Thermal leakage detection on building facades using infrared textures generated by mobile mapping

Ludwig Hoegner, Uwe Stilla
Photogrammetry and Remote Sensing
Technische Universitaet Muenchen
Munich, Germany
Ludwig.Hoegner@bv.tum.de; Stilla@tum.de

***Abstract*—Focus of the paper lies on automated texturing of 3d building models with images recorded with infrared (IR) cameras. Therefore, a data fusion of given 3d models and recorded IR image sequences is performed. A relative orientation of the images of the sequence is generated using Nistér's five-point algorithm and image triplets. This results in a point cloud of correspondence points and a relative camera path. The relative oriented image scene is then matched in two steps first to the recoded GPS camera path and then to the given building with least squares minimum distance. Textures are extracted for every image and its according camera position from all visible surfaces. Because one image does not show complete building facades, the extracted textures of the images are combined to create complete textures for the model surfaces. These textures can be used for feature extraction and object recognition for analyzing buildings and assigned 3d coordinates to found features and objects.**

I. INTRODUCTION

The analysis and refinement of buildings in urban areas has become an imported research field in the last few years. The new GML standard CityGML [1] allows integrating different levels of detail into one GIS database. Buildings in city models normally consist of simple façade structures and optical textures. To refine and automate those processes different solutions have been proposed during the last years [2], [3], [4] AL 2008). Those strategies are extracted geometry and textures together to form a new building or façade and are not merged with an existing city model. Much urban information cannot be extracted from normal optical images but from other optical domains like infrared. Satellites allow the analysis of urban heat islands [5]. Airborne IR-systems are applied for exploration of leakages in district heating systems [6]. Ground cameras are recording the irradiation of building façades [7], for the specification of its thermal behaviour for thermal building passports. IR data of buildings are collected from several photos and analysed directly in the acquired images without any geometric or 3d processing. The focus of this paper is the integration of pre-known building models in the texture extraction process from image sequences to improve the textures' quality.

Thermal cameras record electromagnetic radiation in the invisible infrared (IR) spectra. Thus, surface characteristics of object can be detected, that stay invisible in normal visible spectra. For recognition of objects with little difference in temperature and for identification of small details from distance, thermal cameras must be able to resolve temperatures with an accuracy of 0,01 Kelvin. High-quality infrared cameras are able to record image sequences with standard video frame rate (25 fps) or even higher. Because of special cooling technique, camera optics and the low production numbers, the expenses of infrared cameras are very high compared to normal video cameras. Today, thermal image data is used for many different applications. Typically, IR data of buildings are collected from photos and analyzed directly in the acquired images. Bigger building parts are acquired by combining several images. The results of the analysis are stored in the 2d photos without any geometric or 3d processing. This can be a problem, when images from different cameras or views are combined and stored for further processing.

In contrast to conventional IR inspection of buildings, in this paper an automated strategy is used for texturing an entire building model. Narrow streets and the low resolution and small field of view of IR cameras are serious problems. Only small parts of the building façade are visible in one image. In this case we have to deal with the fact, that structures found in IR images do not always have correspondences in the building model and vice versa. This problem is even getting worse with a moving camera and inaccurate orientation parameters caused by the GPS system. Different strategies for matching of given 3d models and images are well known in computer vision. Single image processing is working with 3 or more correspondences between image and model. An overview over 3-point algorithms is given in [8]. Techniques for 4- and 5-point estimation are published in [9] and [10]. For 6 and more correspondence points the Direct Linear Transformation (DLT) can be used [10]. There are also iterative methods proposed in [11]. Furthermore homographies can be used for plane detection [12]. Longuet-Higgins [13] uses the 8-point algorithm to reconstruct a scene from two different views. When using image sequences, multiple images can be used for pose estimation [2], [3], [4]. Those strategies are based on Nistér's five-point algorithm [14] and constructing a relative oriented scene, where camera orientation, scene geometry and textures are reconstructed together from the same source images. In this paper, those strategies are extended to deal with a given building model in a global coordinate system.

