ANALYSIS OF MEANS TO IMPROVE COOPERATIVE DISPARITY ESTIMATION

Helmut Mayer

Institute for Photogrammetry and Cartography
Bundeswehr University Munich, D-85577 Neubiberg, Germany,
Helmut.Mayer@UniBw-Muenchen.de

KEY WORDS: Cooperative Disparity Estimation, Evaluation, Visualization.

ABSTRACT

(Zitnick and Kanade, 2000) proposes a cooperative approach for disparity estimation from stereo imagery based on support and inhibition in three-dimensional (3D) disparity space. By several means we obtain a significant improvement over the results reported for (Zitnick and Kanade, 2000) in (Scharstein and Szeliski, 2002). The results are in the range of the best approaches in (Scharstein and Szeliski, 2002), while the numerical complexity of our approach compares favorably to these approaches. Our main contribution lies in analyzing the different means for improvement including their performance gain. The most important are the use of symmetric support, the combination of absolute differences and (normalized) cross-correlation weighted by the strength of the horizontal gradient, the use of auto-correlation to estimate the significance of a matching score, the preference for small disparities to obtain more meaningful results for occluded regions, and the enforcement of the alignment of disparity and image gradient. Results for additional images with the same set of parameters show that the means are applicable to a wider range of imagery.

1 INTRODUCTION

Though disparity estimation from stereo imagery has received considerable attention since the first days of computer vision, there is still not one or even a set of standard approaches which can deal with a broad range of imagery. An excellent recent survey (Scharstein and Szeliski, 2002) has grouped existing approaches into a taxonomy and introduced an evaluation metric as well as test data to compare different approaches. For previous work we therefore refer to this survey and introduce only the most recent approaches and compare them to our ideas and results.

(Zitnick and Kanade, 2000), that our work based on, refers to work proposed at the end of the seventies (Marr and Poggio, 1976, Marr and Poggio, 1979). The basic idea is to employ explicitly stated global constraints on uniqueness and continuity of the disparities. While (Marr and Poggio, 1976, Marr and Poggio, 1979) have used two-dimensional (2D) regions to enforce continuity by fusing support among disparity estimates, (Zitnick and Kanade, 2000) employs 3D support regions. Matching scores are calculated for a disparity range (search width) and then stored in a 3D array made up of image width and height as well as disparity range. This array is filtered with a 3D box-filter to obtain the local support for a match from all close-by matches.

Assuming opaque, diffuse-reflecting surfaces, the uniqueness constraint requires that on one ray of view only one point is visible. This implies an inhibition which is realized by weighting down all scores besides the strongest. Support and inhibition are iterated. Thereby, the information is propagated more globally. We have chosen (Zitnick and Kanade, 2000) because it can deal with strong occlusions and large disparity ranges and have extended it by the following means:

- The smoothness of the output is improved by sub-pixel estimation. By a recursive implementation of the 3D box-filter we have sped up the computation. We determine the convergence by calculating a difference image and setting a threshold on its variance.
- Opposed to the original approach, we employ symmetric support. This considerably improves the results.
- As proposed by (Scharstein and Szeliski, 2002), we use for the matching scores besides cross-correlation absolute differences with truncation. We have extended this by combining both. The combination is based on the idea that correlation works best for horizontally textured regions. Therefore, we weight correlation higher for a large horizontal gradient. As we are looking for unambiguous matches, the matching scores are weighted down when a special type of auto-correlation, which is only evaluated outside the matching window and inside the search width, is large.
- It was found that, as expected, using color improves the result. We have introduced a small preference for smaller disparities. This is due to the fact, that occluded regions have a smaller disparity than their occluding regions. By the preference for smaller disparities we increase the probability, that occluded regions, for which no correct match is possible, obtain correct, smaller disparities.
- By combining image gradient and disparity gradient to control the amount of smoothing as proposed by (Zhang and Kambhamettu, 2002), we avoid blurring disparity discontinuities and the elimination of narrow linear structures.
- Finally, determining occlusions and reducing the probabilities for large disparities in these regions is another means to obtain more meaningful, smaller disparities in occluded regions.

The paper is organized as follows. First we give a short account of cooperative disparity estimation as proposed in (Zitnick and Kanade, 2000). Section 3 presents the evaluation metric of (Scharstein and Szeliski, 2002) and our results for the four image pairs obtained with one and the same parameter set. In Section 4 we present the means for improvement in more detail. We analyze them by assessing their performance gain. In Section 5 additional results are presented. The paper ends up with conclusions.

