

Object-based Multi-Image Semi-Global Matching – Concept and first results

F. Bethmann ^a *, T. Luhmann ^b

^a Jade University of Applied Sciences Oldenburg, Ofener Straße 16/19, Oldenburg, Germany - folkmar.bethmann@jade-hs.de
^b Jade University of Applied Sciences Oldenburg, Ofener Straße 16/19, Oldenburg, Germany - thomas.luhmann@jade-hs.de

Commission V, WG V/1

KEY WORDS: image matching, semi-global matching, multi-image matching, accuracy

ABSTRACT:

Semi-Global Matching (SGM) is a widespread algorithm for image matching which is used for very different applications, reaching from real-time applications (e.g. for generating 3D-data for driver assistance systems) to aerial image matching. Originally developed for stereo-image matching, several extensions have been proposed to use more than two images within the matching process (multi-baseline matching, multi-view stereo). Most of these extensions still perform the image matching in (rectified) stereo images and combine the pairwise results afterwards to create the final solution. This paper proposes an alternative approach which is suitable for the introduction of an arbitrary number of images into the matching process and utilizes image matching by using non-rectified images within a closed solution. The proposed approach differs from the original SGM method in two major aspects: Firstly, the cost calculation is formulated in object space within a dense voxel raster by using the grey- (or colour-) values of all images instead of pairwise cost calculation in image space. Secondly, the semi-global (path-wise) minimization process is transferred into object space as well, so that the result of semi-global optimization leads to index-maps (instead of disparity maps) which directly indicate the 3D positions of the best matches. The paper provides a detailed description of the approach and it discusses its advantages and disadvantages. Further on, first results and accuracy analysis are presented.

1. INTRODUCTION

Since its introduction by Hirschmüller (2005) Semi-Global Matching (SGM) has become a widespread matching algorithm which is used for very different applications, reaching from close-range applications in the fields of robotics and computer vision to remote sensing (e.g. for surface model generation from aerial images). SGM offers several advantages in comparison to other image matching approaches: It is a dense image matching technique which can be implemented by using pixel-wise cost functions and therefore yields to good results especially in areas of sharp object boundaries (discontinuities on the object surface). Further on, it is not sensitive to the choice of task-dependent parameters and the structure of the algorithm allows for the use of highly parallelizing hardware (GPU and FPGA) which is important for the implementation of real-time applications (Banz et al., 2010)(Buder, 2012)(Ernst & Hirschmüller, 2008)(Michael et al., 2013).

For a number of applications it is sufficient to use stereo-cameras for image matching. This is especially true for many applications in computer vision (e.g. stereo-cameras in assistance systems) in which the need for real-time results is more important than high accuracies. On the other hand, various tasks focus on the accurate and complete 3D reconstruction of complex scenes (e.g. for aerial image matching, in fields of cultural heritage, archaeology, industrial measurements and so on). For these purposes, dense surface matching has been extended to so-called multi-baseline matching as proposed e.g. in (Hirschmüller, 2005, 2008) or multi-view stereo algorithms as proposed e.g. in (Rothermel et al., 2013)(Wenzel et al., 2013). Multi-baseline matching performs stereo matching by SGM between a base image and all match images. Further on, it removes invalid disparities by consistency check (left-right

check) and combines all stereo matching results by selecting the median value of all disparities for each pixel. Afterwards, it is suggested to calculate a weighted mean of all correct disparities (which are e.g. all disparities within a 1 pixel interval around the median) to increase the accuracy.

The multi-view stereo algorithm in (Rothermel et al., 2013) performs stereo matching for all overlapping image pairs or at least for a selection of overlapping image pairs. After removing outliers by left-right consistency check additional outlier elimination is performed by checking for geometric consistency in object space under consideration of uncertainty ranges that have been derived by error propagation. Finally, all corresponding image coordinates for each object point are used for triangulation to calculate the final 3D coordinates.

However, these approaches are working well but show several disadvantages. For example, SGM in image pairs is typically performed in rectified images, aiming at the simplification of the semi-global optimization to a 2.5D problem with the 2D image coordinates x', y' and one disparity D (or respectively one parallax px') for each pixel. Hence, for a bundle of n images $(n \cdot (n-1))$ images have to be rectified to create $(n \cdot (n-1))/2$ image pairs (e.g. with $n=5$, twenty images have to be rectified). Especially the last may increase the computation time significantly. Besides, the image rectification process induces always a loss of information due to the need of grey- (colour-) value interpolation. Further on, for all subsequently steps the results of pair-wise image matching (which are disparity maps) cannot directly be joined together but have to be fused before. The extended approach for multi-image dense 3D surface matching, which is proposed in this paper, eliminates these disadvantages. It allows for the integration of an arbitrary number of (non-rectified) images into the matching process. The images can either be correlated pair-wise or by using n

* Corresponding author

images. The results of the new approach are index maps (instead of disparity maps) which directly indicate the 3D position of the best matches which simplifies subsequent processing steps (e.g. consistency checks between the results of pair-wise matching) because the index maps are directly comparable to each other. Further advantages of the new approach will be discussed in the following sections. First results are presented in chapter 3.

2. OBJECT-BASED MULTI-IMAGE SEMI-GLOBAL MATCHING (OSGM)

Within this chapter the method for object-based multi-image Semi-Global Matching (OSGM) will be described in detail (section 2.2 to 2.7). In advance, a short review of SGM will be given in section 2.1.

