PFG 2012 / 1, 0051 – 0061
Stuttgart, January 2012

Article

Dense 3D Reconstruction of Low Texture Surfaces Using an Energy Minimization Framework with Smoothness-Based Priors

Ralph Schmidt, Christian Heipke, Manfred Wiggenhagen & Bernd Michael Wolf, Hannover

Keywords: Metrology, 3D reconstruction, matching, surface, texture

Summary: This article describes the dense stereoscopic 3D reconstruction of surfaces which offer only low texture by employing a global matching algorithm with smoothness-based priors in an energy minimization framework. The envisaged application areas are high speed image sequences of dynamic processes where the projection of structured light is not applicable. The lack of depth cues on the measured object normally leads to very sparse and often false reconstructions if common local matching algorithms like cross correlation or least squares matching are employed. Within this AiF funded project an operational photogrammetric stereo measurement system has been developed consisting of a stereo rig with high speed cameras and a global matching algorithm. This system allows for the first time a dense reconstruction of surfaces with low texture in high speed image sequences. Quantitative and qualitative results for two test data sets demonstrate that the determination of a dense point cloud of low texture objects without employing structured light is possible.

Zusammenfassung: *Dichte 3D-Rekonstruktion von Oberflächen mit geringer Textur unter Anwendung eines Energieminimierungsverfahrens mit Glattheitsannahmen.* Dieser Artikel beschreibt die dichte stereoskopische 3D-Rekonstruktion von Oberflächen mit geringer Textur unter Anwendung eines globalen Zuordnungsverfahrens mit Glattheitsannahmen in einem Energieminimierungsverfahren. Der angestrebte Einsatzbereich sind Highspeed-Bildsequenzen dynamischer Vorgänge, bei denen die Projektion von strukturiertem Licht nicht möglich ist. Der Mangel an geeigneten Merkmalen am Messobjekt führt bei der Verwendung von üblichen lokalen Zuordnungsverfahren wie Kreuzkorrelation oder Kleinste-Quadrate-Bildzuordnung normalerweise zu einer dünn besetzten und oft falschen Rekonstruktion. Im Rahmen dieses von der AiF geförderten Projektes wurde ein operationell einsetzbares photogrammetrisches Stereomesssystem bestehend aus einem Stereorack mit Highspeed-Kameras und einem globalen Zuordnungsalgorithmus entwickelt. Dieses System ermöglicht erstmals eine dichte Rekonstruktion von Oberflächen mit geringer Textur in Highspeed-Bildsequenzen. Quantitative und qualitative Ergebnisse für zwei Testdatensätze demonstrieren, dass die Bestimmung einer dichten Punktwolke von Objekten mit geringer Textur möglich ist, ohne strukturiertes Licht zu verwenden.

1 Introduction

The three-dimensional reconstruction of static scenes has been a very important research area over the last three decades. For industrial applications mainly two methods are in use today: a) Structured light with laserline or two-dimensional projector and one or two cameras, b) time-of-flight approaches for laser scanning devices or time-of-flight cameras. While the time-of-flight methods directly measure the depth values respectively distances between the object and the sensor by time measurements, the structured light ap-

www.schweizerbart.de
DOI: 10.1127/1432-8364/2012/0102
1432-8364/12/0102 $ 2.75

proaches obtain depth information from triangulation. Using time-coded light patterns like grey code or time-varying patterns like light stripes an unambiguous correspondence for each image pixel is strived for (Battle et al. 1998, Pagès et al. 2003). These approaches operate in the time domain as several subsequent measurements are necessary for a complete and highly accurate reconstruction of the object or scene. Therefore, these methods are not applicable for dynamic scenes where the object changes its shape and/or position fast in relation to the time that is necessary for recording. Generally, only two dimensional sensors have the potential to capture 3D deformation and motion of a 3D object. An approach operating in the space domain is the conventional photogrammetric setup applying at least two cameras. Often artificial patterns are projected onto the scene to support unambiguous matching even in weakly textured image regions (Kang et al. 1995) which can also be applied to dynamic scenes by reconstructing each frame pair independently. In Davis et al. (2005) and Zhang et al. (2003) a spacetime stereo approach is presented which incorporates the time and space methods into a unifying framework.

In this paper we present an operational photogrammetric stereo measurement system consisting of a stereo rig with high speed cameras and a global matching algorithm which is able to produce a dense point cloud of objects which exhibit low texture. Dense is related here to the aim of determining a disparity value for every pixel of the image. The envisaged application areas are high speed image sequences of dynamic processes. In particular, the surface of an airbag during deployment within a crash test (impact test) should be reconstructed. To our best knowledge no operational projector is available on the market which is able to project a pattern to support correlation because very strong illumination and very high frame rates are needed during the crash tests. Thus, structured light approaches are not applicable. Also especially prepared airbags with imprinted patterns are not feasible since airbags from mass production are used during the tests. Our developed system allows for the first time a dense reconstruction of surfaces with low texture in high speed image sequences. There are no high accuracy requirements as the focus is on a dense and complete reconstruction of the surface of dynamic objects for the analysis of their deformation and motion behaviour. Additionally, for operational reasons the algorithm should be relatively fast which means the user should obtain the results within a few seconds. Also, it should be able to handle image sizes in full HD resolution (1920×1080 pixels). The described work was supported by AiF Arbeitsgemeinschaft industrieller Forschungsvereinigungen "Otto von Guericke" e.V. and thus focuses on technology transfer of innovative solutions.

