IMPROVEMENT OF THE FIDELITY OF 3D ARCHITECTURE MODELING COMBINING 3D VECTOR DATA AND UNCALIBRATED IMAGE SEQUENCES

Hongwei Zheng, Volker Rodehorst, Matthias Heinrichs and Olaf Hellwich

Computer Vision & Remote Sensing, Berlin University of Technology,
Sekretariat FR 3-1, Fanklinstraße 28-29, D-10587 Berlin, Germany -
(fhzheng, vr, matzeh, hellwichg)@cs.tu-berlin.de

KEY WORDS: 3D Modeling, 3D Spatial Database, Dynamic Visualization, Uncalibrated Image Sequences, Multiple View Geometry, 3D Surface Reconstruction, Trinocular, Semi-Global Optimization

ABSTRACT:

The goal of this work is to create a system for improving large architectures models with relatively few uncalibrated image sequences and produces interactively navigable and photorealistic 3D scene. The novel idea is based on the combination of spatial 3D vector data and uncalibrated image sequences to improve the fidelity of 3D architecture models in an extended urban area. First, the spatial 3D vector data is stored in our designed spatial 3D object-relational database. 3D vector data can simplifies the 3D reconstruction problem by supporting and solving directly for the architectural dimensions of the scene. Next, an architecture building is observed from multiple viewpoints by freely moving a video camera around it for taking uncalibrated images. All viewpoints are then linked by controlled correspondence linking for each image pixel. The correspondence linking algorithm allows for accurate depth estimation. Dense surface reconstructions are obtained by adjacent images of the sequence as stereoscopic pairs and computing dense disparity maps for each image pair. A flat wall of the architecture can be computed as a relief. Finally, we displace the surfaces of the model to make them maximally consistent with their appearance across these multiple images. Result shows that it is possible to compute a more realistic 3D architecture model using relatively few image sequences, as long as appropriate 3D vector data is available.

1. INTRODUCTION

With the developments of 3D GIS, a lot of research has been done to provide interactive visualization systems for 3D spatial data models of growing size and complexity. Interactive and high-fidelity 3D spatial data visualization plays an increasing important role in modern 3D GIS. 3D architecture modeling in an urban area in 3D GIS is an interesting case of the general modeling problem since geometries of architectures are typically very structured. People are familiar with what the results should look like, they have higher expectations of the quality for the results.

Currently, since most existing spatial 3D architecture models in urban area are extracted from rectified orthogonal aerial images, we can not achieve high-fidelity and relief-like 3D architecture models in levels of detail, especially walls, doors and windows etc. Moreover, the direct texture mapping can not achieve high-fidelity 3D architecture, shown in Fig. 1. On the other hand, digital video sequences contain a high potential for such photogrammetric applications which is presently not fully used. Projective geometry provides an effective mathematical framework to obtain geometrically precise information from partially calibrated cameras with varying parameters. However, static 3D architecture modeling from pure ground-taken uncalibrated image sequences needs a lot of images and difficult to get a complete 3D architecture model for a large building or a group of 3D architecture models in urban area.

In this paper, we present a novel idea to integrate the advantages of 3D reconstruction using aerial images and ground-taken uncalibrated video sequences. In this approach, we do not directly use two types of images for one 3D architecture reconstruction but using ground-taken uncalibrated image sequences to improve the fidelity of the existing 3D architecture models. First, based on the study of previous work, we have designed and implemented a novel 3D topological and geometric model to manage large 3D spatial vector data and related attribute data, e.g., textures, spatial coordinates etc. Dynamic visualization of these 3D vector data is based on interactive query and the graph rendering pipeline. Second, we reconstruct a certain relief-like surfaces for these existing 3D architecture models using our newly designed trinocularcamera

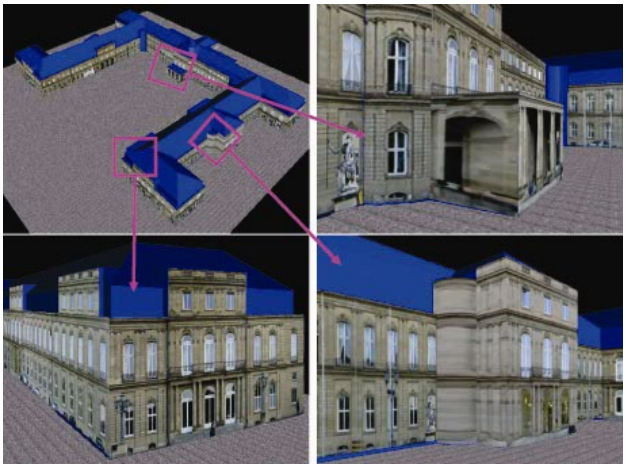

Figure 1: $\begin{array}{c|c} a & b \\ \hline c & d \end{array}$ Some fidelity information can not be reconstructed using pure rectified orthogonal aerial images and 2D texture mapping. (a) The reconstructed 3D architecture models are stored in a spatial 3D database. (b)(c)(d) Without 3D information and distorted 2D texture mapping on these parts.

