Photogrammetrie • Fernerkundung • Geoinformation 6/2007, S. 405–414, 6 figs.

ARTIST: Architectural Model Refinement Using Terrestrial Image Sequences from a Trifocal Sensor

MATTHIAS HEINRICHS, OLAF HELLWICH & VOLKER RODEHORST, Berlin

Keywords: Close-range Photogrammetry, Architectural Model Refinement, Trifocal Video Sensor, Trinocular Rectification, Spatial Stereo, Semi-global Matching, Temporal Tracking

Summary: This paper proposes a high resolution video sensor for the 3D reconstruction of architectural models from multiple image sequences. The hybrid system unifies triangulation methods of spatial stereo with tracking methods of temporal stereo. We describe an efficient spatial image matching algorithm, which is based on trinocular image rectification and semi-global optimization. The motion of the video sensor is estimated using temporal feature tracking and allows the integration of dense point clouds. First experimental results are shown for images of a real scene.

Zusammenfassung: *Verfeinerung von Gebäudemodellen mittels terrestrischer Videosequenzen eines Trifokalsensors*. In diesem Beitrag wird ein hochauflösender Videosensor für die 3D Rekonstruktion von Architekturmodellen aus mehreren Bildfolgen vorgestellt. Das hybride System vereinheitlicht Triangulationsmethoden des räumlichen Stereos mit Verfolgungsmethoden des zeitlichen Stereos. Wir beschreiben ein effizientes Bildzuordnungsverfahren, das auf einer trinokularen Bildkorrektur und semi-globaler Optimierung basiert. Die Bewegung des Videosensors wird durch zeitliche Merkmalsverfolgung geschätzt und erlaubt die Integration dichter Punktwolken. Erste experimentelle Ergebnisse für Bilder einer realen Szene werden vorgestellt.

1 Introduction

Some interesting applications of urban 3D geographic information systems (GIS) require a level of detail (LOD), which is currently not available using airborne data. Therefore, an approach to acquire and/or refine architecture models from terrestrial image sequences is proposed. We develop a fully automated prototype system to recover 3D models of several buildings based on three moving video cameras (trifocal sensor). In general, digital video cameras provide dense image sequences that contain a high potential for photogrammetric application which is presently not fully used. Video sequences for scene modeling and various possible applications are treated by (AKBARZADEH et al. 2006, POLLEFEYS et al. 2004, KOCH 2003). When dense video sequences are used for object reconstruction the correspondence problem does not have to be solved by wide-baseline matching any longer but tracking and motion estimation methods such as affine feature tracking (SHI & TOMASI 1994), visual odometry (NISTÉR 2006), simultaneous localization and mapping SLAM (MONTEMERLO 2003) and optical flow estimation gain importance.

Reconstructing a three-dimensional model from a single video sequence is often conducted with the structure-from-motion SFM approach. In close-range photogrammetry systems have been mounted on vans in order to acquire GIS data semi-automatically (TAO 2000). First attempts on hybrid

1432-8364/07/2007/0405 $ 2.50

algorithms unifying triangulation methods of spatial stereo with tracking methods of temporal stereo are presented by (Neumann & Aloimonos 2002). They propose multi-resolution subdivision surfaces for spatio-temporal stereo. Algebraic projective geometry (Hartley & Zisserman 2003, Faugeras & Luong 2001) provides an effective mathematical framework to obtain geometrically precise information from partially calibrated cameras with varying parameters.

The main novelty of our approach is an exhaustive integration of feature extraction, image matching, orientation for video sequences, as well as modeling of surfaces with their reflectance characteristics. We combine calibrated and relatively oriented trifocal image geometry with temporal tracking in video sequences to generate a photogrammetric model (Zheng et al. 2007). In this scenario, the photogrammetric model is partially reconstructed from the neighboring images of the triplet, partially from the preceding and following images of the sequence. Due to two facts the trifocal video system allows generating a reliable photogrammetric model for each image triplet with little computational effort. First, each candidate triplet of corresponding points has a high potential to be correct as the matching between images of the triplet is stabilized by a tracking approach for points in each of the video sequences. Second, the trifocal tensor allows checking each triplet based on the relative orientation of the cameras. Thus, the system basically acquires a three-dimensional image which is used in a tracking procedure.

