EFFICIENT SEMI-GLOBAL MATCHING FOR TRINOCULAR STEREO

Matthias Heinrichs*, Volker Rodehorst and Olaf Hellwich

Computer Vision & Remote Sensing, Berlin University of Technology, Franklinstr. 28/29, FR 3-1, D-10587 Berlin, Germany – (matzeh, vr, hellwich)@cs.tu-berlin.de

KEY WORDS: Photogrammetry, Area-based Image Matching, Semi-Global Matching, Trinocular Stereo, Similarity Measures

ABSTRACT:

This paper describes an efficient method for dense matching of two or three images. After some investigations in different similarity measures we propose a modification of Semi-Global Matching, which uses a simple energy function that implies piecewise smoothness but no ordering and gives promising results in practice. Our improvements include a symmetric and hierarchical matching strategy and allow an efficient generalization of the stereo matching problem to trinocular surface reconstruction. Finally, we present results for synthetic and for real images.

1. INTRODUCTION

The automatic three-dimensional (3D) reconstruction of an observed scene using digital images has been one of the core challenges in photogrammetry and computer vision for decades. Two or more images may be taken by different cameras at the same time (stereo) or by the same camera at different times (motion).

The key problem is how to find homologous image points which arise from the same physical point in the scene. If this *correspondence problem* is solved, depth information can be derived by triangulation using the orientation parameters of calibrated cameras.

The existing techniques for correspondence analysis can be distinguished by either matching some *features* (or relations between them) producing sparse depth information or matching all *pixels* in the images producing dense depth maps. For surface reconstruction tasks it is essential to compute dense depth maps using every pixel of the entire image.

Stereo image matching remains a difficult problem because of

- *Noise*, which arises from illumination variations and sensor noise during image formation,
- *Untextured regions* or *repetitive patterns*, which introduce ambiguities,
- *Occluded pixels* in one image, which should not be matched with pixels in the other image, as well as
- *Depth discontinuities* at object boundaries, which violate the spatial smoothness and ordering constraints.

1.1 Related Work

For a good overview and quantitative comparison of current two-frame stereo correspondence algorithms, we refer to (Scharstein & Szeliski, 2002), and for multi-view stereo reconstruction algorithms to (Seitz et al., 2006). The simplest algorithm compute stereo correspondence of image windows by searching along the corresponding epipolar line and match with the highest similarity (winner takes all).

The central problem of such *local* matching methods is to determine an optimal support region for each pixel. An ideal window is adaptive and should be bigger in homogeneous regions and smaller at depth discontinuities (Veksler, 2003). To reduce ambiguities, most algorithms make additional assumptions about the scene geometry (see Section 2.1).

Reasonable and commonly made assumptions are that the scene is piecewise smooth and order is preserved. These constraints on the depth map can be formulated into an objective function and directly optimized with various *global* methods.

- *Dynamic Programming* is one of the oldest optimization methods for stereo correspondence. The performance may reach state-of-the-art, if the vertical consistency between the scan lines is enforced (Lei et al., 2006).
- *Graph Cut* is based on the maximum-flow algorithm in graph theory (Kolmogorov & Zabih, 2001). The idea is to construct a specialized graph such that the minimum cut on the graph also minimizes the energy function. This optimization is applied to the whole image and not just one scan line.
- *Belief Propagation* formulates the stereo matching problem as a Markov network and solves it using Bayesian belief propagation to obtain the maximum a posteriori estimation (Klaus et al., 2006).

Unfortunately global methods are often time-consuming and memory exhausting. In addition most of them have problems preserving sharp discontinuities.

In this paper we propose an efficient method that is motivated by some recent work on Semi-Global Matching (Hirschmüller, 2005/2006). This approach uses a simple energy function (see Section 3) that implies piecewise smoothness but no ordering and gives promising experimental results in practice.

Our main contributions are investigations in different similarity measures (Section 2.2), the improvement by symmetric matching (Section 3.6), a hierarchical strategy (Section 4) and the efficient generalization of the stereo matching problem to trinocular reconstruction (Section 5). Finally, we present results for synthetic and for real images.