2.1 Review of SGM

The SGM method as originally described in (Hirschmüller, 2005) proposes an intelligent solution for the approximate minimization of global 2D energy functions as they are used e.g. within global image matching methods. SGM uses the following energy function:

$$E(D) = \sum_p C(p, D_p) + \sum_{q \in N_p} P_1 T \cdot [D_p - D_q = 1] + \sum_{q \in N_p} P_2 T \cdot [D_p - D_q > 1] \quad (1)$$

The first term of (1) sums the matching costs C between a pixel p in image 1 and a potential corresponding pixel in image 2 (at a specific disparity D). The second term adds a penalty P_1 for the current disparity D_p to the cost value C if the difference between D_p and the disparity D_q at a neighbouring pixel q is 1 (the function T returns 1 if $|D_p - D_q| = 1$ and 0 in all other cases). The second term adds a larger penalty to the cost value C if the difference between the disparity D_p to the disparity D_q at a neighbouring pixel q is higher than one (the function T returns 1 if $|D_p - D_q| > 1$ and 0 in all other cases).

First step in SGM is the cost calculation to build up the structure $C(p, D)$ in equation (1).

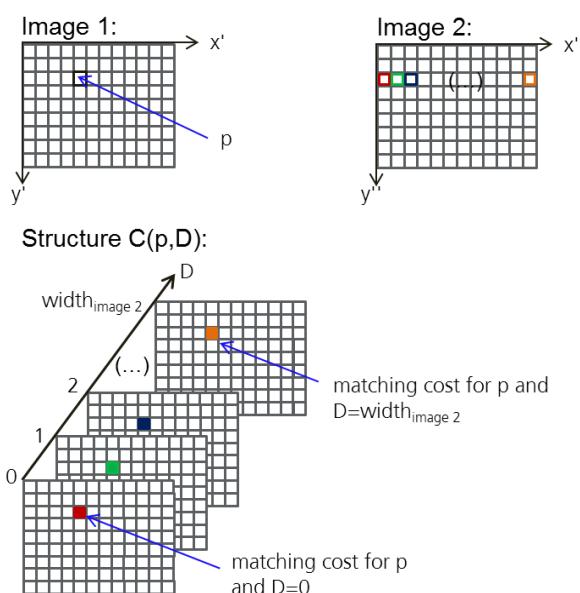


Figure 1. Cost calculation in Stereo SGM

For this, the matching costs between every pixel p in image 1 and all potential corresponding pixels in image 2 (at disparities D) have to be calculated. Since SGM is typically initialized in rectified image pairs the maximum number of possible disparities D is equal to the width of the rectified image 2 (due to horizontal epipolar lines). The cost calculation can be realized by using different cost functions reaching from very simple ones (e.g. differences of absolute intensity values (SAD)) to sophisticated ones (e.g. mutual information as described in (Hirschmüller, 2005)). An analysis of different cost functions is not addressed in this paper but can be found e.g. in (Hirschmüller and Scharstein, 2007).

Second step in SGM is cost aggregation. The main idea of SGM is to utilize cost aggregation not in all directions (which would be necessary for a strength global solution) but in the direction of 16 or at least 8 paths L_r (to perform a “semi-global” solution). Cost aggregation can be done recursively and separately for every path L_r with

$$L_r(p, D) = C(p, D) + \min(L_r(p-r, D), L_r(p-r, D-1) + P_1, L_r(p-r, D+1) + P_1, \min_i L_r(p-r, i) + P_2) - \min_k L_r(p-r, k) \quad (2)$$

In equation (2) p is used as substitution for the x', y' coordinates of a pixel in image 1:

$$L_r(p, D) = L_r(x', y', D) \quad C(p, D) = C(x', y', D). \quad (3)$$

The positions of adjacent pixels are defined separately for each path with $p-r$:

$$L_r(p-r, D) = L_r(x'-u \cdot 1, y'-v \cdot 1, D) \quad (4)$$

(e.g. with $u=1, v=0$ for a path in x' -direction).

The expression in (2) searches the minimum path costs inclusive possibly added penalties P_1 and P_2 at the position of the previous pixel in path direction ($p-r$) and adds this minimum to the cost value $C(p, D)$ at the current pixel p and the disparity D . The last term of (2) subtracts the minimum path cost of the previous pixel to avoid very large values in L_r .

The paths of minimum costs are illustrated for a pixel p at disparity $D=2$ exemplarily for 4 paths in Figure 2:

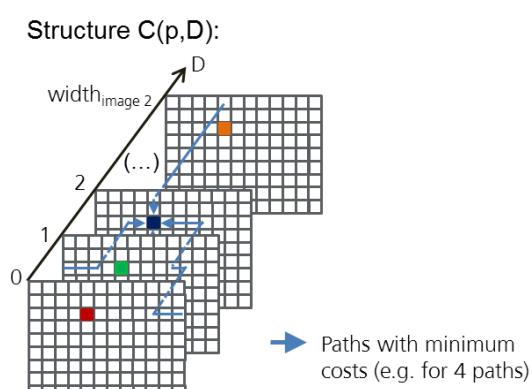


Figure 2. Paths with minimum costs

The results of the cost aggregation for 8 (or 16) paths can be fused with

$$S(p, D) = \sum_{r=1}^{8,16} L_r(p, D) \quad (5)$$

The final disparity D can then be derived from (3) by searching the minimum in $S(p, D)$ for each pixel p with:

$$\min_D S(p, D) \quad (6)$$

The final disparity is equal to the position D for each pixel p on which $S(p, D)$ reaches a minimum. D is stored for each pixel p which leads to the dense disparity map $D(p)$.