Practically useful results can only be derived when the accuracy achieved fulfills photogrammetric standards. At the same time, only an automatic processing would ensure to make use of the full potential of video sequences. However, the computational requirements to deal with hand-held markerless video streams exceed the capabilities of real-time systems. Therefore, the proposed approach is designed for off-line processing of real-time recorded digital video.

The organization of the paper is as follows. In section 2, the trifocal video sensor is briefly introduced. The efficient spatial image matching algorithm, which is based on trinocular image rectification and semi-global optimization, is explained in the subsequent sections 3 and 4. The estimation of the camera motion and the registration of the 3D point clouds using temporal feature tracking are discussed in section 5. First experimental results for images of a real scene are presented in section 6 and finally we conclude and state possible improvements.

2 System Overview

In the actual stage the trinocular stereo rig consists of three Scorpion color cameras from Point Grey Research (PTGrey). The image sensor is based on a progressive scan CCD with square pixels. Each camera is able to acquire and transmit a high resolution digital video sequence (1384×1038 pixels) with up to 19 bayer tiled full frames per second using firewire (IEEE 1394a). The cameras are synchronized with an accuracy of less than 1 ms. The effective data rate of the three cameras is around 80 MB and exceeds the bandwidth of one firewire channel as well as the writing performance of a regular harddrive. Therefore, a desktop PC with three independent 1394a-channels and a RAID-0 array consisting of four SATA disks was assembled. The system is designed to capture video with a maximum data rate up to 200 MB/sec for more than three hours using a battery based power source (see Fig. 1).

For a flexible image acquisition, we selected CCTV-lenses with variable principal distance between 6 mm and 12 mm. We discovered a significant radial lens distortion up to 80 pixels. Therefore, we model radial errors with 3 additional coefficients and resample all images using bicubic interpolation. The undistorted images simplify the geometric imaging model to a line-preserving pinhole camera. In addition to our stationary control point field we developed a mobile calibration rig (see Fig. 2). It allows an on-site calibration in few seconds due to an automatic marker detection and fitting

Fig. 1: a) Trinocular stereo rig with mounted Scorpion cameras and b) the mobile image acquisition system.

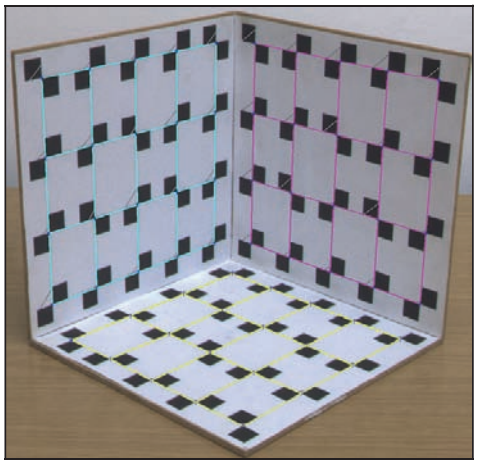

Fig. 2: Image of the mobile calibration rig with color coded results of the 3D model fitting.

algorithm for parameterized 3D models (Lowe 1991).

3 Trinocular Rectification

This section describes the geometric transformation of an uncalibrated image triplet to the stereo normal case (Heinrichs & Rodehorst 2006). In case of a trinocular rectification, the images are reprojected onto a plane, which lies parallel to the projection centers. The proposed trinocular rectification method requires an image triplet with more or less L-shaped camera alignment. The camera configuration is arbitrary, but each projection center must be invisible in all other images. This condition is necessary, since otherwise the epipoles lie in the image and mapping them to infinity will lead to unacceptable distortion of the images.

Furthermore, we assume non-degenerate camera positions, where the camera centers are not collinear, because collinear setups can be rectified by chaining a classical binocular rectification approach. Additionally, a common overlapping area and at least six corresponding image points are necessary, so that the trifocal tensor, the fundamental matrices and the epipoles can be determined. The result consists of three geometrically transformed images, in which the epipolar lines run parallel to the image axes. The resampling due to radial distortion and for trinocular rectification is processed in one step to minimize image blur.