An overview of local and global matching algorithms is given in section 2. A global matching algorithm using smoothness-based priors in an energy minimization framework, which allows a dense reconstruction of surfaces with low texture and which was developed to solve the given task, is presented in section 3. Quantitative and qualitative results for the reconstruction of two test data sets are shown in section 4. Finally, in section 5 conclusions are drawn and an outlook for future developments is given.

2 Matching

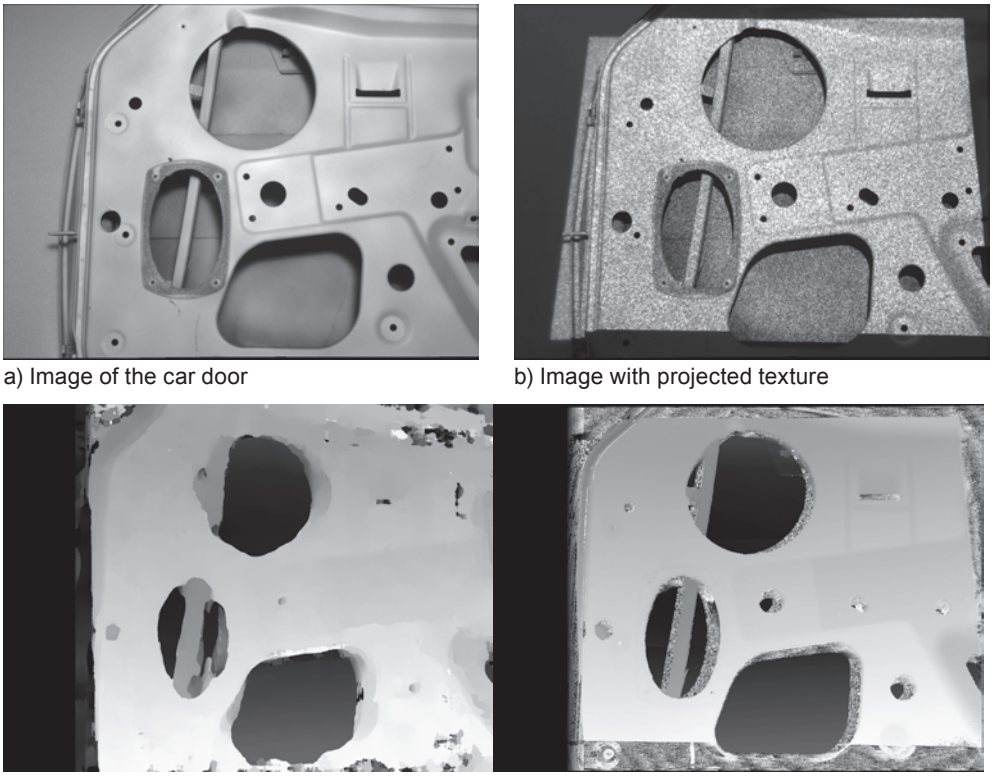
2.1 *Local Matching Algorithms*

Traditional stereo matching methods identify corresponding image points by analyzing the similarity of image patches. Typically, a reference window around the pixel of interest is compared to a search window over a certain search space. In a calibrated setup with two cameras the search space can be reduced to searching along epipolar lines which comply with corresponding image lines in rectified epipolar images. The search space can be further reduced with respect to the expected depth range so that it is not necessary to search along the entire line. One of the simplest similarity measures is the sum of absolute differences (SAD):

$$SAD = \sum_{(x,y)\in W_S} \left| I_r(x,y) - I_l(x,y) \right| \qquad (1)$$

where I_r are the intensity values of the right image, I_l the intensity values of the left image and W_s is the spatial aggregation window around the pixel of interest. Numerous other similarity measures have been developed which mainly aim at accounting for radiometric differences between the two images (Hirschmüller & Scharstein 2009). The final corresponding pixel is selected by simply choosing the pixel in the second image with the lowest SAD which is commonly referred to as "winner-take-all strategy". The calculated pixel offset is called disparity or parallax. However, these local matching approaches impose an implicit smoothness assumption upon the scene which means that no depth discontinuity is allowed to lie within the aggregation window. The reason is that depth discontinuities lead to an unequal object appearance in the images when viewed from different perspectives. Therefore, the matching window has to be as small as possible to reduce the probability of including discontinuities in depth or depth change. On the other hand, small windows may contain too little information for unambiguous matching and are sensitive to noise so that in the matching strategy there is always a tradeoff in choosing an appropriate window size.