2 COOPERATIVE DISPARITY ESTIMATION

The main idea of the cooperative disparity estimation of (Zitnick and Kanade, 2000) is a cooperation between support and inhibition (cf. Fig. 1, left). The support region is a 2D-region or usually a 3D-box. All matching scores, derived, e.g., by (normalized) cross-correlation, in this box corroborate to generate a disparity map which is locally continuous. When employing a 3D-box, sloped regions are modeled implicitly.

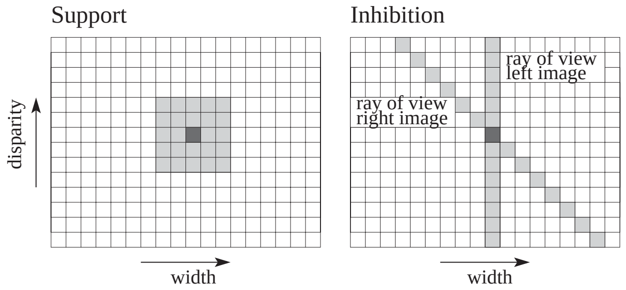


Figure 1: Support and inhibition

Inhibition enforces the uniqueness of a match. Assuming opaque and diffuse-reflecting surfaces, a ray of view emanating from a camera will hit the scene only at one point. The idea is to gradually weight down all matches on a ray of view besides the strongest. For a stereo pair there are two rays (cf. Fig. 1, right). We store the matching scores in a 3D array where for every pixel in the left/reference image the matching score is stored as a voxel. Therefore, for the left image the ray of view is a column in the 2D-slice of width and disparity. Because we work in disparity and not in depth space, the ray of view of the right image consists of the 45° left-slanted diagonal through the pixel of interest. Putting everything together, the support S_n for a pixel at row r and column c with disparity d is defined as

$$S_n(r,c,d) = \sum_{(r',c',d') \in \Phi} L_n(r+r', c+c', d+d') \quad , \qquad (1)$$

with L_n the score for the preceding iteration and Φ the support region. The new score for iteration $n+1$ is obtained as

$$L_{n+1}(r,c,d) \quad = \quad \left(\frac{S_n(r,c,d)}{\sum\limits_{(r'',c'',d'') \in \Psi} S_n(r'',c'',d'')} \right)^{\alpha} \\ * L_0(r,c,d) \quad , \qquad (2)$$

with Ψ the union of the left and right inhibition region and α an exponent controling the speed of convergence. α has to be chosen greater than 1 to make the scores converge to 1. The multiplication with the original matching score L_0 avoids hallucination in weak matching regions. Finally, for each pixel of the left image the disparity is chosen which has the maximum score. Practically, it is important to correct the inhibition value for the fact that on the left and the right side of the image a number of pixels depending on the search width are not matched and, therefore, do not contribute to the inhibition.

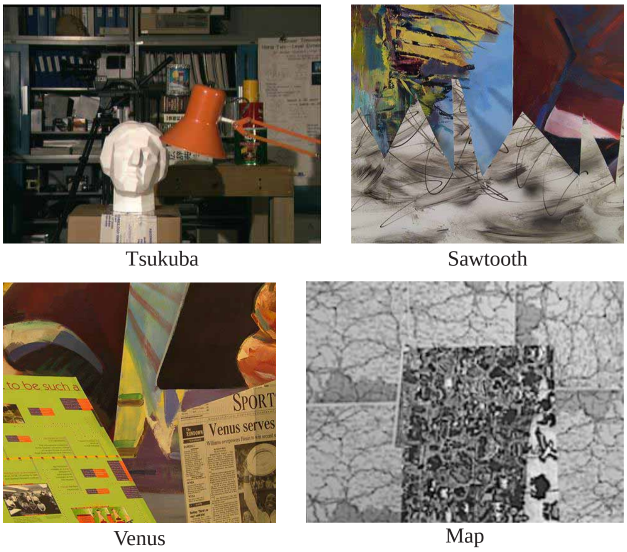
3 EVALUATION

For the evaluation we used the data and the code available at `www.middlebury.edu/stereo` (cf. Figure 2) employing the search widths given there. The measures used in (Scharstein and Szeliski, 2002) and here comprise the number of bad pixels, i.e., pixels which are further away from the ground truth map than a tolerance δ_d. As in (Scharstein and Szeliski, 2002), we also use $\delta_d = 1.0$ and the following measures:

- bad pixels nonocc (all) – $B_{\overline{\mathcal{O}}}$: % bad pixels in non-occluded regions. Used as overall performance measure.
- bad pixels textureless (untex.) – $B_{\overline{\mathcal{T}}}$: % bad pixels in textureless regions.
- bad pixels discont (disc.) – $B_{\mathcal{D}}$: % bad pixels near discontinuities

For sub-pixel estimation a parabola involving the matching scores of the voxels having a smaller $(d-1)$ and larger $(d+1)$ disparity than the given disparity for a pixel d (with $l_n(d) = L_n(r,c,d)$) is used $(-0.5 \leq \Delta d \leq 0.5)$:

$$\Delta d = \frac{l_n(d+1) - l_n(d-1)}{2(2l_n(d) - l_n(d+1) - l_n(d-1))} \quad . \qquad (3)$$


Tsukuba

Sawtooth

Venus

Map

Figure 2: Images (`www.middlebury.edu/stereo`)

The results presented in Figures 3, 4, and 5, for the first and the last image also compared to their ground-truth, give an indication of the quality obtained.