3.1 *Camera Setup*

A given image triplet consists of the original images b (base), h (horizontal) and v (vertical). Subsequently, we denote the rectified images $\tilde{b}$, $\tilde{h}$ and $\tilde{v}$. The rectification tries to fit any image triplet to an L-configuration. This setup has the following properties:

- The epipolar lines of image $\tilde{b}$ and image $\tilde{h}$ correspond with their image rows.
- The epipolar lines of image $\tilde{b}$ and image $\tilde{v}$ correspond with their image columns.

- The epipolar lines of image $\tilde{h}$ and image $\tilde{v}$ have a slope of minus unity.

The last property has the advantage, that the disparities between corresponding points in $\tilde{b} \leftrightarrow \tilde{h}$ and $\tilde{b} \leftrightarrow \tilde{v}$ are equal. The basic idea of rectification is to map the epipoles e between the images b, h and v to infinity.

3.2 *Linear Rectification*

The initial task is to determine the relative orientation of the images. The fully automatic approach uses interest point locations from a modified FÖRSTNER operator (RODEHORST & KOSCHAN 2006) in combination with the SIFT descriptor (LOWE 2004) for matching. We have implemented a robust estimation of the trifocal tensor **T**, which describes the projective relative orientation of three uncalibrated images. It is based on a linear solution derived from six points seen in three views (HARTLEY & ZISSERMAN 2003, MAYER 2003) followed by a non-linear bundle adjustment over all common points. To handle the large number of high resolution images the computationally intensive RANSAC algorithm for robust outlier detection has been replaced by a faster evolutionary approach called Genetic Algorithm Sampling Consensus GASAC (RODEHORST & HELLWICH 2006). Note that the fundamental matrices derived from this tensor are not independent and have only 18 significant parameters in total. Let $\mathbf{H}_b$, $\mathbf{H}_h$ and $\mathbf{H}_v$ be the unknown 3×3 homographies between the original and rectified images. These primitive rectifying homographies can linearly be determined from the compatible fundamental matrices with 6 degrees of freedom left. A detailed derivation is given in (HEINRICHS & RODEHORST 2006).

3.3 *Imposing Geometric Constraints*

We recommend to calculate values of the remaining 6 degrees of freedom in the following order:

- Finding proper shearing values
- Determine a global scale value
- Finding right offset values

The shearing of images can be minimized by keeping two perpendicular vectors in the middle of the original image perpendicular in the rectified one. This results in quadratic equations which have two solutions. The result with the smaller absolute value is preferred. On the one hand, the global scale should preserve as much information as possible, but on the other hand produce small images for efficient computation. Therefore, we adjust the length of the diagonal line through $\tilde{b}$ to the original length in b. The

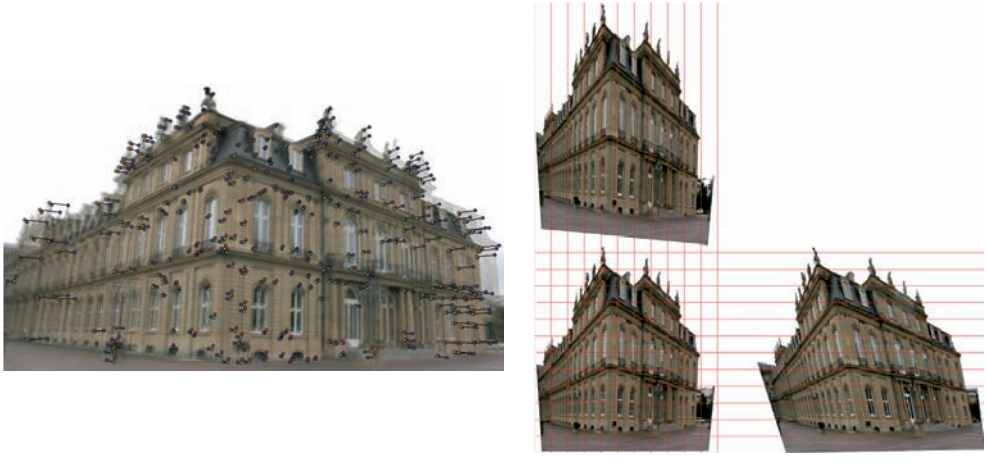

Fig. 3: a) Overlaid image triplet of Stuttgart palace with robust feature matches and b) rectified normal images.

offsets shift the image triplet in the image plane. To keep the absolute coordinate values small, the images should be shifted to the origin.

