17th International Conference on the Application of Computer
Science and Mathematics in Architecture and Civil Engineering
K. Gürlebeck and C. Könke (eds.)
Weimar, Germany, 12–14 July 2006

PRECISE 3D MEASUREMENT WITH STANDARD MEANS AND MINIMAL USER INTERACTION — EXTENDED SINGLE-VIEW RECONSTRUCTION

M. Trummer*, J. Denzler, and H. Süße

**Chair for Computer Vision,*
Department of Computer Science,
Friedrich-Schiller-University Jena
D-07737 Jena, Germany
(www.inf-cv.uni-jena.de)
E-mail: trummer@informatik.uni-jena.de

Keywords: 3D measurement, 3D reconstruction.

Abstract. *The paper outlines a new method for 3D measurement explicitly targeting at practical applications. In particular, in order to realize this intention, robust mathematical techniques, the usage of standard or even low resolution cameras and minimal user interaction have to be claimed. The extended single-view reconstruction satisfies these claims by the use of special knowledge and a clear problem separation. The algorithm expects a planar quadrangle with known geometry inside the scene as well as the intrinsic and radial distortion camera parameters to be noted. From two or more digital images of the scene, the algorithm in general returns the 3D coordinates of the world points (up to an unknown Euclidean transformation), whose mapped points have been selected. With the marked image mappings of two world points, the spatial distance between these points can simply be calculated from the reconstructions. With more than two photographs error compensation is achieved (errors due to noise and possibly imprecise inputs).*

1 INTRODUCTION AND PROBLEM STATEMENT

The task of 3D reconstruction from digital images is, in theory, well understood, and many procedures, e.g. self-calibration and stratisfied reconstruction (see [1]), have been proposed. Nonetheless, attempts at practical applications suffer from the necessities of generality, simplicity **and** accuracy.

Here, *generality* means, the application can be used within any scene of a defined field of application. So, for general 3D reconstruction/measurement the stratisfied approach does not fit very well, since the step from projective towards affine reconstruction needs to determine the plane at infinity (or equivalent), and this is quiet a hard claim. One compatible field of application for stratisfied reconstruction is the reconstruction of adequate architectural scenes, where vanishing points can be estimated from imaged parallel lines in orthogonal directions and, thus, the plane at infinity. Other applications are based on the computation of nothing more than homographies (cf. [2]). Therefore, the generality of these methods is apparently restricted in sense of considering only a 2D subspace (plane) in 3D space, and being limited to this subspace.

When looking at *simplicity*, firstly, the term of simplicity shall be understood as a low amount of all technical, financial and procedural expenses that have to be carried out in order to apply a certain method and yield a result. Then, the reciprocal influence of the mentioned necessities becomes obvious. Homography-based applications such as the one described in [2] can be easy to perform: If you want to survey a street segment, just take a picture of it with four noted points and go ahead. But, problems become clear when considering the well-known fact that a plane (or the homography between real and image plane) to be estimated should be photographed from close to above the plane (i.e. looking at the plane in a direction parallel to the plane normal). So, preferred vantage points are on bridges, large ladders or in helicopters above the street. Along the way, real street surfaces are never perfectly plane (but this belongs to ”generality”). In order to tame errors arising from these facts, more than four point correspondencies can be used to give a least-squares estimation. It requires an accordant amount of user interaction (procedural expenses) for marking, measuring the points on the street and selecting within the image. If you decide to use a self-calibration approach for 3D measurement/reconstruction, then the whole opposite of simplicity is at hand. The application in [3] needs at least 20 point correspondencies in at least four images. This gives at least 80 points to select within the images, but recommended are $40 * 6 = 240!$ This seems to be far from applicable for many purposes.

The number of marked correspondencies can heavily affect the *accuracy* of the computed result. In addition, the quality of the result of course depends on the computational model and the numerical procedures, since they influence the stability with respect to noisy input.

