\mathbf{p},d) \quad (3)$$

*http://vision.middlebury.edu/stereo/data/
†http://cvlabwww.epfl.ch/data/multiview/denseMVS.html
‡http://www.cvlibs.net/datasets/kitti/eval_stereo.php

The optimal disparity for each pixel is chosen by the WTA strategy on *S*, thus creating the final disparity map $D_L(\mathbf{p})$:

$$D_L(\mathbf{p}) = \underset{d}{argmin}\big(S(\mathbf{p},d)\big) \tag{4}$$

Since the introduction of SGM several variations or extensions have emerged, aiming at improving its performance, computational efficiency, or both. SGM is also implemented in real-time on a variety of platforms, i.e. FPGA or GPU. Moreover, thanks to its implementation in OpenCV, many algorithms use SGM as part of their stereo matching procedure. Recently, non-local methods (Huang et al., 2016) have also introduced cost-aggregation approaches similar to that of SGM; two iterations are needed for the image-guided non-local matching cost computation, and afterwards the estimated cost is optimized via SGM.

Regarding the definition of cost penalties, a class of SGM variations is dedicated to the development of functions for the adjustment of penalty P_2, which is imposed on disparity changes between neighbouring pixels larger than 1 pixel; they have been reviewed in detail by Stentoumis et al. (2015). These penalty functions are based on the fact that, if the intensity change between pixel **p** and the preceding one in path *L* is high and the disparity change between them is larger than 1, the existence of actual edges or object boundaries is highly probable. Hirschmüller (2005) has firstly introduced an adaptive penalty function. The function was created by dividing P_2 with the intensity gradient of neighbouring pixels in the reference image for each path, while checking that $P_2 \geq P_1$. Besides, Banz et al. (2012) evaluated the performance of three more penalty functions for P_2 and the case of constant penalty, which is fixed to an empirically defined value. The proposed penalty functions were: negatively (P_{2n}) and inversely (P_{2i}) proportional to the absolute intensity gradient of the currently processed pixels along the path; and negatively proportional (P_{2v}) to the variance of intensity in a local window. A lower bound P_{2min} was introduced to guarantee that $P_2 \geq P_1$.

A challenging aspect of SGM implementation is obviously the selection of values for the penalties. If parameters have not been properly tuned, the performance of the algorithm may not be as efficient as expected. In fact, penalty adjustment is needed for every different pair of images or, if a different matching cost method is used, even for the same stereo-pair. In this paper, we introduce a method for automatically estimating penalties P_1 and P_2. This follows the computation of certain simple statistical properties from the DSI volume which is created in the previous step of cost calculation. Therefore, penalties are considered as being *self-adjusted* to the particular stereo-pair, in relation to the cost function used.

To our knowledge, no method for the automatic estimation of penalties of SGM has been proposed up to now – with the exception of Chuang et al. (2016), where however a specific cost function was used, the penalties were extracted after the creation of an initial disparity map from only two costs of each pixel (the lowest and the second lowest), and the evaluation was based on only four image pairs.

3. SELF-ADJUSTING PENALTY VALUES

The idea behind extracting the values for the SGM penalties from the DSI itself originates from the fact that penalties P_1, P_2 are actually costs that influence the pixel-wise matching cost *C*. Cost penalties are added to each pixel's initial cost C(**p**,d) depending on the disparity *d*, so their values should be related to this initial cost. In the proposed method penalties are derived from the DSI S(x,y,l) representation of the initial cost C(**p**,d); (x,y) are the image coordinates of a pixel **p** and *l* is the label that maps a disparity *d* to the DSI, $l = l(\mathrm{d})$:

$$S_{\min}(x,y) = \min_{l=1:N_d} S(x,y,l) \tag{5}$$

$$P_1 = \frac{\sum_{y=1}^{H}\sum_{x=1}^{W}\sum_{l=1}^{N_d}\big(S(x,y,l) - S_{\min}(x,y)\big)}{N \cdot N_d} \tag{6}$$

$$P_2 = \max_{y=1:H,\,x=1:W,\,l=1:N_d}\big(S(x,y,l) - S_{\min}(x,y)\big) \tag{7}$$