4 Trinocular Image Matching

After geometric transformation of the given image triplets using the proposed rectifying homographies, the correspondence problem must be solved using dense stereo matching. To find corresponding image points which arise from the same physical point in the scene, we suggest a modified semi-global matching (SGM) technique (Heinrichs et al. 2007, Hirschmüller 2005/2006). The normal images substantially simplify and accelerate the time-consuming computation. After rectification,

$$\begin{aligned} &\tilde{b}(x,y) \approx \tilde{h}(x + D(x,y), y) \quad \text{and} \\ &\tilde{b}(x,y) \approx \tilde{v}(x, y + D(x,y)) \end{aligned} \tag{1}$$

hold where x is the column coordinate, y the image row coordinate and D is called disparity map. The disparity at the current position (x, y) is inversely proportional to the depth of the scene. In addition to that, we assume that an estimation of the smallest and largest displacement is roughly given. This defines the search range $[d_{\min}, d_{\max}]$ for a reference point in $\tilde{b}$ along the corresponding rows and columns in $\tilde{h}$ and $\tilde{v}$ respectively.

4.1 *Local Similarity Measures*

Area-based matching is a widely used method for dense stereo correspondence. The similarity is computed statistically on the rectangular neighborhood (matching window) around the examined pixel. The algorithm searches at each pixel in reference image $\tilde{b}$ for maximum correlation in the horizontal image $\tilde{h}$ and the vertical image $\tilde{v}$ by shifting a small window pixel-by-pixel along the corresponding epipolar line. The normalized cross correlation NCC measures the linear relation between two image windows a and b normalizing over all intensity changes. The modified NCC (MNCC)

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

handles homogeneous areas better by adding the two denominator variances instead of multiplying them (Egnal 2000). We precalculate the means and the means of squared intensities to accelerate the computation significantly. Finally, we transform the correlation coefficient range to $[0, 1]$. Due to the proposed trinocular rectification, the disparities in the horizontal and vertical image pairs are identical. Thus, the position of a matching candidate in image $\tilde{b}$ and $\tilde{h}$ is linked exactly to a position in $\tilde{v}$. Now, the local matching costs $1 - \rho$ can simply be averaged. This increments the computational effort for the additional third image only by a linear factor. Additionally, the matching is more robust, because the linked cost function has less local minima than the individual cost functions for the image pairs.

4.2 *Semi-Global Optimization*

Generally, matching based on local costs only is ambiguous and therefore a piecewise smoothness constraint must be added. In (Hirschmüller 2005/2006) a very simple and effective method of finding minimal matching costs is proposed. SGM tries to determine a disparity map D such that the energy function

$$\begin{aligned} E(D) = &\sum_{x,y \in I} ((1 - \rho(a(x,y), b(x + D(x,y), y))) \\ &+ Q_1 \sum_{i,j=-1}^{1} \mathrm{T}[|D(x,y) - D(x+i, y+j)| = 1] \\ &+ Q_2 \sum_{i,j=-1}^{1} \mathrm{T}[|D(x,y) - D(x+i, y+j)| > 1]) \\ &\text{for } i \neq j \end{aligned} \tag{3}$$

is minimal. The first term calculates the sum of all local matching costs using the inverse correlation coefficient ρ of the image windows a and b around the current position (x, y) and the related disparity in D. The subsequent terms require a Boolean function T that return 1 if the argument is true and 0 otherwise. Explained intuitively, $E(D)$ accumulates the local matching costs with a

small penalty $Q_1 = 0.05$, if the disparity varies by one from the neighbored disparities. If the disparity differs by more than 1, a high penalty $Q_2 \in [0.06, 0.8]$ is added. The actual value of Q_2 depends on the intensity gradient in the original image. Long gradients result in a low Q_2 while short gradients result in a high Q_2. This prevents depth changes in homogeneous regions. There are only two different penalties for the depth changes. First, Q_1 ensures that regions with a slightly changing depth are not penalized too hard. Second, if depth changes in the scene occur, the size of the discontinuity is not correlated to the penalty.