2. AREA-BASED MATCHING

Area-based matching is a widely used method for dense stereo correspondence. The similarity is computed statistically on the rectangular neighborhood (matching window) around the examined pixel. The algorithm searches at each pixel in reference image I_1 for maximum correlation in the horizontal image I_2 (and/or vertical image I_3) by shifting a small window pixel-by-pixel along the corresponding epipolar line.

2.1 Geometric Constraints

The time-consuming computation can be substantially simplified and accelerated by utilizing geometric constraints.

2.1.1 Disparity Limit

Assuming that the relative image orientation is known and that the smallest and highest displacement are roughly given, the search range $[d_{\min}, d_{\max}]$ of an image point in the reference view can be reduced to line segments in the corresponding images.

2.1.2 Normal Images

We assume that the epipolar lines of a horizontal image pair are parallel to the *x*-axis (*stereo normal case*) and those of a vertical image pair parallel to the *y*-axis. For the trinocular case using three images we recommend an L-shaped configuration. If the baselines of both stereo pairs have the same length, the horizontal and vertical displacements are identical.
Except for critical configurations, most images can be geometrically transformed so that the epipolar lines coincide with the same image rows and/or columns. This process is called *trinocular rectification* (Heinrichs & Rodehorst, 2006). After rectification it holds

$$\begin{aligned} I_1(x,y) &\approx I_2\big(x + D(x,y), y\big) \\ I_1(x,y) &\approx I_3\big(x, y + D(x,y)\big) \end{aligned} \tag{1}$$

where I_1, I_2, I_3 are rectified normal images, x is the column coordinate, y the image row coordinate and D is called disparity map. Please note that for differing image setups the horizontal and vertical disparities need to be scaled appropriately. The disparity at the current position (x,y) is inversely proportional to the depth of the scene.

2.2 Local Similarity Measures

In the following, popular similarity measures for calculating local image matching costs will be briefly described.

2.2.1 Difference Correlation (SSD, SAD)

A very simple but effective matching metric is the difference correlation. For each color pixel in $I = (R,G,B)$ the red, green and blue channels should be normalized by its intensity

$$I' = \frac{I}{R+G+B} \tag{2}$$

to compensate brightness and contrast variations. The color difference of two $n \times n$ image windows $a(x,y)$ and $b(x,y)$ with $N = 3n^2$ pixels can be defined by the *sum of squared differences* (SSD)

$$\rho_{SSD}(a,b) = \frac{1}{N}\sum_{i,j=1}^{n}\sum_{k=R,G,B}\big(a'_k(i,j) - b'_k(i,j)\big)^2 \tag{3}$$

or the slightly faster *sum of absolute differences* (SAD)

$$\rho_{SAD}(a,b) = \frac{1}{N}\sum_{i,j=1}^{n}\sum_{k=R,G,B}\big|a'_k(i,j) - b'_k(i,j)\big|. \tag{4}$$

Each measure should be normalized to a range between unity and zero, whereas unity means maximum similarity. The normalized correlation coefficient can be derived using

$$\rho'(a,b) = \frac{1}{1+\rho(a,b)}. \tag{5}$$

2.2.2 Cross-Correlation (NCC, MNCC)

The statistically based *normalized cross-correlation* (NCC) measures the linear relation between two image windows normalizing over all intensity changes.

$$\rho_{NCC}(a,b) = \frac{\sigma_{ab}}{\sqrt{\sigma_a^2 \cdot \sigma_b^2}}$$

$$\text{Covariance: } \sigma_{ab} = \frac{1}{N}\left(\sum_{i,j=1}^{n}\sum_{k=R,G,B} a_k(i,j)\cdot b_k(i,j)\right) - \bar{a}\cdot\bar{b}$$

$$\text{Variance: } \sigma_a^2 = \frac{1}{N}\left(\sum_{i,j=1}^{n}\sum_{k=R,G,B} a_k(i,j)^2\right) - \bar{a}^2 \tag{6}$$