In the above equations, *W* and *H* are the width and height of the base image; N_d is the number of disparity labels; and *N* is the number of image pixels. The minimum matching cost $S_{min}(x,y)$ of a pixel over all labels *l* is subtracted from all potential costs S(x,y,l) in order to normalize the DSI values per pixel. Finally, the mean value cost per all pixels corresponds to penalty P_1, while the maximum value cost per all pixels corresponds to penalty P_2. Of course, in this way it is ensured that $P_2 > P_1$. *The importance of such a definition for the penalties is that they are estimated, without user intervention, from the DSI itself*; thus, the self-adjusting penalties remove the need for the conventional, time-consuming tuning step. Appropriate penalty values will be automatically derived, regardless of the stereo-pair, or the matching cost used. Furthermore, no training datasets of stereo-pairs will be needed for penalties estimated from them to be applied to testing stereo-pairs under an assumed scenario of many similar images. Moreover, these self-adjusting penalties are not computationally expensive. In conclusion, for every stereo-pair the penalties for SGM, or every SGM-like method, can be estimated solely from the DSI, regardless of the matching cost employed and the existence of a ground truth disparity map or of multiple data for training.

In the cost calculation step, common cost functions such as Absolute Difference (AD) of intensities, or Census transform with a 7x7 window were used. Next, SGM was used for cost optimization. The WTA strategy is adopted during the disparity optimization step for acquiring the initial disparity map. Finally, disparity refinement is possible, e.g. with the use of photo-consistency, sub-pixel disparity interpolation, or median filtering.

4. RESULTS

The presented algorithm has been evaluated on the 15 training stereo image pairs of quarter-size resolution from Middlebury Stereo Evaluation – Version 3 (Scharstein et al., 2014), and also on the 21 quarter-size stereo image pairs of 2006 datasets from Middlebury College. The algorithm has also been tested using an EPFL multi-view dataset with external architectural scenes (Strecha et al., 2008). All processes have been implemented in the Matlab programming environment.

4.1 Middlebury 2014 datasets

The Hamming distance on Census transformed images was used as the initial matching cost. Next, penalties P_1 and P_2 for SGM were estimated via the suggested method and were employed to the SGM algorithm. In Table 1 the computed penalties for each stereo-pair are shown.

Stereo-pairs	Penalties	
	P_1	P_2
Adirondack	15.2	47
ArtL	15.0	47
Jadeplant	15.2	47
Motorcycle	16.0	47
MotorcycleE	16.1	47
Piano	14.2	47
PianoL	14.2	47
Pipes	15.5	47
Playroom	15.8	47
Playtable	15.6	47
PlaytableP	16.0	47
Recycle	15.7	47
Shelves	13.2	47
Teddy	15.0	47
Vintage	15.7	47

Table 1. Penalties estimated by the SGM-SAP method for each stereo-pair of the Middlebury 2014 dataset.

The initial disparity map was derived by the WTA strategy. Finally, sub-pixel disparities are estimated by a sequential disparity interpolation and 7x7 median filtering for smoothing with outlier tolerance. The error percentage is computed by comparing each resulting disparity value of non-occluded pixels with the corresponding ground truth value, while applying an error threshold of 0.5 pixel. This threshold value was chosen because the default value used in Middlebury online evaluation platform is 2.0 pixels for full image resolution, which corresponds to a threshold of 0.5 pixel for quarter-size images.

In Fig. 1 some results of the method are seen for three representative stereo pairs as far as the size of matching error is concerned. The strong impact of sub-pixel interpolation on the disparity map can be noted, which is mainly due to the fact that the raw algorithm estimates integer disparities, whereas a 0.5 pixel error threshold is used. A considerable effect is also achieved by denoising via a median filter with large kernel.

The estimated disparity maps of training images were submitted to the Middlebury benchmark evaluation page (Fig. 2), resulting in an error of 22.8% and the 34th position for non-occluded pixels and a 2.0 pixel error threshold (date of evaluation: January 22, 2017). The image pairs displaying the best performance were *Playtable* (28th position) and *Vintage* (32th position), whereas those of poorest performance were *ArtL* (43th position), *Pipes* and *PlaytableP* (42th position).

