3D LEAST-SQUARES-BASED SURFACE RECONSTRUCTION

Duc Ton, Helmut Mayer

Institute of Photogrammetry and Cartography, Bundeswehr University Munich, Germany
Duc.Ton|Helmut.Mayer@unibw.de

KEY WORDS: Surface Reconstruction, 3D, Least-Squares Matching, Robust Estimation

ABSTRACT:

This paper addresses extensions of the classical least-squares matching approaches of (Wrobel, 1987, Ebner and Heipke, 1988) particularly in the direction of full three-dimensional (3D) reconstruction. We use as unknowns the movement in the direction of the normals for a triangulation of the surface. To regularize the ill-posed inverse reconstruction problem, we smooth the surface by enforcing a low curvature in terms of that the vertices of the triangulation are close to the average plane of their direct neighbors. We employ a hierarchy of resolutions for the triangulation linked to adequate levels of image pyramids, to expand the range of convergence, and robust estimation, to deal with occlusions and non-Lambertian reflection. First results using highly precise and reliable, but sparse points from the automatic orientation of images sequences as input for the triangulation show the potential of the approach.

1 INTRODUCTION

The goal of this paper is to generate a dense three-dimensional (3D) model from given orientations of cameras and reliable but sparse points obtained by an automatic orientation procedure.

In a recent survey for two images (Scharstein and Szeliski, 2002) the four steps (1) Matching cost computation (2) Cost (support) aggregation (3) Disparity computation / optimization and (4) Disparity refinement are named for a typical stereo algorithm. The test described in (Scharstein and Szeliski, 2002) has sparked a large interest into stereo matching. Here we report shortly only about approaches that deal with more than two images.

Our work is inspired by (Fua and Leclerc, 1996) which also employ 3D triangular facets for the surface. Opposed to them, we only focus on stereo, we optimize the vertices of the 3D facets along their normals, and we employ robust least-squares optimization to deal with occlusions.

In recent work on 3D reconstruction such as (Lhuillier and Quan, 2005) or (Strecha et al., 2004) points from the image orientation are used as starting point for dense surface reconstruction. In the former case a bounded regularization approach is employed for surface evolution by level-set methods. The approach is different from ours as it is not focusing on wide-baseline scenarios and it therefore can use a very dense set of points stemming from the orientation. Wide baselines are the scope of the latter approach. As we they use the 3D points as starting points, but they formulate the 3D reconstruction problem in terms of an Bayesian approach and use the EM-algorithm to solve it.

A computationally very efficient approach is presented in (Hirschmüller, 2006). It employs a semi-global matching in the form of dynamic programming in 16 directions. This together with a mutual information based computation of the matching costs results into well regularized results and still a high performance allowing to work with very large images.

Opposed to the above approaches we decided to extend the classical least-squares matching approaches of (Wrobel, 1987, Ebner and Heipke, 1988) in the direction of full 3D reconstruction from wide-baseline image sequences in a similar way as (Schlüter, 1998). We move the vertices of a triangulation resulting from a densification of a triangulation obtained from our initial reliable but sparse 3D points in the direction of their normals. To deal with occlusions, we employ robust estimation. Regularization is based on additional observations modeling the local curvature of the surface.

According to the above four steps of (Scharstein and Szeliski, 2002) we do matching cost computation by squaring brightness differences between transformed values for individual images and an average image. The latter can be considered as an orthophoto of the surface. The costs are aggregated over the whole surface consisting of planar triangles and the computation of disparities or in our case of 3D coordinates is done together with the refinement in the least-squares estimation.

The potential of the least-squares approach lies in its high possible accuracy. Yet, least-squares matching is known to converge to local minima and thus good approximations are necessary. They are obtained here by using as basis sparse but highly precise and reliable points. They stem from a multi-image matching and robust bundle adjustment approach suitable for large baselines (Mayer, 2005) extended by the five point algorithm of (Nistér, 2004). The radius of convergence is additionally expanded by a coarse-to-fine optimization for different levels of resolution for the triangles.

The paper is organized as follows. After sketching basic ideas and giving an overview of the algorithm we detail the ideas in the following sections. Finally we give results and end up with conclusions.