$$\text{Mean: } \bar{a} = \frac{1}{N}\sum_{i,j=1}^{n}\sum_{k=R,G,B} a_k(i,j)$$

We precalculate the means $\bar{a}, \bar{b}$ and the means of squared intensities in order to significantly accelerate the computation. The definition of the *modified* NCC (MNCC)

$$\rho_{MNCC}(a,b) = \frac{2\cdot\sigma_{ab}}{\sigma_a^2 + \sigma_b^2} \tag{7}$$

handle homogeneous areas better by adding the two denominator variances instead of multiplying them (Egnal, 2000). Finally, we use

$$\rho'(a,b) = 0.5\cdot\big(\rho(a,b)+1\big). \tag{8}$$

to transform the correlation coefficient range to [0,1].

2.2.3 Mutual Information (MI)

Mutual information is a popular matching metric for images from airborne cameras (Hirschmüller, 2005) as well as inhomogeneous sensors like magnetic resonance imaging (MRI) and computerized axial tomography (CAT) scanners (Viola & Wells, 1997). A good description of MI can be found in (Kim, 2003; Pluim, 2000; Maes, 1997) and a comparison with MNCC in (Egnal, 2000). The major advantage of MI is its robustness against radiometric differences, i.e. non-Lambertian reflection properties and different gamma nonlinearities (Hirschmüller & Scharstein, 2007).
In (Hirschmüller, 2005) an improvement of pixel-wise matching costs based on MI is introduced. It requires a rough correspondence map, which can be computed hierarchically from the previous resolution. It suggests a random map for the lowest resolution to compute the MI for the next resolution. The intensity probabilities P_1 and P_2 are computed over all intensities i, j of all corresponding pixels $(\mathbf{x_1}, \mathbf{x_2})$ in images I_1, I_2.

$$\begin{aligned} P_1(i) &= \frac{1}{M}\sum^{\mathbf{x}_1} T\big[I_{1\mathbf{x}_1} = i\big] \\ P_2(j) &= \frac{1}{M}\sum^{\mathbf{x}_2} T\big[I_{1\mathbf{x}_2} = j\big] \end{aligned} \tag{9}$$

M is the number of correspondences and T is a Boolean function which returns 1 if the argument is true and 0 otherwise. The joint probability P_{ab} of the corresponding intensities is computed by

$$P_{12}(i,j) = \frac{1}{M}\sum^{\mathbf{x}_1\mathbf{x}_2} T\Big[\big(I_{1\mathbf{x}_1} = i\big) \wedge \big(I_{2\mathbf{x}_2} = j\big)\Big]. \tag{10}$$

Now the pixel-wise MI can be defined by

$$mi_{I_1,I_2}(i,j) = h_{I_1}(i) + h_{I_2}(j) - h_{I_1,I_2}(i,j) \tag{11}$$

where the entropy values h are defined by:

$$h_{I_1}(i) = -\frac{1}{M}\log\left(P_1(i) \otimes g(i)\right) \otimes g(i)$$
$$h_{I_2}(j) = -\frac{1}{M}\log\left(P_2(j) \otimes g(j)\right) \otimes g(j) \quad . \qquad (12)$$
$$h_{I_1,I_2}(i,j) = -\frac{1}{M}\log\left(P_{12}(i,j) \otimes g(i,j)\right) \otimes g(i,j)$$

The convolution $\otimes$ with a Gaussian g is applied for Parzen estimation (Kim 2003). This leads to a lookup table of 256×256 values for 8bit intensities. The MI matching costs can be computed by looking for the min-max values in this table:

$$\rho_{MI}(i,j) = \frac{mi_{I_1,I_2}(i,j) - mi_{\min}}{mi_{\max} - mi_{\min}}. \qquad (13)$$

For color images the three channels are computed separately and the MI costs for every pixel are calculated by the mean values of the three colors. This results in more stable matching than using only the joint intensity of all color channels.

3. SEMI-GLOBAL OPTIMIZATION

Generally, the calculation of local matching costs is ambiguous and a piecewise smoothness constraint must be added. In (Hirschmüller, 2005/2006) a very simple and effective method of finding minimal costs is proposed.