system and related novel algorithms. These reconstructed 3D relief-like surfaces or partial architecture models with textures are then registered on their related 3D architecture models to improve the fidelity of these 3D architecture models. Finally, the combination of existing 3D vector data and reconstructed partial 3D models are presented. The main contribution of this work is the intensive use of 3D vector data information of architectures at every stage of the process. Vector data information is used for quality control as reference for building the accurate architecture 3D model. The trinocular-camera system and related novel algorithms can produce high-accuracy 3D surfaces in an efficient and robust manner. The output of the whole system is an extended urban area where the 3D buildings with realistic stereo texture maps are accurately located. This approach can also be considered as a proper way to improve the accuracy of large 3D architecture modeling as well as updating the existing 3D GIS databases. The results of the method are presented for a downtown area in a city in Germany.

The paper is organized as follows. In Section 2, we discuss some related work on modeling 3D spatial data and 3D reconstruction using uncalibrated image sequence. In Section 3, the proposed 3D spatial topological data model, geometric model and dynamic visualization are presented. 3D surface reconstruction using uncalibrated image sequences is presented in Section 4. Section 5 presents the improved 3D architecture models. Conclusions and future work are summarized in Section 6.

2. RELATED WORK

2.1 Modeling 3D Spatial Vector Data

The difficulties in realizing 3D GIS or 3D geo-spatial systems result from these different side. Although spatial data can be modeled in different ways, the first difficult is still the construction of a conceptual model of 3D data. The conceptual 3D model integrates information about semantics, 3D geometry and 3D spatial relationships (3D topology). The conceptual model provides the methods for describing real-world objects and spatial relationships between them. The design of a conceptual model is a subject of intensive investigations and several 3D models have already been reported.

2.1.1 3D Modeling Based on Topological Spatial Relationships. The topological model is closely related to the representation of spatial relationships, which are the fundament of a large group of operations to be performed in GIS, e.g. inclusion, adjacency, equality, direction, intersection, connectivity, and their appropriate description and maintenance is inevitable. Several 3D models have already been reported in the literature. Each of the models has strong and weak points for representing spatial objects. (Carlson, 1987) proposed a model called the simplified complex. The simplex is the simplest representation of a cell. The author uses the simplexes to denote spatial objects of node, line, surface, and volume. The model can be extended to *n* dimensions. (Molenaar, 1992) presents a 3D topological model called 3D Formal Vector Data Structure (3DFDS). The model maintains nodes, arcs, edges and faces that are used to describe four types of features named points, lines, surfaces and bodies. The Formal Data Structure (FDS) is the first data structure that considers the spatial object an integration of geometric and thematic properties. Tetrahedral Network was introduced by (Pilouk, 1996) to overcome some difficulties of 3DFDS in modelling objects with indiscernible boundaries (such as geological formations, pollution clouds, etc.). The Simplified Spatial Model was designed to serve web-oriented applications with predominance of visualization queries (Zlatanova, 2000). The Urban Data Model (UDM) represents the geometry of a body or a surface by planar convex faces (Coors, 2002).

2.1.2 3D Geometric Data Structures. Some data structures are suitable for managing large 3D scenes. *R*-trees, R^+-tree and oct-tree are very versatile and frequently used spatial data structures. Traditional indexing methods (*B*-trees etc.) are only useful for one dimensional data. Guttman (1984) invented the *R*-tree in order to handle spatial data efficiently, as required in computer aided design and geo-data applications. *R*-trees consist of overlapping rectangles that either contains geometrical objects or *R*-tree rectangles of the next deeper level. R^+-trees are the modification of R^+-trees, avoids overlap at the expense of more nodes and multiple references to some objects. Qing Zhu (2002) proposed to select the *R*-tree data structure to accelerate spatial retrieving for the integrated 3D databases of large Cyber City. The oct-tree is a generic name for all kinds of tree that are built by recursive division of space into eight subspaces.

2.2 3D Reconstruction using Uncalibrated Video Sequences

Digital video cameras provide dense image sequences that contain a high potential for photogrammetric application which is presently not fully used. Image and video sequences for scene modeling and various possible applications are treated by (Akbarzadeh and Pollefeys, 2006, Pollefeys and Koch, 2004). 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 flow tracking (Shi and Tomasi, 1994), *visual odometry* (Nistér and Bergen, 2006), *simultaneous localization and mapping* SLAM (Montemerlo, 2003), optical flow estimation, rigid body motion estimation from small baselines, interrupted feature point tracking (Sugaya and Kanatani, 2004) as well as motion segmentation techniques (Shashua and Levin, 2001) gain importance.

Reconstructing a three-dimensional model from a single video sequence is often conducted with the *structure-from-motion* SFM algorithm. First attempts on hybrid algorithms for *spatio-temporal stereo* are presented by (Neumann and Aloimonos, 2002). Algebraic projective geometry (Hartley and Zisserman, 2004, Faugeras and Luong, 2001) provides an effective mathematical framework to obtain geometrically precise information from partially calibrated cameras with varying parameters. Particular problems occur when focus and zoom functionality is used - a problem which can be tackled by *auto-calibration* with time-varying camera intrinsic parameters (Pollefeys, 1999).