2.2 Cost calculation in object space

As already mentioned above the new approach (OSGM) differs in two major aspects from standard SGM. Firstly, cost calculation is formulated in object space instead of cost calculation in image space. Therefore, the object space is subdivided into a voxel raster in a first step. Each voxel may be a cube or a cuboid. The size of the cuboids (ΔX , ΔY , ΔZ) defines the resolution in object space (in X-, Y-, Z- direction of the global coordinate system, see Figure 3). The definition of the cuboids' size should be done under consideration of the mean GSD to ensure an adequate sampling rate. In order to provide hierarchical approaches the resolution of the images may be reduced (image pyramids).

In a second step, the central coordinate of each voxel is re-projected into all images by using the collinearity equations. Further on, the grey (or colour) values of the corresponding image coordinates are used for cost calculation. Thus, the structure $C(p, D)$ in (1) which equals to the more detailed description $C(x', y', D)$ with the image coordinates x', y' , is modified to $C(X, Y, Z)$ in which the coordinates X, Y, Z indicate the 3D position of a voxel. By doing so, the matching costs for each voxel can be calculated. Figure 3 illustrates the cost calculation in object space exemplarily for 3 images:

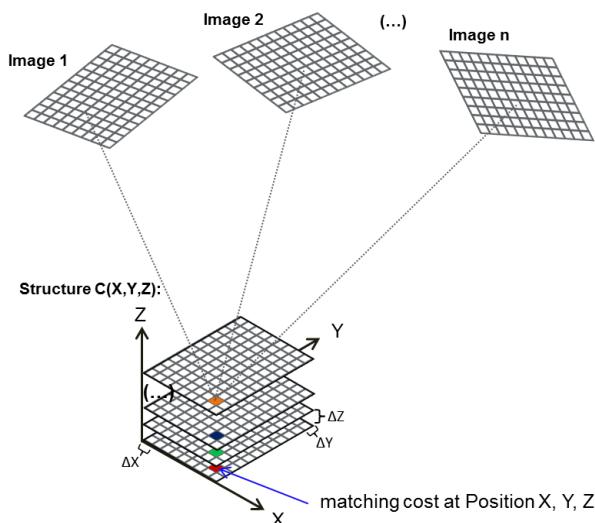


Figure 3. Multi-image cost calculation

Since the re-projection of the voxel coordinates X, Y, Z leads to sub-pixel coordinates within the images it is necessary to use interpolated grey- (or colour-) values for cost calculation.

Therefore, the cost values in $C(X, Y, Z)$ belong to sub-pixel image coordinates. Hence, the subsequently described SGM in object space (see section 2.4) leads directly to 3D points with sub-pixel accuracy (see section 3.5). This is one advantage compared to the standard SGM in which sub-pixel accuracy is typically achieved by interpolating between neighbouring cost-values in disparity space, e.g. by quadratic curve fitting as suggested in (Hirschmüller, 2008).

2.3 Cost calculation for n images

Most of the common cost or similarity functions (e.g. Census or normalized cross-correlation (NCC)) are designed for the calculation of the (dis)similarity between two signals (or respectively two images) and therefore well-suited for pair-wise image matching. Thus, for a combined cost calculation for n images it is necessary to think about sensibly extensions of cost or similarity functions for multi-image correlation. However, since pair-wise image matching in multi-image bundles can be used for consistency checks and can therefore be regarded as an important tool for the reliable detection of occlusions and other disturbances, both strategies (pair-wise image-matching and combined multi-image matching) should be considered within the new approach. We distinguish three possible procedures for cost calculation:

- (A): Pair-wise cost calculation for all possible image pairs and initialization of one structure $C(X, Y, Z)$ with the minimum cost value
- (B): Pair-wise cost calculation for all possible image pairs and initialization of one structure $C_i(X, Y, Z)$ for each image pair i
- (C): One structure $C_j(X, Y, Z)$ for every possible number of rays to one voxel but minimum of 2 rays (e.g. $j=4$ for 5 images) and initialization of $C_j(X, Y, Z)$ with the combined costs out of $2..j$ images

Procedure (A) describes the simplest way of cost calculation for multi-image bundles. Its main advantage is that the structure $C(X, Y, Z)$ has to be built up only for one time. Further on, a consistency check is done implicitly by searching the minimum cost value between all image pairs. A disadvantage of (A) is that it is impossible to detect voxels that are not visible or visible just in one image so that a possibly high number of outliers may remain in the data.

To reduce the number of outliers strategy (B) can be used. Compared with (A), the main advantage of (B) is the possibility of extensive consistency checks because $(n \cdot (n-1))/2$ matching results can be compared to each other (with n = number of images). A disadvantage of (B) is its memory consumption because the structure $C(X, Y, Z)$ has to be generated for $(n \cdot (n-1))/2$ times (for every possible image pair). But since the matching can be done one after another for every image pair, the latter argument is not a criterion for exclusion.

For multi-view stereo approaches sophisticated strategies for selecting sufficient image pairs for the pair-wise matching has been described e.g. in (Wenzel et al., 2013). These strategies can be adapted for OSGM to reduce efforts in pair-wise cost calculation with strategy (B).