978-1-4244-3461-9/09/$25.00 ©2009 IEEE

II. Overview

In contrast to conventional IR inspection of buildings, for texturing the whole building or a building complex in dense urban areas and their narrow streets, the low resolution and the small field of view avoid a direct line matching of edges in the images and edges of the given building model. The proposed concept is based on the assumption that a 3d model of the recorded building is given containing 3d vertex coordinates and triangulated polygon surfaces with given texture coordinates. The coordinates should be given in a national coordinate system like Gauss-Krueger. GPS coordinates are often inaccurate and a direct line matching of the images and the projection of the model's edges fails because of the lack of visible façade edges in many of the images. In former works we introduced a line matching strategy for infrared image sequences [15]. The usage of continuous image sequences taken from a moving car allows performing a relative orientation of the images of a sequence to extract estimated façade planes and a relative camera path [2], [3]. Instead of generating an isolated only relative oriented model, this estimation is matched with the pre-known building model and measured camera path. Several matching steps are performed: At first, the estimated camera path is matched onto the measured GPS path using the time stamp stored both in the measurement and the images. Next, the minimum distance and rotation of the estimated surfaces and the model is calculated to overcome the errors of the measured camera path. After this step, the corrected camera path and orientation are used to project the images of the sequence into the 3d model space. This is done using backwards projection techniques. The resolution of the outgoing texture is defined first. For every pixel of the texture, its coordinates in the 3d model are calculated and used to find the corresponding interpolated pixel value of the input image. Caused by the oblique partial occlusion of facades has to be taken into account. This procedure is repeated for every visible façade in every input image of the sequence. Additionally, for every pixel of the texture, the quality of the pixel's value is stored. This quality is given with respect to the distance of the interpolated intensity value to the given pixel values of the input image.

In the end, for every façade a couple of textures are generated. The resulting complete texture is calculated by merging the pixel values of the partial textures. This is done by copying one over another in the order of the time stamps. Because of the oblique view every visible pixel of a texture has a higher resolution than the texture before. To overcome remaining positioning errors, a line matching between the partial textures is included in this step.

III. Data Acquisition

Current IR cameras cannot reach the optical resolution of video cameras or even digital cameras. The camera used for the acquisition of the test sequences offers an optical resolution of 320x240 pixels with a field of view (FOV) of only 20°. The FLIR SC3000 camera is recording in the thermal infrared (8 - 12 µm). On the top of a van, the camera was mounted on a platform which can be rotated and shifted. Like in the visible spectrum, the sun affects infrared records. In long-wave infrared the sun's influence appears only indirect, as the sun is not sending in the long wave spectrum, but of course is affecting the surface temperature of the building.

Caused by the small field of view and the low optical resolution it was necessary to record the scene in oblique view to be able to record the complete facades of the building from the floor to the roof and an acceptable texture resolution. The image sequences were recorded with a frequency of 50 frames per second. To minimize holes in the textures due to occlusion caused by the oblique view, every façade was recorded with a view forward looking and a view backward looking. The viewing angle related to the along track axis of the van was constant. An example of a recorded sequence is shown in figure 1. The position of the camera was recorded with GPS and, for quality measurements from tachymeter measurements from ground control points.

Figure 1. Example images from one sequence along a building.

IV. Position Estimation

As mentioned in chapter II, the position estimation is split into two parts, first the relative orientation of the recorded images with their estimated relative camera orientations and, second the matching of this relative oriented estimated scene to the given GPS camera path and given 3d model.

A. Relative Orientation of the Image Sequence

Mayer [2] has introduced an approach for wide-baseline image sequences. Given the known inner orientation of the camera, SIFT features [16] are matched via cross-correlation. RANSAC [17] is used to choose SIFT features for the estimation of the fundamental matrix F and trifocal tensor T of triplets of adjacent images of the sequence. The found inliers are used for a robust bundle adjustment [12]. To orient the whole image sequence, the triplets are linked based on homographies and already known 3d points of the already

978-1-4244-3461-9/09/$25.00 ©2009 IEEE

oriented sequence part. The orientation of the points is fixed to vertical planes as points on facades are aligned in vertical planes. The point correspondences are transposed to a 3d point cloud in the relative oriented object space. The camera orientation for every input image is estimated. This approach for wide-baseline images is also applicable for image sequences with the same viewing direction and moving camera. Images for the triplets have to be chosen so that the difference of the position of found SIFT features is big enough for a depth estimation. This is checked by calculating the moving direction of the features between two images. The first image is fixed. The second image is chosen in ascending order from the following images of the sequence, until the movement of the points has reached a specified threshold (Fig. 2).

Figure 2. IR image with selected SIFT features, that have correspondences in the following image. Arrows show the moving direction of the points and numer of pixels they move.