3. SPATIAL MODELING AND DYNAMIC VISUALIZATION

3.1 Design of 3D Spatial Topological Data Model

To satisfy the demands of a 3D GIS graphics interface, the data structure must comply with the following requirements. The 3D data structure should have real 3D support, fast spatial queries, level of detail, and compatibility with different storage architecture. While these requirements are mostly oriented towards efficient visualization, the data structure must also be versatile enough to fulfill typical demands of traditional GISs. It must be possible to query for items of a certain type, to query for objects that meet conditions, to query for adjacent objects etc.

Inspired by the UDM (Coors, 2001) model, shown in Fig. 2, we concentrate on the representation and storage of spatial data, including the spatial attributes, spatial operations, spatial constrains and spatial relationships between of geo-entities. Based on our designed diagram in Fig. 3, we define several basic terms in this system. The *TopoNode* represents the 0-dimensional primitive expressing point coincidence, it is the base of topology relationships. The role of the *TopoNode* is in the boundary of *TopoEdges* and in the support of topological point features. The orientation attribute expresses the sense in which the included node is used e.g. “start” or ”end” node. The optional *coboundary* of a node is a set of directed edges which are incident on this node. Edges emanating from this node appear in the node *coboundary* with a negative orientation.

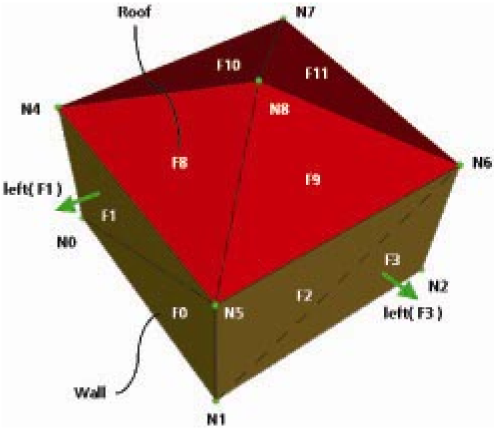

Figure 2: A single architecture model with nodes (*N*1~ *N*9), edges (two nodes form an edge), and faces (3 nodes form a face).

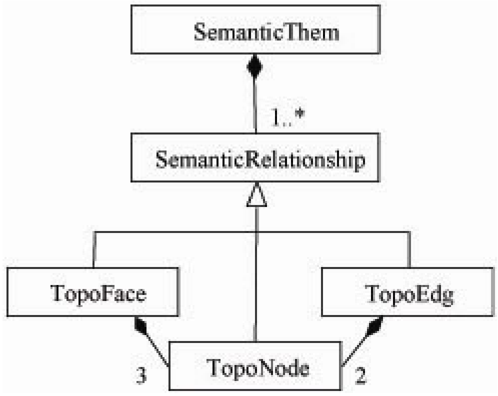

Figure 3: Our proposed diagram of spatial topology classes.

The *TopoEdge* represents the 1-dimensional primitive expressing linear coincidence. The topological boundary of an *TopoEdge* consists of a negatively directed start *TopoNode* and a positively directed end *TopoNode*. The optional *coboundary* of an edge is a circular sequence of directed faces which are incident on this edge in document order. The *TopoFace* represent the 2D dimensional topology primitive expressing surface overlap. The topological boundary of a face consists of a set of directed edges. The role of the *TopoFace* is in the *coBoundary* of topology edges and in the support of surface features. The orientation attribute expresses the sense in which the included face is used e.g. “inward ”or “outward” with respect to the surface normal in any geometric realization. There is strong symmetry in the relationships between topology primitives of adjacent dimensions. Topology primitives are bounded by directed primitives of one lower dimension. The *coboundary* of each topology primitive is formed from directed topology primitives of one higher dimension.

3.2 Design of 3D Spatial Geometric Classes

Dealing with spatial feature is the most important work of 3D GIS database. We proposed the spatial data model that has been used in our project here, which mainly bases on OpenGIS Simple Features Specification for SQL query. In fig. 4, we define these spatial geometric features and classes. The data model supports several primitive geometry types and the spatial object composed by the instances of these types: points, lines, polygons, circles, arc Polygons Polyhedron.

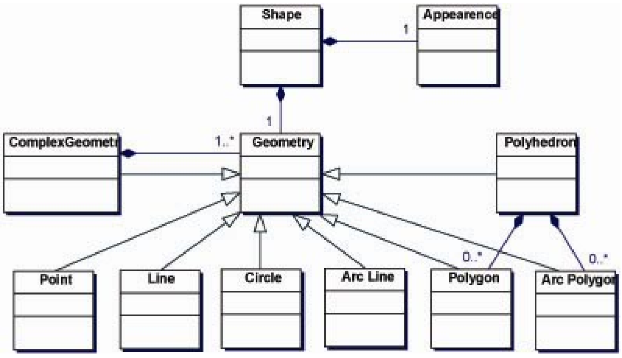

Figure 4: Our proposed spatial geometry classes.