In Fig. 1 the influence of different window sizes is shown using the example of a scene depicting the inner side of a car door without interior paneling. The reference result in Fig. 1d was produced by projecting different noise patterns on the door to support correspondence and employing a spacetime matching algorithm like in Davis et al. (2005). A 1×1 space window, i. e. the single pixels and no spatial aggregation window were used with 12 time windows, i. e. 12 different noise patterns. Without the projected texture the aggregation window has to be very large (45×45



a) Image of the car door

b) Image with projected texture

c) Disparity map of the car door image using a 45×45 pixels window

d) Reference disparity map using a $1 \times 1 \times 12$ pixels spacetime window

Fig. 1: Influence of different aggregation window sizes.

pixels in this case) to gather enough grey value variation to allow a reliable correspondence search (Fig. 1c). The determined disparities are encoded as grey values and stretched to the range 0–255 so that the shown disparity maps are obtained. It can be observed that the large aggregation window oversmoothes small features. Especially at the edges of holes – which mark disparity jumps – it can be observed that the locations of the edges are displaced. This phenomenon is usually referred to as "foreground fattening" (Scharstein & Szeliski 2002) where pixels from the far surface are wrongly assigned the same disparity as pixels on the near surface.

2.2 Global Matching Algorithms

To avoid the problems of the local matching techniques global strategies have been developed which calculate the disparity map for the entire image by formulating the problem as an energy minimization framework (Terzopoulos 1986):

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \tag{2}$$

The data term $E_{data}(d)$ measures the pixelwise similarity, which means how well the disparity function d is consistent with the input images:

$$E_{data}(d) = \sum_{(x,y)} C(x,y,d(x,y)) \tag{3}$$

Pixelwise similarity means that no aggregation window like in eq. 1 is necessary to compute the matching cost C though in principle this is also possible. The term $E_{smooth}(d)$ in eq. 2 explicitly encodes smoothness assumptions about the disparity function which is weighted by factor λ. These smoothness priors support the generation of reasonable disparity values over noisy and low texture areas. The smoothness term can e. g. be the sum of spatially varying smoothness costs for all 4-connected direct neighbours:

$$E_{smooth}(d) = \sum_{\{p,q\} \in N} w_{p,q} \cdot V(|\, d_p - d_q \,|) \tag{4}$$

where q is a vertical or horizontal direct neighbour to pixel position $p = (x,y)$ and N is the set of all neighbouring pixel pairs. The smoothness term is the product of a function V of disparity difference and optional spatially varying weights $w_{p,q}$ for each pixel pair. The weights can for instance be computed by an edge detector because disparity discontinuities often coincide with intensity edges. Disparity differences between neighbouring pixels (first derivative) are penalized by this energy function where bigger differences lead to higher penalties. Second order smoothness terms which penalize the curvature of the disparity function require triple cliques rather than pairs, i. e. three consecutive neighbouring pixels (Woodford et al. 2009). Unfortunately, this algorithm only works for very small images and is computationally very demanding.

The energy can be viewed as an objective function of a Markov random field (MRF) (Geman & Geman 1984) where a minimum of the energy corresponds to a maximum a posteriori (MAP) estimate. Finding the best result (global minimum) requires an exhaustive search over all candidate solutions which is not applicable here because the problem is NP-hard (Boykov et al. 2001), i. e. the computational complexity of the problem takes non-deterministic polynomial-time to solve. Therefore, algorithms look for a good approximate solution, i. e. a local minimum. On the other hand, the global minimum may not be the best result anyway because the energy function might not model the real world correctly. Early algorithms for energy minimization like iterated conditional modes (ICM) (Besag 1986) or simulated annealing (Barnard 1989) were comparatively ineffective. Besides having a very slow convergence behaviour these algorithms only find a good solution if a single dominant global minimum exists which is usually not the case with the energy functions employed for stereo matching. During the last years significant progress has been made in developing powerful algorithms like graph cuts (Boykov et al. 2001) or message passing algorithms like belief propagation (Sun et al. 2003). These global matching strategies produce excellent disparity maps and mark the state-of-the-art (Scharstein & Szeliski 2007).

An interesting alternative is semiglobal matching (Hirschmüller 2008) which approximates the global 2D smoothness constraints of the MRF by combining several 1D optimization paths. Thus, a significant reduction of the computational time is achieved. Another principle of incorporating smoothness constraints provide variational methods which also minimize an energy function with a data and a smoothness term (Poggio et al. 1985).

3 Selected Matching Algorithm

In this section we present our global matching algorithm with smoothness-based priors which support the propagation of reasonable disparity values over low texture areas. Initial tests with semiglobal matching and a variational approach using the energy functional of Slesareva et al. (2005) with a multigrid method for energy minimization did not achieve satisfactory results on our special data sets. Therefore, we focus on the MRF formulation.