Compared to the original SGM algorithm (Hirschmüller, 2008) and its results submitted in the Middlebury platform, our method presents an error higher only by 1.8%, and it is only 3 positions lower in the evaluation list. *Playtable* and *Vintage* show lower errors (35% to 38.8% and 40.6% to 41.1%, respectively), whereas *Jadeplant* and *ArtL* show the highest errors compared to those of Hirschmüller (31.9% to 26.4% and 18.8% to 15%, respectively). The errors of stereo the pairs for both methods as well as their ranking are presented in Fig. 3.

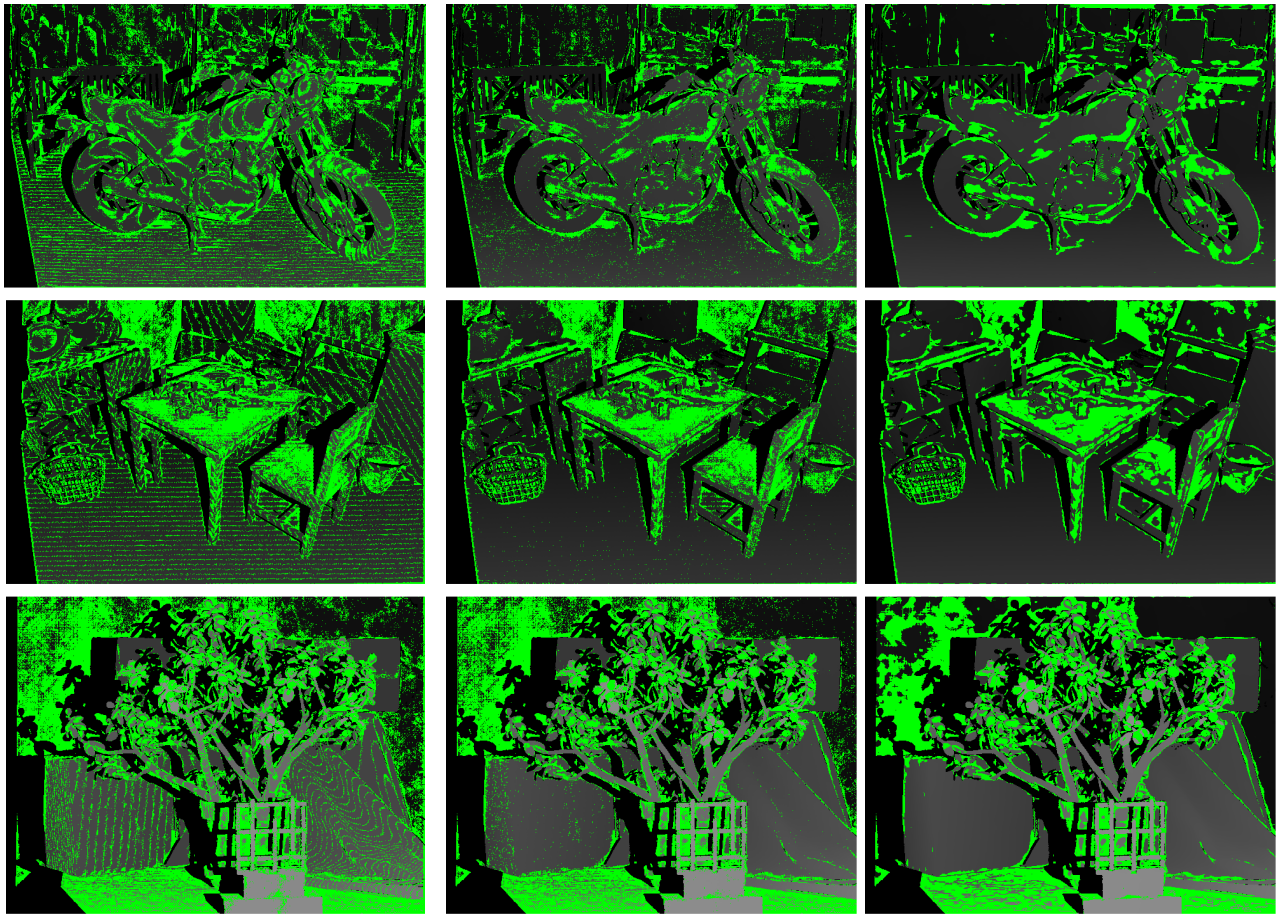

Figure 1. Estimated disparity maps using SGM-SAP. *Left*: disparity maps without any refinement; *centre:* sub-pixel interpolation; *right:* median filtering. Differences above 0.5 pixel from the ground truth are highlighted in green. *Top* to *bottom*: *Motorcycle*, *PlaytableP* and *Jadeplant* stereo-pairs.

bad 2.0 (%)

Date	Name	Res	Weight Avg	Adiron MP: 5.7 nd: 290 im0 im1 GT nonocc	ArtL MP: 1.5 nd: 256 im0 im1 GT nonocc	Jadepl MP: 5.2 nd: 640 im0 im1 GT nonocc	Motor MP: 5.9 nd: 280 im0 im1 GT nonocc	MotorE MP: 5.9 nd: 280 im0 im1 GT nonocc	Piano MP: 5.4 nd: 260 im0 im1 GT nonocc	PianoL MP: 5.4 nd: 260 im0 im1 GT nonocc	Pipes MP: 5.7 nd: 300 im0 im1 GT nonocc	Playrm MP: 5.3 nd: 330 im0 im1 GT nonocc	Playt MP: 5 nd: 290 im0 im1 GT nonocc	PlaytP MP: 5 nd: 290 im0 im1 GT nonocc	Recyc MP: 5.6 nd: 260 im0 im1 GT nonocc	Shelvs MP: 5.9 nd: 240 im0 im1 GT nonocc	Teddy MP: 2.7 nd: 256 im0 im1 GT nonocc	Vintge MP: 5.5 nd: 760 im0 im1 GT nonocc