The spatial data model is a hierarchical structure consisting of elements and geometries, which correspond to representations of spatial data. Geometries are made up of these elements. An *element* is the basic building block of geometry. The supported spatial element types are points, lines, Circles, Arc lines, polygons and Arc polygons. Each coordinates in an element is stored as an *X, Y, Z* triple. *Shape* is a container which collects all the attributes in spatial dimension, responding to each entity. It is composed by a pair of components called Geometry and Appearance. A *geometry* (or geometry object) is the representation of a spatial feature, modeled as an ordered set of primitive elements. Geometry can consist of a single element, which is an instance of one of the supported primitive types, or a homogeneous or heterogeneous collection of elements. A *multi-polygon* such as one used to represent a set of islands, is a homogeneous collection. A heterogeneous collection is one in which the elements are of different types, e.g., a point and a polygon. The *appearance* class encapsulates the graphical attributes which defines the representation of geometry data on visualization. It stores the color, size, rotation, texture, etc, which are some visualization styles related to the shape but not to the layer or workspace. The details are referred to (Zheng and Hahn, 2003).

3.3 Dynamic Visualization

Visualization is a general term to denote the process of extracting data from the model and representing them on the screen. The number of polygons that can pass the graphics pipeline per second is limited. The two most important approaches to improve the rendering speed, culling and the concept of levels of detail, try to reduce the number of polygons without reducing the image quality. (Clark, 1976) presented culling algorithms avoid rendering objects that are not currently visible through a rendering pipeline. The concept of Levels of Detail (LOD) (Clark, 1976) has been introduced to facilitate visualization of large scenes. Furthermore, the switch of LOD can be controlled by one parameter (Lindstrom et al., 1996), i.e. the distance between the viewing point and the object. Most of the algorithms developed assume that the LOD are pre-computed different representations of objects. In GIS applications, it is not unusual that more than 99.9 percent of the entire scene is invisible! According to (Zheng and Hahn, 2003, Zheng and Hellwich, 2003), in this system, we use VRML to visualize the query 3D data. VRML specifications present a flexible organization of geometry and texture LOD based on separate *a priori* designed VRML document. VRML is a convenient tool for testing the visualization result of 3D data which can get from 3D local or remote database server, shown in Fig. 5.

4. 3D SURFACE RECONSTRUCTION

In this section, we describe the refinement of architectural models from multiple image sequences using a highly resolving video sensor (Heinrichs et al., 2007b). The hybrid system unifies triangulation methods of spatial stereo with tracking methods of temporal stereo. We explain 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. The main novelty is an intensive integration of feature extraction, image matching as well as orientation for video sequences. In this scenario, the photogrammetric model is partially reconstructed from the neighboring images of the triple, partially from the preceding and following images of the sequence.

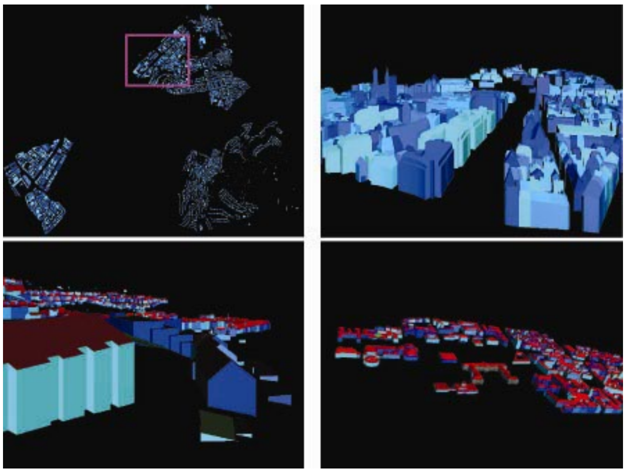

Figure 5: $\begin{array}{c|c} a & b \\ \hline c & d \end{array}$ Dynamic Visualization based on 3D spatial query. In (a), black regions among building groups are mountain regions. (b)(c)(d) These buildings stand on a real DEM model and their attributes can be easily changed (color).

4.1 System Design

The trinocular stereo rig consists of three color cameras based on progressive scan CCDs with a resolution of 1384 × 1038 pixels. Each sensor is able to acquire digital video with up to 19 frames per second. The cameras are synchronized with an accuracy of less than 1 ms via firewire. A desktop PC with three independent 1394a-channels and a RAID-0 array was assembled to capture video with a maximum data rate up to 200 MB/sec. The mobile image acquisition system is capable to record video sequences for more than three hours using a battery based power source.

For a flexible image acquisition, we selected CCTV-lenses with variable principle distance (6-12 mm). We have developed a mobile calibration rig for on-site calibration using an automatic marker detection and fitting algorithm for parameterized 3D models (Lowe, 1991). If necessary, varying intrinsic camera parameters can be recovered with an automatic self-calibration procedure, based on the direct linear estimation of the dual absolute quadric (Pollefeys, 1999). 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.

