A Census-Based Stereo Vision Algorithm Using Modified Semi-Global Matching and Plane Fitting to Improve Matching Quality[*]

Martin Humenberger, Tobias Engelke, Wilfried Kubinger
AIT Austrian Institute of Technology
Donau-City-Strasse 1, 1220 Vienna, Austria
martin.humenberger@ait.ac.at, tobias.engelke@ait.ac.at, wilfried.kubinger@ait.ac.at

Abstract

This paper introduces a new segmentation-based approach for disparity optimization in stereo vision. The main contribution is a significant enhancement of the matching quality at occlusions and textureless areas by segmenting either the left color image or the calculated texture image. The local cost calculation is done with a Census-based correlation method and is compared with standard sum of absolute differences. The confidence of a match is measured and only non-confident or non-textured pixels are estimated by calculating a disparity plane for the corresponding segment. The quality of the local optimized matches is increased by a modified Semi-Global Matching (SGM) step with subpixel accuracy. In contrast to standard SGM, not the whole image is used for disparity optimization but horizontal stripes of the image. It is shown that this modification significantly reduces the memory consumption by nearly constant matching quality and thus enables embedded realization. Using the Middlebury ranking as evaluation criterion, it is shown that the proposed algorithm performs well in comparison to the pure Census correlation. It reaches a top ten rank if subpixel accuracy is supposed. Furthermore, the matching quality of the algorithm, especially of the texture-based plane fitting, is shown on two real-world scenes where a significant enhancement could be achieved.

1. Introduction

3D data perception of the surrounding environment of a robot platform or an autonomous vehicle is essential for reliable operation. Common sensors are based on laser, radar, or time-of-flight. These techniques enable high quality 3D perception with the drawback of low resolution and high costs. For a number of robot applications such as people or scene recognition as well as robot navigation digital cameras are used. Stereo vision is technology that uses two in parallel mounted digital cameras to determine the depth of a scene. Advantages are the low price, the high resolution and the fact that the images can be used for any other application as well. For home applications it is also quite useful because it is purely passive technology and thus does not effect the surrounding environment.

For depth calculation the so called correspondence problem (stereo matching), which is the search for corresponding projections of the same scene point onto both camera planes, has to be solved. The horizontal displacement of corresponding pixels is denoted as disparity. Area-based stereo matching algorithms try to calculate the complete disparity map, which is an image of the same size as the camera images with the disparity instead of the intensity value for each pixel. The advantage is that with a single capture a huge number of surrounding 3D points can be determined. The matching process is based on similarity comparison of areas of the images (correlation), thus textureless areas are a difficult challenge. Pixels visible in only one of the images are called occlusions and obviously cannot be found by correlation.

In general, area-based matching algorithms calculate the costs for each matching candidate and optimize them afterwards to find the correct matches. Once the local costs are calculated, a minimum search (*winner takes all*, WTA) can be used to find the best matching pixels. Another strategy is to apply global optimization to the local costs to enhance the probability of correct matching. Here, not only the pixels' neighborhoods are used to calculate the costs, but the whole scanline or even the whole image. With these techniques, especially on textureless areas better results can be achieved. The drawback of global optimizing algorithms is the huge processing time and memory consumption. To the authors' knowledge, no implementation of a global optimization is commercially available for purely embedded real-time platforms without dedicated hardware such as field programmable gate arrays (FPGA).

[*]This work was has been supported by the European Union project ROBOTS@HOME under grant FP6-2006-IST-6-045350.

The goal of this work is to enhance the matching quality of local matching approaches by the use of global optimization techniques. The challenge is to keep the computational effort and the memory consumption low to enable embedded and real-time processing.

2. Related Work

Lots of research has been done in stereo vision, thus a large number of stereo matching approaches exists. A good comparison of many different stereo matching algorithms can be found in [14, 4]. Lots of algorithms use a local costs function such as *Sum of Absolute Differences* (SAD) or *Sum of Squared Differences* (SSD). A good evaluation of costs functions for stereo matching can be found in the work of Hirschmueller and Scharstein [10, 11]. Other methods use the *Census transform*, introduced by Zabih and Woodfill [18], where the costs are calculated using the *Hamming distance* of two Census transformed pixels. The Census transform itself is a non-parametric local transform that uses intensity differences within a pixel's neighborhood to determine a bit string representing that pixel.

Well known global optimization techniques are e.g. *Dynamic Programming* [2, 8], *Graph Cuts* [13], *Belief Propagation* [17, 16, 7], or *Semi-Global Matching* [9, 6]. To overcome the problem of occlusions, approaches based on image segmentation [3, 1, 15] came up. The goal is to determine initial disparities for each segment and fit a model onto them. The model can then be used to refine the disparities inside the segment with the assumption that all pixels inside the segment follow the assumed model.