The resulting relative oriented model shows a 3d point cloud of the SIFT features and the estimated camera position for every image of the sequence (Fig. 3). The structure of the facades is already visible. Most of the points are located in the edges of the windows and grouped in lines.

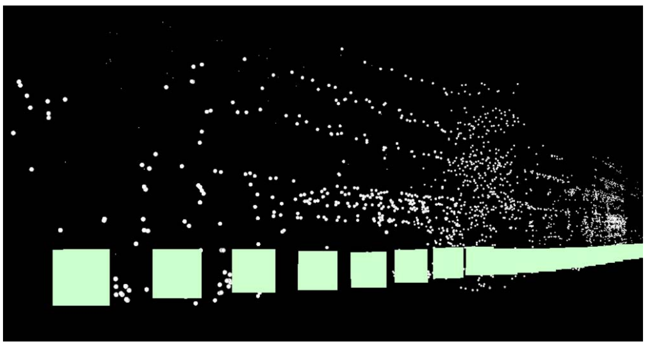

Figure 3. Point cloud of one image sequence along a group of facades. The squares are representing the estimated camera positions.

B. Transformation of the Estimated Model to the Global Coordinate System of the Given Building Model

The model generated in the previous section is given in local relative coordinates where the first and second camera position define the ground plane x-y with the first camera position as origin and the second as x direction. The z axis is chosen so that the point cloud defines vertical planes. Before textures can be extracted and assigned to the given building model, it is necessary to transform the estimated model to the global coordinate system.

The first transformation is using the given GPS path of the recording camera. Every measured camera position is given with a time stamp. The images of the sequence contain a time stamp, too and so the relative estimated camera path contains a time stamp for each camera position as it belongs to one image. We perform a least squares matching of the GPS path and the estimated camera path to calculate the transformation matrix for the estimated model.

After this transformation, the estimated model is approximately congruent to the corresponding surfaces. A further refinement is achieved by performing a least squares matching of the point cloud and the facades of the building model. A grouping of the points is done before the matching to remove non façade and wrong points i.e. of trees. Every point is assigned to the surface with the smallest distance. Points with a distance that differs beyond a threshold are rejected for this façade. Random faces are generated by combining three RANSAC chosen points assigned to the same surface. Their face's normal is compared to the normal of the assigned façade. This leads to a mean normal vector for the points belonging o one façade. Points that only belong to faces with a normal that differs beyond a threshold to that mean normal are rejected. The remaining points are now used for the least squares matching. A result is shown in figure 4.

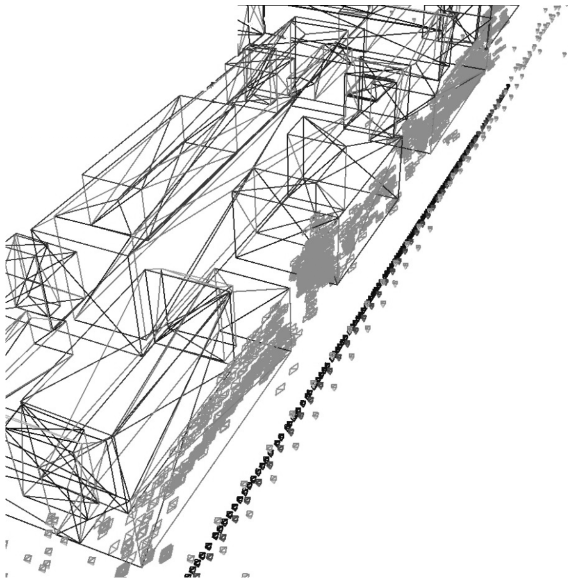

Figure 4. Grid model of the building with point cloud (light grey) and camera path: light grey: GPS path, dark grey transformed estimated camera positions