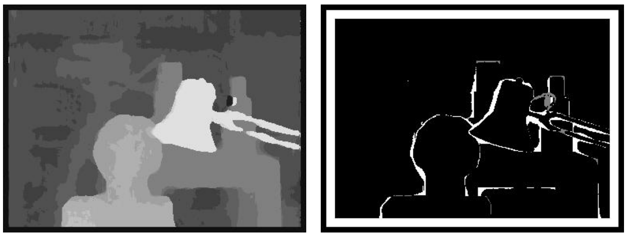


Figure 3: Disparities for Tsukuba (left) and differences to ground truth (right) – white: disparities more than 1 pixel too large; grey: disparities more than 1 pixel too small.

From Table 1 it can be seen that the evaluation of the results compares favorably with other approaches presented in the online version of (Scharstein and Szeliski, 2002) at

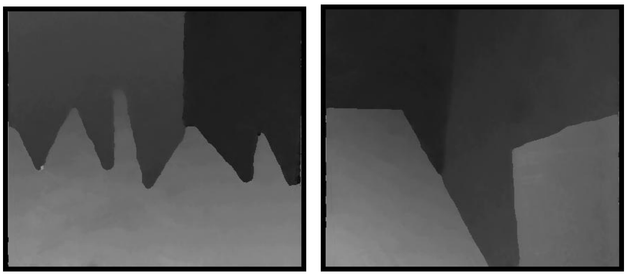


Figure 4: Disparities for Sawtooth (left) and Venus (right) including sub-pixel estimation

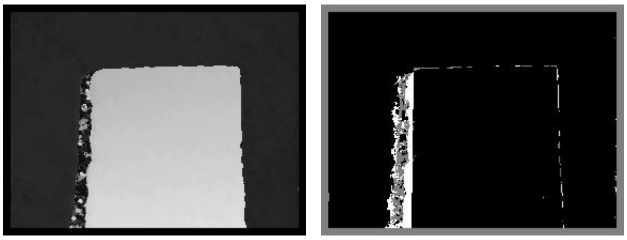


Figure 5: Disparities for Map including sub-pixel estimation (left) and differences to ground truth (right) – white: disparities more than 1 pixel too large; grey: disparities more than one pixel too small.

`www.middlebury.edu/stereo`. As required there, only one parameter set given in Table 3 was used. When summing up the rank for all images and all evaluation criteria we obtain 63 for the pixel precise disparities and are ranked number three in the individual result page of the online version of (Scharstein and Szeliski, 2002) as of April 3, 2003. The first ranked approach has a rank of 52, the second of 55, and the fourth to seventh approach a rank of 68, 72, 74, and 75. Run time for all images is about 102 seconds on a 2.5 GHz PC. This time is better than those reported for the seventh (2706 seconds) and fifth (528 seconds) performing algorithms in (Scharstein and Szeliski, 2002). The times for Tsukuba, Sawtooth, Venus, and Map are 22, 28, 36, and 16 seconds, respectively.

For the interpretation of the results with sub-pixel estimation, which were obtained with the same parameters as above, one needs to consider that while the ground-truth for Tsukuba is pixel precise, the ground-truth for the rest is sub-pixel precise. As the distance for the evaluation is fixed to one pixel and as we restrict $|\Delta d| \leq 0.5$, for Tsukuba sub-pixel estimation can only result into an equal or lower performance. For the other three images the performance can improve, as it does. The same is also true for the root mean square (RMS) error given in Table 2. That the result for Tsukuba can only degrade for sub-pixel precise estimation is the reason why we concentrate on pixel precise disparity estimation for the rest of the paper.

				subpixel		
	all	untex.	disc.	all	untex.	disc.
Tsukuba	1.67_4	0.77_3	9.67_6	2.24_7	1.58_7	11.70_7
Sawtooth	1.21_7	0.17_6	6.90_8	0.72_5	0.03_3	6.82_8
Venus	1.04_3	1.07_3	13.68_{12}	0.78_2	0.68_2	10.66_8
Map	0.29_5	0.00	3.65_6	0.24_4	0.00	3.36_6

Table 1: Percentage of bad pixels with all means for improvement included (right three columns: sub-pixel precise results). The subscripts of the percent values indicate the rank of each value according to the online version of (Scharstein and Szeliski, 2002).