In the work of Humenberger et al. [12], a very fast real-time implementation of a Census-based, local optimizing stereo matching algorithm was introduced. The processing time was evaluated for several platforms (central processing unit, digital signal processor, and graphics processing unit) reaching real-time capability on all of them. This algorithm was especially designed for home-robot applications and thus has to cope with textureless areas such as white walls. The used local Census transform with a large window size of 16 × 16 can deal quite well with it but has its limitations. Obviously, at textureless areas larger than the Census window no reliable matching is possible.

3. Proposed Algorithm

Figure 1 shows the workflow of the proposed algorithm. First, the images are captured with the calibrated stereo camera, the lens distortion is corrected, and the image pair is rectified. Second, the initial costs are calculated using the sparse Census correlation of [12]. Then, a modified semi-global matching (SGM) is applied to increase the confidence of the matches, and thus to determine the initial disparity map. Afterwards, a segmentation is done on either the left stereo image or the texture map. A planar model is fitted onto the segments which is used to finally determine the refined disparity map.

3.1. Census Correlation

For fitting a planar model onto the image segments, an initial disparity map is needed. Census correlation proved to be a good choice for reliable costs calculation and is thus used in this work. The Census transform uses local intensity differences ($n \times m$) around each pixel to transform the intensity value to a bit string with

$$T(u,v) := \bigotimes_{i=-n'}^{n'} \bigotimes_{j=-m'}^{m'} \xi(I(u,v), I(u+i, v+j)) \ , \quad (1)$$

where $I(u,v)$ is the intensity of pixel (u,v), $n' := \lfloor \frac{n}{2} \rfloor$, $m' := \lfloor \frac{m}{2} \rfloor$, and $\bigotimes$ denotes a bit-wise catenation. The auxiliary function ξ is defined as

$$\xi(x,y) := \begin{cases} 0 & \text{if } x \leq y \\ 1 & \text{if } x > y \end{cases} \ . \quad (2)$$

The costs of two Census transformed pixels are determined with the Hamming distance of the two bit strings and are calculated with

$$C(u,v,d) := \text{Hamming}(T_r(u,v), T_l(u+d,v)) \ , \quad (3)$$

where $T_r(u,v)$ and $T_l(u,v)$ are the bit strings of the pixel (u,v) in the left and right images, respectively.

3.2. Modified Semi-Global Matching

Semi-global matching was first introduced by Hirschmueller [9]. This technique minimizes the global energy in horizontal, vertical, and diagonal directions. Hereby, an eight or sixteen neighborhood can be used. The costs-path $L_r(p, d_p)$ of the pixel $p := (u,v)$ at disparity d_p in direction r is calculated recursively with

$$\begin{aligned} L_r(p,d_p) := C(p,d_p) + \min(&L_r(p-r, d_p), \\ &L_r(p-r, d_p-1) + P_1, \\ &L_r(p-r, d_p+1) + P_1, \\ &\min_{k \in \mathcal{D}} L_r(p-r, k) + P_2) \ , \end{aligned} \quad (4)$$

where P_1 is a penalty, which is added if the disparities differ by one and the penalty P_2 is added if the disparities differ by more than one ($P_1 < P_2$). $\mathcal{D}$ is the set of all possible disparities. Afterwards the costs S are summed up over all paths in all directions r

$$S(p, d_p) := \sum_r L_r(p, d_p) \ . \quad (5)$$

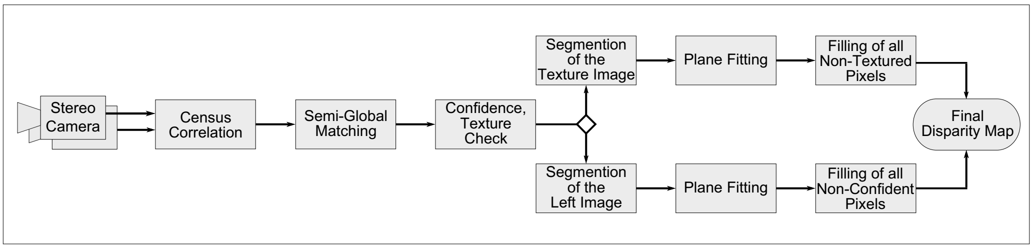

Stereo Camera
Census Correlation
Semi-Global Matching
Confidence, Texture Check
Segmention of the Texture Image
Segmention of the Left Image
Plane Fitting
Plane Fitting
Filling of all Non-Textured Pixels
Filling of all Non-Confident Pixels
Final Disparity Map

Figure 1. The workflow of the proposed algorithm.

For each pixel the disparity with the lowest costs is selected to be the initial disparity (WTA).