Computing the minimum energy of $E(D)$ leads to NP-hard complexity, which is difficult to solve efficiently. Following (Hirschmüller 2005), a linear approximation over possible disparity values $d \in [d_{\min}, d_{\max}]$ is suggested by summing the costs of several 1D-paths L through the search space towards the actual image location (x_1, y_1). A path L with $i = n \ldots 1$ steps is recursively defined and the number of accumulated paths should be at least eight. We introduce a threshold for the local matching costs, to penalize dissimilar candidates. If the correlation coefficient ρ is lower than a certain threshold the local matching costs is set to a high constant value. If the minimal costs for the best matching candidate is higher than this value the match is marked as invalid.

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

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

$$D_1(x, y) + D_2(x + D_1(x, y), y) \leq 1. \qquad (4)$$

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

4.3 *Hierachical Approach*

Based on the original image resolution a number of reduced images are computed using a scale factor f_i. The search range can be scaled as well by f_i, so that the computational complexity drops dramatically from the actual scale level to the next smaller one. The images are processed from the lowest resolution to the highest one. Only image points of the first layer have to be checked at every possible location within the search range. To reduce the number of candidates in the succeeding layers the potential information of the previous layer is used and refined. If displacement information from a previous layer is available, the number of candidates can be reduced by restricting the possible range. The valid candidates fulfill at least one of three criteria:

- **Accuracy improvement:** The information from the previous layer has an accuracy of $\pm s \cdot f_i$, where s is the distance from one

candidate to the next one and f_i is the scale factor from the previous layer to the actual one. Possible matches within this accuracy range must be checked.

- **Unmatched points:** Points in the target image, which are already matched in a previous layer, should be excluded from further matching to avoid double matches.
- **Edge preservation:** If points of the previous layer lie on a surface edge, the depth value of the associated points in the actual layer is bounded by the depths of the two neighboring points. It might happen that not all of this information is available.

Fig. 4 illustrates this technique. The diagonal strip represents the search space of the original layer. Every column is the search space for a pixel position. The red line represents the selected correspondence. The thin diagonal stripe around the red line is the accuracy improvement from the first criterion. Vertical lines are unmatched positions in the previous layer. Therefore, the search space at these positions has to be analyzed completely to find possible new matches.

Horizontal lines represent unmatched points from the second criterion. The small vertical strips are caused by the edge preservation of the third criterion. Candidates in the black area are excluded by the hierarchical approach. This shows the efficiency of the proposed method. The search space is reduced to approximately 25 % of the original size. After calculating the local costs for each candidate our modified version of SGM calculates the best match for the given position.

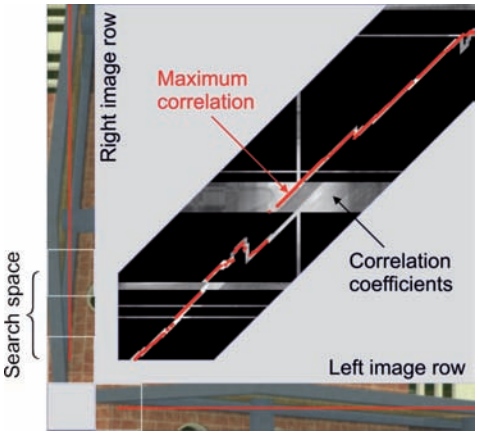

Fig. 4: Sample of the reduced search space of corresponding image rows.

5 Camera Path Estimation and 3D Point Cloud Registration

The temporal image correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$ are determined using the KLT tracker (Shi & Tomasi 1994). The feature-based approach uses local similarity measures and the temporal epipolar geometry. With a hierarchical approach using image pyramids the estimation of the orientation on a coarse level allows to improve the matching on a finer level. We extend the approach by filtering outliers using a temporal trifocal geometry (see Section 3.2). After the matching process we are able to orient the images fully automatically. Furthermore, the minimal 5-point algorithm (Nistér 2004) computes the essential matrix $\mathbf{E}$ of a camera pair even from correspondences on critical surfaces, i. e. planes. However, the presence of false correspondences in the tracking data and the unstable computation of eigen vectors (Batra et al. 2007) require a robust computation of the essential matrix via GASAC. This procedure results in a set of succeeding camera pairs with a uniform base length.