Revising the mentioned applications, a better method for 3D measurement should meet the demands of generality, simplicity and accuracy at the same time. Since Euclidean measurement from photographs always claims a known distance to be imaged, it is likely to place a simple reference object within the scene and, hence, affirming generality without constraining the other demands. Furthermore, the necessary user interaction (considering in-place interaction and postprocessing with computer) should be minimal, and the errors of results should stay inside known (and tight) error boundaries.

As an attempt to reach these goals, this paper presents the extended single-view reconstruction for 3D point reconstruction and 3D measurement (by computing the distance between reconstructed points). The following section shows the theoretical background. In section 3 an application example lights up practical issues. Test results are placed in section 4. The paper concludes by summarizing important features of the procedure in section 5.

2 EXTENDED SINGLE-VIEW RECONSTRUCTION

This section shows the theory of the extended single-view reconstruction. It basicly performs a partial 3D reconstruction of the scene, using a plane quadrangle with known geometry and the camera's internal and radial distortion parameters as inputs. Namely, the corner points of the reference quadrangle and the 3D points defining the spatial distance are reconstructed from two or more digital images.

The procedure is characterized by clear and semantical problem separation, i.e. things belonging together are kept together. For understanding, an opposite example would be the splitted computation of camera orientation in [6]. Here, all entities, in particular a camera's external parameters, are calculated within one single and direct step without any nonlinear optimization. The partial results are optimal in terms of least squares.

Initially, each image is considered separately. Making use of the mentioned special knowledge allows to get an Euclidean reconstruction of the corner points in 3D space from their image points. Hence, the position of the plane reference quadrangle is yielded in 3D space, but with respect to each camera center. This is the *single-view reconstruction* from [4].
The *extension* concerns the usage of all available (but at least two) pictures of the scene. So, the position of the reference points (corners of the reference quadrangle) in 3D space is given per picture, but in different coordinate frames, that is with the respective camera centre as origin. Since the single-view reconstruction is Euclidean, it is known that each two 3D reconstructions of the quadrangle (and, hence, each two coordinate frames) can be equated by pure 3D translation and rotation, assuming no errors or noise. By the determination of these transformations with the method from [5], the extrinsic parameters of each camera in relation to the first one are achieved.
Now, with complete camera matrices known, triangulation (see [1]) is possible.

2.1 Single-View Reconstruction

Only one image is considered. Let $\mathbf{y^{(i)}} = (y_1^{(i)}, y_2^{(i)})^T$ (with $i = 1, ..., 4$) be the known 2D CAD coordinates of the corner points (derived from quadrangle geometry) of the reference quadrangle and $\mathbf{x^{(i)}} = (x_1^{(i)}, x_2^{(i)})^T$ the corresponding image coordinates. Then $\mathbf{X^{(i)}} = (X_1^{(i)}, X_2^{(i)}, X_3^{(i)})^T$ are the related points in 3D space, and $\tilde{\mathbf{y}}/\tilde{\mathbf{x}}/\tilde{\mathbf{X}}$ shows the notation in homogeneous coordinates, $\tilde{\mathbf{y}}^{\mathbf{(i)}} \sim (y_1^{(i)}, y_2^{(i)}, 1)^T$ etc. The writing $\sim$ shows equity up to an unknown factor. Vectors are always column vectors here.

During this first step, the 3D coordinates of the points $\mathbf{X^{(i)}}$ are calculated by considering two transformations between different representations of the reference points.

It is noted that the points $\mathbf{X^{(i)}}$ all lie on a plane in 3D space, that is the rotated and translated CAD plane. Therefore, the representation

$$\mathbf{X^{(i)}} = \mathbf{d} + y_1^{(i)}\mathbf{e^{(1)}} + y_2^{(i)}\mathbf{e^{(2)}} \qquad (1)$$

is suitable, where $\mathbf{d}$ is the origin of the CAD plane rotated and translated in 3D space, and $\mathbf{e^{(1)}}$, $\mathbf{e^{(2)}}$ the normalized vectors of the coordinate frame's axes. The points $\mathbf{x^{(i)}}$ are the image points of the $\mathbf{X}^{(\mathbf{i})}$ (less the radial distortion), therefore

$$\tilde{\mathbf{x}}^{(\mathbf{i})} \sim P\tilde{\mathbf{X}}^{(\mathbf{i})} \tag{2}$$