At first the similarity measure for the data term $E_{data}(d)$ is chosen. Absolute differences (AD) are one of the simplest matching costs which penalize dissimilarity. However, they presume photo consistency which means that corresponding points should have the same grey value in the stereo image pair. In our experiments we found that AD gave poor results because of the violation of this assumption due to different viewing directions of the cameras resulting in unequal illumination conditions and reflectance behaviour. Therefore, we use the census transform as matching cost (Zabih & Woodfill 1994) which has been shown to have the best performance in the presence of local brightness differences (Hirschmüller & Scharstein 2009). The census transform is a non-parametric matching cost which is based on the local order of the grey values. A bit string is generated where each bit corresponds to one of the neighbourhood pixels around the current pixel of interest. If the neighbourhood pixel has a lower grey value than the centre pixel the bit is set. In our implementation the bit strings are calculated over a 9×7 local window and are stored in a 64 bit integer variable. The final matching cost value is determined by calculating the Hamming distance (Hamming 1950) between the bit strings of the left and the right image. The Hamming distance is simply the number of positions where the corresponding bits are different. The computation involves a "bitwise exclusive or" step and the counting of the nonzero bits which can be implemented very efficiently on modern CPUs.

For the smoothness term $E_{smooth}(d)$ several different forms are possible. A truncated version of the smoothness penalty V is used here which ensures that the penalty cannot be higher than V_{max}:

$$V(|d_p - d_q|) = \min(|d_p - d_q|^k, V_{\max}) \tag{5}$$

A truncated term is discontinuity-preserving and does not favour depth discontinuities with large disparity differences over smaller differences because neither is assumed to be more probable than the other. We use the L_1 distance with $k = 1$ but in principle a quadratic version with $k = 2$ or the Potts model ($V_{max} = 1$) is possible, too. The energy function can also be extended with an additional term for handling occlusions (Kolmogorov & Zabih 2001) but as occlusions did not pose a problem in our experiments we did not employ such a term.

For energy minimization we use an optimized version of the tree-reweighted message passing (TRW) (Wainwright et al. 2005) algorithm called sequential TRW (TRW-S) (Kolmogorov 2006). While the original TRW was motivated by maximizing a lower bound of the energy it is not guaranteed that this bound be increased over the iterations. The algorithm also may not converge, a problem which can be observed with belief propagation, too. TRW-S guarantees not to decrease the lower bound, it always converges and it has been shown that it often achieves superior results compared to graph cuts or belief propagation (Szeliski et al. 2008). Another advantage of practical importance is that it only consumes half as much memory as traditional message passing algorithms. A detailed description of the message passing scheme is outside the scope of this paper; it can be found in the relevant literature (Pearl 1988, Wainwright et al. 2005, Kolmogorov 2006).

4 Experiments

4.1 *Hardware Setup*

The application area is the processing of high speed image sequences recorded by high performance cameras like pco.dimax (www.pco.de/). A stereo rig shown in Fig. 2a allows capturing high speed stereoscopic image sequences at up to 1,300 fps with full 2000×2000 pixels resolution with 12bit dynamic. The SOLVing3D.titan system uses only one camera and a beamsplitter (s3d.camsplitter) (Fig. 2b) which provides a cost effective alternative and avoids the problem of synchronization but image size is reduced by half. Higher image rates of up to 150,000 fps are achievable at the cost of a reduced image size, e. g. in a typical beamsplitter setup two 500×500 pixels images at 9,000 fps are acquired.

For the evaluation of the algorithm described above two test data sets have been created. The first data set consists of a stereo pair of an inflated airbag. As can be seen in Fig. 3a, the images show only very little grey value variations. For technical reasons only one stereo pair could be recorded rather than a complete sequence. Additionally, another stereo pair of the same scene with a projected speckle pattern was recorded which was used to obtain a reference disparity map as ground truth (Fig. 3b). This disparity map was generated using a local matching algorithm with a window size of 5×5 pixels. Because of the technical limitations to project structured light in order to capture ground truth data with high speed image sequences – in fact this was the motivation for this project – still images were used. For the second data set a white styrofoam head was used which also hardly exhibits any texture but in contrast to the first data set contains small features like eyes, nose and mouth (Fig. 4). In order to generate a high quality reference data set the head was captured by employing a spacetime stereo approach (Davis et al. 2005) using a 3×3 space window and 12 different speckle patterns.

4.2 *Results*

To assess the quality of the stereo reconstruction, first a visual inspection of the resulting disparity map was made. The result for the airbag obtained without using projected texture is displayed in Fig. 5a and from the stereo pair with projected speckle pattern in Fig. 5b. The search range in the rectified epipolar images amounts to about 100 pixels. The black area in the left part of the disparity map stems from a constant parallax offset. The parameter λ was set to 20 and V_{max} to 7. Because there are no disparity discontinuities that coincide with intensity edges on the smooth surface of the airbag the influence of the weights $w_{p,q}$ were set to 0. These parameters were determined in an iterative procedure by visually inspecting the result. The TRW-S algorithm has been iterated only five times. In principle a stop criterion for energy minimization could be defined, but we found that typically after five iterations the change is negligible. Remarkably, we were able to reconstruct the whole surface of the airbag without any gaps, holes or other gross errors. The shape of the airbag is clearly