3.1 Energy Function

Semi-global matching (SGM) tries to determine a disparity map D such that the energy function

$$E(D) = \sum_{x,y \in I}\left(\left(1-\rho\left(a(x,y), b(x+D(x,y),y)\right)\right)\right.$$
$$+ Q_1 \sum_{i,j=-1}^{1} \mathrm{T}\left[\left|D(x,y) - D(x+i,y+j)\right| = 1\right] \qquad (14)$$
$$\left. + Q_2 \sum_{i,j=-1}^{1} \mathrm{T}\left[\left|D(x,y) - D(x+i,y+j)\right| > 1\right]\right) \quad \text{for} \quad i \neq j$$

is minimal. The first term calculates the sum of all *local matching costs* using the inverse correlation coefficient ρ (see Section 2.2) of the image windows a and b around the current position (x, y) and the related disparity in D. Explained intuitively, $E(D)$ accumulates the local matching costs with a small penalty $Q_1 = 0.05$ if the disparity varies by one from the neighboring disparities. If the disparity differs by more than 1, a high penalty $Q_2 \in [0.06, 0.8]$ is added. The actual value of Q_2 depends on the intensity gradient in the original image. Long gradients result in a low Q_2 while short gradients result in a high Q_2. This prevents depth changes in homogeneous regions. There are only two different penalties for the depth changes. First, Q_1 ensures that regions with a slightly changing depth are not penalized too hard. Second, if depth changes in the scene occur, the size of the discontinuity is not correlated to the penalty.

3.2 Optimization Strategy

Computing the minimum energy of $E(D)$ leads to NP-hard complexity, which is difficult to solve efficiently. Following (Hirschmüller, 2005), a linear approximation over possible disparity values $d \in [d_{\min}, d_{\max}]$ is suggested by summing the costs of several 1D-paths L towards the actual image location (x_1, y_1). A path L with $i = n \ldots 1$ steps is recursively defined as

$$\begin{aligned} L^n(x_n, y_n, d) &= 1 - \rho\left(a(x_n, y_n), b(x_n + d, y_n)\right) \\ L^i(x_i, y_i, d) &= 1 - \rho\left(a(x_i, y_i), b(x_i + d, y_i)\right) + \\ &\min\left(L^{i+1}(x_{i+1}, y_{i+1}, d-1) + Q_1,\right. \\ &\quad L^{i+1}(x_{i+1}, y_{i+1}, d), \\ &\quad L^{i+1}(x_{i+1}, y_{i+1}, d+1) + Q_1, \\ &\quad \left. L^{i+1}_{\min} + Q_2\right) \end{aligned} \qquad (15)$$

where the minimal costs over all disparities

$$L^i_{\min} = \min\left(L^i(x_i, y_i, d)\right) \qquad (16)$$

of the previous step are constant and must be computed only once. Changes of image positions from one recursive step to the next depend on the path direction $\mathbf{r} = (r_x, r_y)$ and the actual position:

$$\begin{aligned} x_i &= x_1 + (i-1) \cdot r_x \\ y_i &= y_1 + (i-1) \cdot r_y \end{aligned} \qquad (17)$$

Now, the final disparity map can be estimated using

$$D(x,y) = \min_d\left(\sum_{\mathbf{r}} L_{\mathbf{r}}(x,y,d)\right), \qquad (18)$$

where the number of accumulated paths should be ≥ 8.

3.3 Thresholding

We introduce a threshold for the local matching costs in order to penalize dissimilar candidates. If the correlation coefficient ρ is lower than a certain threshold, the local matching costs are set to a high constant value. If the minimal costs for the best matching candidate are higher than this value, the match is marked as invalid. Since MI provides statistical values which depend on the intensity distribution of the image a global threshold is not applicable to MI cost values.