04/08/15	REAF	H	20.6 30	19.7 36	9.93 24	20.5 23	12.6 32	11.0 27	20.8 31	32.1 29	13.9 29	28.6 29	48.5 39	20.8 38	16.5 29	50.3 37	8.19 34	46.9 42
07/25/14	SGM	Q	21.0 31	14.9 27	15.0 40	26.4 34	14.3 37	13.2 33	22.7 34	36.6 33	15.6 37	26.3 26	38.8 32	20.0 36	16.9 30	47.7 30	8.00 33	41.1 34
11/06/16	SPS	F	21.1 32	15.1 28	13.1 34	26.6 35	11.2 26	11.6 28	27.2 40	38.0 35	15.5 35	32.3 39	26.3 22	22.3 44	22.7 41	46.2 27	7.07 27	40.3 31
07/28/14	SGM	F	22.1 33	28.4 45	6.52 9	20.1 22	13.9 35	11.7 29	19.7 27	33.2 31	15.5 36	30.0 34	58.3 52	18.5 31	23.8 42	49.5 34	7.38 29	49.9 45
01/22/17	SGM-SAP	Q	22.8 34	16.2 31	18.8 43	31.9 38	15.8 40	15.0 35	22.8 35	37.6 34	17.1 42	29.4 33	35.0 28	21.5 42	18.4 35	51.0 39	10.8 37	40.6 32
08/25/14	LPS	H	23.2 35	11.4 24	32.3 54	21.7 27	5.72 14	68.5 54	19.0 25	50.2 44	8.89 18	25.9 25	20.9 17	18.0 30	16.5 28	39.7 17	6.96 25	26.6 16
07/28/14	SGBM1	H	23.3 36	25.2 41	13.7 37	26.1 33	11.8 29	21.3 39	25.6 37	48.5 41	11.4 22	31.4 37	40.2 33	19.9 35	17.7 33	47.3 29	11.8 39	45.1 39

Figure 2. Results from the Middlebury evaluation platform for the training images of the 2014 dataset. The results of the self-adjusting penalty method (SGM-SAP) are highlighted in red.