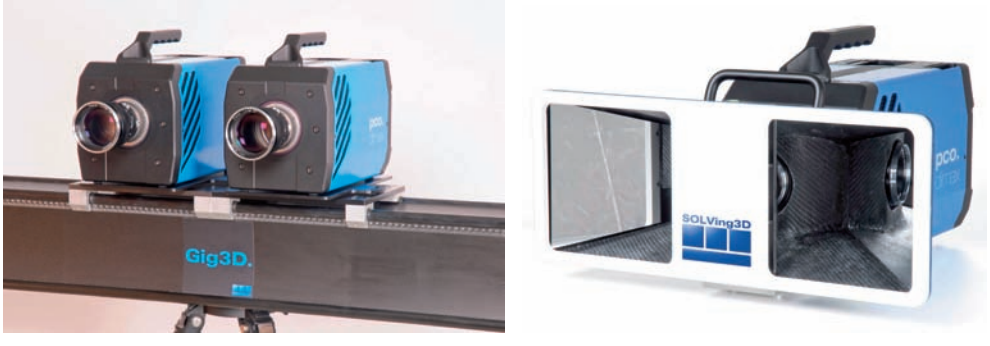


Fig. 2: a) pco.dimax in a stereo rig setup, b) With mounted beamsplitter.

visible. On the other hand the disparity map does not display a surface as smooth as the reference but exhibits small stepwise disparity jumps. The reason is that the prior of the form given in eq. 5 enforces a bias towards fronto-parallel surfaces over which the disparity is constant. The costs for the reconstruction of fronto-parallel planes are always lower

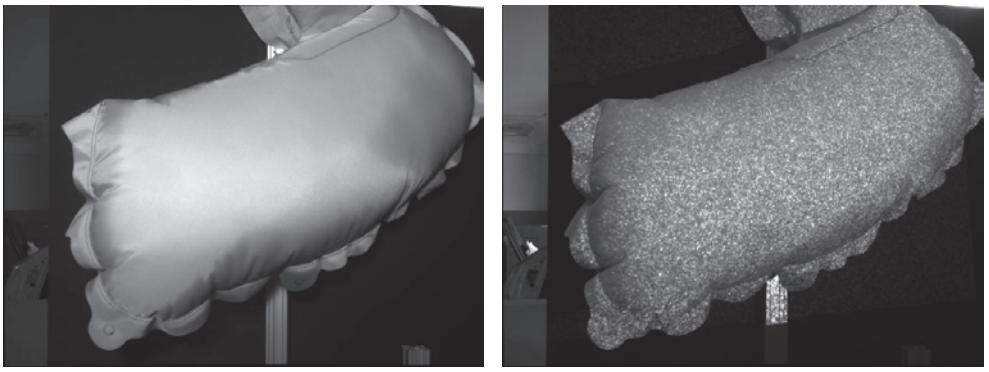


Fig. 3: a) Airbag test image, b) Airbag test image with projected speckle pattern.

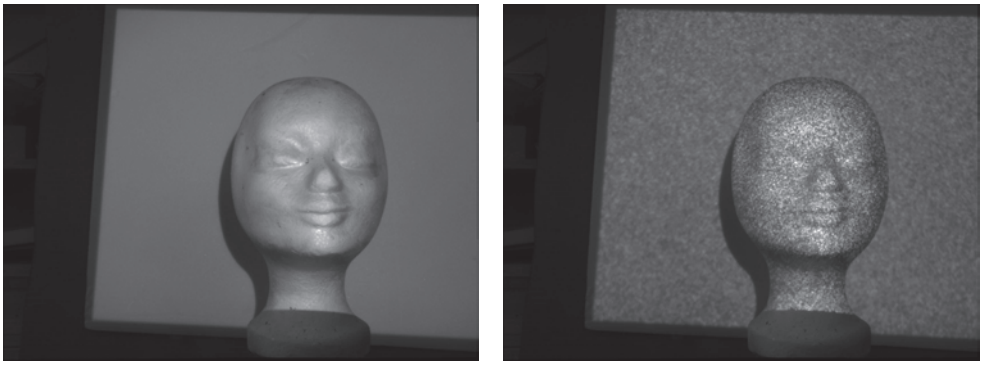


Fig. 4: a) Styrofoam head test image, b) Styrofoam head test image with projected speckle pattern, twelve different patterns were used for spacetime stereo.