Semi-global matching determines the optimal paths through the whole image for each pixel, thus the costs of the path have to be stored for the whole image. Zinner et al. [19] showed that an optimized high-speed implementation of a Census-based stereo matching approach benefits from a line-by-line processing of the images. Only a number of lines equal to the aggregation block size has to be stored at once. Especially for embedded systems this approach is advantageous because the data can then be processed in the fast on-chip memory. To keep the benefit of line-by-line processing, a modified SGM technique is introduced in this work. It uses the assumption that a part of the image is enough for each pixel to benefit from the SGM. Therefore the initial costs matrix is divided into horizontal stripes with a range of n_r pixels (the last stripe may be smaller). The stripes are treated like the whole image and the paths are calculated with Equ. 4 as well. For determining the optimal paths through the stripes a number equal to the range of the initial costs has to be stored. Thus, the memory consumption depends on the size of the range and the number of disparities. The stripes are then processed separately and the resulting disparity map (DM) is stored as a combination of the total number of stripes. The influence of this modification in terms of matching quality, processing time, and memory consumption is described in detail in Sec. 4.1.

3.3. Confidence and Texture

Even if SGM increases the reliability of the matches, a number of false positives remain. To determine them, a confidence value is calculated for each match during costs optimization. As mentioned above, large textureless areas are difficult to match even if SGM is done over the whole image. To identify them a texture image is calculated.

The confidence is calculated as the relation of the costs difference between the best two matching candidates and the maximum possible costs with

$$\mathrm{CM}(u,v) := \min\left\{255, 1024\left(\frac{\Delta(u,v)}{\mathrm{MaxCosts}}\right)\right\} \quad , \qquad (6)$$

where $\Delta(u,v)$ is the difference between the best two matching candidates for pixel (u,v). The texture is the result of a variance filter over an $n \times m$ window with

$$\begin{aligned}\mathrm{TM}(u,v) := \frac{1}{nm}\sum_{i=-n'}^{n'}\sum_{j=-m'}^{m'} I(u+i,v+j)^2 \\ -\left(\frac{1}{nm}\sum_{i=-n'}^{n'}\sum_{j=-m'}^{m'} I(u+i,v+j)\right)^2 \quad .\end{aligned} \qquad (7)$$

Non-confident pixels and pixels in textureless areas are then determined by the use of the two thresholds τ_1 and τ_2. Only pixels which pass the confidence and texture check

$$\mathrm{DM}_{\mathrm{init}}(u,v) := \begin{cases} \mathrm{DM}(u,v) & \text{if } \mathrm{CM}(u,v) \geq \tau_1 \\ & \wedge\ \mathrm{TM}(u,v) \geq \tau_2 \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

are used for the initial disparity map.

3.4. Segmentation and Plane Fitting

Once the initial disparity map is calculated, textureless areas and non-confident pixels are optimized with segmentation and plane fitting. The segmentation can either be done by color on the left input image (mean-shift [5]) or binary on the texture image. The texture image (TI) is derived from the texture map with

$$\mathrm{TI}(u,v) := \begin{cases} 0 & \text{if } \mathrm{TM}(u,v) \leq t_{texture} \\ 255 & \text{otherwise} \end{cases} \quad , \qquad (9)$$

where $t_{texture}$ is the used threshold. The segmentation process on the binary texture image is straight forward. All white pixels are united to one segment and all connected black pixels are joint to single segments.

An advantage of the texture segmentation is that monochrome input images can be used as well as color images. On the one hand, monochrome cameras deliver images of higher quality than color cameras and on the other hand, for this kind of segmentation the focus exactly lies on textureless areas which are the main regions of interest for optimization. The advantage of color segmentation is that the segments are more accurate and that occlusions can better be optimized. Section 4 shows that color segmentation proved to be more suitable for the Middlebury datasets and texture segmentation for real-world scenes.

Both segmentations have in common that only pixels which successfully passed the confidence check are used for the plane fitting step. A plane is represented by three parameters a, b, and c of equation

$$d(u,v) := au + bv + c \ . \tag{10}$$

These parameters can be estimated with the method of least squares by solving the linear equation system

$$\begin{bmatrix} \sum_{i=1}^{m} u_i^2 & \sum_{i=1}^{m} u_i v_i & \sum_{i=1}^{m} u_i \\ \sum_{i=1}^{m} u_i v_i & \sum_{i=1}^{m} v_i^2 & \sum_{i=1}^{m} v_i \\ \sum_{i=1}^{m} u_i & \sum_{i=1}^{m} v_i & \sum_{i=1}^{m} 1 \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} u_i d_i \\ \sum_{i=1}^{m} v_i d_i \\ \sum_{i=1}^{m} d_i \end{pmatrix} , \tag{11}$$

where m is the number of confident pixels in the segment. Unfortunately this is not robust against outliers, thus the method described by Bleyer and Gelautz [3] is used. The problem is solved by iteratively eliminating outliers until the calculated plane has reached its final state.