4.2 Relative Orientation

The initial task is to determine the relative orientation of the images. The fully automatic approach use interest point locations from the *Förstner operator* (Rodehorst and Koschan, 2006) in combination with the *SIFT descriptor* (Lowe, 2004) for matching. We have implemented a robust estimation of the trifocal tensor $\mathcal{T}$, which describes the projective relative orientation of three uncalibrated images. It is based on a linear solution of six points seen in three views (Hartley and Zisserman, 2004) followed by a non-linear bundle adjustment over all common points. To handle the large number of highly resolving 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 and Hellwich, 2006). The results of the feature based matching are shown in Fig. 6.

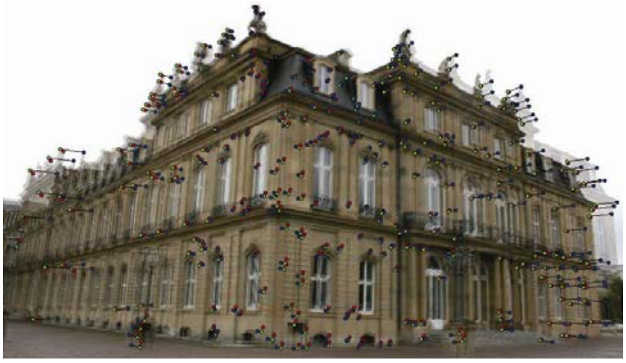

Figure 6: Overlaid image triplet with robust feature matches to determine the relative orientation.

4.3 Trinocular Rectification

This section describes the geometric transformation of a given image triplet to the stereo normal case (Heinrichs and

Rodehorst, 2006). The images are reprojected onto a plane, which lies parallel to the projection centers. The result consists of three geometrically transformed images, in which the epipolar lines run parallel to the image axes. 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 fundamental matrices F derived from $\mathcal{T}$ are not independent and fulfill the additional constraint

$$\mathbf{e}_{hv}^{\top}\mathbf{F}_{hb}\mathbf{e}_{bv} = \mathbf{e}_{vb}^{\top}\mathbf{F}_{vh}\mathbf{e}_{hb} = \mathbf{e}_{vh}^{\top}\mathbf{F}_{vb}\mathbf{e}_{bh} = 0.$$

The epipoles **e** are the left and right null-vectors of the compatible fundamental matrices and can be determined simultaneously using singular value decomposition (Hartley and Zisserman, 2004). Now, the unknown 3×3 homographies between the original and rectified images are given by

$$\mathbf{H}_b = \begin{bmatrix} 1 & 0 & s_1 \\ 0 & 1 & s_2 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_3 & 0 & 0 \\ 0 & \alpha_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_2 & 0 & 0 \\ 0 & \alpha_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} w_{b1}F_{33}^{bv} - F_{31}^{bv} & w_{b2}F_{33}^{bv} - F_{32}^{bv} & 0 \\ F_{31}^{bh} & F_{32}^{bh} & F_{33}^{bh} \\ w_{b1} & w_{b2} & 1 \end{bmatrix}$$

$$\mathbf{H}_v = \begin{bmatrix} 1 & 0 & s_1 \\ 0 & 1 & s_2+s_3 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_3 & 0 & 0 \\ 0 & \alpha_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_2 & 0 & 0 \\ 1-\alpha_2 & 1 & F_{33}^{bv}(\alpha_2-1) \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} F_{13}^{bv} & F_{23}^{bv} & F_{33}^{bv} \\ w_{v1}(F_{33}^{bv} - F_{33}^{hv}) + F_{13}^{hv} - F_{13}^{bv} & w_{v2}(F_{33}^{bv} - F_{33}^{hv}) + F_{23}^{hv} - F_{23}^{bv} & 0 \\ w_{v1} & w_{v2} & 1 \end{bmatrix}$$

$$\mathbf{H}_h = \begin{bmatrix} 1 & 0 & s_1+s_3 \\ 0 & 1 & s_2 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha_3 & 0 & 0 \\ 0 & \alpha_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1-\alpha_1 & F_{33}^{bv}(\alpha_2-1) \\ 0 & \alpha_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} w_{h1}(F_{33}^{bv} - F_{33}^{bh}) + F_{13}^{bh} - F_{31}^{hv} & w_{h2}(F_{33}^{bv} - F_{33}^{bh}) + F_{23}^{bh} - F_{32}^{hv} & F_{33}^{bv} - F_{33}^{hv} \\ w_{h1}F_{33}^{bh} - F_{13}^{bh} & w_{h2}F_{33}^{bh} - F_{23}^{bh} & 0 \\ w_{h1} & w_{h2} & 1 \end{bmatrix}$$

where we abbreviate the cross products of the epipole pairs with $w_b = e_{bh} \times e_{bv}$; $w_h = e_{hb} \times e_{hv}$ and $w_v = e_{vb} \times e_{vh}$. We recommend to calculate the remaining 6 parameters in the following order:

- Finding proper *shearing* values for α_1 and α_2
- Determine a global *scale* value α_3
- Finding right *offset* values for s_1, s_2 and s_3

The shearing of images *h* and *v* 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 forα_1 and α_2 which have two solutions, where the smaller $|\alpha_x|$ is preferred. On the one hand, the global scaleα_3 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 *b* with its projection in $\tilde{b}$. The offsets s_1, s_2 and s_3 shift the image triplet in the image plane. To keep the absolute coordinate values small, the images should be shifted to the origin. The normal images of the trinocular rectifi- cation are shown in Fig. 7b-c.