The strategies of (A) and (B) both just combine the grey- or colour values of two images and can therefore not really be regarded as multi-image matching. However, especially (B) can be used in a first step to create a robust result and get information about which voxel is visible in which image. The results of (B) can afterwards be used to initialize a sophisticated

matching with (C) by re-calculating the costs for all voxels that are visible in pairs, triples, quadruples and so on.

2.4 Cost aggregation in object space

The second essential difference of our approach compared to standard SGM is that the cost aggregation as well as the cost calculation is transferred to the object space. The global energy function of SGM in (1) is extended to

$$E(Z) = \sum_p C(X, Y, Z_p) + \sum_{q \in N_p} P_1 T \cdot [Z_p - Z_q \models 1] + \sum_{q \in N_p} P_2 T \cdot [Z_p - Z_q \succ 1] \quad (7)$$

The equation (7) can be interpreted in analogy to (1) but with the difference that changes in Z-direction of the global coordinate system are penalised with P_1 and P_2 for adjacent voxels instead of penalising disparity changes between adjacent pixels. Hence, the smoothness constraint controls the smoothness in Z-direction of the global coordinate system and therefore equation (7) can be regarded as a 2.5D realization of object-based SGM.

For the minimization of (7) by adapting the semi-global approach the path-wise cost aggregation can be done recursively for every path L_r with

$$\begin{aligned}
L_r(v, Z) = & C(v, Z) + \min(L_r(v - r, Z), \\
& L_r(v - r, Z - \Delta Z) + P_1), \\
& L_r(v - r, Z + \Delta Z) + P_1), \\
& \min_i L_r(v - r, i \cdot \Delta Z) + P_2)) \\
& - \min_k L_r(v - r, k \cdot \Delta Z) + P_2))
\end{aligned} \tag{8}$$

The expression in (8) is an extension of (2) in which v is used as substitution for the X,Y-coordinate of a voxel:

$$\begin{aligned} L_r(v, Z) &= L_r(X, Y, Z) \\ C(v, Z) &= C(X, Y, Z). \end{aligned} \tag{9}$$

The X,Y-position of adjacent voxels are defined separately for each path with $v-r$:

$$L_r(v-r, Z) \equiv L_r(X - u \cdot \Delta X, Y - v \cdot \Delta Y, Z) \quad (10)$$

(e.g. with $u=1$, $v=0$ for path $t=1$, see Figure 4)

The expression in (8) searches the minimum path costs including possibly added penalties P_1 and P_2 at the position of the previous voxel in path direction ($v-r$) and adds this minimum to the cost value $C(X, Y, Z)$ of the current voxel. The penalty P_1 is added if the difference in Z-direction between the current voxel and the adjacent voxel is equal to ΔZ and P_2 is added if the difference in Z-direction is larger than ΔZ . The last term of (8) subtracts the minimum path cost of the previous voxel to avoid very large values in L_v .

The paths of minimum costs are illustrated for a voxel with $Z=2\Delta Z$ exemplarily for 8 paths in Figure 4.

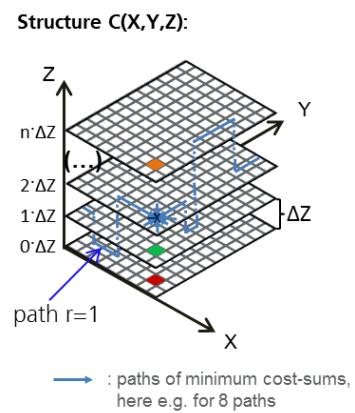


Figure 4. Paths with minimum costs

Analogue to (5) the results of the cost aggregation for 8 (or 16) paths can be fused with

$$S(v, Z) = \sum_{r=1}^{8,16} L_r(v, Z) \quad (11)$$

The matching result can then be derived from (11) by searching the minimum in $S(v, Z)$ for each v :

$$\min_Z S(v, Z) \quad (12)$$

The final Z-coordinate for each voxel v is equal to the position Z on which $S(v, Z)$ reaches a minimum. The final value can then be stored in a index map $Z(v)$ for each voxel v (instead of a disparity map $D(p)$).

2.5 Consistency checks

If pair-wise cost calculation has been performed so that the structure $C_i(X, Y, Z)$ has been built up for i times (see section 2.3) the semi-global minimization of (7) can be done for i times as well (with $i = (n \cdot (n-1)) / 2$ and $n = \text{number of images}$). Hence, i index maps $Z_i(v)$ can be calculated and afterwards be fused e.g. by testing the differences $Z_j(v) - Z_i(v)$ against a threshold t and calculating a mean value $Z_{mean}(v)$ if the test is positive for all pairs:

A consistency check with (13) eliminates all voxels which are not visible in all images and therefore object areas which are partly occluded would be removed. To avoid this, enhanced consistency checks are possible, e.g. clustering all index maps which lead to equal Z-values and choose the Z-value which has been estimated most frequently.

Generally, the more complex the object surface the more sophisticated the consistency check should be. Sophisticated strategies for pair-wise image selection are proposed e.g. in (Wenzel et al., 2013).