A problem of segmentation regarding real-time capability is that the processing time strongly depends on the number of segments found. This work tries to deal with this problem by limiting the number of possible segments. The authors know that this is just a first step towards real-time segmentation because also the absolute number of confident pixels inside the segments influences the processing time.

After plane fitting, the last step is to optimize and refine the initial disparity map with the calculated planar model.

3.5. Disparity Map Refinement

In contrast to traditional model-based segmentation optimization, in this work only non-confident pixels (which failed the confidence check) or pixels in textureless areas (which failed the texture check) are refined with the calculated planes. The others are taken from the initial disparity map. Additionally, only reliable segments are used for refinement because in difficult areas the initial data may be not good enough for a correct model estimation. The reliability of the planes is differently determined for color and texture segmentation.

For color segments the function

$$\Omega_c(C) := \begin{cases} \text{true} & \text{if } \frac{n_c}{n_p} \leq t_{confidence} \\ \text{false} & \text{otherwise} \end{cases} \tag{12}$$

is used where C is the segment, n_c the number of non-confident pixels and n_p the number of pixels in C. If the segment is reliable, thus the fraction of confident pixels in the segment is higher than the given threshold $t_{confidence}$, Ω_c is true and false otherwise.

In large textureless areas often a low number of confident pixels exists. The use of Ω_c would not be advantageous because the percentage of confident pixels in textureless areas varies with the segment size. To overcome this, another reliability metric,

$$\Omega_t(C) := \begin{cases} \text{true} & \text{if } \delta \leq t_{plane} \\ \text{false} & \text{otherwise} \end{cases} , \tag{13}$$

is introduced to measure the quality of the estimated plane where t_{plane} is the used threshold. The criterion is the average distance

$$\delta := \frac{1}{m} \sum_{i=0}^{m} |d_i - (au_i + bv_i + c)| \ , \tag{14}$$

between the points and the estimated plane, where m is the number of confident pixels in the segment.

Summarizing, the last step of the proposed algorithm is the refinement of the initial disparity map. For color segmentation only non-confident pixels and for texture segmentation only pixels in textureless areas are refined. The reliability of the estimated planes is determined and only reliable planes are used for this final optimization.

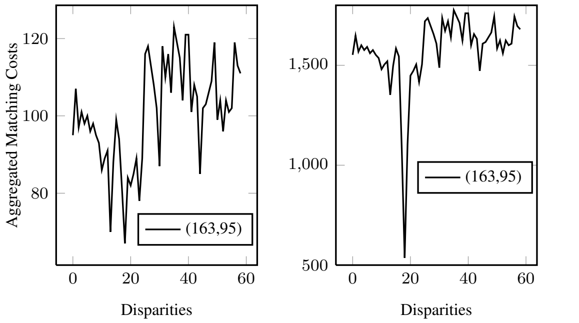
4. Evaluation

This section presents the results of the proposed algorithm. First, the matching quality, the processing time, and the memory consumption of the modified semi-global matching is evaluated. Then, on the one hand, for evaluation of the matching quality the well known Middlebury ranking is used. The main advantage is the possibility of comparing the stereo vision algorithm with many others online. The datasets used for this evaluation are not realistic representatives for the target application, thus results for real-world scenes are shown on the other hand.

4.1. Modified Semi-Global Matching

Semi-Global Matching optimizes the disparities in either 8 or 16 directions with the use of two penalties $P_1 = 54$ and $P_2 = 99$. The use of 16 directions showed no considerable enhancement of the results so 8 directions are used because of the shorter processing time. The optimal penalties were determined by evaluation of all meaningful combinations.

Figure 2 shows an evaluation of matching quality and memory consumption for the modified SGM approach. As can be seen in Fig. 2(a), the average percentage of matched pixels over the four main ranking Middlebury datasets with ranges $n_r = 5(5)190$ is very similar to the original approach (straight black line). The enhancement of the modified SGM is the reduced memory consumption. Original SGM has a memory consumption of about 40 MB for the



(a) Census (b) SGM

Figure 3. Illustration of the confidence improvement by using modified SGM with $n_r = 55$. The confidence value using pure Census is $\mathrm{CM}(u, v) = 13, 65$ and with SGM the maximum of $\mathrm{CM}(u, v) = 255$.

Teddy dataset. As can be seen in Fig. 2(b) if a range of about 55 is used, the memory consumption is about 5 MB. If a very small range of 5 is used, the memory consumption even is about 300 KB which makes it very suitable for embedded realization. For better visualization Fig. 2(c) shows the memory consumption and the percentage of correct matches for the Tsukuba dataset plotted in one chart.