4.4 Semi-Global 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 homologous image points which arise from the same physical point in the scene, we suggest a modified semi-global matching (SGM) technique (Heinrichs et al., 2007a, Hirschmüller, 2006, Hirschmüller, 2005). After rectification it hold

$$\tilde{b}(x,y) \approx \tilde{h}(x + D(x,y), y) \approx \tilde{v}(x, y + D(x,y))$$

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 highest displacement is roughly given. This defines the search range [d_{min}, d_{max}] for a reference window in $\tilde{b}$ along the corresponding rows in $\tilde{h}$ and columns in $\tilde{v}$. As local similarity measure between two image windows *a* and *b* we use the statistically based *modified normalized cross correlation* (MNCC)

$$\rho_{MNCC}(a,b) = 0.5 \cdot \left(\frac{2 \cdot \sigma_{ab}}{\sigma_a^2 + \sigma_b^2} + 1 \right)$$

where $\bar{a}$ denote the mean, σ_a^2 the variance and σ_{ab} the covariance. We precalculate the means $\bar{a}$, $\bar{b}$ and the means of squared intensities to accelerate the computation significantly. Due to the proposed trinocular rectification, the disparities in the horizontal and vertical image pairs are identical, so that the correlation coefficients can simply be averaged. This increments the computational costs 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.

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

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

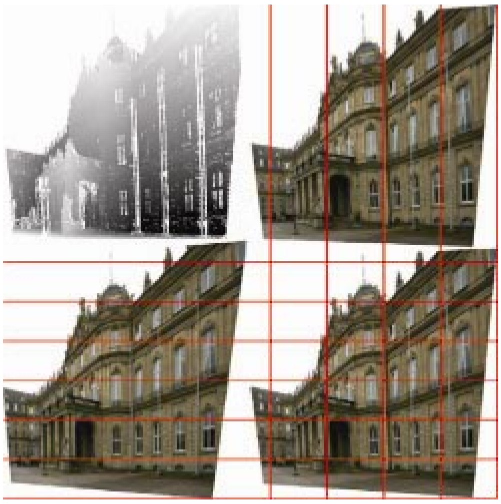

Figure 7: $\frac{a|b}{c|d}$ (a) Disparity map of the spatial image matching and (b-d) normal images of the trinocular rectification.

for $i \neq j$ is minimal. The first term calculates the sum of all local matching costs using the inverse correlation coefficient 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 cost with a small penalty $Q_1 = 0.05$, if the disparity varies by one from the neighbored disparities and a high penalty $Q_2 \in 2$ [0.06, 0.8], if the disparity differ more than 1. Furthermore, the higher penalty is even amplified, if the current location in the original image has no gradient. This prevents depth changes in homogeneous regions. Computing the minimum energy of *E*(*D*) leads to NP complexity, which is hard 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* towards the actual image location (*x; y*). See (Heinrichs et al., 2007a) for further details. The final disparity map can be estimated using

$$D(x, y) = \min_{d} \left(\sum_{\mathbf{r}} L_{\mathbf{r}}(x, y, d) \right)$$

To enforce stability, we check the *left/right consistency* (LRC) of the bidirectional correspondence search. 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

$$D_1(x, y) + D_2\left(x + D_1(x, y), y\right) \leq 1.$$

Otherwise, 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.

The images are processed hierarchically 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. If displacement information from a previous layer is available, the number of candidates can be reduced by restricting the possible range to approximately 25% of the original size. A disparity map of the spatial image matching using global optimization is shown in Fig. 7a.

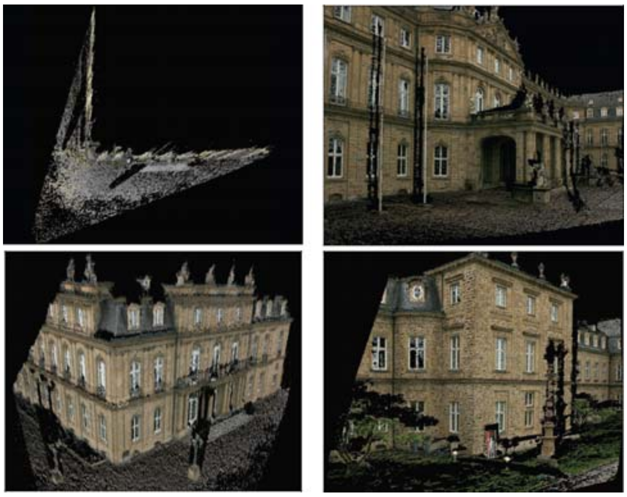

Figure 8: $\frac{a|b}{c|d}$ (Reconstructed partial 3D architecture models with texture projected 3D relief-like surfaces. These models are related to Fig.1. (a) Reconstructed 3D point clouds with texture information of (c). (b) (d) Reconstructed partial 3D models.