3.4 Sub-pixel Matching

If the cost function has a local extreme at position $D(x, y)$, a sub-pixel disparity can be estimated by fitting a parabola curve to the correlation coefficients of the best match s_0 and its two neighbors s_{-1} and s_{+1}:

$$s_i = \rho_{SSD}\left(a(x,y), b(x + D(x,y) + i, y)\right) \qquad (19)$$

The interpolation is only valid for cubic functions like SSD (Hirschmüller, 2005). Therefore, s_i must be computed again if

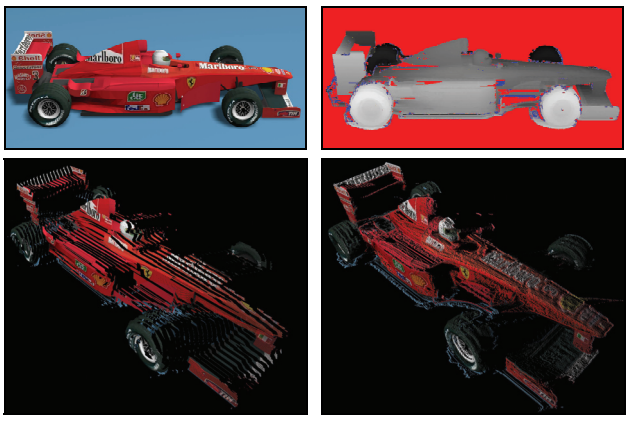


Figure 1. Sub-pixel improvements on the 3D reconstruction of a synthetically rendered Formula One car

different correlation functions were used before. Since the distance between the three coefficients is unity, they can be shifted symmetrically around the origin. The equation for sub-pixel disparity in x-direction is defined as

$$D_{sub}(x,y) = D(x,y) + \frac{s_{-1} - s_{+1}}{2 \cdot (s_{+1} + s_{-1} - 2 \cdot s_0)}. \quad (20)$$

3.5 Optimizing Memory Requirements

One disadvantage of SGM is the required space for all correlation values, which is needed to compute all non-horizontal trails $L_{\mathbf{r}}$. The memory for this buffer is $O(n^3)$ depending on the image width, height and disparity search range. We save memory by reducing the length of the trails.
Since the influence of previous L after a disparity discontinuity is very low, we need the complete path only for homogeneous areas. Except for trails along the epipolar lines, we limit the length of L to a small value (e.g. five). Therefore, the buffer size reduces to $O(n^2)$, which allows to process larger images.

3.6 Symmetric Matching

An important issue for image matching is the stability of the found correspondence. A correspondence can be unstable either due to an occlusion or because the image significance is very low, e.g. in homogeneous regions or periodic patterns. To enforce stability, we check the *left/right consistency* (LRC) of the bidirectional correspondence search.
A robust matching process should produce a unique result. On one hand, LRC detects most stereo errors and does not depend critically on thresholds. On the other hand, LRC does not report an error if the two matching directions mistakenly agree and it requires one extra matching process. Nevertheless, the computational expense is tolerable for many applications.
LRC leads to two disparity maps D_i, one for each image permutation. If the matched point in the second image points back to the original one in the first image

$$D_1(x,y) + D_2\left(x + D_1(x,y), y\right) \le 1, \quad (21)$$

the match is validated. Otherwise it is invalidated or in case of multi-image stereo matching, other permutations of the disparity map must verify this match. In addition, using the reverse direction guarantees that all matched points are one-to-one correspondences, because doubly matched points can verify only one location.

4. HIERARCHICAL APPROACH

This section describes the hierarchical approach of the matching process using image pyramids.

4.1 Building an Image Pyramid

Based on the original image resolution a number of reduced images are computed using a scale factor f_i. The search range can be scaled by f_i too, so that the computational complexity drops dramatically from the actual scale level to the next smaller one.

4.2 Hierarchical Processing

For the hierarchical approach the original images are resampled to different resolution layers. The layers are processed from the lowest resolution to the highest one. Only image points of the first layer have to be checked at every possible location within the search range.