2 BASIC IDEAS AND OVERVIEW OF ALGORITHM

The problem of surface reconstruction is formulated in terms of least-squares adjustment. To be able to work in full 3D, we triangulate the surface and move the vertices of the triangulation along a path independent from the definition of the coordinate system, namely the direction of the normal at the vertex of the triangulated surface. The direction of the normal in the vertex N_u for the unknown number u is estimated as the average of the normal vectors of the planes attached to the vertex. E.g., for Figure 1 this means $N_u = \frac{N_1+N_2+N_3+N_4+N_5}{5}$. It is normalized by $N_u^{norm} = \frac{N_u}{\|N_u\|}$.

The basic ideas of our approach can be summarized as follows:

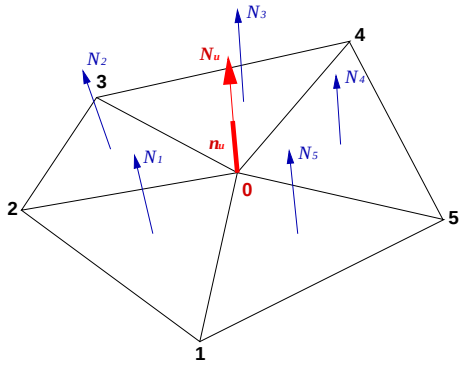

Figure 1: Relation of normal N_u at (center) vertex to normals of neighboring planes and unknown size of movement n_u

- It is based on a triangulated 3D surface.
- The vertices of the triangles move along their respective normals. The sizes of movement are the unknowns n_u (cf. Figure 1).
- Points inside the 3D triangles are projected into the images resulting into the observations.
- The goal of moving the vertices along the normal vectors is to obtain as small a squared gray value difference as possible between the back-projected points in the images and their average value supposed to be the reflectance value of the surface in a least-squares sense.
- Additionally to the image observations representing the data term the surface is regularized by observations enforcing its local smoothness in terms of curvature.
- To deal with outliers, e.g., in the form of local occlusions, robust estimation is used.

The algorithm consists of:

- Creation of triangulated surface from the given sparse 3D points
- Densification of the triangulation at different resolution levels by splitting the triangles of the surface. This results in the unknowns for whom the initial values are interpolated from the neighboring given 3D points.
- By splitting of the triangles of the unknowns and projection of the resulting points into the images the image observations are obtained. The analysis of a local neighborhood of the unknowns leads to additional smoothness observations.
- Robust least-squares adjustment to estimate improved values for the unknowns at the different levels of resolution

Before describing the steps of the algorithm, we detail the contents of the design matrix of the least-squares estimation problem which will be constructed in the course of the algorithm.

3 PARTIAL DERIVATIVES FOR THE DESIGN MATRIX

The image observations are devised to describe how well the intensities in all images showing a 3D point fit to an average intensity computed from all these images by taking the difference between the individual values and the average. Unfortunately, the lighting might be different for the images, the camera might have used a different gain, or the surfaces have a non-Lambertian bidirectional reflection distribution function (BRDF). Therefore, to reduce the bias of the estimation, the overall brightness of the images is estimated at the beginning from a small neighborhood of all given sparse, but reliable 3D points seen in an image.

For the given non-linear problem the design matrix consists of the partial derivatives of the intensity value I_i of observation i in an image according to the change of the size of movement n_u of unknown u. They are given by

$$\begin{aligned}\frac{\partial I_i}{\partial n_u} &= \frac{\partial I}{\partial x}\frac{\partial x}{\partial n_u} + \frac{\partial I}{\partial y}\frac{\partial y}{\partial n_u} \\ &= \frac{\partial I}{\partial x}\left(\frac{\partial x}{\partial X}\frac{\partial X}{\partial n_u} + \frac{\partial x}{\partial Y}\frac{\partial Y}{\partial n_u} + \frac{\partial x}{\partial Z}\frac{\partial Z}{\partial n_u}\right) \\ &\quad + \frac{\partial I}{\partial y}\left(\frac{\partial y}{\partial X}\frac{\partial X}{\partial n_u} + \frac{\partial y}{\partial Y}\frac{\partial Y}{\partial n_u} + \frac{\partial y}{\partial Z}\frac{\partial Z}{\partial n_u}\right)\end{aligned}$$

with

- $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}$ the image gradients in x and y direction which can, e.g., be estimated by the Sobel operator,
- $\frac{\partial x|\partial y}{\partial X|\partial Y|\partial Z}$ describing how the position in x- or y-direction in the image is affected by changing the 3D point coordinates X, Y, or Z corresponding to observation i, and
- $\frac{\partial X|\partial Y|\partial Z}{\partial n_u}$ the derivative of the 3D point coordinates X, Y, or Z according to the size of the unknown movement. The points move in the direction of N_u. The size of their movement depends on the distance of the 3D point from the line connecting the other two vertices of the triangle the point is lying in.