				subpixel		
	all	untex.	disc.	all	untex.	disc.
Tsukuba	0.83	0.63	1.74	0.87	0.56	1.90
Sawtooth	0.61	0.31	1.70	0.56	0.24	1.67
Venus	0.47	0.44	1.31	0.38	0.35	1.27
Map	0.99	0.42	3.44	0.94	0.26	3.36

Table 2: RMS error with all means for improvement included (right three columns: sub-pixel precise results)

Size matching	$5 \times 5 \times 1$
Size support	$11 \times 11 \times 3$
Truncation Value	4 gray values
Threshold for convergence	0.005 * $search\_width$
Threshold for mixing scores	45 gray values
Preference for larger disparities	0.05 * $search\_width$
Number iterations for occlusion	2

Table 3: Parameters for the results in the figures and tables above

4 MEANS FOR IMPROVEMENT

In the remainder of the paper we illustrate the means and their performance gain. The two basic means presented in the first subsection only speed up the processing. The rest of the means are explained using Tsukuba as the running example in separate subsections. Their gain is assessed in the final subsection by comparing the evaluation results when excluding the respective means from the processing to the result when all means are used.

4.1 Recursive 3D Box-filter and Convergence Determination

Filtering with a 3D box-filter based on simple summation is highly redundant. To get rid of this redundancy, we use a standard recursive filter. For it we separate the filter into one-dimensional (1D) staffs and 2D sheets. By adding pixels on top of each other we generate staffs (cf. Fig. 6). From them we build sheets and finally from the sheets the box. The update is done recursively. To filter with a translated box, instead of adding sheets we add a (new) sheet on one side and subtract the (old) sheet on the other side. The same is done for the sheets and the staffs. By this means the complexity becomes independent of the size of the box.

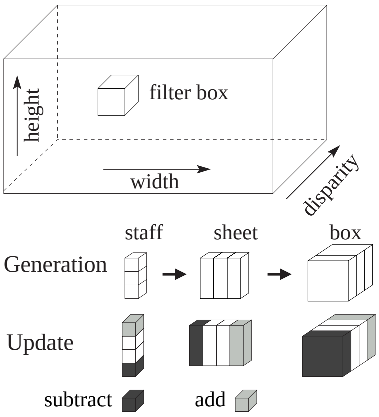


Figure 6: Separation of $3 \times 3 \times 3$ box: Generation and update

The performance gain depends on the size of the 3D-box, but is considerably large for meaningful box sizes. For Tsukuba of size 384×288 pixels and a search width of 15 pixels, one iteration of the simple algorithm on a 2.5 GHz PC takes 3.2 seconds for a

$11 \times 11 \times 3$ box. The separated algorithm needs 0.10 seconds. It is interesting to compare this with the times for the inhibition. For Tsukuba the inhibition takes 0.30 seconds per iteration. If one substitutes the square, i.e., $\alpha = 2$, for the general exponential, it reduces to 0.13 seconds. Because we found that this gives also the best results in nearly all cases, we have only used $\alpha = 2$ in our experiments.

The meaningful number of iterations varies for different images. It proved useful to decide about the number of iterations by convergence determination. For the latter also a parameter is needed, but empirical investigations have shown that it is relatively independent of the images at hand. To determine the convergence, we compute the difference image of the disparity maps from the last two iterations and compute the standard deviation σ. Empirically we found that a good threshold for σ is 0.005 of the search width. This results in 34 iterations for Tsukuba, 23 for Sawtooth, 30 for Venus, and 28 for Map.

4.2 Symmetric Support

Older experiments conducted by us showed that weighting down the right inhibition improves the performance. Recently, we were hinted that this asymmetry might stem from asymmetric support. It was immediately confirmed that this is correct. Our experiments verify that a symmetric support, where a box and a tilted box are added as shown in Figure 7, considerably improves the performance (cf. below). Also the tilted box is implemented recursively by adding / subtracting staffs from the box.

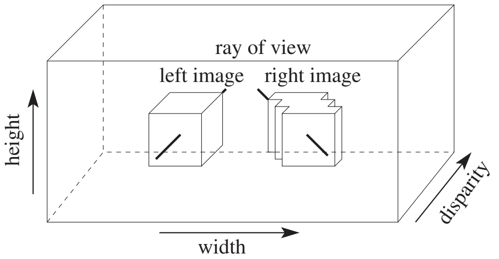


Figure 7: Symmetric support

4.3 Combination of Absolute Differences and Correlation

As suggested by (Scharstein and Szeliski, 2002), we have based the correlation scores on absolute differences. Experiments showed that the performance for squared differences was in nearly all cases worse than for absolute differences. We also found that the use of a minimum filter mimicking shiftable windows degraded the performance. This is probably due to the small window sizes that we use (cf. Table 3). For the absolute differences we truncate the difference value with a value $trunc$. The matching score for absolute differences is $score_{abs\_diff} = 1 - abs\_diff/trunc$, with $0 \leq score \leq 1$.