with the camera transformation matrix P. But, since world and camera coordinate frame are assumed equal, the camera mapping simplifies to

$$\tilde{\mathbf{x}}^{(\mathbf{i})} \sim K\tilde{\mathbf{X}}^{(\mathbf{i})} \tag{3}$$

with the upper triangular matrix K containing the camera's intrinsic parameters,

$$K = \begin{pmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

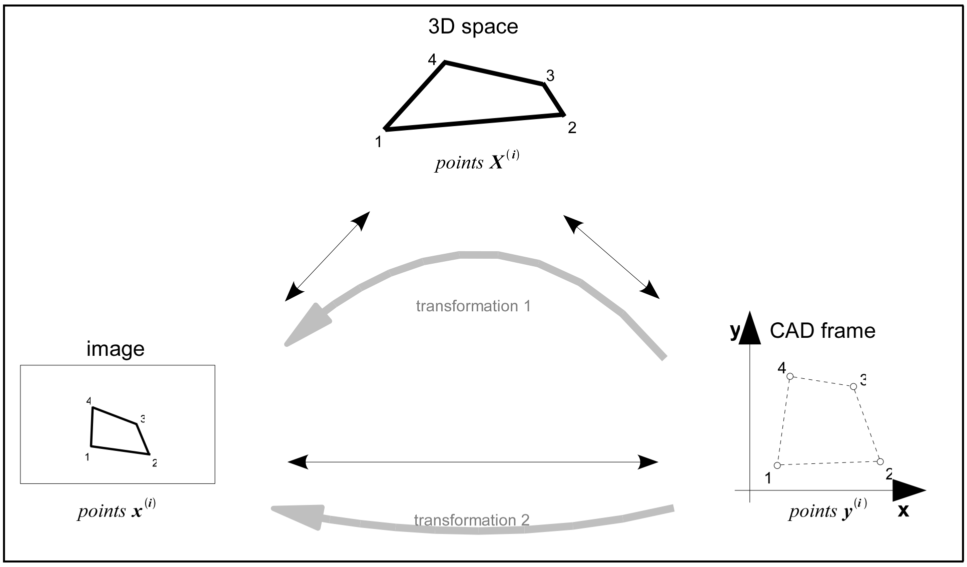

Fig. 1: Considered transformations within the single-view reconstruction.

Now, looking at equation (3) and taking (1) into account, one can easily derive

$$\tilde{\mathbf{x}}^{(\mathbf{i})} \sim K\mathbf{d} + Ky_1^{(i)}\mathbf{e}^{(\mathbf{1})} + Ky_2^{(i)}\mathbf{e}^{(\mathbf{2})}. \tag{5}$$

This is transformation 1 from fig. 1.

On the other hand, there is a homography H mapping CAD coordinates to image coordinates, hence

$$\tilde{\mathbf{x}}^{(\mathbf{i})} \sim H\tilde{\mathbf{y}}^{(\mathbf{i})}. \tag{6}$$

This homography is transformation 2 from fig. 1. Now, let H be written as $H = (\mathbf{h_1}, \mathbf{h_2}, \mathbf{h_3})$, then it is easily checked that

$$\tilde{\mathbf{x}}^{(\mathbf{i})} \sim \mathbf{h_1}y_1^{(i)} + \mathbf{h_2}y_2^{(i)} + \mathbf{h_3} \tag{7}$$

applies. Within equations (5) and (7) we compare coefficients related to $y_j^{(i)}$ $(j = 1, 2)$ and get

$$\begin{pmatrix} K\mathbf{e}^{(\mathbf{1})} \\ K\mathbf{e}^{(\mathbf{2})} \\ K\mathbf{d} \end{pmatrix} \sim \begin{pmatrix} \mathbf{h_1} \\ \mathbf{h_2} \\ \mathbf{h_3} \end{pmatrix}. \tag{8}$$

So, there are three linear equation systems that are very simple to solve, since K is an upper triangular matrix. Because of the basic projective relations, the equity only applies up to a common factor. But this one is easy to resolve for, making use of the claims $|\mathbf{e^{(1)}}| = 1$ or $|\mathbf{e^{(2)}}| = 1$ or, better in practice, $|\mathbf{e^{(1)}}||\mathbf{e^{(2)}}| = 1$.