We model the projection of (homogeneous) 3D points $\mathbf{X}$ to image points $\mathbf{x}$ by (Hartley and Zisserman, 2003)

$$\mathbf{x}' = \mathbf{P}\mathbf{X} \tag{1}$$

with the projection matrix $\mathbf{P}$

$$\mathbf{P} = \mathbf{K}\boldsymbol{R}(\boldsymbol{I}| - \boldsymbol{X}_0)$$

describing the collinearity equation consisting of the calibration matrix $\mathbf{K}$ comprising principal point, principal distance, scale difference and sheer as well as translation described by the Euclidean vector $\boldsymbol{X}_0$ and rotation by the matrix $\boldsymbol{R}$.

We additionally employ quadratic and quartic terms to model the radial distortion to obtain an accuracy in the range of one fifth to one tenth of a pixel, but we will not include this issue in the further discussion, to make the paper more readable.

For improved flexibility we work in a relative coordinate system which can be obtained from images alone. The first camera position is used as the origin of the coordinate system. The rotation of the first camera is fixed and is supposed to point to the negative z-direction. The distance of the first and the second camera is set to one.

4 TRIANGULATION OF GIVEN SPARSE 3D POINTS

We assume that the given sparse 3D points stemming from a highly accurate bundle adjustment using possibly many images are precise and reliable. We thus fix their 3D position.

One basic problem for a full 3D approach is the linking of triangles. It is at least difficult, often even impossible to link points in 3D just based on proximity. E.g., consider a thin surface, where points on both sides of the surface should not be linked, but might be much closer than points on the same side of the surface.

To avoid the above problem, we split the images into overlapping triplets. For them we assume that the topology of the 3D points can in essence be modeled in two dimensions (2D) in the images. We therefore can triangulate the points for the triplets in one of the images. To obtain compact triangles, we employ Delaunay triangulation. This reduces problems with elongated thin triangles.

First, we project via equation (1) and the known camera parameters the given 3D points into the central image of the triplet. There they are triangulated. After this, triangulations for different triplets can be stitched together which leads to full 3D triangulations. Yet, for us this is subject of further work. All following steps can now work on this basic global triangulation in as many images as available. The given 3D points are shown as (black) numbers in Figure 2.

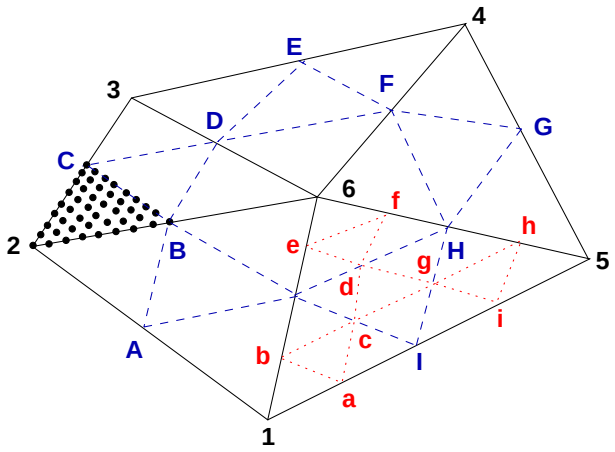

Figure 2: Creation of the vertices of the triangulation corresponding to unknowns and observations – The given 3D points are considered as control points and are marked as (black) numbers. The first level of unknowns are denoted by (blue) capital letters and the second level, which is detailed only for one original triangle, by (red) small letters. Observations are sketched for triangle 2CB (left) as black dots. They are denser because the threshold employed is $\frac{1}{10}$ of the side length for the unknowns.

5 CREATION OF UNKNOWNS AND COARSE-TO-FINE STRATEGY

The vertices of the triangulation corresponding to the unknowns are generated by splitting the sides of the triangles obtained from the given 3D points (cf. preceding Section) at their center if the length of the side is beyond a given threshold. The new unknowns receive their 3D position by linear interpolation. This leads to the first level of unknowns marked by blue capital letters in Figure 2. If the length of the sides should still be above the given threshold, i.e., the triangle is rather big as no 3D point could be found inside it, we split the triangle again along the sides and obtain a second level of unknowns marked by red small letters in Figure 2. This is done recursively until all side lengths are below the threshold.