When looking at results based on (normalized) cross-correlation compared to results where absolute differences were employed, we got the idea that the failure modes seem to be different and that it might be useful to combine both. The combination is done by

$$score_{comb} = \frac{score_{abs\_diff} + weight * score_{corr}}{1 + weight} \tag{4}$$

$$\text{with} \quad weight = \frac{horiz\_grad}{threshold\_for\_mixing\_scores} \ . \tag{5}$$

A large horizontal gradient $horiz\_grad$ (cf. Figure 8 left) increases the probability for a good match for cross-correlation, because cross-correlation works best for strongly textured regions and the matching is done in horizontal direction.

In addition to the combination, a special type of auto-correlation $auto\_corr$ (cf. Figure 8 right) is used to indicate potentially false matches. It is determined as the maximum value of correlation along the horizontal line ranging from outside the matching window to the search width. If this auto-correlation is large, it means that there are similar structures already in the reference image and, therefore, the match is highly likely to be ambiguous also in the other image. The auto-correlation is used to weight down the matching score by $score = score * (1 - 0.5 * auto\_corr)$. Both, horizontal gradient and auto-correlation are smoothed with a Gaussian.

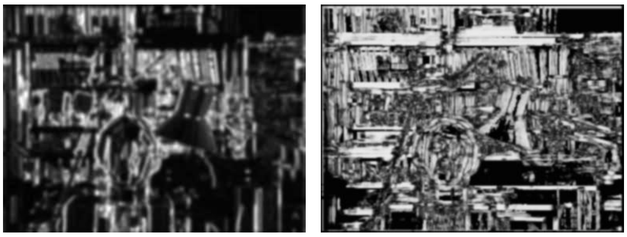


Figure 8: Horizontal gradient (left); Maximum auto-correlation along the horizontal line ranging from outside the matching window to the search width (right). Both images are slightly smoothed with a Gaussian.

4.4 Use of Color and Preference for Smaller Disparities

As some of the images of the dataset of (Scharstein and Szeliski, 2002) are colored, it is useful to employ this information. For the absolute differences we take the average of the individual results for the three colors. As we found that color does not help too much for the correlation, we correlated only the average images of the colors.

As noted in (Zhang and Kambhamettu, 2002), there is a tendency of the cooperative approach to fatten regions with larger disparities. We counteract this by reducing the matching scores by (d is disparity)

$$\begin{aligned} score_{red} = \quad & score * (1 - \frac{d * preference}{search\_width} \\ & *(1 - 0.5 * auto\_corrr)) \ . \end{aligned} \tag{6}$$

The reduction of the matching score is motivated as follows: Occluded regions have a smaller disparity than their occluding regions. As there is no correct matching possible for occlusions, introducing a slight bias towards smaller disparities increases the probability, that occluded regions obtain correct, smaller disparities. For $preference$ a value of 0.05 was found suitable. By reducing the matching score there is a tendency for regions with a large auto-correlation (cf. above) to obtain a wrong, too small disparity value. Therefore, we reduce the preference with the same factor as above.

4.5 Enforcing the Alignment of Image and Disparity Gradients

In many cases the materials or surface characteristics are considerably different at both sides of a disparity discontinuity. This results in a typical alignment of large disparity gradients with strong image gradients. While in (Zhang and Kambhamettu, 2002) the image is segmented into several regions and the support is restricted to these regions, we take a more conservative policy. Additionally to the support size given in Table 3 (box_{supp_size}) we smooth the image with a $3 \times 3 \times 3$ box filter (box_{333}) and mix the results according to the combined strength of $gradient_{comb} = gradient_{image} * gradient_{disparity}/255$. This combined gradient is then smoothed by a Gaussian. Both gradients are determined by 3×3 Sobel-filters and scale from 0 to 255. An adequate weight and threshold was empirically found to be the combination of the combined gradient with half the search width:

$$weight = \frac{gradient_{comb}}{0.5 * search\_width} \ . \qquad (7)$$

We truncate values below 1. The combined blurring reads

$$r(box_{comb}) = \frac{r(box_{supp\_size}) + weight * r(box_{333})}{1 + weight} \ , \qquad (8)$$

where $r(box_{xxx})$ stands for the result of filtering with the respective box-filter. The smaller box_{333} is only employed for values above the half search width. These are the white regions in Figure 9. As can be seen, the regions with reduced smoothing fit better to the actual disparity continuities after convergence (right).