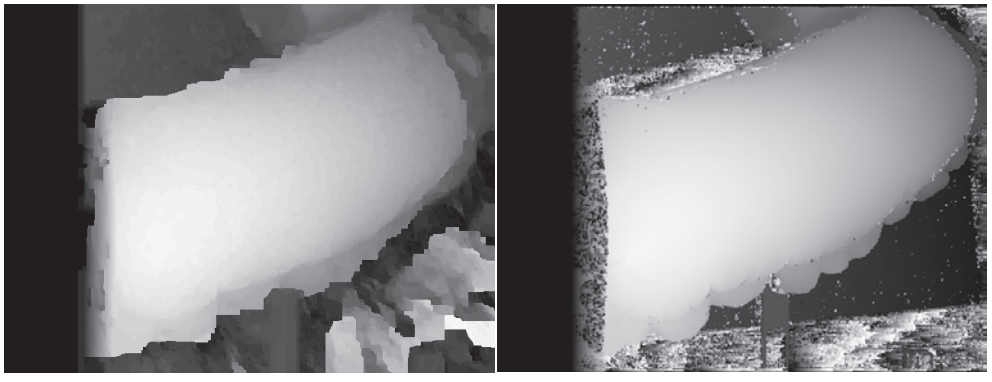


Fig. 5: a) Disparity map airbag test image, b) Reference disparity.

than for slanted surfaces with several small disparity jumps. Therefore, there is a tendency that slanted surfaces are reconstructed as a sequence of several small fronto-parallel surface patches.

The result for the styrofoam head is displayed in Fig. 6. The search range amounts to about 50 pixels because the background has not been incorporated into the range. This fact results in a noisy pattern but in our application the background is not of interest. The influence of the smoothing term was reduced a little bit by setting λ to 10 in order to preserve more details; V_{max} was set to 20. The computed disparity map is also not as smooth as the reference which is again an effect of the smoothness term's tendency to favour fronto-parallel surfaces. Nevertheless, the details of the face like eyes, nose and mouth are clearly visible. Owing to the edge preserving property of the smoothness term of the energy function such small details could also be reconstructed. Generally, the parameters had to be tweaked carefully to our class of textureless data sets to achieve good results. Nevertheless, once set only small adaptations of λ and V_{max} are necessary in order to achieve optimal results. The processing took about 15 to 25 seconds on a current CPU which is an acceptable time in our use case.

In order to obtain a quantitative measure of the quality of the resulting disparity map a difference image was calculated between the computed and the reference disparity values. The resulting error map shows the absolute differences >1 of the disparity values (Fig. 7).

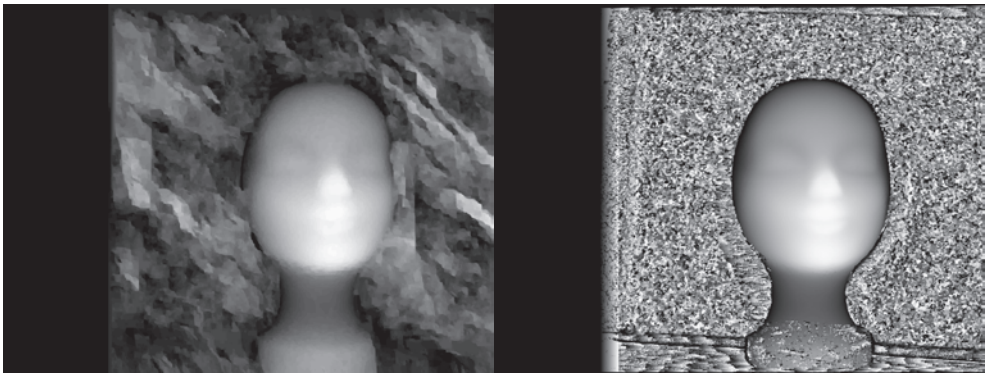


Fig. 6: a) Disparity map styrofoam head test image, b) Reference disparity.

a) Airbag error map b) Styrofoam head error map

Fig. 7: Error maps displaying the absolute differences to ground truth >1.

Only the errors at the object have been considered, the background was masked out. The differences due to the fronto-parallel surface patches cannot be observed in the airbag test image because they lie in the subpixel range. Larger errors can only be found at the border of the object. Additionally, there is a small area in the back near the border which shows larger differences. The reason for this behaviour is currently unclear. In principle, for the styrofoam head test image the same observations can be made: The surface is not as smooth as the reference and shows errors at the border of the object.

In order to obtain global quantitative quality measures root mean square error (RMSE) and the percentage of erroneously matched pixels are given (Scharstein & Szeliski 2002):

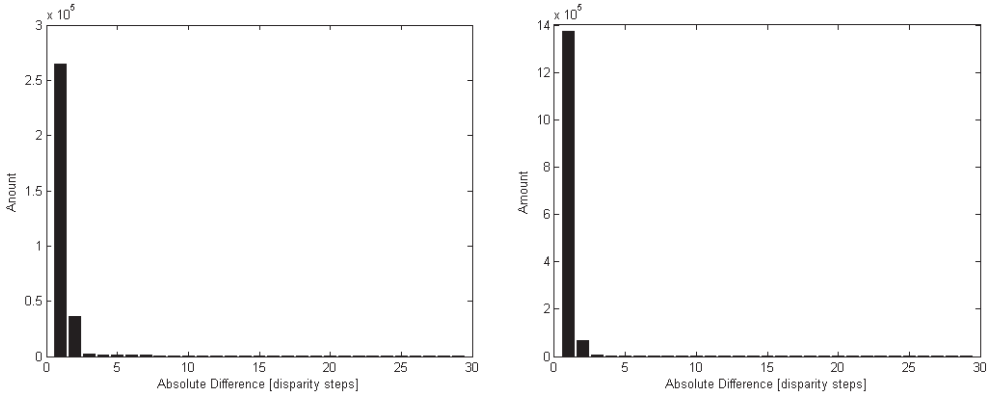
$$RMSE = \sqrt{\frac{1}{N}\sum_{(x,y)} |\, d_C(x,y) - d_R(x,y)\,|^2} \qquad (6)$$