Finally, the vectors $\mathbf{d}$, $\mathbf{e^{(1)}}$ and $\mathbf{e^{(2)}}$ are calculated, and with help of equation (1) the 3D coordinates of the points $\mathbf{X^{(i)}}$ are yielded. This solution is achieved by not more than solving four simple linear equation systems.

2.2 An Optimal Estimation of a 3D Euclidean Transformation from Point Correspondencies

Within this step, the information of all available n $(n > 1)$ images is merged. Up to now, we got n different 3D reconstructions of the reference quadrangle in independent coordinate frames with points ${}^{\mathbf{(j)}}\mathbf{X^{(i)}}$ (point i, $i = 1, ..., 4$, of the reconstruction from image j, $j = 1, ..., n$). But, since the single-view reconstruction is Euclidean, these n reconstructions only differ by 3D translation and rotation, i.e. a 3D Euclidean Transformation. To calculate this transformation, at least three 3D point correspondencies are needed. But in this case, four such correspondencies are given that are affected by noise as well as every data in practical application. Hence, for images k and l and the corresponding reconstructions of the reference quadrangle, the optimal rotation matrix $R_{l,k}$ and translation vector $t_{l,k}$ are to be computed. This is the solution for the least squares problem

$$\sum_{i=1}^{4} |{}^{\mathbf{(k)}}\mathbf{X^{(i)}} - (R_{l,k}{}^{\mathbf{(l)}}\mathbf{X^{(i)}} + t_{l,k})|^2 \rightarrow \text{minimum}, R \text{ rotation matrix.} \tag{9}$$

With respect to the nonlinear restrictions of the rotation matrix R this is a hard problem, and the straightforward attempt to solve it would end up in a bulky nonlinear optimization problem. But this one can be evaded by using the approach of Walker, Shao and Volz [5] for solving the LSE in (9). When the 3D Euclidean transformation is represented by a dual union quaternion, the least squares problem in (9) can be put down to an eigenvalue problem for a matrix built up from the point correspondencies. Following this way, the homogeneous transformation matrix

$$[\tilde{\mathbf{RT}}]_{\mathbf{(l,k)}} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{10}$$

is yielded as least squares solution by not more than solving an eigenvalue problem. And if this transformation is now applied to the former origins (the camera centres) of the different coordinate frames, the cameras can be located in one single coordinate frame, and therefore the extrinsic parameters for each camera (e.g. with respect to the first one) are known.

2.3 3D Reconstruction

Let now the world coordinate frame match the one of the first camera. Then for other cameras j $(1 < j \leq n)$ the extrinsic paramters with respect to the first can be determined using the method mentioned in the last paragraph. So we get the homogeneous transformation matrix $[\tilde{\mathbf{RT}}]_{\mathbf{(j,1)}}$ for camera j. With

$$\tilde{K} = \begin{pmatrix} \alpha_x & \gamma & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{11}$$

and

$$ {}^{(j)}P = \tilde{K}[\widetilde{\mathbf{RT}}]_{(\mathbf{j},\mathbf{1})} \tag{12} $$

the complete camera matrix of each camera is known.

For image points ${}^{(\mathbf{j})}\tilde{\mathbf{x}}$ from camera j and there preimages $\mathbf{X}$ applies the mapping

$$ {}^{(\mathbf{j})}\tilde{\mathbf{x}} \sim {}^{(j)}P\tilde{\mathbf{X}}, \tag{13} $$

hence

$$ {}^{(\mathbf{j})}\tilde{\mathbf{x}} \times ({}^{(j)}P\tilde{\mathbf{X}}) = 0. \tag{14} $$