As a well-known feature of least-squares matching is its rather restricted radius of convergence, we employ a coarse-to-fine strategy. It comprises

- different levels of densification of the triangles by setting the thresholds for the lengths of the sides of the triangles differently and
- use of image resolutions adapted to the sizes of the triangles by selecting a corresponding level of an image pyramid.

6 DETERMINATION OF IMAGE OBSERVATIONS

The coordinates of the 3D points corresponding to the observations are generated similarly as above for the unknowns except that a smaller threshold, namely 10% of the threshold of the unknowns is used. The resulting 3D points are sketched as black dots on the left hand side of Figure 2. The 3D points are projected into all images they can be seen from. The intensity value I_i of observation i at the projected homogeneous image point $\mathbf{x} = \alpha(xy1)^\mathsf{T}$ in an image is given by $I_i = g(x, y)$, with g the bilinear interpolation function.

For the design matrix A an unknown is affected only by the observations belonging to neighboring triangles. This leads to a sparse design matrix. We employ this by only computing those parts belonging to the actual observations, i.e., which are non-zero.

Yet, it also means that only unknowns in the normal equations are correlated which have common triangles. To obtain a banded normal equation matrix, for which efficient solutions are available, with a band-width as small as possible, we traverse the triangles along the shorter side of the given area for 3D reconstruction. For the example in Figure 3 this leads to a banded normal equation matrix sketched in Figure 4. One can regard the first unknowns to belong to the triangles marked in red in the lower left corner of the triangulation in Figure 3, the next unknowns to the triangles marked in green right of it, the next the blue, etc. All vertices of the triangles, i.e., the unknowns, are linked only to two layers of triangles which leads to a normal equation matrix with just one band parallel to the main diagonal. The width of the band depends on the length of the layer. Thus, it is useful to traverse the triangulation along the shorter side.

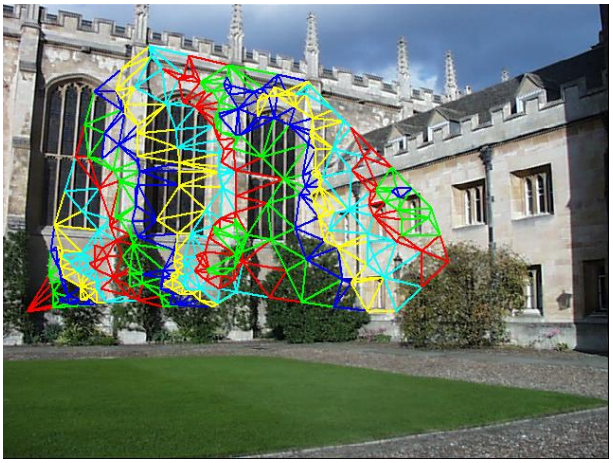

Figure 3: Traversal of triangles along the shorter side. The different colors correspond to different layers of the traversal. The traversal starts in the lower left corner (red triangles) – image Trinity from web-page Criminisi and Torr

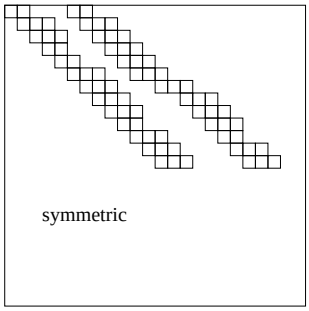

Figure 4: Banded structure of normal equation matrix resulting from traversal of triangulation along the shorter side in Figure 3.

7 REGULARIZATION BY SMOOTHNESS OBSERVATIONS

Due to noise and occlusions 3D surface reconstruction is an ill-posed problems which has to be regularized. One way to accomplish this is via additional observations enforcing the smoothness of the surface. Their influence is controlled via the ratio between the weights for the image and the smoothness observations.

We describe smoothness in terms of the deviation $h_{change}$of a vertex from an average plane derived from the neighboring vertices in the direction of its normal N.