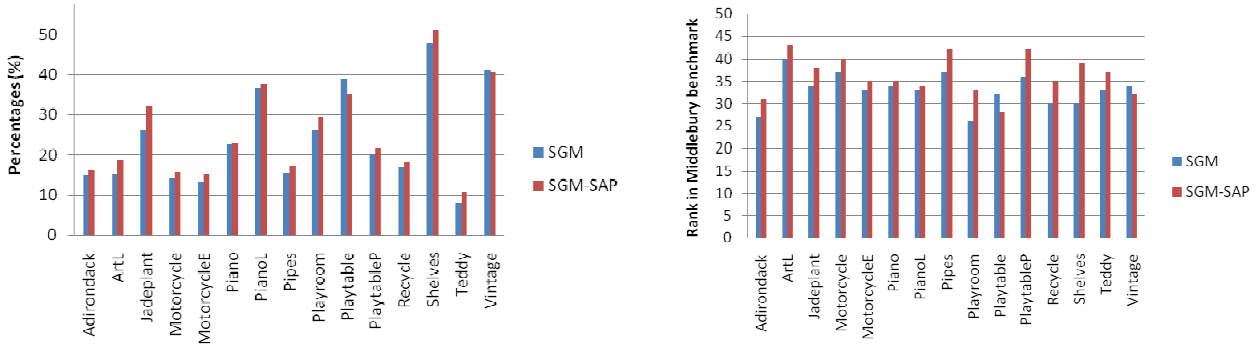

Figure 3. Errors for all stereo-pairs (left) and ranking in Middlebury benchmark (right) for SGM and our method (SGM-SAP).

Finally, it is noted that, compared to our method, in the original SGM algorithm additional refinements are being used, such as left-right consistency or removal of disparity segments smaller than 100 pixels, whereas median filtering is not applied.

In Fig. 4 disparity maps are seen, in which differences in errors when compared to ground truth between the original SGM and our method are highlighted. Pixels whose disparity difference against ground truth is larger than 0.5 pixel if original SGM is applied but less than 0.5 if our method is used are highlighted in blue. Pixels whose disparity difference compared with ground truth is above 0.5 pixel if our method is applied but below 0.5 if the original SGM is used are highlighted in red. It is observed that, ignoring small artifacts produced by either method, our method outperforms SGM in slightly slanted surfaces (e.g. floor of *Playtable* or *Motorcycle*), but performs less well in areas of low texture (e.g. in the background of the *Jadeplant* stereo-pair).

Finally, additional experiments regarding the suggested method were conducted and evaluated on the Middlebury 2014 datasets. In particular, the median instead of the mean value was used as far as the estimation of penalty P_1 is concerned. The differences in the total error of 15 pairs regarding the initial method were negligible (0.01%). Furthermore, when a 9x7 window is used for Census transform (as in the original SGM algorithm) before the penalty adjustment, the error is the same regarding the initial disparity maps and by 0.4% higher after refinements. Besides, after the automatic estimation of both penalties by our method, a penalty function proposed by Hirschmüller (2005) was tested for the adjustment of penalty P_2 to the intensity gradient. The estimated disparity map appeared as noisier and initially showed an error higher by 2%, which after disparity refinement was reduced to 0.8%.

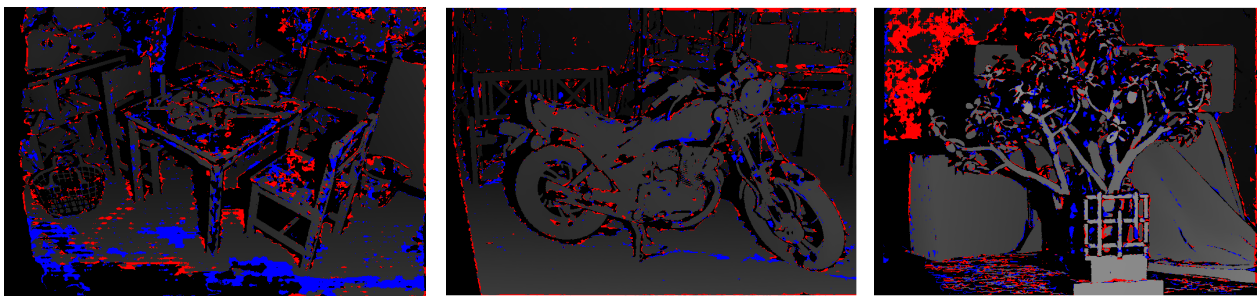

Figure 4. Disparity maps of the suggested method (from left to right: *Playtable*, *Motorcycle* and *Jadeplant* stereo pairs). Pixels in blue indicate errors of SGM algorithm which do not exist in our method; pixels in red depict errors of our method not present in SGM.