This vector product gives three equations that are linear in the unknown homogeneous coordinates $\tilde{\mathbf{X}}$. But, since each image point has only two coordinates, there can be not more than two linear independent equations in (14) that are useful for triangulation. Let the projection matrix of camera j now be written as

$$ {}^{(j)}P = \begin{pmatrix} {}^{(\mathbf{j})}\mathbf{p_1^T} \\ {}^{(\mathbf{j})}\mathbf{p_2^T} \\ {}^{(\mathbf{j})}\mathbf{p_3^T} \end{pmatrix}, \tag{15} $$

then two equations from (14) are chosen, and the homogeneous linear equation system

$$ A\tilde{\mathbf{X}} = 0 \tag{16} $$

is yielded with

$$ A = \begin{pmatrix} {}^{(\mathbf{1})}\mathbf{p_1^T} - {}^{(1)}\tilde{x}_1{}^{(\mathbf{1})}\mathbf{p_3^T} \\ {}^{(1)}\tilde{x}_2{}^{(\mathbf{1})}\mathbf{p_3^T} - {}^{(\mathbf{1})}\mathbf{p_2^T} \\ \vdots \\ {}^{(\mathbf{n})}\mathbf{p_1^T} - {}^{(n)}\tilde{x}_1{}^{(\mathbf{n})}\mathbf{p_3^T} \\ {}^{(n)}\tilde{x}_2{}^{(\mathbf{n})}\mathbf{p_3^T} - {}^{(\mathbf{n})}\mathbf{p_2^T} \end{pmatrix}. \tag{17} $$

Desired is the optimal solution $\tilde{\mathbf{X}} \neq (0,0,0,0)^T$ with respect to least squares. This is achieved by singular value decompostion of A,

$$ A = UDV^T, \tag{18} $$

and the algebraically optimal $\tilde{\mathbf{X}}$ is the column of V corresponding to the smallest singular value.

So, within this step an optimal solution for the 3D coordinates of $\mathbf{X}$ based on the information from all available images is achieved. From the reconstructed 3D points the Euclidean distance can be computed easily.

3 APPLICATION

An application example paying attention to user interaction (cf. simplicity) and the field of applications (cf. generality) is given in this section.

The necessary equipment for applying the proposed procedure is a plane reference quadrangle with known geometry and a standard handheld digital camera, that is calibrated with respect to internal and radial distortion parameters. The reference quadrangle can be, as seen in fig. 2, a simple self-made one, or it can already be part of the scene. So, whenever one of these possibilities applies, the procedure can be used for 3D reconstruction/measurement. This is the constraint for generality.

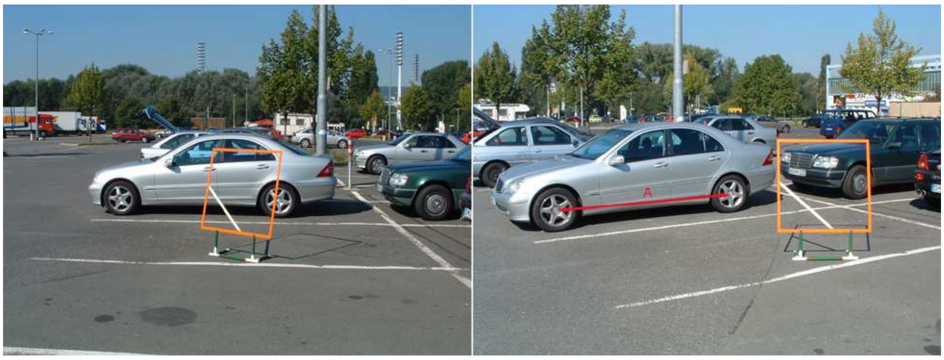

Fig. 2: Application example, measuring the wheel base of the car (A).

User interaction is inevitable, in-place as well as during postprocessing with a computer. But here, in-place user interaction is nothing more than placing the reference quadrangle within the scene (when not using a ”natural” quadrangle) and taking two or more pictures from different vantage points. Thus, it seems to be close to minimal. The postprocessing step comprises the selection of image points, namely the corner points of the reference quadrangle and the points to reconstruct (fig. 2: wheel centers). For a reference quadrangle being placed by hand, an automatic detection of the corner points with sub-pixel accuracy is intended, but not realized yet. In terms of simplicity, there should be as few as possible reference points. But, using a triangle as reference figure would lead to a system of nonlinear equations when doing the monocular reconstruction (see [4]). These could be solved, but not as stable and robust as performed here.