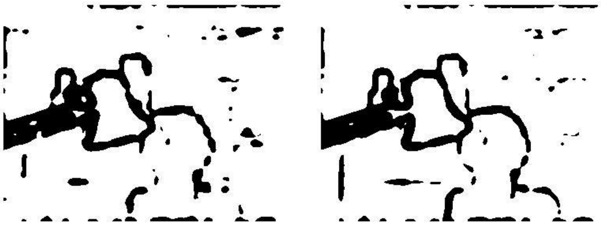


Figure 9: Regions where the smoothing of the image function is reduced due to large image and disparity gradients after 5 (left) and after 30 (right) iterations

4.6 Determination of Occlusions and Sharpening of Disparity Discontinuities

The tendency to estimate too large disparities (cf. also Section 4.4) is especially true for occluded regions. (Egnal and Wildes, 2002) describes different approaches to determine occlusions. An idea is to use one of these approaches and reduce the probability of larger disparities for occluded regions, for which no matching is possible, and which have a smaller disparity than their occluding regions. The determination of occlusions works best when the result is already cleaned from gross errors. Empirically, the optimum procedure was found to let the basic algorithm converge first, and then to multiply $L_0(r, c, d)$ by $(search\_width - d)/search\_width$. This reduces the energy or probability of the original matching scores for larger disparities. They influence the process via equation (2). The reduction of the original matching scores is done several times. For the experiments in this papers it is done two times. After the reduction the algorithm runs until σ falls again below the given threshold.

Because the disparities are already smooth when the algorithm has converged for the first time, it is sufficient to compute an indication for an occluded region by what in (Egnal and Wildes, 2002) is termed occlusion constraint. Here it is determined by the predicate $((d - d_{occ}) + (c - c_{occ})) < 0$. d and c are the disparity and the column coordinate of the point under investigation. d_{occ} and c_{occ} are the disparity and the column coordinate of the preceding point when starting from the left side of the image if no occluding point was found yet. If an occluding point was found, it is only updated to be the preceding point when the above predicate is not true any more. To obtain compact regions, morphological opening and closing with circular structuring elements with a radius of 2.5 pixels are used. In Figure 10 the occlusions determined for the final convergence of the algorithm are shown.

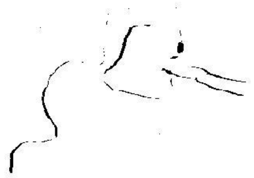


Figure 10: Occluded regions

4.7 Assessment of the Gain of the Means

Table 4 gives in the first row as reference the results when all means are employed. In the other rows results which are considerably worse than the reference are shown in bold, while results which are considerably better are marked in italics.

For the symmetric support in the second row, the result is clear-cut. Apart from the textureless regions in Venus, there is an improvement nearly everywhere.

The third row shows the results for absolute differences with the optimum truncation value of 4 gray values. As can be seen, the performance gain is considerable for the combination for all images besides Tsukuba. Our interpretation of this is as follows: Absolute differences make use of brightness differences even for weakly textured regions. This is useful only for constant lighting conditions, similar viewing angles, and well-behaved reflection functions. Yet, it is an advantage compared to (normalized) correlation which is invariant to differences in brightness and contrast. It can therefore produce a high score when matching a smooth bright to a smooth or even textured dark region when the weak texture happens to be similar, even though this is practically implausible. On the other hand, by restricting ourselves to relatively small truncation values, we do not make full use of heavily textured regions by the absolute differences, where correlation works best.

From the fourth row it can be seen, that auto-correlation helps, though mostly for Tsukuba and Venus. Both have strong repetitive textures in the form of the books for Tsukuba and the rows of letters for Venus.

The fifth row shows, that color helps. Yet, for Venus there is ample room for improvement. This might stem from the fact that Venus is partly relatively greenish and we only sum up the color information without weighting it according it to contrast.

From the sixth row one can see that no preference for small disparities results in a noticeable degradation of the overall results especially for Tsukuba, while for Map there is only a small improvement and for the other images there is none.