4.2 Middlebury 2006 datasets

For this case, Absolute Differences of intensities and Hamming distance on Census-transformed images were used as cost metrics. Subsequently, penalties P_1 and P_2 for SGM were computed from the proposed method and were applied to the SGM algorithm. The initial disparity map has been created in the WTA mode.

The overall error for the 21 pairs was compared against the errors obtained from the same method without automatic penalty estimation, namely by using the optimal parameters of a tuning process (Stentoumis et al., 2015). The error percentage is calculated after the comparison of each resulting disparity value in non-occluded areas with the corresponding ground truth value, while an error threshold of 1 pixel is applied. The error of our method was higher by only 0.87% (11.89% to 11.02%) when the Census metric served as cost function and by 2.27% higher (25.72% to 23.45%) when Absolute Differences were applied. Therefore, it is concluded that the proposed method is expected to work well for any matching cost function.

In Table 2 the estimated (SGM-SAP) penalty values for two matching costs are seen against the values derived by the tuning process for each individual stereo-pair. Optimal penalties [P_1, P_2] which lead to the minimum of the mean errors over all stereo-pairs are [10,100] from the tuning of AD-SGM method, while for Census-SGM method these are [25,100].

Stereo-pair	AD				Census			
	SGM-SAP		Tuning		SGM-SAP		Tuning	
	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
Aloe	40	243	15	70	17	47	10	50
Baby1	21	203	10	100	15	47	10	50
Baby2	24	242	20	250	15	47	25	100
Baby3	20	203	20	130	15	47	10	50
Bowling1	32	250	20	70	14	47	25	100
Bowling2	31	224	10	250	15	47	10	100
Cloth1	32	220	15	100	18	47	25	250
Cloth2	32	251	15	70	17	47	10	50
Cloth3	37	246	15	70	17	47	10	50
Cloth4	43	246	20	100	17	47	10	50
Flowerpots	28	181	20	250	14	47	10	50
Lampshade1	25	236	10	70	13	47	40	150
Lampshade2	24	243	5	70	13	47	55	150
Midd1	28	234	15	70	12	47	70	150
Midd2	27	228	15	100	12	47	70	150
Monopoly	29	229	35	220	13	47	25	100
Plastic	21	207	15	220	11	47	25	150
Rocks1	26	189	15	100	16	47	10	50
Rocks2	31	227	25	160	17	47	10	50
Wood1	21	180	10	160	16	47	5	200
Wood2	19	219	30	70	14	47	10	50

Table 2. Penalty values extracted from the tuning process and the values estimated by the suggested SGM-SAP method for each stereo-pair of Middlebury 2006 (using two cost metrics).