2.6 Discontinuities in X- and Y-direction

The approach for OSGM as described in section 2.4 is a 2.5D solution because for every raster-node X, Y exactly one Z -value can be estimated. For a lot of applications in which objects with low geometric complexity have to be reconstructed a 2.5D solution is sufficient. For more complex object geometries it may be necessary to develop a 3D solution. Therefore, the cost aggregation can principally be adapted for applying the smoothness constraint in X - or respectively in Y -direction instead of in Z -direction. Hence, the equations (7) to (12) can be modified by changing Z and X (or Z and Y).

2.7 Hierarchical computation

As described in section 2.2 the matching costs have to be calculated for every voxel within the voxel raster (Figure 3). Since the number of voxels may increase especially for large objects, the process of cost calculation may increase the computation time significantly. To reduce this loss of performance the algorithm can be implemented hierarchically by using image pyramids. A convenient approach for a hierarchical implementation of SGM has been proposed by (Rothermel et al., 2012) which can be adapted for the new approach as well. It is proposed to initialize the matching in a high level of an image pyramid (images with low resolution) and to use the matching result to limit the number of possible disparities for the next pyramid level by searching the minimum and maximum disparity for each pixel e.g. within a 7×7 neighbourhood. Since the new approach estimates Z -values directly instead of disparities the approach for hierarchical computation have to be adapted to limit the range of possible Z -values rather than limiting the disparity range from one pyramid level to the next.

Further on, in (Rothermel et al., 2012) decreases the number of possible disparities implicitly by reducing the resolution of the images (since the disparity map has the same size as the rectified image). In the OSGM the interval ΔZ has to be decreased e.g. by using the main GSD for the images with reduced resolution.

3. EXPERIMENTS AND RESULTS

Sections 3.1 and 3.2 describe the test object and used image data for first investigations on the new approach. In section 3.3 different the cost functions which were used for the first implementations are discussed. Finally, in section 3.5 first results are presented.

3.1 Test object

For the investigations a test object with a sinusoidal surface was chosen which was originally developed for investigations of optical measurement systems for area scanning. The surface was designed by using a 2.5D sinus function which leads to an object surface of continuous curvature. Since the SGM should be well-suited also for the robust measurement of depth discontinuities, the original shape of the test object was extended by adding a wedge (see Figure 5). On the one hand, the wedge allow for the investigation of the algorithms' behaviour at sharp object boundaries. On the other hand, the limits of the achievable spatial resolution can be investigated by analysing the matching results at the peak of the wedge. The latter one is of special interest for the comparison of pixel-wise cost functions to window-based cost functions.

The object surface was textured with a stochastic pattern which is assumed to be well-suited for image matching. Since the SGM should lead to robust matching results also in areas with no or low texture two areas without any texture were added: one in a valley and one on a plane on top of the wedge (Figure 5).

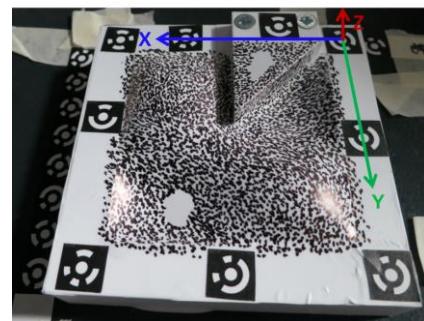


Figure 5. Test object

For the generation of 3D-reference data the test object was measured with a fringe projection system. This was done for the white surface (before texturing the surface). The accuracy of the fringe projection system was determined before by applying the VDI/VDE 2634 guideline part 2 and can be specified with a probing error $R=0.08\text{mm}$ and a sphere-spacing error $\Delta l=0.05\text{mm}$.

3.2 Image data

The object was captured with a Nikon D2x camera with 24mm lens. The images were orientated by bundle block adjustment and the camera was calibrated simultaneously. For first investigations three images out of the bundle were selected (Figure 6) which are used for image matching.

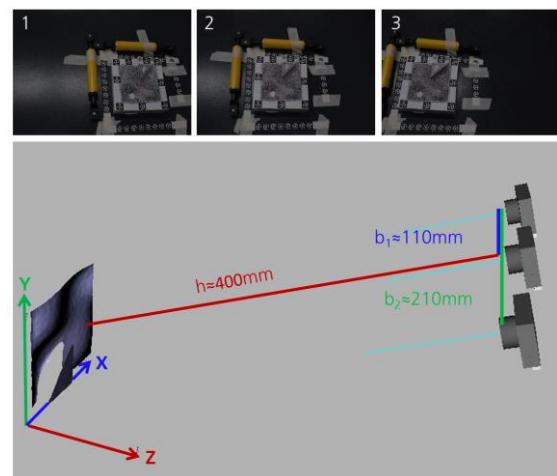


Figure 6. Images of the image triple and camera positions in 3D space

The configuration in Figure 6 yields to an approximate GSD on the object surface of around 0.1mm (with a distance to the object of $h \approx 400\text{mm}$, a camera constant of $c \approx 24\text{mm}$, a pixel size of $px \approx 0.0055\text{mm}$).