As a reminder, the confidence of the matches is essential for successful plane fitting. Figure 3 shows the improvement of the confidence when modified SGM is used. Both costs functions show the same correctly matched pixel (disparity is at the lowest costs) for Census on the left side and for Census with modified SGM on the right side. The difference between the two best matching candidates is very low for Census, thus the confidence is very low. SGM highly increases the difference and thus the confidence as well. The more correct matches are marked as confident, the more can be used for the plane fitting step what again increases the quality of the final disparity map.

4.2. Middlebury Ranking

Scharstein and Szeliski [14] have developed an online evaluation platform, the Middlebury stereo evaluation [14], which provides about 40 stereo image datasets. The main feature is an online comparison of submitted area-based stereo matching algorithms. To evaluate an algorithm on this website, disparity maps of four datasets have to be generated and uploaded. In the proposed algorithm, due to occlusions or non-confident areas, not all pixels have a corresponding match. For the Middlebury evaluation a completely dense disparity map is mandatory so the missing pixels have to be extrapolated. For this evaluation, the color-based segmentation approach is used because it has the big advantage that many occluded areas (if Ω_c is true) are filled with the calculated planes rather with extrapolation. Outliers are reduced with a final median filter.

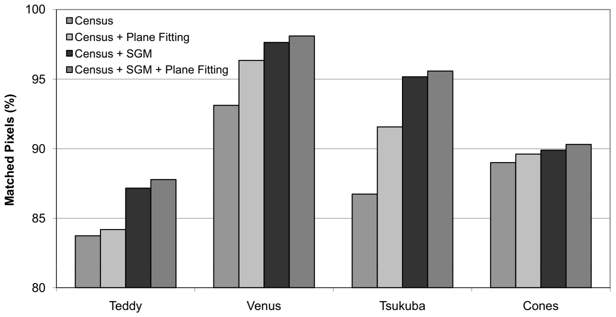


Figure 5. The percentage of correctly matched pixels with Census correlation, plane fitting (color segmentation) and SGM.

The resulting disparity maps are compared with the ground truth, which is the reference disparity map of the scene. Figure 4 shows the resulting disparity maps of the proposed algorithm. Figure 5 shows the resulting improvements of the different algorithm steps.

Table 1 compares different algorithm configurations in the Middlebury evaluation framework. The best result in the ranking (rank 37) could be achieved with a combination of Census correlation, SGM and plane fitting. Additionally to the proposed algorithm steps, the results of standard SAD for local costs calculation are shown.

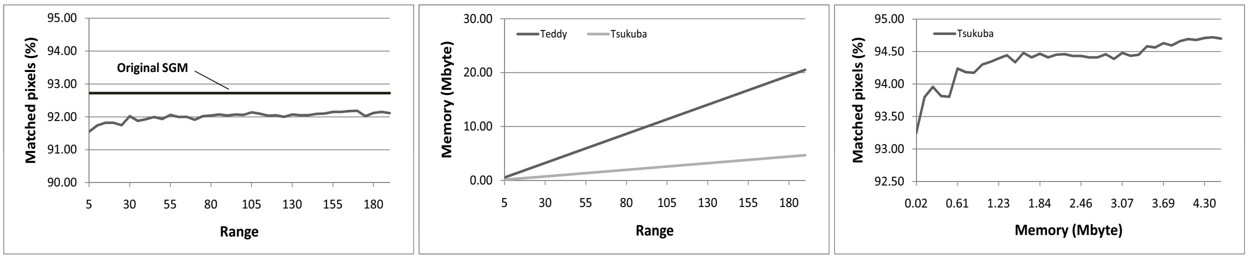
As can be seen, SGM clearly improves the quality of the matches. When using the proposed modified SGM technique the rank shrinks a few places. This is caused by the fact that the entries in the Middlebury ranking are very close together so little worse results may cause significant degradation in the ranking. A meaningful metric is the average bad matches percentage. It shows that the overall performance of original SGM and the modified version is quite similar. Also interesting is, when using the average bad matches as criterion, that SGM produces nearly the same matching quality for Census correlation and SAD.

An important factor in Table 1 is the confidence threshold. As mentioned in the previous section, SGM significantly increases the confidence of matched pixels. In comparison to SAD and Census, for SGM a much higher confidence threshold can be used without eliminating too many true positives. This increases the number of confident matches which is essential for the plane fitting step.

The evaluation of the processing time shows that the use of multi-core central processing units (CPU) reasonably accelerates the processing. The only performance optimization is the parallel processing of the functional behavior with OpenMP[1], thus an optimized implementation can for sure achieve a lower processing time.

In the main ranking of the Middlebury website, all matches within an error threshold of 1 are valid. If the error threshold is set to 0.5, which means that subpixel accuracy

[1] http://www.openmp.org



(a) (b) (c)

Figure 2. Evaluation of different ranges n_r for modified SGM: (a) Percentage of correct matched pixels (average over the Middlebury datasets), (b) memory consumption, and (c) a combined chart of memory consumption and correct matches.