	Tsukuba			Sawtooth			Venus			Map		
Means	all	untex.	disc.	all	untex.	disc.	all	untex.	disc.	all	untex.	disc.
everything included	1.67	0.77	9.67	1.21	0.17	6.90	1.04	1.07	13.68	0.29	0.00	3.65
no symmetric support (4.2)	**2.06**	**1.19**	**11.90**	**1.49**	**0.43**	**10.29**	0.97	*0.67*	**14.67**	**0.39**	0.00	**4.63**
absolute differences only (4.3)	1.73	0.69	9.84	**1.42**	**0.20**	6.97	**1.32**	1.14	*10.38*	**2.53**	**1.19**	**16.82**
without use of auto-correlation (4.3)	**1.99**	**1.01**	**11.42**	1.22	0.16	6.82	1.17	**1.37**	**15.34**	0.29	0.00	3.68
no color used (4.4)	**2.15**	**1.14**	**12.39**	**1.40**	**0.44**	7.01	*0.82*	*0.71*	*10.16*	0.29	0.00	3.65
no preference for small disparities (4.4)	**2.05**	**1.18**	**11.78**	1.23	0.19	7.06	1.18	1.34	14.24	0.27	0.00	*3.24*
no alignment of gradients (4.5)	**2.82**	**1.53**	**16.07**	1.20	0.18	6.92	1.08	1.09	13.80	**0.51**	**0.24**	**6.63**
no occlusion modeling (4.6)	1.67	0.77	9.67	1.22	0.17	6.87	1.05	1.10	13.76	**0.39**	**0.71**	3.68

Table 4: Comparison of Different Means for Improvement: Percentage of bad pixels of results without using the respective means (worse results are marked in bold and better results in italics)

The improvement by means of the enforcement of the alignment of image and disparity gradient in the sixth row is extremely large for Tsukuba and considerable for Map.

Modeling occlusion in the last row only helps for Map. Our experience shows that for other parameter settings the modeling can result in a small, yet noticeable improvement for all images. Though, the importance of this means is still not absolutely clear.

5 ADDITIONAL RESULTS

To show that our means for improvement and the parameters used are not only valid for the data set at `www.middlebury.edu/stereo`, we experimented with other image pairs with the same parameters as given in Table 3. The only modification employed was to make the absolute differences invariant against a different average brightness of the image windows. This had to be done, because, opposed to the data set at `www.middlebury.edu/stereo`, many other image pairs have a significantly different gray value for homologous windows.

For the image pair Sport (cf. Figures 11 and 12) from INRIA's Syntim image database one can see that the approach works reasonably well for a relatively large disparity range (45 pixels search width for the epipolar resampled image Sport reduced to 267 × 271 pixels). The image pair Kitchen (cf. Figures 13 and 14) stems from the web page http://research.microsoft.com/virtuamsr/virtuatour.html at Microsoft maintained by Antonio Criminisi and Phil Torr. The results show the high quality achievable with the improved approach. Similar results were obtained also for a larger number of other images.

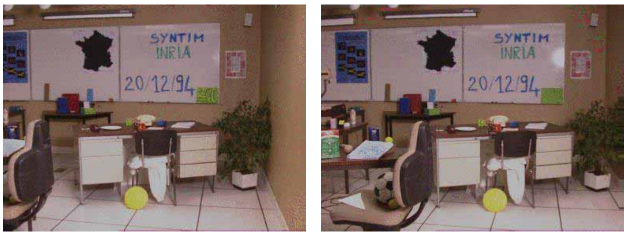


Figure 11: Input images for Sport from INRIA's Syntim image database

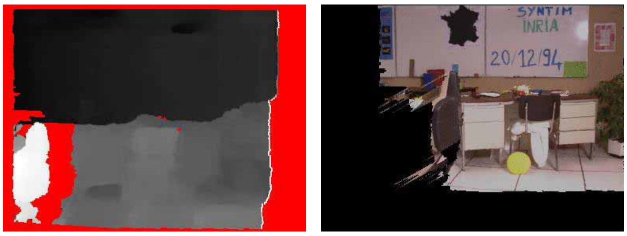


Figure 12: Result (occluded regions in red / light gray) with the same parameters as in Table 3 (left) and visualization (occluded regions in black) with trifocal tensor according to (Avidan and Shashua, 1998) (right)

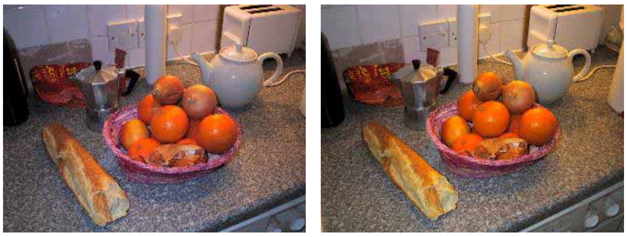


Figure 13: Input images for Kitchen from web page Torr and Criminisi

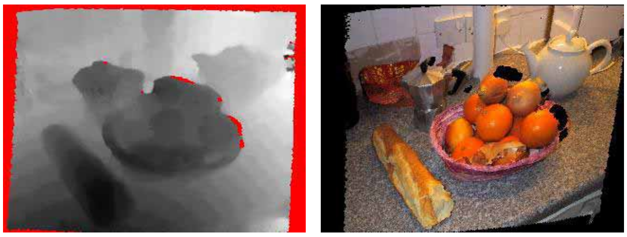


Figure 14: Result (left) and visualization (right); occlusions and parameters cf. Fig. 12.