978-1-4244-3461-9/09/$25.00 ©2009 IEEE

V. Texture Generation

To generate a bitmap texture for a façade from an image the image coordinates have to be transferred to pixel coordinates of a façade texture. The scheme for a complete image sequence is shown in figure 5. At first, the resolution of the texture in pixels per meter is set. Now, the texture size is chosen according to the defined texture resolution and texture coordinates are generated for every pixel. These 2d texture coordinates are transformed to 3d coordinates in global coordinates from the known 3d coordinates and texture coordinates of the vertices of the façade. These points have to be transformed to camera coordinates and further on image coordinates of the input infrared image. To reduce the number of operations, the number of facades to be explored for one input image is reduced using standard computer graphics techniques well known from ray casting [18]. All facades with a distance longer than the far plane threshold are removed because the optical texture resolution for those facades would be very low. The remaining facades are divided into visible front faces and invisible back faces and the back faces are removed. Facades that are complete out of the field of view of the virtual camera are removed as well. For the remaining facades, every 3d point generated for the surface texture is projected into the input image if and only if the projection ray does not intersect any other façade polygon. Otherwise, the point is covered by another façade and invisible from the camera. For those pixels, the intensity value is set to -1 for blank. For pixel without any intersection of the projection ray, the intensity value at the calculated image coordinates is interpolated bilinear from the intensity values of the neighboring pixels of the input image.

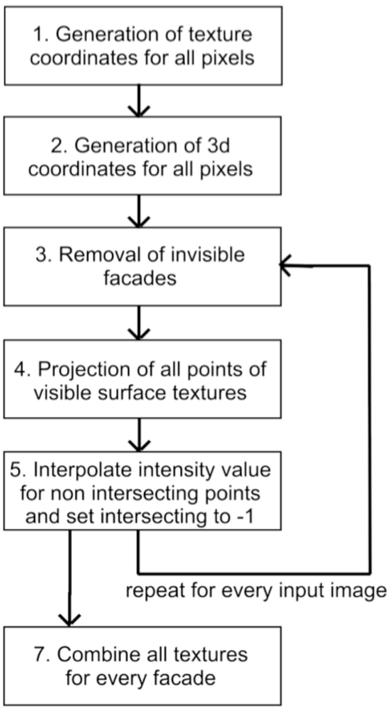

Figure 5. Scheme for the texture extraction

This strategy, known as backwards projection, avoids holes in the generated textures. The texture coordinates and global 3d coordinates of every pixel of the texture is only computed once for every façade and stored for the whole process of projecting images onto the facades of the building model. Figure 6 shows one example texture.

Figure 6. Partial texture of one façade extracted from one IR image

As this texture is extracted from an oblique view, the spatial resolution is not constant. It is getting lower from the left to the right (Fig 7a). Due to this the texture is blurred at the right. For composing a complete texture for a building façade, it is necessary to combine several partial textures. The recording strategy with oblique viewing camera offers the possibility for a quite simple combination strategy of all textures extracted from the image sequence and belonging to one façade. The texture resolution is decreasing to the right (Fig. 7a) and all invisible parts of the texture are marked out. So we copy every following texture onto the already composed. This new texture only overwrites parts where it has a higher resolution than the existing composed texture. A correlation is done for every added texture to remove remaining positioning errors. This correlation is repeated with several small different angels to find a tilt in the texture to add. The result is a texture with a high resolution area in the first floor and decreasing resolution to the ground and particularly to the roof (Fig. 7b and Fig. 8)

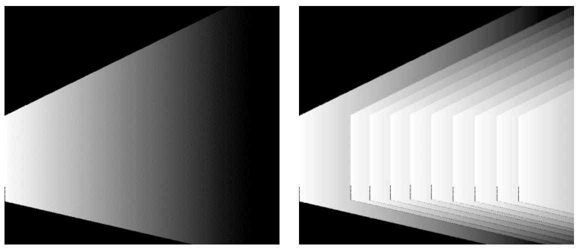

Figure 7. Resolution images of the texture in Fig. 6. White means high resolution, black is low resolution, a) Resolution image of a single partial texture, b) Resolution image of a combination of several partial textures.

978-1-4244-3461-9/09/$25.00 ©2009 IEEE

Figure 8. Surface texture generated from the complete image sequence

Because the appearance of IR signatures is changing over time and depends for example on the weather conditions, the combination of textures to create a new combined texture is very difficult. That is why normally only textures of one record session should be used to create new textures. So, for example from one record session the images from the terrestrial sequence with oblique forward view and oblique backward view can be combined to fill gaps or find errors caused by occlusion.