$$B = \frac{1}{N}\sum_{(x,y)} (|\, d_C(x,y) - d_R(x,y)\,| > \delta_d) \qquad (7)$$

where $d_C(x,y)$ is the computed and $d_R(x,y)$ is the reference disparity map. N is the number of pixels in the image and δ_d is the error threshold which is set to 1 here. A differentiation into textureless, occluded and depth discontinuity regions like in Scharstein & Szeliski (2002) was not performed because we mainly focus on textureless objects. For the airbag test data set we obtained a RMSE of 1.72 and an error percentage of 6.6 %. The values for the head data set are slightly worse with a RMSE of 2.85 and an error percentage of 8.7 %. The RMSE values seem to be quite high but in consideration of the fact that we reconstructed two nearly textureless objects we did not expect to reach subpixel accuracy. Compared to the Middlebury evaluation ranking (Scharstein & Szeliski 2009) our error percentage did not achieve a top score but again in consideration of our nearly textureless objects the values are very good.

In addition to the overall error, in Fig. 8 histograms show the distribution of the amount of the absolute differences to ground truth. As can be seen, the average error amounts to 1 disparity step with a few pixels having a difference of two steps. A negligible amount of pixels exhibit outliers which mostly can be found at the edges of the object or at the erroneously reconstructed small area on the airbag. However, experience with the results of the Middlebury evaluation ranking show that without projecting structured light or a speckle pattern it is hardly possible to avoid such effects.

Overall, it can be stated that both test objects could be accurately reconstructed without projecting structured light. Hence, for the first time it is possible to stereoscopically analyze the deformation and motion behaviour of objects in high speed image sequences. The requirements of a short processing time and the handling of images in HD size are fulfilled. However, high precision measurements with subpixel accuracy are not possible with



a) Airbag errors b) Styrofoam head errors

Fig. 8: Histogram displaying the absolute differences to ground truth.

this method but for the envisaged application areas like the analysis of the inflation behaviour of an airbag the obtained accuracy is sufficient.

5 Conclusions and Outlook

We presented an operational photogrammetric stereo measurement system which allows for the first time a dense reconstruction of surfaces with low texture in high speed image sequences. It has been shown that it is possible to generate reasonable disparity values over noisy and low texture areas of an object by employing a global matching strategy with smoothness prior. Matching approaches using small local windows usually fail at this task because they do not contain sufficient unique and discriminating information for reliable correlation. Very large windows are not an option because of the foreground fattening effect. The prior helps in finding the correct solution taking into account the smoothness assumption. However, this smoothness assumption results in favouring fronto-parallel surfaces over slanted or highly curved surfaces. Obviously, this circumstance does not always model real objects or scenes correctly which has been shown in our experiments. Nevertheless, for our application area the algorithm delivers useful results. The average deviation to the two ground truth data sets is only about one pixel. This does not constitute a high accuracy measurement but the quality is sufficient for the analysis of the deformation and motion behaviour of objects.

The matching algorithm could be improved by using second order smoothness constraints like in Woodford et al. (2009) in order to model slanted surfaces more accurately without favouring fronto-parallel surfaces. Unfortunately, this algorithm has limitations in operational systems because it only works for very small images and is computationally very demanding.

The MRF used in our investigation assumes conditional independence of the variables which might not result in a realistic posterior distribution. In principle this assumption can be relaxed by employing conditional random fields (CRFs) which learn the free parameters of the MRF from ground truth data sets (Pal et al. 2011). However, like models with second order smoothness constraints this method has not reached an operational status yet.

Acknowledgments

The project was supported by Bundesministerium für Wirtschaft und Technologie (BMWi) within the scope of Zentrales Innovationsprogramm Mittelstand (ZIM) as a KF cooperation (FuE-Kooperationsprojekt zwischen Unternehmen und Forschungseinrichtungen) between the Institute of Photogrammetry and GeoInformation of Leibniz Universität Hannover (No. KF2182801KM9) and SOLVing3D GmbH (No. KF2182901KM9). This support is greatly acknowledged. We also would like to thank the anonymous reviewers who helped to improve the manuscript.

References

Barnard, S., 1989: Stochastic Stereo Matching over Scale. – International Journal of Computer Vision **3** (1): 17–32.