4.5 Temporal Feature Tracking

The temporal image correspondences for the partially calibrated cameras are determined using the KLT tracker (Shi and Tomasi, 1994). The feature-based approach uses local similarity measures and the 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. After the matching process we are able to orient the images full automatically. The standard reconstruction method will fail for some regular objects, which are common in urban areas (i.e. dominant planes from facades). To solve this problem we improved our algorithm with the *Geometric Robust Information Criterion* (GRIC) published in (Pollefeys and Gool, 2002).

Furthermore, the minimal 5-point algorithm (Nistér, 2004) computes the essential matrix **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 eigenvectors (Batra and Hebert, 2007) requires 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 these temporal image pairs is realized following (Fitzgibbon and Zisserman, 1998) using the essential matrix E and a set of common image points x ↔ x'. The goal is to register the spatial reconstructions into the same coordinate system by determining a spatial homography **H** which results in the best overlap of the two reconstructions. A spatial homography has 15 degrees of freedom and a projection matrix only 11. The remaining four parameters can be found by minimizing the algebraic distance *d* (**X, HX**') of the triangulated object points **X** subject to the constraint **PH** = **P**'. The direct solution minimize an algebraic error with no direct geometric or statistical meaning, so we recommend to refine the reprojection error using bundle adjustment.

5. RESULTS AND DISCUSSION

5.1 3D Spatial Modeling and Dynamic Visualization

Following the designed spatial topological data model and geometric classes, we convert and construct a 3D spatial databases server which includes nearly all the architecture buildings in a city. Furthermore, the 3D spatial database is constructed using the object-relational database structure. The advantage of using such integration has several aspects. Firstly, although the 3D spatial database is very large, we can still get fast query of 3D spatial data and related attributes. Second, due to the 3D spatial database is object-oriented, we can easily construct object-oriented query and visualization. Fig. 5 and Fig. 1 are the results of remote query based dynamic visualization. This system can be directly used for mobile, internet 3D GIS systems.

5.2 3D surface reconstruction

The results of using the trinocular system and related methods for 3D surface reconstruction are presented for the Stuttgart palace in Germany (see Fig. 8). The results show that it is possible to compute a realistic 3D architecture models utilizing highly resolving video sequences. For these experiments, we use uncalibrated image triplets but eliminated radial distortion in advance. The hierarchical approach reduces the computational cost up to one quarter without significant lost 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.

5.3 Improved 3D Architecture Models

The challenging work is how to update the existing 3D architecture model using these high-fidelity reconstructed 3D surfaces. We have designed a reasonable approach to achieve this target. First, we can directly register the 3D surface in Fig. 8 on the 3D architecture model in Fig. 1. However, due the the larger size of 3D surface (The size of the reconstructed point clouds is around 50-150 Mb), we can not directly triangulate such 3D point clouds into vector data format and match it to the existing 3D architecture model. Therefore, we reduce the size of point clouds of 3D surfaces in most homogeneous regions, but keep point clouds in edge and discontinuous regions. We normalize the coordinate systems for existing 3D architecture models and 3D surface point clouds and match the point clouds to the existing 3D architecture model in a supervised manner.

6. CONCLUSIONS AND FUTURE WORK

Architectural models in extended urban area are always well structured which consist of planes, polyhedrons and freely formed surfaces. At this occasion, first, the well-designed 3D spatial data structure and dynamic visualization support a robust platform to integrate existing 3D spatial architecture model extracted from rectified aerial images and reconstructed high-accuracy 3D surface from terrestrial uncalibrated video sequences. It is also an efficient and promised approach to reconstruct and dynamic visualize a complete or a group of 3D architecture models with high-fidelity. Secondly, for the refinement of architectural models we introduce a highly resolving video sensor. The hybrid system unifies triangulation methods of spatial stereo with tracking methods of temporal stereo. A linear method for trinocular rectification of uncalibrated images can be solved in closed form with 6 degrees of freedom. The motion of the video sensor is estimated using temporal feature tracking and allows the integration of dense point clouds. For the future work, based on the characteristic properties of architecture buildings and further development of the potential of trifocal views, we train geometric primitives, like planes and polyhedrons which can be fitted to the large point clouds to increase the accuracy and spare memory. The existing 3D architecture model can thus be updated using these geometric primitive point clouds, respectively. Also, Such combination and update can be trained in a semi-supervised and unsupervised manner.

ACKNOWLEDGEMENTS

This work was partially supported by grants from the German Research Foundation DFG.

REFERENCES

Akbarzadeh, A., J.-M. F. P. M. B. C. C. E. D. G. P. M. M. P. S. S. B. T. L. W. Q. Y. H. S. R. Y. G. W. H. T. D. N. and Pollefeys, M., 2006. Towards urban 3d reconstruction from video. In: 3rd Int. Symp. on 3D Data Processing, Visualization and Transmission (3DPVT).