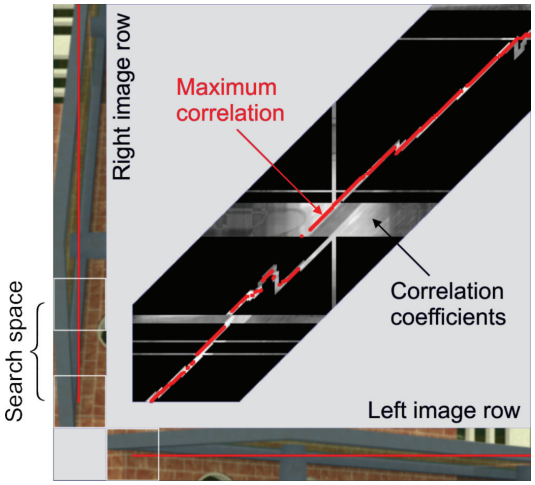


Figure 2. Sample of the reduced search space

To reduce the number of candidates in the succeeding layers, the potential information of the previous layer is used and refined. If displacement information from a previous layer is available, the number of candidates can be reduced by restricting the possible range. The valid candidates fulfill at least one of the following three criteria:

1. **Accuracy improvement:** The information from the previous layer has an accuracy of $\pm s \cdot f_i$, where s is the distance from one candidate to the next one and f_i is the scale factor from the previous layer to the actual one. Possible matches within this accuracy range must be checked.
2. **Unmatched points:** Points in the target image which are already matched in a previous layer should be excluded from further matching in order to avoid double matches.
3. **Edge preservation:** If points of the previous layer lie on a surface edge, the depth value of the associated points in the actual layer is bounded by the depths of the two neighboring points. It might happen that not all information is available.

Figure 2 illustrates this technique. The diagonal strip represents the search space of the original layer. Every column is the search space for a pixel position. The red line represents the selected correspondence. The thin diagonal stripe around the red line is the accuracy improvement from criterion 1. Vertical lines are unmatched positions in the previous layer. Therefore, the search space at these positions has to be analyzed completely to find possible new matches.
Horizontal lines represent unmatched points from criterion 2. The small vertical strips are caused by the edge preservation of criterion 3. Candidates in the black area are excluded by the hierarchical approach, which shows the efficiency of the proposed method. The search space is reduced to approximately 25% of the original size. After calculating the local costs for each candidate, our modified version of SGM calculates the best match for the given position.

5. TRINOCULAR MATCHING

In (Hirschmüller, 2005/2006), a multiple image technique is suggested which matches every image pair independently and selects the median of all disparities. This does not take advantage of the fact that every match of a single image pair leads to only one possible position in all remaining images. Thus, the position of a match candidate in a single image of the set is linked to certain computable positions in all other images

of the set. It is not necessary to search the whole range in every image pair independently. Since every candidate is linked to a specific candidate in all other images, the local matching costs of every pair between the candidate and the reference point can be simply averaged. This leads to two major advantages: First, the computing costs for all images are increasing only by a linear factor, while matching every pair would lead to a quadratic increase of the computational costs. Second, the cost function is more robust because local minima in a single set do not lead to outliers in the median selection technique. In addition, symmetric patterns parallel to one camera baseline are very unlike to be symmetric to the other baseline, too. Therefore, only the third image can provide sufficient information to find the right correspondence. More images are not integrated in this approach because the camera configuration gets too special and the common image region gets smaller with every additional image. We consider the trinocular case as a good compromise between stability and generality for most cases.

6. EXPERIMENTAL RESULTS

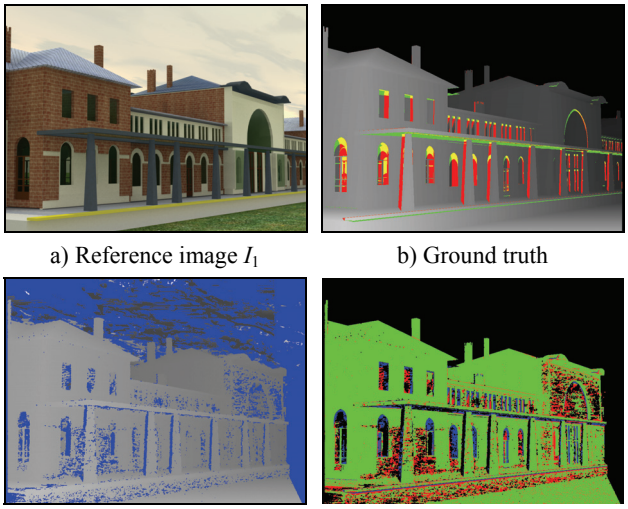
The proposed algorithm is tested on synthetic image triplets with ground truth and real image triplets from a regular 5MP digital camera.