Battle, J., Mouaddib, E. & Salvi, J., 1998: Recent Progress in Coded Structured Light as a Technique to Solve the Correspondence Problem: A Survey. – Pattern Recognition **31** (7): 963–982.

Besag, J., 1986: On the Statistical Analysis of Dirty Pictures. – Journal of the Royal Statistical Society, Series B **48** (3): 259–302.

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

Davis, J., Nehab, D., Ramamoorthi, R. & Rusinkiewicz, S., 2005: Spacetime Stereo: A Unifying Framework for Depth from Triangulation. – IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2): 296–302.

Geman, S. & Geman, D., 1984: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. – IEEE Transactions on Pattern Analysis and Machine Intelligence **6** (6): 721–741.

Hamming, R.W., 1950: Error Detecting and Error Correcting Codes. – Bell System Technical Journal **29** (2): 147–160.

Hirschmüller, H., 2008: Stereo Processing by Semiglobal Matching and Mutual Information.

– IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2): 328–341.

Hirschmüller, H. & Scharstein, D., 2009: Evaluation of Stereo Matching Costs on Images with Radiometric Differences. – IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (9): 1582–1599.

Kang, S.B., Webb, J., Zitnick, L. & Kanade, T., 1995: A Multibaseline Stereo System with Active Illumination and Real-Time Image Acquisition. – Proceedings Fifth International Conference on Computer Vision: 88–93.

Kolmogorov, V. & Zabih, R., 2001: Computing Visual Correspondence with Occlusions using Graph Cuts. – Proceedings IEEE International Conference on Computer Vision **2:** 508–515.

Kolmogorov, V., 2006: Convergent Tree-Reweighted Message Passing for Energy Minimization. – IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (10): 1568–1583.

Pagès, J., Salvi, J., García, R. & Matabosch, C., 2003: Overview of Coded Light Projection Techniques for Automatic 3D Profiling. – Proceedings IEEE International Conference on Robotics and Automation **1:** 133–138.

Pal, C., Weinman, J., Tran, L. & Scharstein, D., 2011: On Learning Conditional Random Fields for Stereo. – International Journal of Computer Vision, doi: 10.1007/s11263-010-0385-z.

Pearl, J., 1988: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. – Morgan Kaufmann Publishers, San Francisco, California.

Poggio, T., Torre, V. & Koch, C., 1985: Computational Vision and Regularization Theory. – Nature **317** (6035): 314–319.

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

Scharstein, D. & Szeliski, R., 2007: http://vision.middlebury.edu/stereo/ (13.11.2011).

Scharstein, D. & Szeliski, R., 2009: http://vision.middlebury.edu/stereo/eval/ (13.11.2011).

Slesareva, N., Bruhn, A. & Weickert, J., 2005: Optic Flow Goes Stereo: A Variational Method for Estimating Discontinuity-Preserving Dense Disparity Maps. – Lecture Notes in Computer Science **3663:** 33–44.

Sun, J., Zheng, N.N. & Shum, H.Y., 2003: Stereo Matching using Belief Propagation. – IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (7): 787–800.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M. & Rother, C., 2008: A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors. – IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (6): 1068–1080.

Terzopoulos, D., 1986: Regularization of Inverse Visual Problems Involving Discontinuities. – IEEE Transactions on Pattern Analysis and Machine Intelligence **8** (4): 413–424.

Wainwright, M., Jaakkola, T. & Willsky, A., 2005: MAP Estimation via Agreement on Trees: Message-Passing and Linear Programming. – IEEE Transactions on Information Theory **51** (11): 3697–3717.

Woodford, O., Torr, P., Reid, I. & Fitzgibbon, A., 2009: Global Stereo Reconstruction under Second-Order Smoothness Priors. – IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (12): 2115–2128.

Zabih, R. & Woodfill, J., 1994: Non-parametric Local Transforms for Computing Visual Correspondence. – Proc. 3rd European Conf. on Computer Vision **2:** 151–158.

Zhang, L., Curless, B. & Seitz, S.M., 2003: Spacetime Stereo: Shape Recovery for Dynamic Scenes. – Proceedings IEEE Conference on Computer Vision and Pattern Recognition: 367–374.

Addresses of the Authors:

Dr.-Ing. Ralph Schmidt, Prof. Dr.-Ing. habil. Christian Heipke, Dr.-Ing. Manfred Wiggenhagen, Leibniz Universität Hannover, Institut für Photogrammetrie und GeoInformation, Nienburger Str. 1, 30167 Hannover, Germany, Tel.: +49-511-762-2484, Fax: +49-511-762-2483, e-mail: surname@ipi.uni-hannover.de

Dr.-Ing. Bernd Michael Wolf, SOLVing3D GmbH, Osteriede 8-10, 30827 Garbsen, Germany, Tel.: +49-5131-907972-0, Fax: +49-5131-907972-9, e-mail: bernd-m.wolf@solving3d.de

Manuskript eingereicht: Juli 2011
Angenommen: Oktober 2011