3.3 Cost functions

Since the first aim of the investigation is to test the new matching approach, until now the focus was set on the

development of the matching more than on implementing sophisticated cost functions (e.g. like mutual information). A simple cost function which is often used for SGM is given by census (Zabih and Woodfill, 1994). Census is highly invariant against radiometric differences between the images and therefore leads to robust matching results. The cost parameter of census is given by the hamming distance between two image windows. Hence, the maximum number of distinguishable cost values is equal to the maximum hamming distance h_{max} which depends on the window size (e.g. for a 5x5 window $h_{max}=25$). Since changes of the centre coordinates of the voxels in Z-direction by small increments ΔZ (see Figure 7) lead to sub-pixel movements of the matching windows within the images, it is necessary to use a cost function that allows for the distinction between these sub-pixel movements. First investigations by using census have shown that it does not fulfil this requirement due to its limited resolution as described above. Another popular and well-known similarity function which is invariant against radiometric differences, is normalized cross correlation (NCC) which is able to detect sub-pixel movements within certain limits. Hence, for the following investigations NCC was used. The correlation coefficient is defined by

$$\rho_{fg} = \frac{\sigma_{fg}}{\sigma_f \cdot \sigma_g} = \frac{\left(\sum (f_i - \bar{f})(g_i - \bar{g}) \right) / n}{\sqrt{\sum (f_i - \bar{f})^2 / n} \cdot \sqrt{\sum (g_i - \bar{g})^2 / n}} \quad (14)$$

In (14) σ_{fg} is the covariance between the grey-values within the two image windows f and g and σ_f and σ_g are the variances of the grey-values in the image windows. Since the coefficient ρ_{fg} is a measure of the similarity and SGM typically uses cost values for the description of the *dissimilarity*, (14) is modified with

$$\rho = 1 - \rho_{fg} \quad (15)$$

In (15) ρ is the cost parameter which leads to cost values within the interval 0.0 (low matching costs, high similarity) to 2.0 (high matching costs, low similarity). One disadvantage of NCC compared to census is the need for a higher bit depth for representing the real numbers in ρ (e.g. by using 32bit floating point data types). Since for census a depth of 8 bit (which allows for the use of matching windows up to 16x16 pixels) is sufficient for a lot of applications, the use of 32 bit data types for NCC leads to a four times larger memory requirement to build up the cost structure $C(X,Y,Z)$ in (7) as well as for each structure $L_r(X,Y,Z)$ in (9) for the aggregated costs. Another issue concerning the NCC is that it is not invariant against image rotations and different image scales. To achieve invariance for rotations and scales the matching window is defined in object space instead of in image space by defining a squared point raster around each voxel centre. The point raster is oriented parallel to the XY-plane of the global coordinate system.

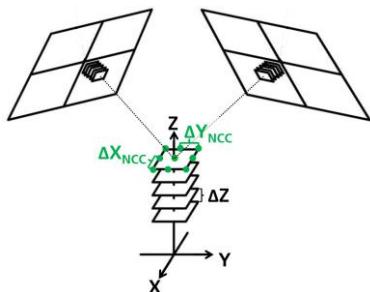


Figure 7. Correlation with vertical line locus

Thus, cost calculation is similar to the well-known vertical line locus approach for image correlation. A disadvantage of this approach is that for object areas that are sloped with respect to the XY-plane the correlation coefficient may decrease significantly.

Further on, NCC as described in (14) allows for the correlation of image pairs and is therefore suitable for cost calculation with procedure (A) or (B) as described in section 2.3 but not for real multi-image correlation.

3.4 Parameter settings

For all subsequently presented matching results the following parameters were used (if not otherwise specified): The size of the measurement volume is adapted to the size of the test object which is about 120mm in X- and Y-direction and about 30mm in Z-direction, starting in the origin of the coordinate system (see Figure 6). The voxel size which also defines the spatial resolution of the measurement was set to $\Delta X = \Delta Y = 1.0\text{mm}$ and $\Delta Z = 0.025\text{mm}$. This leads to about 9 million voxels within the measurement volume. If 32 bit data types for the cost values in $C(X,Y,Z)$ and the aggregated costs in $L_r(X,Y,Z)$ are used (as described in the foregoing section) the required memory for each structure $C(X,Y,Z)$ is about 35 Mbytes and for the 8 structures $L_r(X,Y,Z)$ about 280 MByte (=8·35MByte) which is far away from hardware limits.

The window size for NCC is 40x40 points with a point distance of $\Delta X_{NCC} = \Delta Y_{NCC} = 0.1\text{mm}$ (see Figure 7) which is adapted to the mean GSD (section 3.2). Hence, the window size on the object surface is about 4x4mm.

The penalties P_1 and P_2 for SGM are not tuned automatically by using gradient information as proposed in Hirschmüller (2008) but set to fixed values with $P_1=0.1$ and $P_2=0.6$. The cost aggregation is done by using 8 paths.

All subsequently presented matching results were generated without any pre-processed image filtering and no post-process filtering of the resulting index maps.

For the comparison of the matching results with respect to the fringe projection measurement best fit transformations were applied by using the software Geomatic Qualify.