6.1 Synthetic Images with Ground Truth

The rendered image in Figure 3 shows a train station with a resolution of 1280×1024 pixels. The image has many occlusions, homogenous regions and a lot of regular patterns like bricks. One of the three original images is presented in (a). Figure 3(b) illustrates the ground truth with horizontal occlusions (red), vertical occlusions (green) or both (yellow). The computed disparity map can be seen in (c), where blue pixels indicate inconsistencies or no match at all. The last picture (d) shows the color-coded evaluation: green pixels have an error less than one and false matches are marked red. Blue pixels represent false matches in occluded areas. Table 1 contain the quantitative evaluation results excluding the sky, which was not available in the ground truth data. The results describe the completeness, correctness and identified occlusions. The mean errors and standard deviations are given in the first two columns. The trifocal method is applied with bifocal fallback in regions which are seen only by two cameras. Most of the mismatched areas are due to interpolation of the dark and transparent windows. The unmatched areas result from homogeneous wall parts in the original image.

Method	Error [pix.]	Std. dev.	Match [%] Total	Match [%] Correct	Occ. [%]	Time [sec.]
SAD	1.17	2.72	79.9	78.8	39.9	89.5
SSD	0.95	2.40	80.6	83.2	44.3	98.8
NCC	1.06	2.73	80.7	83.2	42.5	93.6
MNCC	0.98	2.63	83.1	83.7	42.5	92.4
MI	0.78	2.13	83.5	85.8	50.0	86.5
One way	1.17	2.88	99.6	77.5	0.1	53.6
Symmetric	0.96	2.43	81.0	82.9	44.0	97.4
Bifocal	0.94	3.24	78.5	82.9	58.6	90.0
Trifocal	0.78	2.18	84.3	86.1	49.6	103.0

Table 1: Quantitative evaluation of the matching quality



a) Reference image I_1 b) Ground truth

c) Disparity map D d) Color-coded evaluation

Figure 3. Matching result using the synthetic image triplet with ground truth

6.2 Real Image Triplets

For real data, ground truth is often not available. The quality is therefore difficult to evaluate. However, some properties like perpendicularity or parallelism can be measured quite well. In Figure 4 the statue of King Friedrich I. of Prussia is reconstructed. Figure 4(a) shows the rectified reference image (2311×2200 pixel) and (b) the disparity map using a search range of 400 pixels. The illustrations 4(c) to 4(e) represent different views of the reconstructed 3D point cloud using parallel projection to enable the verification of perpendicularity and parallelism. Figure 4(f) shows a perspective projection to give an overview of the scene. The matching of this 5MP image triplet was completed in 12 minutes on a 2,4GHz dual core CPU.

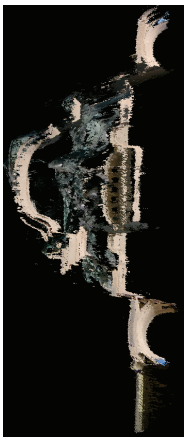
7. CONCLUSIONS

The proposed algorithm is a fast and effective adaptation of the SGM for multiple images. The hierarchical approach reduces the computational time by up to one quarter without any significant lost in accuracy. Additionally, we analyzed different cost functions. SAD and SSD perform very well on data with low noise and SSD outperforms SAD in every aspect, except speed. The use of the more robust (M)NCC is preferred for image data with higher noise and illumination changes, e.g. from a video sensor or scanned analog image, but generally has a weaker performance than MI. The optimized version of MNCC is even faster and detects more pixel than SAD/SSD with a minimal loss of accuracy. The advantage of MI is its ability to perform even in extreme circumstances where (M)NCC fails. Additionally, the pixel-wise matching of MI does not lead to blurred edges and it is the fastest technique. On the other hand this pixel-wise matching leads to some uncertainties in homogeneous regions, where window-based functions lead to more stable results. The matching of complex real images shows that stable matches are found in homogenous regions even if the path for SGM is limited to a short length. Combining the local costs of an image triplet to a single value stabilizes the matching especially in regions with repetitive patterns like bricks, grids or stripes.