6 CONCLUSIONS

Ranking our results in the frame of the online version of (Scharstein and Szeliski, 2002) at `www.middlebury.edu/stereo` shows that we have obtained a relatively good performance also compared to the run time of our algorithm. On one hand, we have to admit that we have fine-tuned our approach for an optimum performance with the given data set. On the other hand, the last section shows that we also obtain reasonable results for other image pairs using the same parameters.

The results reported in (Sun et al., 2002) are partly better than that presented in this paper. Though, it takes 288 seconds on a 500 MHz PC for Tsukuba, i.e., more than double as long when scaled to 2.5 GHz. Graph cuts (Boykov et al., 2001, Kolmogorov and Zabih, 2001) with and without the handling of occlusions also have a similar or better performance than our algorithm especially in combination with the fast max-flow algorithm. Yet, an interesting question would be if it might be possible to reach an improvement by some of our means for these algorithms. Especially the combination of correlation and absolute differences as well as using the auto-correlation function to characterize probably unreliable regions might be fruitful in terms of performance as well as speed. A similar reasoning applies for fast though not absolutely high quality algorithms such as (Hirschmüller et al., 2002). (Zhang and Kambhamettu, 2002) has an advantage for depth discontinuities due to a more advanced modeling of the image function, but it could also benefit from our more wide range of means of improvements.

Ways to proceed are for instance to make use of highly reliable points as in (Lhuillier and Quan, 2002) to initialize the cooperation process, the use of more images as in (Koch et al., 1999, Kolmogorov and Zabih, 2002) or (Leloğlu et al., 1998), where are merging of pairs is done in object space based on relaxation, or to locally optimize the window size and shape (Veksler, 2002). We have started to project the results into a third image by means of the trifocal tensor to obtain more evidence especially for occluded regions.

REFERENCES

Avidan, S. and Shashua, A., 1998. Novel View Synthesis by Cascading Trilinear Tensors. IEEE Transactions on Visualization and Computer Graphics 4(4), pp. 293–306.

Boykov, Y., Veksler, O. and Zabih, R., 2001. Fast Approximate Energy Minimization via Graph Cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(11), pp. 1222–1239.

Egnal, G. and Wildes, R., 2002. Detecting Binocular Half-Occlusions: Empirical Comparison of Five Approaches. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(8), pp. 1127–1133.

Hirschmüller, H., Innocent, P. and Garibaldi, J., 2002. Real-Time Correlation-Based Stereo Vision with Reduced Border Errors. International Journal of Computer Vision 47(1), pp. 229–246.

Koch, R., Pollefeys, M. and Van Gool, L., 1999. Robust Calibration and 3D Geometric Modeling from Large Collections of Uncalibrated Images. In: Mustererkennung 1999, Springer-Verlag, Berlin, Germany, pp. 413–420.

Kolmogorov, V. and Zabih, R., 2001. Computing Visual Correspondence with Occlusions Using Graph Cuts. In: Eighth International Conference on Computer Vision, pp. 508–515.

Kolmogorov, V. and Zabih, R., 2002. Multi-Camera Scene Reconstruction via Graph Cuts. In: Seventh European Conference on Computer Vision, Vol. III, pp. 82–96.

Leloğlu, U., Roux, M. and Maître, 1998. Dense Urban DEM with Three or More High-Resolution Aerial Images. In: International Archives of Photogrammetry and Remote Sensing, Vol. (32) 4/1, pp. 347–352.

Lhuillier, M. and Quan, L., 2002. Match Propagation for Image Based Modeling and Rendering. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(8), pp. 1140–1146.

Marr, D. and Poggio, T., 1976. Cooperative Computation of Stereo Disparity. Science 194, pp. 209–236.

Marr, D. and Poggio, T., 1979. A Computational Theory of Human Stereo Vision. In: Proceedings Royal Society London B, Vol. 204, pp. 301–328.

Scharstein, D. and Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision 47(1), pp. 7–42.

Sun, J., Shum, H.-Y. and Zheng, N.-N., 2002. Stereo Matching Using Belief Propagation. In: Seventh European Conference on Computer Vision, Vol. II, pp. 510–524.

Veksler, O., 2002. Stereo Correspondence with Compact Windows via Minimum Ratio Cycles. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(12), pp. 1654–1660.

Zhang, Y. and Kambhamettu, C., 2002. Stereo Matching with Segmentation-Based Cooperation. In: Seventh European Conference on Computer Vision, Vol. II, pp. 556–571.

Zitnick, C. and Kanade, T., 2000. A Cooperative Algorithm for Stereo Matching and Occlusion Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(7), pp. 675–684.