The average plane is computed as weighted average of the heights of the vertices h_i above the plane through the given vertex and perpendicular to the normal N at the given vertex. For this the vertices are projected along the normal N, resulting in the primed (blue) numbers in Figure 5. The weighting is done according to the inverse distance d_i of the points. The average height of the vertices is at the same time the height of the given vertex at height 0 above or below this plane:

$$h_{smooth} = \sum_{i=1}^{n} \frac{h_i}{d_i} / \sum_{i=1}^{n} \frac{1}{d_i}$$

h_{smooth} is combined with the average inverse distance to an observation describing the curvature at the vertex:

$$l_{smooth} = h_{smooth} * \frac{\sum_{i=1}^{n} \frac{1}{d_i}}{n}$$

8 ROBUST LEAST-SQUARES ADJUSTMENT

To solve the least-squares adjustment for the unknowns x, we must factorize the normal equation matrix $A^{\mathsf{T}}PA$, with the design matrix A and the weight matrix P. As there might be thousands or even tens of thousands of unknowns, the factorization of the matrix requires special attention. We basically employ that as detailed above in Section 6 we obtain a (symmetric positive definite) band matrix. We then use a Cholesky factorization for banded symmetric matrices and solve for x.

To stabilize the solution, we employ the Levenberg-Marquardt algorithm. I.e., we multiply the elements on the main diagonal with a factor ranging from 1.0001 to 1.1 and take the result with the smallest average standard deviation σ_0. The non-linear optimization is done a couple of times until the ratio of the σ_0 between two iterations falls below 1.01.

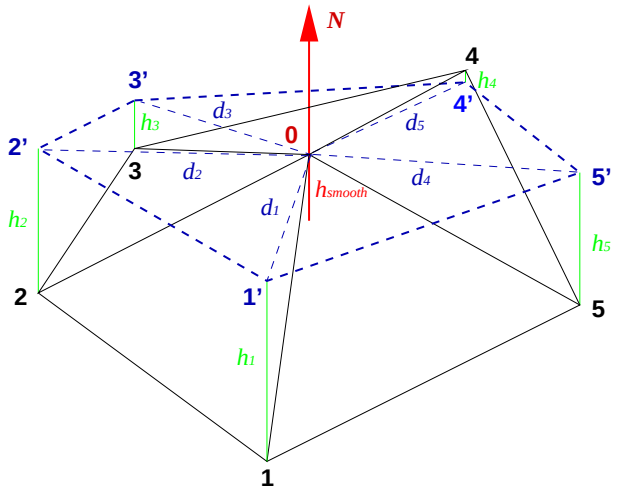

Figure 5: Smoothing – The (black) numbers denote the original vertices. The (blue) primed numbers show their projection on the plane perpendicular to the normal N through the given point 0 with height h_i. h_{smooth} corresponds to the height of the given vertex above or below the (weighted) average plane.

As weight matrix P we use a diagonal matrix. It is normalized to unity before being multiplied with the design matrix or the vector of the observations l. Initially all weights are set to one besides a scaling factor weighing image and smoothness observations against each other as explained in Section 7.

Because there might be bad or wrong matches due to occlusions or non-Lambertian behavior of the surface, robust estimation is used. We particularly base robust estimation on standardized residuals $\bar{v}_i = v_i/\sigma_{v_i}$ involving the standard deviations σ_{v_i} of the residuals, i.e., the differences between observed and predicted values. As the computation for the individual observation is computationally costly, we substitute it by an estimate of the average standard deviation of the gray value, particularly 3 gray values. We then do reweighting of the elements of P with $w_i = 1/\sqrt{2 + \bar{v}_i{}^2}$ (McGlone et al., 2004).

9 RESULTS

In this section we report about initial results. In all cases we compare the initial triangulation on the first level consisting of the reliable but sparse points from the orientation with the final densified result to show the improvement obtained by our approach. The output is done in VRML – virtual reality modeling language format.

Figures 6 and 8 give results for two scenes derived from image triplets from the web-page of Antonio Criminisi and Phil Torr while in Figure 7 we present a 3D reconstruction for an image triplet showing a part of the Zwinger in Dresden.

The final results in Figures 6 and 7 demonstrate that the densification of the triangulation leads to a better reconstruction of the details of the scene. This can be even better appreciated in Figures 8 and particularly 9, where one can see that, e.g., for the front part of the baguette or the left rim of the red basket on the right hand side, the densified triangulation is accompanied with improved normals.