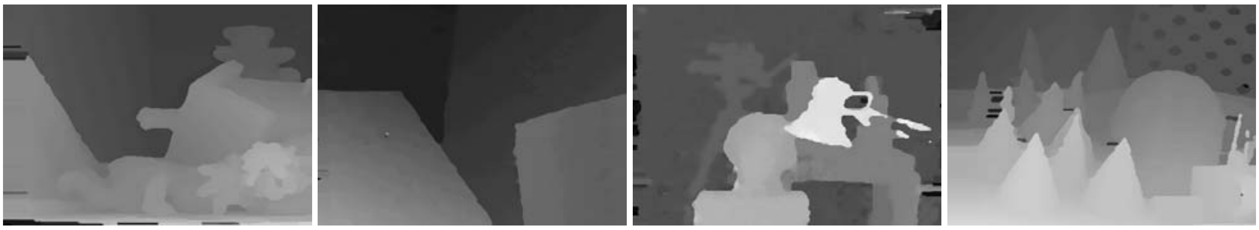


Figure 4. The results of the proposed algorithm (Census correlation with SGM and color-based segmentation) for the Middlebury datasets.

is supposed, the best position of the proposed algorithm increases to rank 10.

4.3. Real-World Scenes

The Middlebury stereo database gives a good idea of the matching quality in comparison to other approaches. The drawback is that the datasets are created under very controlled conditions with high quality digital cameras which cannot be found in real-world applications. To show the power of plane fitting with texture segmentation, two realistic scenes for robot applications are evaluated.

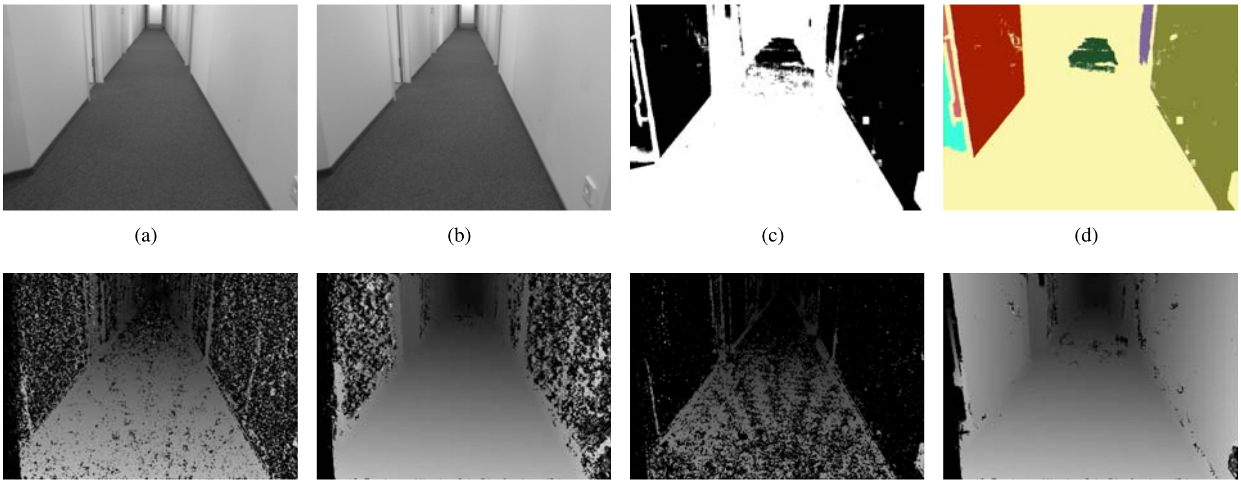
Figure 6 shows a floor scene which is difficult for area-based stereo matching approaches. In general, randomly patterned surfaces, such as the carpet in this scene, can be matched well. The most difficult areas for stereo matching here are the monotone white walls (marked black in the texture image in Fig. 6(c)). The pure Census correlation in Fig. 6(e) can deal with the carpet well but has its problems with the walls. The same for the combination of Census and SGM in Fig. 6(f) with the enhancement that the carpet is completely dense. The walls are in both disparity maps reduced to noise. To deal with this, the confidence check was introduced which eliminates obviously wrong matches. The result of Census correlation with confidence check in Fig. 6(g) shows that the disparity map is very sparse and almost all matches of the walls are eliminated. The proposed algorithm was developed exactly to optimize such scenes. The resulting disparity map in Fig. 6(h) shows that areas of the image with enough texture (white in the texture image) are kept original and areas with low texture (black in the texture image) are used for the segmentation-based optimization. The quality of the planes strongly depends on the data used for fitting, so a high confidence threshold of $\tau_1 = 200$ is used. To show the quality of the 3D data the 3D point clouds for three algorithm configurations are given in Fig. 7. As can be seen, the walls are completely wrong when no optimization is used. Only the planes in Fig. 7(e) are good estimates of the walls in the scene. Especially Fig. 7(c) shows the impact of the higher confidence of Census in comparison to SAD. Not all textureless areas can be optimized using texture-based segmentation. Figure 8 shows an example where an estimated plane does not fit correctly. Here, the estimated plane in Fig. 8(c) of the wall behind the door is obviously wrong. Therefore the threshold function Ω_t was introduced to eliminate such planes as shown in Fig. 8(d). A limitation of the approach is that the fitted planes are only estimations of the real world. Problematic are curved surfaces because a plane cannot be fitted on there. However, most curved surfaces are not textureless in the images because of different light reflections on the surfaces. Additionally, the probability that such a surface would be eliminated by Ω_t is high because the distance of the points from the curved surface to the estimated plane is large. Nevertheless, in indoor home robot applications, the assumption that textureless areas are planar in many cases can be made.