3.5 Results

The cost structure $C(X,Y,Z)$ in (7) can be analysed by searching the minimum cost value in Z-direction (“the winner takes it all” approach) denoted as NCC result in the following sections. The first investigation focusses on the comparison of the NCC result to the OSGM result. The point clouds for both results overlaid with the TIN of the fringe projection measurement are displayed in Figures (8) and (9):

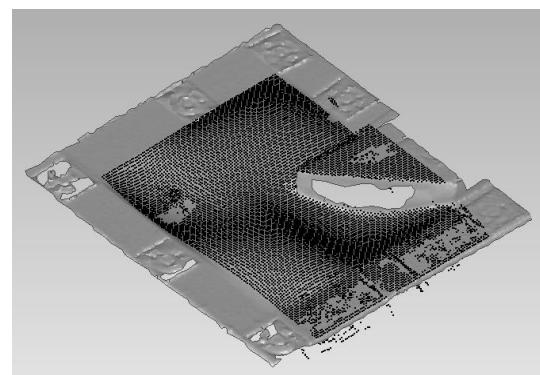


Figure 8. NCC result

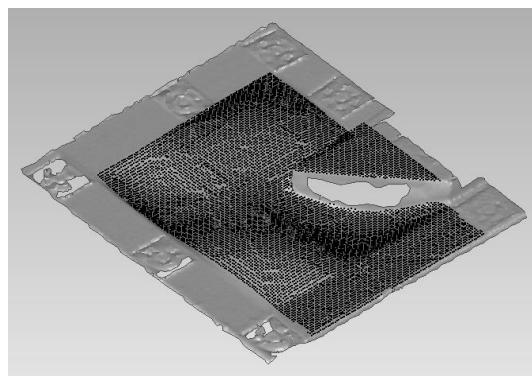


Figure 9. OSGM result

Both approaches lead to good results at the border of the wedge. The NCC result shows a high number of outliers in areas with no texture (on top of the wedge, in the “valley” and on the right-hand side of the object). In contrast, the OSGM expectedly leads to a more smooth result without outliers in untextured areas. For an extended accuracy evaluation both results were compared to the fringe projection measurement. The results of this comparison are illustrated in Figures (10) and (11):

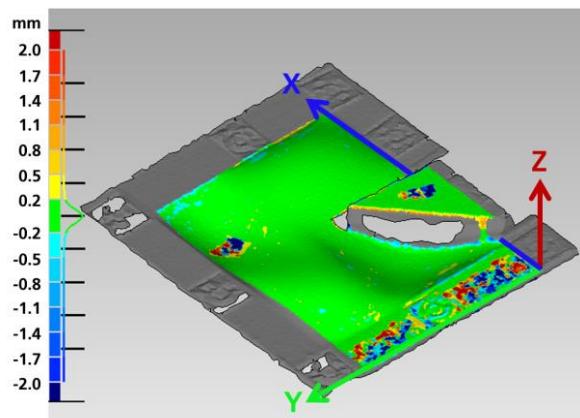


Figure 10. Comparison of the NCC result to fringe projection result

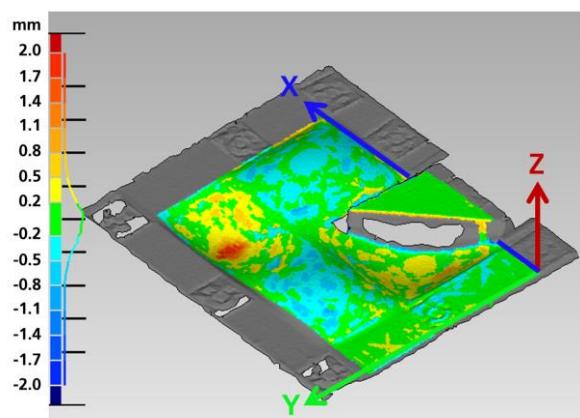


Figure 11. Comparison of the OSGM result (with $P_1=0.1$, $P_2=0.6$) to fringe projection result

Apart from the areas with outliers the NCC result is more accurate than the OSGM result. Most deviations are within the interval of -0.2 mm to 0.2 mm and the histogram of the deviations equals the Gaussian distribution (Figure 10). In

contrast, the OSGM result shows systematic deviations in areas with continuous curvature and a clearly wider distribution within the histogram. This is caused by the smoothness constraint of SGM which allow for changes in Z-direction between adjacent voxels only if the cost value plus the penalty P_1 (or P_2) is lower than the cost values of the adjacent voxels. However, the smoothness constraint induces a very robust result in areas without any texture.

Since the smoothness constraint causes a smoothing in Z-direction OSGM leads to a result with only small deviations in non-textured areas that are parallel to the XY-plane of the global coordinate system (e.g. in areas on top of the wedge or on the right hand side of the object). In contrast, in the non-textured areas of continuous curvature the smoothing of SGM leads to significant systematic deviations up to 1.7 mm e.g. in the valley on the left-hand side of the object (Figure 11).

The effect of smoothing can be reduced if the penalties for P_1 and P_2 are modified. For example, if the penalties are modified to $P_1=0$ and $P_2=0.1$ the matching result of OSGM can be improved as illustrated in Figures (12) and (13).

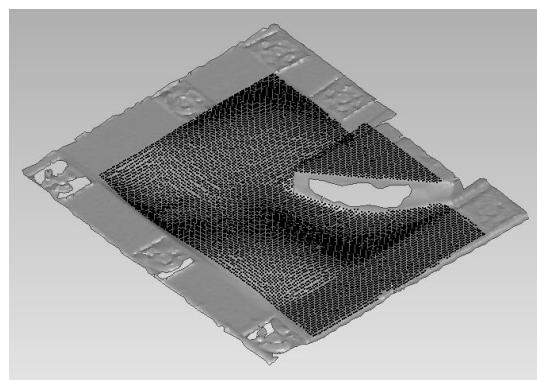


Figure 12. OSGM result with $P_1=0$ and $P_2=0.1$

A penalty for $P_2=0.1$ is obviously sufficient to avoid outliers in non-textured areas (Figure 12). Further on, the modification of the penalties reduces the smoothing and leads to less systematic deviations in well-textured areas of continuous curvature (Figure 13). Merely in the non-textured areas in the valley (on the left-hand side of the object), the systematic deviations remain due to missing information for correct matches. In this area the smoothness constraint avoids outliers but the 3D points deviate systematically from the correct shape of the object.