10 CONCLUSIONS

We have shown an extension of the classical least-squares matching approaches of (Wrobel, 1987, Ebner and Heipke, 1988) which

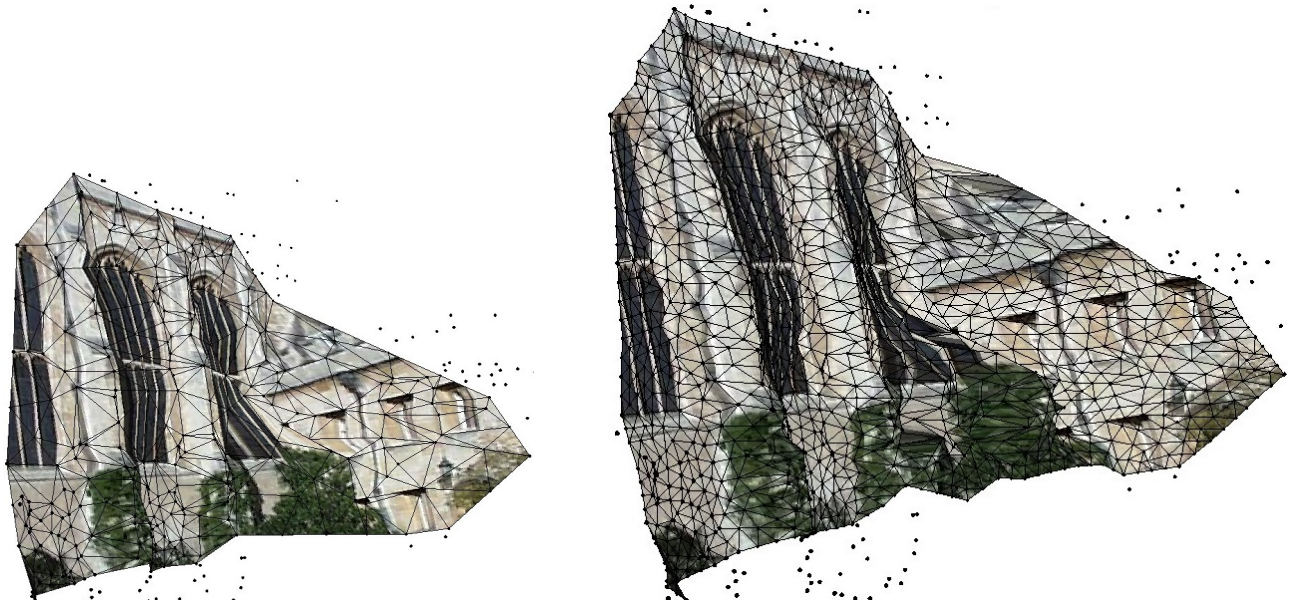

Figure 6: Result for triplet Trinity (from web-page Criminisi and Torr) at first level of resolution (left) and on third level after optimization (right). Please particularly look at the drainpipe on the right facade.