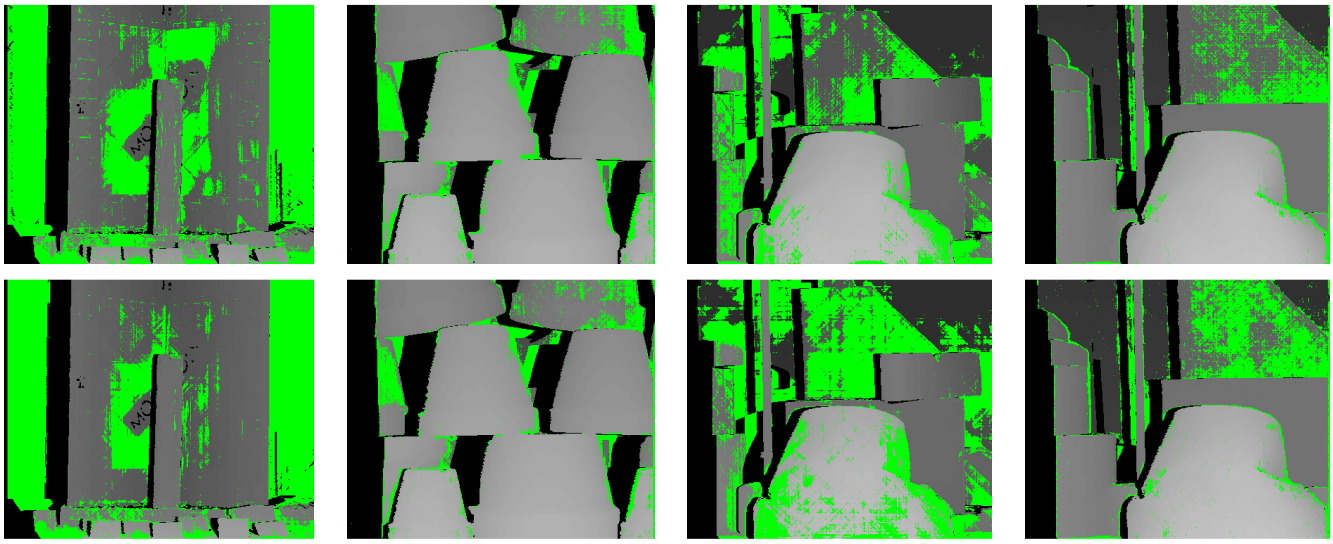

Figure 5. Disparity maps derived from a tuning process (top row) and the suggested method (bottom row). From left to right: stereo-pairs *Monopoly*, *Flowerpots*, *Lampshade1*, *Lampshade2*. In the first two our method (with the use of AD and Census as matching costs) has the best performance; in the other two the optimal parameters of tuning perform best concerning the estimated errors.

In Fig. 5 representative results are seen. In particular, the disparity maps of pairs in which lower errors are obtained with the penalties of our method and the corresponding disparity maps which use the optimal parameters of tuning are shown. On the other hand, the disparity maps of pairs in which lower errors are achieved with the estimated parameters of tuning and the corresponding disparity maps which use the penalties of the proposed method are displayed. Both methods employ Absolute Differences and Census as matching costs. As it may be observed, our method performs better in slanted surfaces with adequate texture (e.g. the *Monopoly* board or the surface of a flowerpot in the corresponding pair). However, its performance lags behind SGM when matching surfaces are of low texture (e.g. the magazine box and the foreground object in the *Lampshade1* and *Lampshade2* pairs).

4.3 *Herz-Jesu-K7* dataset

The *Herz-Jesu-K7* pair (6 Mpixel images: 0006.png, 0007.png) in quarter-size resolution was used as a scenario of an architectural scene. In the cost calculation step the Hamming distance on Census-transformed images was computed. Penalties P_1 and P_2 for SGM were then estimated with the proposed method and were used for the SGM algorithm. In Table 3 the computed penalties for each pair are seen. The initial disparity map was derived by WTA strategy. Finally, erroneous disparities are identi-

fied via the left-right consistency check. Fig. 6 shows the epipolar images of *Herz-Jesu-K7* and the estimated disparity map.

Stereo-pair	Penalties	
	P_1	P_2
Left-to-right	12.6	47
Right-to-left	13.1	47

Table 3. Penalty values for the stereo-pair of *Herz-Jesu-K7* estimated by the suggested method.

Figure 6. Results of SGM-SAP on the Herz-Jesus stereo-pair. *Top*: epipolar images of the stereo-pair; *2nd row*: estimated disparity map of the base image; *3rd row*: registration of the reconstructed point cloud onto the ground truth data; *bottom*: detail of the registration between the laser scanner point cloud and the image-based reconstructed model.

The accuracy of reconstruction can be estimated after registration of the generated point cloud onto the ground truth data (obtained by laser scanning) via the ICP algorithm. It is noted that first some minor pre-processing of the point cloud was conducted (only the object of interest was kept). The overall mismatch is represented by an average distance of 25 mm and a standard deviation of 20 mm. If reduced to mean image scale, these values correspond to ~1.1 and ~1.1 pixel, which are considered as satisfactory. In Fig. 6 an image of the result of the registration is shown, while a detail of the reconstruction is also illustrated.