The registration of temporal image pairs using the estimated camera path is realized following (Fitzgibbon & Zisserman 1998). We are given temporal image pairs, each with an estimated essential matrix $\mathbf{E}$ and a set of homologous image points $\mathbf{x} \leftrightarrow \mathbf{x}'$. The goal is to register the spatial reconstructions into the same coordinate system by determining a spatial homography $\mathbf{H}$ which results in the best overlap of the two reconstructions. Since a spatial homography has 15 degrees of freedom and a projection matrix only 11, two corresponding cameras $\mathbf{P}$ and $\mathbf{P}'$ for a common image can be exactly registered by $\mathbf{P} = \mathbf{P}'\mathbf{H}^{-1}$. This constrains eleven parameters of $\mathbf{H}$. The remaining four

parameters can be found by minimizing the algebraic distance of the triangulated object points $d(\mathbf{X}, \mathbf{HX'})$ subject to the constraint $\mathbf{PH} = \mathbf{P'}$. The solution is a member of the 4-parameter family of homographies

$$\mathbf{H(v)} = \mathbf{P}^{+}\mathbf{P'} + \mathbf{hv}^{\mathsf{T}} \tag{5}$$

where $\mathbf{h}$ is the nullvector and $\mathbf{P}^{+}$ the pseudoinverse of $\mathbf{P}$. A direct solution for $\mathbf{v}$ can be obtained using a system of 3 equations per object point:

$$\mathbf{bX'}^{\mathsf{T}}\mathbf{v} = \mathbf{c} \tag{6}$$

where the 3-vectors $\mathbf{b}$ and $\mathbf{c}$ are defined by $b_k = h_k X_4 - h_4 X_k, \quad c_k = X_k a_4 - X_4 a_k$ and $\mathbf{a} = \mathbf{P}^{+}\mathbf{P'X'}$. The direct solution minimizes an algebraic error with no direct geometric or statistical meaning, so we refine the reprojection error using bundle adjustment.

6 Experimental Results

First results of the proposed method are presented for the Stuttgart palace in Germany. Figs. 5 and 6 show that it is possible to compute a realistic 3D architecture model utilizing high resolution video sequences. For

Fig. 5: a) The resulting disparity map of the Stuttgart palace using the modified SGM and b) top-view on the point cloud of the reconstructed corner.

Fig. 6: Available architecture model of the Stuttgart palace from aerial images improved by manual texture mapping (top) and 3D point clouds acquired automatically with the proposed trifocal sensor system using one image triplet (bottom).

these experiments, we use partially calibrated image triplets where radial distortion has been eliminated in advance. The hierachical approach reduces the computational costs to 25 percent without significant loss in accuracy. The dense matching of one image triplet with 1.4 mega pixels was completed in 2 minutes on a 2.4 GHz dual core CPU.

7 Conclusions and Future Work

In this paper, we introduce a high resolution video sensor for the 3D reconstruction of architectural models. The hybrid system unifies triangulation methods of spatial stereo with tracking methods of temporal stereo. We presented a linear method for trinocular rectification of uncalibrated images, which can be solved in closed form with 6 degrees of freedom. In a post processing stage, proper geometric constraints are selected to minimize projective distortion. The proposed dense matching algorithm is a fast and effective adaption of the SGM for image triplets. Combining the local costs of an image triplet to a single value stabilizes the matching, especially in regions with repetitive patterns like bricks, grids or stripes. The motion of the video sensor is estimated using temporal feature tracking and allows the integration of dense point clouds. First experimental results for image sequences of a real scene demonstrate the potential performance. The resulting photogrammetric models are combined as a metric model of the scene. The approach proposed is generic from a methodological point of view and allows various applications. Nevertheless, architectural models consist of planes, polyhedrons and freely formed surfaces. At this occasion geometric primitives, like planes and polyhedrons, should be fitted to the large point cloud to increase the accuracy and spare memory. Such systematic generation of a photogrammetric model from several trifocal views gives a digital video system its full potential.

Acknowledgements

This work was partially supported by grants from the German Research Foundation DFG. We would like to thank Hongwei Zheng and Nicole Bouvier for the textured model of the palace as well as Dirk Mehren for the acquisition of the trifocal images in Stuttgart.

References

Akbarzadeh, A., Frahm, J.-M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., Phelps, M., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., Towles, H., Nistér, D. & Pollefeys, M., 2006: Towards Urban 3D Reconstruction From Video. – 3rd Int. Symp. on 3D Data Processing, Visualization and Transmission (3DPVT).