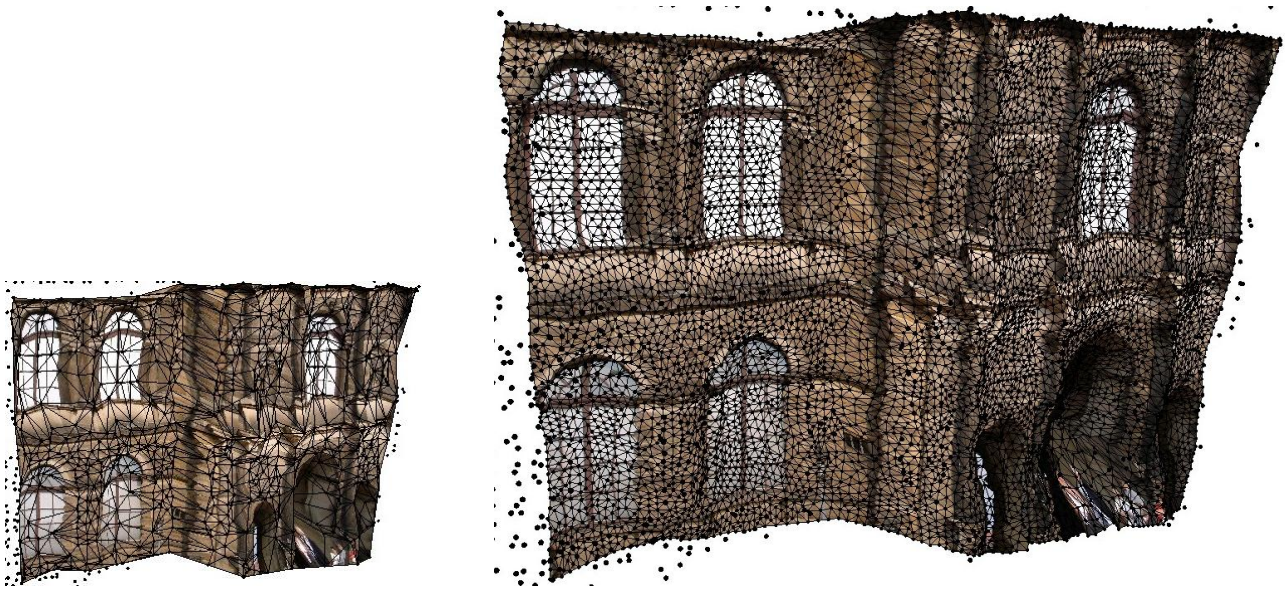

Figure 7: Result for triplet Zwinger at first level of resolution (left) and on third level after optimization (right)

were confined to 2.5D surfaces to 3D by employing the normals of a triangulation similarly as (Schlüter, 1998). Opposed to the latter, our approach is focusing on wide-baseline settings and we employ robust estimation to deal with occlusions.

First results show the potential but also the shortcomings of the approach. We still need to extend it by linking the triangulation of image triplets into triangulations for larger number of images and many parts of the algorithm need to be refined. We also consider to move the vertices towards edges in the image, as the latter tend to give hints on break-lines of the surface, though we note that this problem is mitigated as the initial points are at corners by definition. Finally, we want to check the fit in terms of least-squares error of each triangle before subdividing them to avoid small triangles in homogeneous areas.

Very recently, (Pons et al., 2007) have presented an approach with similarities to ours, though they link surface reconstruction with scene flow estimation over time. They employ graphics hardware to speed up processing. This idea could also help to speed up our algorithm as the determination of the observations entails large numbers of projections from 3D space into the images which could very well be solved by graphics hardware.

ACKNOWLEDGMENTS

We want to thank the republic of Vietnam for supporting Duc Ton by a PhD scholarship

REFERENCES

Ebner, H. and Heipke, C., 1988. Integration of Digital Image Matching and Object Surface Reconstruction. In: International Archives of Photogrammetry and Remote Sensing, Vol. (27) B11/III, pp. 534–545.

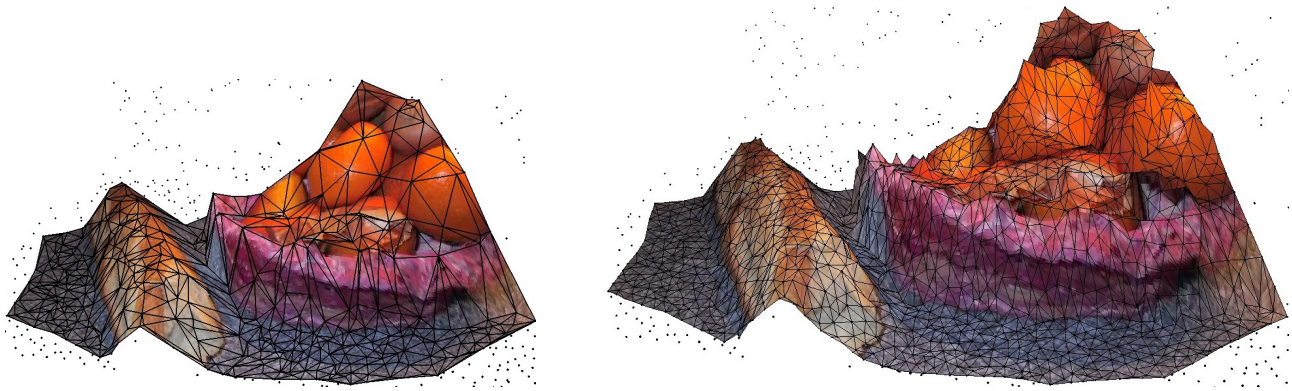

Figure 8: Result for triplet kitchen (from web-page Criminisi and Torr) at first level of resolution (left) and on third level after optimization (right)