5. CONCLUSIONS

This work has presented a novel approach (SGM-SAP) aiming at the self-adjustment of penalty values of Semi-Global Matching for any image pair for any matching cost method. This is achieved by the automatic estimation of the penalties through a simple process with low computational requirements, relying on the Disparity Space Image (DSI) volume, which has been already computed in the previous step of the matching process. Therefore, no tuning of penalties is needed and no dataset of similar images with corresponding ground truth disparity maps has to be available. The proposed method has been evaluated on the challenging Middlebury-Version 3 stereo-pairs, as well as on Middlebury 2006 datasets. Results show that the percentages of errors of the estimated disparity maps from SGM-SAP are competitive to the results from the typical SGM approach (in essence they differ by only ~2%). The significance of the proposed method of self-adjusting penalties is that in existing applications of SGM the values of these penalties are generally being estimated after a time-consuming tuning process.

Future work includes attempts for further improvements of the method and testing it with the use of other matching cost methods or SGM-like approaches. Furthermore, evaluation of the suggested method on more complex or outdoor scenes, e.g. on the KITTI dataset, will be conducted in the near future.

REFERENCES

Banz C., Pirsch P., Blume H., 2012. Evaluation of penalty functions for Semi-Global Matching cost aggregation. ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXIX-B3, pp. 1–6.

Chuang T.Y., Ting H.W., Jaw J.J., 2016. Hybrid-based dense stereo matching. ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B3, pp. 495–501.

Georgousis S., Stentoumis C., Doulamis N., Voulodimos,A., 2016. A hybrid algorithm for dense stereo correspondences in challenging indoor scenes. IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, pp. 460–465.

Hirschmüller H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. Proc. Computer Vision and Pattern Recognition, pp. 807–814.

Hirschmüller H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30, pp. 328–341.

Hu X., Mordohai P., 2012. A quantitative evaluation of confidence measures for stereo vision. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34, pp. 2121–2133.

Huang X., Zhang Y., Yue Z., 2016. Image-guided non-local dense matching with three-steps optimization. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, III-3, pp. 67–74.

Li X., Liu J., Chen G., Fu H., 2016. Efficient methods using slanted support windows for slanted surfaces. IET Computer Vision, 10(5), pp. 384–391.

Luo W., Schwing A.G., Urtasun R., 2016. Efficient deep learning for stereo matching. IEEE Conference on Computer Vision and Pattern Recognition, pp. 5695–5703.

Scharstein D., Hirschmüller H., Kitajima Y., 2014. High-resolution stereo datasets with subpixel-accurate ground truth, German Conference on Pattern Recognition. Springer, pp. 31–42.

Spyropoulos A., Mordohai P., 2015. Ensemble classifier for combining stereo matching algorithms. IEEE International Conference on 3D Vision, Lyon, pp. 73–81.

Stentoumis C., Karkalou E., Karras G., 2015. A review and evaluation of penalty functions for Semi-Global Matching. IEEE International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, pp. 167–172.

Strecha C., von Hansen W., van Gool L., Fua P., Thoennessen U., 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. IEEE Computer Vision and Pattern Recognition.

Szeliski R., 2011. Computer Vision, Texts in Computer Science. Springer, London.

Tombari F., Mattoccia S., di Stefano L., Addimanda E., 2008. Classification and evaluation of cost aggregation methods for stereo correspondence. IEEE Computer Vision and Pattern Recognition.

Wang Z.-F., Zheng Z.-G., 2008. A region based stereo matching algorithm using cooperative optimization. Computer Vision and Pattern Recognition.

Yang Q., 2012. A non-local cost aggregation method for stereo matching. IEEE Conference on Computer Vision and Pattern Recognition. pp. 1402–1409.

Yang Q., Wang L., Yang R., Stewénius H., Nistér D., 2009. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, pp. 492–504.

Zbontar J., LeCun Y., 2015. Computing the stereo matching cost with a convolutional neural network. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1592–1599.

Zhang K., Fang Y., Min D., Sun L., Yang S., Yan S., Tian Q., 2014. Cross-scale cost aggregation for stereo matching. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1590–1597.