a) Reference image I_1

b) Disparity map D

c) Textured point cloud



d) Top view

e) Side view

f) Perspective view

Figure 4. 3D reconstruction results on real image triplet of Friedrich I

REFERENCES

Egnal, G., 2000: "Mutual Information as a Stereo Correspondence Measure", Computer and Information Science MS-CIS-00-20, University of Pennsylvania, PA, USA, 8 p.

Heinrichs, M. and Rodehorst, V., 2006: "Trinocular Rectification for Various Camera Setups", Symp. of ISPRS Commission III - Photogrammetric Computer Vision PCV'06, Bonn, Germany, pp. 43-48.

Hirschmüller, H., 2005: "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information", IEEE Conf. on Computer Vision and Pattern Recognition CVPR'05, Vol. 2, San Diego, CA, USA, pp. 807-814.

Hirschmüller, H., 2006: "Stereo Vision in Structured Environments by Consistent Semi-Global Matching", IEEE Conf. on Computer Vision and Pattern Recognition CVPR'06, Vol. 2, New York, NY, USA, pp. 2386-2393.

Hirschmüller, H and Scharstein, D., 2007: "Evaluation of Cost Functions for Stereo Matching", IEEE Conf. on Computer Vision and Pattern Recognition CVPR'07

Kim, J., Kolmogorov, V. and Zabih, R., 2003: "Visual Correspondence Using Energy Minimization and Mutual Information", IEEE Int. Conf. Computer Vision, 2003, Vol. 2, pp. 1033- 1040

Klaus, A., Sormann, M. and Karner, K., 2006: "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure", IEEE Int. Conf. on Pattern Recognition ICPR'06, Hong Kong, China, pp. 15-18.

Kolmogorov, V. and Zabih, R., 2001: "Computing Visual Correspondence with Occlusions via Graph Cuts", Int. Conf. on Computer Vision ICCV'01, Vol. 2, Vancouver, Canada, pp. 508-515.

Lei, C., Selzer, J. and Yang, Y.H., 2006: "Region-Tree based Stereo using Dynamic Programming Optimization", IEEE Conf. on Computer Vision and Pattern Recognition CVPR'06, New York, NY, USA, pp. 2378-2385.

Maes, F., Collignon A., Vandermeulen D., Marchal G., and Suetens P., 1997: "Multimodality image registration by maximization of mutual information", IEEE Trans. Med. Imag., vol. 16, no. 2, pp. 187–198.

Pluim J.P.W., Antoine Maintz J.B., and Viergever, M. A., 2000: "Image Registration by Maximization of Combined Mutual Information and Gradient Information", IEEE Trans. Med. Imag., vol. 19, no. 8, pp. 809-814

Scharstein, D. and Szeliski, R., 2002: "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", Int. Journal of Computer Vision, 47(1-3), pp. 7-42.

Seitz, S., Curless, B., Diebel J., Scharstein, D. and Szeliski R., 2006: "A comparison and evaluation of multi-view stereo reconstruction algorithms", IEEE Conf. on Computer Vision and Pattern Recognition CVPR'06, Vol. 1, New York, NY, USA, pp. 519-526.

Veksler, O., 2003: "Fast Variable Window for Stereo Correspondence using Integral Images," IEEE Conf. on Computer Vision and Pattern Recognition CVPR '03, Vol. 1, pp. 556-561.

Viola, P. and Wells, W.M., 1997: "Alignment by maximization of mutual information", International Journal of Computer Vision, 24(2), pp. 137-154.