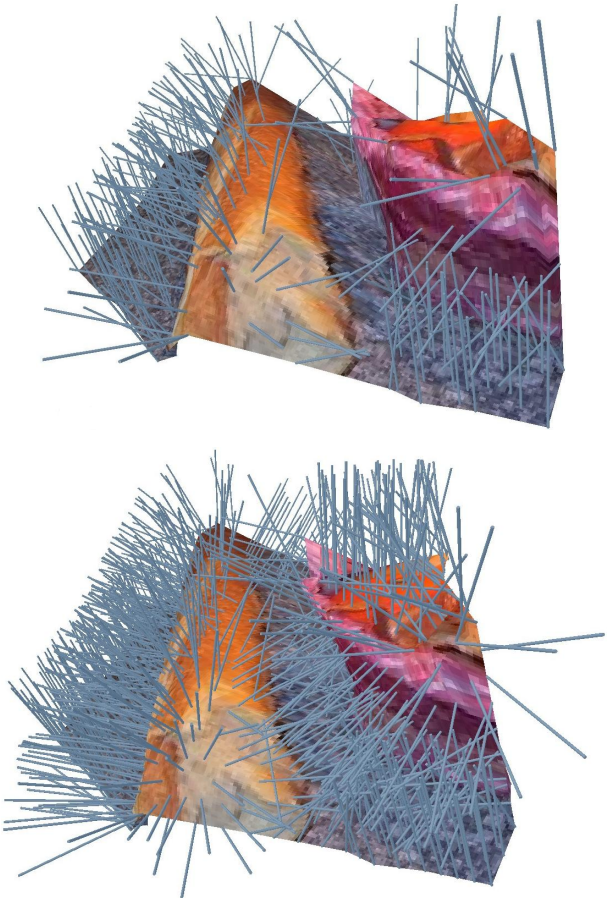

Figure 9: Detail of result for triplet kitchen (cf. Fig. 8) with normal vectors

Fua, P. and Leclerc, Y., 1996. Taking Advantage of Image-Based and Geometry-Based Constraints to Recover 3–D Surfaces. Computer Vision and Image Understanding 64(1), pp. 111–127.

Hartley, R. and Zisserman, A., 2003. Multiple View Geometry in Computer Vision – Second Edition. Cambridge University Press, Cambridge, UK.

Hirschmüller, H., 2006. Stereo Vision in Structued Environments by Consistent Semi-Global Matching. In: Computer Vision and Pattern Recognition, Vol. 2, pp. 2386–2393.

Lhuillier, M. and Quan, L., 2005. A Qasi-Dense Approach to Surface Reconstruction from Uncalibrated Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(3), pp. 418–433.

Mayer, H., 2005. Robust Least-Squares Adjustment Based Orientation and Auto-Calibration of Wide-Baseline Image Sequences. In: ISPRS Workshop in conjunction with ICCV 2005 "Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images" (BenCos), Beijing, China, pp. 1–6.

McGlone, J., Bethel, J. and Mikhail, E. (eds), 2004. Manual of Photogrammetry. American Society of Photogrammetry and Remote Sensing, Bethesda, USA.

Nistér, D., 2004. An Efficient Solution to the Five-Point Relative Pose Problem. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(6), pp. 756–770.

Pons, J.-P., Keriven, R. and Faugeras, O., 2007. Multi-View Reconstruction and Scene Flow Estimation with a Global Image-Based Matching Score. International Journal of Computer Vision 72(2), pp. 179–193.

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

Schlüter, M., 1998. Multi-Image Matching in Object Space on the Basis of a General 3-D Surface Model Instead of Common 2.5-D Surface Models and its Application for Urban Scenes. In: International Archives of Photogrammetry and Remote Sensing, Vol. (32) 4/1, pp. 545–552.

Strecha, C., Fransen, R. and Van Gool, L., 2004. Wide-Baseline Stereo from Multiple Views: A Probabilistic Account. In: Computer Vision and Pattern Recognition, pp. 552–559.

Wrobel, B., 1987. Digital Image Matching by Facets Using Object Space Models. In: Symposium on Optical and Optoelectronic Applied Science and Engineering – Advances in Image Processing, Vol. 804, SPIE, pp. 325–333.