5. Conclusion and Future Work

This paper introduced a stereo matching approach consisting of a combination of Census-based correlation, SGM disparity optimization, as well as segmentation-based plane

Table 1. The proposed algorithm in different configurations evaluated with the Middlebury framework with the processing time for a multi-core CPU and the used confidence threshold.

	Threshold = 1.0		**Threshold = 0.5**		**Processing time (ms)**						
		Av. bad		Av. bad	1 core		2 cores		4 cores		**Confidence**
Algorithm	Rank	matches	Rank	matches	Teddy	Tsukuba	Teddy	Tsukuba	Teddy	Tsukuba	**threshold**
Census	56	9.86	16	14.40	582	129	348	83	230	58	30
Census + SGM	40	8.35	9	12.10	6931	834	3820	489	2142	262	95
Census + SGM + Plane Fitting	37	8.19	10	12.20	17526	4604	11708	4019	8362	3622	95
SAD	66	13.20	57	22.20	505	99	345	76	252	59	5
SAD + SGM	47	8.62	19	10.50	6839	824	3851	476	2171	267	10
SAD + SGM + Plane Fitting	46	8.35	19	15.20	18078	4653	11730	4019	9398	3906	10
Census + SGM ($n_r = 10$)	52	9.19	14	13.70	6175	946	4585	545	3010	481	95
Census + SGM ($n_r = 55$)	55	9.70	14	13.90	5159	689	2877	461	2069	368	95
Census + SGM ($n_r = 180$)	51	9.05	11	12.90	5661	757	3146	641	2978	571	95
Census + SGM ($n_r = 10$) + Plane Fitting	48	8.84	12	13.70	16123	4853	12448	4086	9561	3561	95
Census + SGM ($n_r = 55$) + Plane Fitting	52	9.32	14	14.00	14096	4311	9907	3971	8326	3726	95
Census + SGM ($n_r = 180$) + Plane Fitting	46	8.90	11	13.10	15206	4513	10298	4282	10248	4177	95



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 6. The results of the floor scene with large textureless areas: (a) original left image, (b) original right image, (c) texture image ($\tau_2 = 20$), (d) texture-based segmentation, (e) disparity map for pure Census correlation, (f) disparity map for Census correlation and SGM, (g) disparity map for Census correlation with confidence check, (h) disparity map for Census correlation, SGM, and plane fitting.

fitting for enhancements on textureless and occluded areas. The algorithm is designed for robot applications such as navigation or scene interpretation and the single steps are capable for real-time implementation. A modification of original SGM makes the approach capable for embedded realization as well. The image is divided into stripes that may fit into fast on-chip memory of digital signal processors. Semi-Global Matching significantly increases the confidence of the matches. It is shown that the segmentation-based plane fitting performs well with the Census-based correlation method. The main advantage is the improvement of the matching quality in occluded and textureless areas. Furthermore it is shown that the texture-based segmentation approach makes it possible to match large textureless areas very well which are a significant problem for standard area-based stereo matching approaches.

In future research more plane fitting techniques i.e. the Random Sample Consensus (RANSAC) algorithm, will be evaluated. Furthermore, a different set of models that can deal with curved surfaces as well will be developed. To prove the real-time capability an optimized implementation on a GPU of the proposed algorithm as well as an embedded realization of the modified SGM approach are planned.

References

[1] K. Andreas, S. Mario, and K. Konrad. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. pages 15–18, 2006.

[2] S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision*, 35(3):269–293, 1999.