VI. LEAKAGE DETECTION

As visible in figure 8, there remain many structures in the texture. We want to find leakages like heating pipes. Unfortunately especially the windows show nearly the same behavior with horizontal and vertical lines. We find and remove the windows by combining the forward and backward view texture. The window edges are moving between the two textures because the windows are not in the surface plane of the façade. The same is visible for the vehicle standing in front of the building (lower left). When the backward view texture and the forward view texture are subtracted, these geometrical errors remain (Fig. 9). Together with a region growing on the strong image gradients, most of the windows are located (Fig. 10). For all detected windows, the smallest rectangle is searched that includes all pixels assigned to that window. Minimum and maximum size is defined to avoid groups of windows to be connected as one single window.

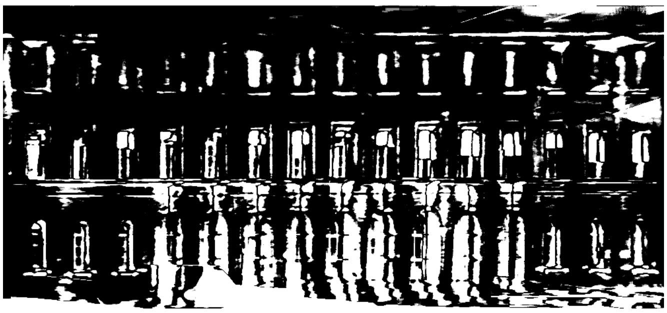

Figure 9. Filter mask. White parts are ignored for the leakage search

Figure 10. Region growing result. All grey parts belong to disturbing objects. Only the white parts are used in the further processing

In the remaining image parts a new gradient search is done to find remaining edges. After a region growing these areas are detected as leakages. For every leakage its center of gravity on the texture is calculated and transformed to 3d coordinates like explained in section 5. Figure 11 shows the image parts marked as leakages.

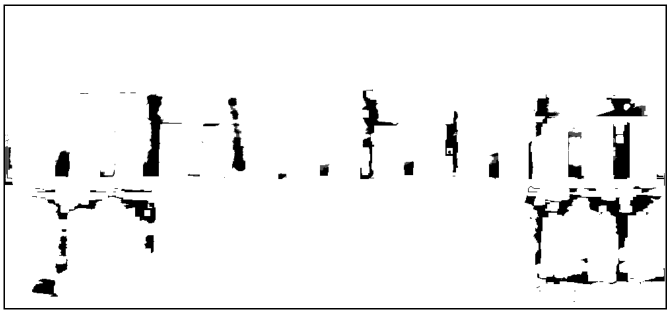

Figure 11. Leakage areas marked in black

VII. DISCUSSION

Infrared light has special characteristics that lead to many problems. The physical behavior of the IR spectrum causes camera systems with lower resolution than normal video or photo cameras. Many things to be seen in one infrared image, cannot be seen in another because the conditions have changed and many façade structures that can be seen in visible cannot be seen in infrared and vice versa. In addition, the small viewing angle and resolution of the camera does not allow the record of a complete building in dense urban areas with narrow streets in one image. As mentioned in chapter 4, the position estimation of the camera is a major task. The extraction of the estimated camera path and point cloud is limited to facades without disturbing objects like trees which generate wrong 3d points and can cause a wrong matching to the building model. The oblique view leads to different speeds of the feature points. At the right side, they move very slowly and so there position is estimated only very imprecise. The points on the left are moving quite faster and thus are estimated much more adequate. Points in the first floor thus are much more adequate than points in the upper floors. This can cause errors in the vertical alignment and camera orientation. This leads to errors in the texture projection. In the later steps the oblique view causes differences in the texture resolution that influence the quality of the window removement and leakage detection.

Most of the leakages were detected for the samples. Problems occur for inhomogeneous façade structures and for leakages in direct neighborhood to windows. Both areas are marked out and not taken into account for the leakage search. Outer influences like sun shadows can cause additional edges in the images and outshine relevant structures. Best results can be achieved before sunrise without snow and the trees without foliage.

Further improvements are possible to remove the windows from the texture. One possible solution is given in [19]. A more stable detection of the leakages is possible by combining several textures of the same façade. Even if the temperature of the façade itself is changing, there remains a radiation difference between the background of the façade and the leakages. The actual state of the quality of the textures is shown in figure 12.