Batra, D., Nabbe, B. & Hebert, M., 2007: An Alternative Formulation for Five Point Relative Pose Problem. – IEEE Workshop on Motion and Video Computing: 21–26.

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

Faugeras, O.D. & Luong, Q.-T., 2001: The geometry of multiple images. – The MIT Press, Cambridge, Massachusetts.

Fitzgibbon, A.W. & Zisserman, A., 1998: Automatic Camera Recovery for Closed or Open Image Sequences. – European Conference on Computer Vision: 311–326.

Hartley, R. & Zisserman, A., 2003: Multiple view geometry in computer vision. – Cambridge University Press, 2nd edition.

Heinrichs, M. & Rodehorst, V., 2006: Trinocular Rectification for Various Camera Setups. – Photogrammetric Computer Vision PCV'06, Bonn, 43–48.

Heinrichs, M., Hellwich, O. & Rodehorst, V., 2007: Efficient Semi-Global Matching for Trinocular Stereo. – Photogrammetric Image Analysis PIA'07, Munich, 185–190.

Hirschmüller, H., 2005: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. – Computer Vision and Pattern Recognition **2:** 807–814.

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

Koch, R., 2003: 3D-scene modeling from image sequences. – Photogrammetric Image Analysis PIA'03: 3–9.

Lowe, D., 1991: Fitting parameterized three-dimensional models to images. – IEEE Transactions on Pattern Analysis and Machine Intelligence **13**(5): 441–450.

Lowe, D., 2004: Distinctive image features from scale invariant keypoints. – International Journal of Computer Vision **60**(2): 91–110.

Mayer, H., 2003: Robust Orientation, Calibration, and Disparity Estimation of Image Triplets. – Pattern Recognition – DAGM'03, Springer: 281–288.

Montemerlo, M. 2003: FastSLAM – A Factored Solution to the Simultaneous Localization and Mapping Problem with Unknown Data Association. – PhD thesis, CMU-RI-TR-03-28, Robotics Institute, Carnegie Mellon University.

Neumann, J. & Aloimonos, Y., 2002: Spatio-Temporal Stereo using multi-resolution subdivision surfaces. – International Journal of Computer Vision **47**(1/2/3): 181–193.

Nistér, D., 2004: An efficient solution to the five-point relative pose problem. – IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(6): 756–777.

Nistér, D., Naroditsky, O. & Bergen, J., 2006: Visual odometry for ground vehicle applications. – Journal of Field Robotics **23**(1): 3–20.

Pollefeys, M., Gool, L.V., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J. & Koch, R., 2004: Visual modeling with a hand-held camera. – International Journal of Computer Vision **59**(3): 207–232.

Rodehorst, V. & Hellwich, O., 2006: Genetic Algorithm SAmple Consensus (GASAC) – A Parallel Strategy for Robust Parameter Estimation. – Int. Workshop "25 Years of RANSAC" in conjunction with CVPR'06, New York.

Rodehorst, V. & Koschan, A., 2006: Comparison and Evaluation of Feature Point Detectors. – Proc. of 5th Turkish-German Joint Geodetic Days, Berlin.

Shi, J. & Tomasi, C., 1994: Good features to track. – Computer Vision and Pattern Recognition: 593–600.

Tao, C.V., 2000: Semi-Automated object measurement using multiple-image matching from mobile mapping image sequences. – Photogrammetric engineering and remote sensing **66**(12): 1477–1485.

Zheng, H., Rodehorst, V., Heinrichs, M. & Hellwich, O., 2007: Improvement of the Fidelity of 3D Architecture Modeling Combining 3D Vector Data and Uncalibrated Image Sequences, ISPRS Workshop on Updating Geospatial Databases with Imagery and on DMGISs, Urumqi, 127–134.

Address of the Authors:

Dipl.-Ing. Matthias Heinrichs, Prof. Dr.-Ing. Olaf Hellwich, Dr.-Ing. Volker Rodehorst, Computer Vision & Remote Sensing, Berlin University of Technology, Franklinstr. 28/29, Sekr. FR 3-1, D-10587 Berlin, e-mail: matzeh@cs.tu-berlin.de, hellwich@cs.tu-berlin.de, vr@cs.tu-berlin.de

Manuskript eingereicht: Mai 2007
Angenommen: Juni 2007