The effects of using only direct and well manageable mathematical methods are described within the next section.

4 TEST RESULTS

This section is dedicated to the accuracy of the extended single-view reconstruction. It gives mainly qualitative information about error values and simple rules for keeping the error small. Nonetheless, comprehensive test series with real image data and noisy input have been accomplished, and for more numbers the interested reader may be referred to [7].

Since the marking of all relevant image points is handcraft up to now, every measurement has been executed at least five times in order to get error values with some significance. All tests have been conducted with the reference square from fig. 2 (side length 1m) and a Fuji Finepix S304 (3 megapixels).

To start with the example from fig. 2, the distance between camera and target object was about 8m, between camera and reference object about 5m. The datasheet of the manufacturer states the wheele base as 2.715m, and the result of the survey from the two images was 2.728m. This gives a relative error of 0.5%.

This relative error value is convenient for most of the tested scenes, when keeping in mind some simple constraints that all arise from well known photogrammetric facts. At first, more resolution, of course, leads to smaller errors. Secondly, the vantage points, from which pictures are taken, should be away from each other. The tests showed that the angle between the optical axes of the cameras should be larger than 30° to give stable results. Thirdly, the angle between the camera’s optical axes and the plane of the reference object should be large enough to keep the discretization error small. Again, the tests stated angles above 30° as noncritical. Paying attention to these rules, the relative measurement error could be kept smaller than 1%. This result also applies to pictures from VGA cameras.

Conclusively, it might be interesting that an attached nonlinear optimization (minimizing backprojection error with bundle-adjustment by the method of Levenberg-Marquardt), using the yielded result as initial solution, did not improve the result in general. This optimization showed the typical unpredictable behaviour, possibly due to improper error modeling (backprojection error with constraints on reconstructed reference points) or noise.

5 CONCLUSION – BENEFIT AND BOUNDARIES

This paper presents the extended single-view reconstruction, a method for general 3D point reconstruction from two or more digital images with the help of a plane quadrangle as reference object. 3D measurement, thus, is realized by reconstructing the two 3D points defining the spatial distance. The required inputs are the geometry of the quadrangle as well as the camera's internal and radial distortion parameters.

The procedure shows a clear separation (monocular reconstruction of reference quadrangle – estimating 3D Euclidean transformation – triangulation), while the according partial solutions are optimal in terms of least squares. Furthermore, these partial solutions are computed by simple, direct and well manageable mathematical methods (solving linear equation systems, Eigenvalue problem, singular value decomposition). Thus, the overall result is optimal and stable. The relative measurement error is between 0.5% an 1%.

All kinds of expenses are kept low, especially the necessary user interaction is minimized.

Since the procedure is a photogrammetric application, the typical limits concerning image resolution, object distance and observation angle apply. Furthermore, the presence of a plane quadrangle with known geometry is crucial.

REFERENCES

[1] R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision*, Second Edition, Cambridge University Press, Cambridge, 2002

[2] *DAKO-MESS*, Measurement System for Traffic Crash Scenes, www.dako.de

[3] *V-STARS System*, Picture Measurement, NTI-MEASURE, www.nti-measure.com

[4] K. Voss, R. Neubauer, M. Schubert: *Monokulare Rekonstruktion für Robotvision*, Verlag Shaker, Aachen, 1995, p. 76

[5] M. W. Walker, L. Shao, R. A. Volz: *Estimating 3D location parameters using dual number quaternions*, Computer Vision, Graphics and Image Processing, vol. 54, 1991, p. 358–367

[6] Z. Zhang: *A flexible new technique for camera calibration*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, 2000, p. 1330–1334

[7] M. Trummer: *Metrische 3-D-Vermessung einer Straßenszene mit Spezialwissen unter minimaler Interaktion – Erweiterte monokulare Rekonstruktion*, diploma thesis, Chair for Computer Vision, Friedrich-Schiller-University Jena, 2005