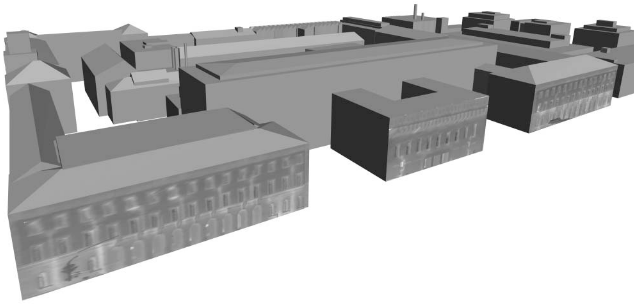

Figure 12. View of a group of textured facades

ACKNOWLEDGMENT

Our work is part of the DFG (German Research Society) research project "Enrichment and Multi-purpose Visualization of Building Models with Emphasis on Thermal Infrared Data"as part of the bundle project "Interoperation of 3D Urban Geoinformation (3DUGI)". The authors thank Dr. Clement, Dr. Schwarz and Mr. Kremer of FGAN-FOM, Department Optronic Systems Analyses, Ettlingen, for their assistance during the recording campaign.

REFERENCES

[1] G. Gröger, T.H. Kolbe, A. Czerwinski, C. Nagel, "OpenGIS City Geography Markup Language (CityGML) Encoding Standard, Version 1.0.0", International OGC Standard Open Geospatial Consortium, Doc. No. 08-007r1 2008.

[2] H. Mayer, "3D Reconstruction and Visualization of Urban Scenes from Uncalibrated Wide-Baseline Image Sequences", Photogrammetrie – Fernerkundung – Geoinformation, 3/07:167–176, 2007

[3] M. Heinrichs, O. Hellwich, V. Rodehorst, "Robust Spatio-Temporal Feature Tracking", XXI Congress of the Int. Society for Photogrammetry and Remote Sensing (ISPRS), Int. Archives of Photogrammetry and Remote Sensing, vol. XXXVII, Beijing, China, July 2008, 2008

[4] C.P. Lo, D.A. Quattrochi, "Land-Use and Land-Cover Change, Urban Heat Island Phenomenon, and Health Implications: A Remote Sensing Approach", Photogrammetric Engineering & Remote Sensing, vol. 69(9), pp. 1053–1063, 2003.

[5] M. Pollefeys, D. Nistér, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, and H. Towles, "Detailed real-time urban 3d reconstruction from video", International Journal of Computer Vision (IJCV), 78(2-3):143-167, 2008.

[6] L. Koskeleinen, "Predictive maintenance of district heating networks by infrared measurement", Proc. SPIE, vol. 1682, pp. 89–96, 1992.

[7] M. Klingert, "The usage of image processing methods for interpretation of thermography data ",17th International Conference on the Applications of Computer Science and Mathematics in Architecture and Civil Engineering, Weimar, Germany, 12-14 July 2006.

[8] R.M. Haralick, C.N. Lee, K. Ottenberg, M. Nolle, „Review and analysis of solutions of the 3-point perspective pose estimation problem", IJCV, vol.13(3), pp. 331-356, 1994.

[9] L. Quan, L.D. Lan, "Linear n-point camera pose determination", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.21(8), pp. 774-780, 1999.

[10] B. Triggs, "Camera pose and calibration from 4 or 5 known 3d points", Proc. International Conference on Computer Vision (ICCV'99), 1999.

[11] R.M. Haralick, H. Joo, C.N. Lee, X. Zhuang, V.G. Vaidya, M.B. Kim, "Pose estimation from correspondence point data", SMC, vol.19(6), pp. 1426-1446, 1989.

[12] R.L. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, ISBN: 0521623049, 2000.

[13] H.C. Longuet-Higgins, "A computer algorithm for reconstruction a scene from two projections", Nature, vol.239, pp. 133-135, 1981.

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

[15] L. Hoegner, H. Kumke, L. Meng, U. Stilla, "Automatic extraction of textures from infrared image sequences and database integration for 3D building models", PFG Photogrammetrie Fernerkundung Geoinformation. Stuttgart: Schweizerbartsche Verlagsbuchhandlung. 2007(6): 459-468, 2007

[16] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision 60(2):91-110, 2004

[17] M. Fischler, R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", Communications of the ACM 24: 381–395, 1981

[18] J.D. Foley, "Computer Graphics: Principles and Practice", Addison-Wesley 1995, ISBN 02-018-4840-6

[19] S. Reznik, H. Mayer, "Implicit Shape Models, Model Selection, and Plane Sweeping for 3D Facade Interpretation", Stilla U et al (Eds) PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (3/W49A): 173-178, 2007