Batra, D., B. N. and Hebert, M., 2007. An alternative formulation for five point relative pose problem. In: IEEE Workshop on Motion and Video Computing WMVC ’07, pp. 21–26.

Carlson, E., 1987. Three-dimensional conceptual modelling of subsurface structures. Technical Papers of ASPRS/ACSM Annual Convention 4, pp. 188–200.

Clark, J., 1976. Hierarchical geometric models for visible surface algorithm. In Communications of the ACM 19(10), pp. 547–554.

Coors, V., 2001. 3D-GIS in networking environments. In: International Workshop on ”3D Cadastres”.

Faugeras, O. and Luong, Q.-T., 2001. The geometry of multiple images. In: The MIT Press, Cambridge, Massachusetts, p. 644.

Fitzgibbon, A. and Zisserman, A., 1998. Automatic camera recovery for closed or open image sequences. In: Proc. ECCV’98, pp. 311–326.

Hartley, R. and Zisserman, A., 2004. Multiple view geometry in computer vision, 2nd edition. In: Cambridge University Press, p. 672.

Heinrichs, M. and Rodehorst, O., 2006. Trinocular rectification for various camera setups. In: ISPRS Symp. Photogrammetric Computer Vision PCV’06, Bonn, pp. 43–48.

Heinrichs, M., Hellwich, O. and Rodehorst, O., 2007a. Efficient semi-global matching for trinocular stereo. In: Photogrammetric Image Analysis.

Heinrichs, M., Hellwich, O. and Rodehorst, V., 2007b. Architectural model refinement using terrestrial image sequences from a trifocal sensor. (submitted).

Hirschmüller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: IEEE Conf. on CVPR’05, San Diego, Vol. 2, pp. 807–814.

Hirschmüller, H., 2006. Stereo vision in structured environments by consistent semi-global matching. In: IEEE Conf. on CVPR’06, New York, Vol. 2, pp. 2386–2393.

Lindstrom, P., Koller, D., Ribarsky, W. and Hodges, L.F.and Faust, N. T. G., 1996. Real-time continuous level of detail rendering of height fields. Siggraph 96 Computer Graphics Proceedings pp. 109–118.

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

Lowe, D., 2004. Distinctive image features from scale invariant keypoints. Int. J. of Computer Vision 60(2), pp. 91–110.

Molenaar, M., 1992. A topology for 3d vector maps. ITC Journal 1, pp. 25–33.

Montemerlo, M., 2003. FastSLAM - A Factored Solution to the Simultaneous Localization and Mapping Problem with Unknown Data Association. PhD thesis, Carnegie Mellon University.

Neumann, J. and Aloimonos, Y., 2002. Spatio-temporal stereo using multi-resolution subdivision surfaces. Int. J. of Computer Vision 47(1-3), pp. 181–193.

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

Nistér, D., O. N. and Bergen, J., 2006. Visual odometry for ground vehicle applications. J. of Field Robotics 23(1), pp. 3–20.

Pilouk, M., 1996. Integrated Modelling for 3D GIS. PhD thesis, ITC.

Pollefeys, M., 1999. Self-calibration and metric 3D reconstruction from uncalibrated image sequences. PhD thesis, Catholics University Leuven.

Pollefeys, M., F. V. and Gool, L. V., 2002. Surviving dominant planes in uncalibrated structure and motion recovery. In: Proc. ECCV02, Copenhagen, Vol. 2, pp. 837–851.

Pollefeys, M., L. V. G. M. V. F. V. K. C. J. T. and Koch, R., 2004. Visual modeling with a hand-held camera. Vol. 59 Number 3, pp. 207–232.

Rodehorst, V. and Hellwich, O., 2006. Genetic algorithm sample consensus (gasac) - a parallel strategy for robust parameter estimation. In: Int.Workshop ”25 Years of RANSAC” in conjunction with CVPR’06, New York, p. 8.

Rodehorst, V. and Koschan, A., 2006. Comparison and evaluation of feature point detectors. In: Proc. of 5th Turkish-German Joint Geodetic Days, Berlin, p. 8.

Shashua, A. and Levin, A., 2001. Multi-frame infinitesimal motion model for the reconstruction of (dynamic) scenes with multiple linearly moving objects. In: Proc. ICCV’01, Vol. 2, pp. 592– 599.

Shi, J. and Tomasi, C., 1994. Good features to track. In: Int. Conf. on CVPR’94, pp. 593–600.

Sugaya, Y. and Kanatani, K., 2004. Extending interrupted feature point tracking for 3-d affine reconstruction. In: Proc. ECCV’04, Vol. 1, pp. 310–321.

Zheng, H. and Hahn, M., 2003. Data modeling and dynamic visualization in 3D mobile GIS. Technical report, Stuttgart.

Zheng, H. and Hellwich, O., 2003. Towards automatic 2D texture mapping of large 3D architecture modeling. Technical report, Berlin University of Technology.

Zlatanova, S., 2000. 3D GIS for urban development. PhD thesis, ITC, The Netherlands.