[3] M. Bleyer and M. Gelautz. A layered stereo matching algorithm using image segmentation and global visibility con-

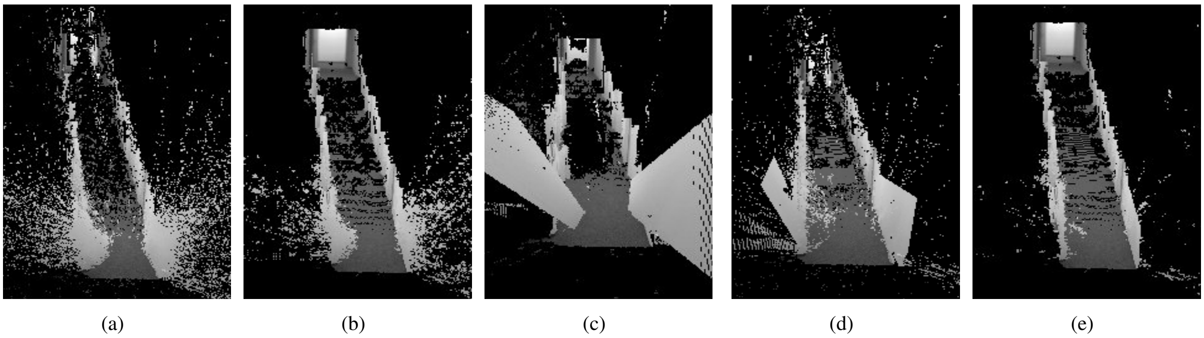


Figure 7. 3D point cloud of the floor scene with (a) pure Census correlation (without confidence and texture check), (b) Census with SGM, (c) SAD with SGM and plane fitting, (d) Census with plane fitting, (e) Census with SGM and plane fitting.

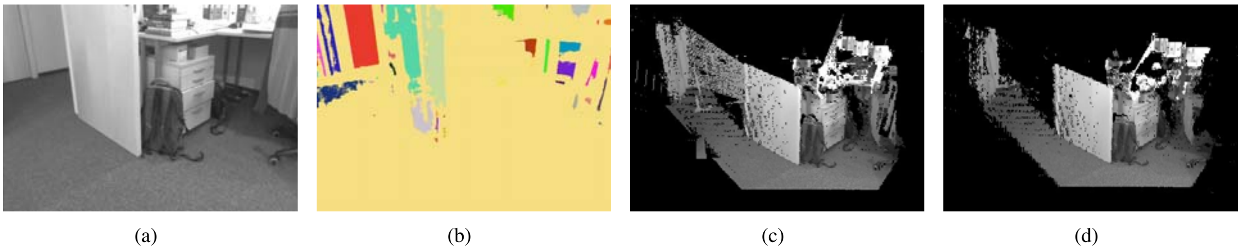


Figure 8. Results of the desk scene: (a) Left stereo image, (b) the segmentation of the texture image, (c) 3D point cloud with SGM and texture-based optimization without plane check, and (d) with plane check and a threshold of $t_{plane} = 0.2$.

straints. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59:128–150, 2005.

[4] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:993–1008, 2003.

[5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, 2002.

[6] I. Ernst and H. Hirschmueller. Mutual information based semi-global stereo matching on the gpu. pages 228–239, 2008.

[7] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *Int. J. Comput. Vision*, 70(1):41–54, 2006.

[8] S. Forstmann, Y. Kanou, J. Ohya, S. Thuering, and A. Schmitt. Real-time stereo by using dynamic programming. *Computer Vision and Pattern Recognition Workshop*, 3:29, 2004.

[9] H. Hirschmueller. Stereo vision in structured environments by consistent semi-global matching. pages 2386–2393, 2006.

[10] H. Hirschmueller and D. Scharstein. Evaluation of cost functions for stereo matching. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8, 2007.

[11] H. Hirschmuller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1582–1599, Sept. 2009.

[12] M. Humenberger, C. Zinner, and W. Kubinger. Performance evaluation of a census-based stereo matching algorithm on embedded and multi-core hardware. In *Proccedings of the International Symposium on Image and Signal Processing and Analysis*, volume 6, 2009.

[13] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. pages 508–515, 2001.

[14] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2001.

[15] Q. Yang, C. Engels, and A. Akbarzadeh. Near real-time stereo for weakly-textured scenes. *British Machine Vision Conference (BMVC)*, 2008.

[16] Q. Yang, L. Wang, R. Yang, S. Wang, M. Liao, and D. Nister. Real-time global stereo matching using hierarchical belief propagation. *The British Machine Vision Conference*, 2006.

[17] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. pages 239–269, 2003.

[18] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Proceedings of 3rd European Conf. Computer Vision*, pages 151–158, Stockholm, 1994.

[19] C. Zinner, M. Humenberger, K. Ambrosch, and W. Kubinger. An optimized software-based implementation of a census-based stereo matching algorithm. *Lecture Notes in Computer Science, Advances in Visual Computing*, 5358:216–227, 2008.