All in all the accuracy increases significantly due to the modifications of the penalties.

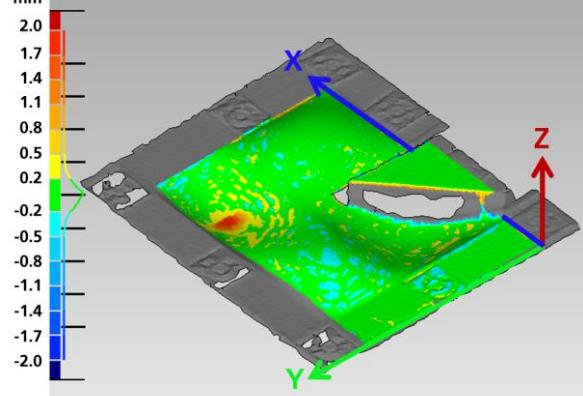


Figure 13. Comparison of the OSGM result (with $P_1=0$ and $P_2=0.1$) to fringe projection result

4. SUMMARY AND OUTLOOK

The presented extension of SGM to Object-based Semi-Global Matching (OSGM) is mainly characterized by transferring the process of cost calculation and path-wise cost aggregation from image space into the object space. Instead of estimating dense disparity maps, index maps are generated which directly indicate the best matches in 3D space.

The new approach was tested under laboratory conditions by using a test object with reference data of a fringe projection measurement. The tests show very promising results. OSGM maintains the benefits of SGM (e.g. robustness in non-textured areas, good result at sharp object boundaries) and adds several advantages:

In opposite to most multi-baseline or multi-view stereo approaches the new approach works without rectified images and therefore reduces the efforts for pre-processing (no need for image rectification) and for post-processing (no need for the fusion of disparity maps). Further on, the new method allows for the integration of more than two images into the matching process and is therefore suitable for real multi-image correlation. All in all, the OSGM algorithm has a clearly simplified structure compared to SGM in multi-view stereo approaches.

Further developments will focus on the implementation of sophisticated pixel-wise cost functions to fully exploit the advantages of SGM. Furthermore, the implementation should be extended to a hierarchical approach (as described in section 2.7) to increase the computational performance.

Based on proposals in Rothermel et al. (2012) further investigations and developments should be carried out concerning the strategies for an optimal selection of image-pairs for a pair-wise cost calculations following strategy (B) as described in section 2.3.

Furthermore, the new approach will be investigated by using other (close-range- and aerial-) test datasets.

Since the structure of the new approach separates the process of cost calculation from special properties of image sensors, extensions for the integration of other sensors (e.g. aerial or satellite sensors) should be considered. Furthermore, the integration of colour- or multi-spectral information into the matching should be considered which possibly may add helpful information for stabilizing the matching process.

ACKNOWLEDGEMENTS

The research has been supported by the Lower Saxony program for Research Professors, 2013-2016.

REFERENCES

- Banz, C., Hesselbarth, S., Flatt, H., Blume, H. and Pirsch, P., 2010. Real-time stereo vision system using semi-global matching disparity estimation: Architecture and fpga-implementation. In: IEEE Conference on Embedded Computer Systems: Architectures, Modeling and Simulation.
- Buder, M., 2012. Dense realtime stereo matching using a memory efficient Semi-Global-Matching variant based on FPGAs. In: Kehtarnavaz, N., Carlsohn, M. (Eds.): Real-Time Image and Video Processing 2012. SPIE Proceedings Vol. 8437.

- H. Hirschmüller, 2005. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 807-814, June 2005.
- H. Hirschmüller and D. Scharstein 2007. Evaluation of Cost Functions for Stereo Matching. Computer Vision and Pattern Recognition. IEEE Conference on, pp. 1-8.
- Hirschmüller, H., 2008. Stereo processing by semi-global matching and mutual information. IEEE TPAMI 30(2), pp. 328-341.
- Ernst, I. and Hirschmüller, H., 2008. Mutual information based semi-global stereo matching on the gpu. In: ISVC, Vol. LNCS 5358, Part 1, Las Vegas, NV, USA, pp. 228-239.
- Hirschmüller, H. and Scharstein, D., 2009. Evaluation of stereo matching costs on images with radiometric differences. IEEE TPAMI 31(9), pp. 1582-1599.
- Michael, M., Salmen, J., Stallkamp, J., Schlippsing, M. 2013. Real-time Stereo Vision: Optimizing Semi-Global Matching. 2013 IEEE Intelligent Vehicles Symposium (IV) June 23-26, 2013, Gold Coast, Australia.
- Rothermel, M., Wenzel, K., Fritsch, D., Haala, N. 2012. SURE: Photogrammetric surface reconstruction from imagery. Proceedings LowCost3D Workshop 2012, 04th – 05th Decembre 2012, Berlin.
- Wenzel, K., Rothermel, M., Fritsch, D., Haala, N. 2013. Image acquisition and model selection for multi-view stereo. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5/W1, pp. 251-258.
- Zabih, R. and Woodfill, J. 1994. Non-parametric local transforms for computing visual correspondance. In *Proc. ECCV*, pages 151-158.