CASE STUDY OF THE 5-POINT ALGORITHM FOR TEXTURING EXISTING BUILDING MODELS FROM INFRARED IMAGE SEQUENCES

Hoegner Ludwig, Stilla Uwe

Technische Universitaet Muenchen, GERMANY - Ludwig.Hoegner@bv.tu-muenchen.de, stilla@tum.de

Commission III, WG III/4

KEY WORDS: Urban, Image, Acquisition, Texture, Extraction, Thermal, Infrared

ABSTRACT:

Today, thermal inspections of buildings are normally done in recorded single infrared images directly. Thus, no 3d references of found objects and features like i.e. heating pipes or leakages is possible. In computer vision several techniques for the extraction of building surfaces and surface textures from optical images have been developed during the last years. Those algorithms like i.e. 3-point matching, surface estimation via homography or the 5-point algorithm introduced by Nistér are specialised for optical images with their strong edges and high resolution. In this paper, the 5-point algorithm introduced by Nistér is adopted for the extraction of textures from infrared images sequences for an already given building model. Special problem caused by the physical behaviour of the infrared spectrum and the technical limitations of the cameras will be discussed including their influence on the usability of the matching algorithm.

1. INTRODUCTION

One focus in today's discussion of global warming and climate lies on thermal inspection of single buildings on the one hand and urban environment on the other hand. With ground cameras the irradiation of building facades can be investigated (Klingert, 2006) analyzing infrared images. Airborne IR-systems are applied for vehicle detection (Hinz and Stilla, 2006, Stilla and Michaelsen, 2002) or exploration of leakages in district heating systems (Koskeleinen, 1992). Satellite images are used for the analysis of urban heat islands (Lo and Quattrochi, 2003). Janet Nichol (Nichol and Wong, 2005) first introduced a method to integrate 3d information and infrared images. Satellite IR data are combined with simplified block models of building in a 3d city models. The satellite images however only allow to assign a building roof a temperature but not to search for structures on the roof. Façades remain almost invisible.

To inspect and analyze the thermal behaviour of building façades in detail, it is necessary to record them with ground based cameras. In difference to airborne and satellite images, ground images normally do not contain a complete building in a single image. Therefore, it is necessary to combine several images to extract the complete texture for a façade. This combination needs the knowledge of the parameters of the camera used for the record to correctly project the images into the scene.

Techniques for position estimation, matching and scene reconstruction have been in use in image processing of optical images for a couple of years. The estimation of exterior orientation from a single image works with at least 3 correspondences (3-point algorithm) between image and model (Haralick et al, 1994). Techniques for 4- and 5-point estimation are elicited by Quan (Quan and Lan, 1999) and Triggs (1999). For 6 and more correspondence points the Direct Linear Transformation (DLT) can be applied (Triggs, 1999). For homogenious façade structures that approximately form a plane, homography can be adopted to detect planes in image pairs and the relative exterior orientation of the camera in relation to these planes (Hartley and Zisserman, 2000). Another popular strategy working on image pairs is Nistér's 5-point position estimation (Nistér 2004). This algorithm searches for pairs of points of interest in image pair like the homography, but can handle several planes visible in the image pair. Due to the small field of view, the low spatial resolution of the IR images and the low level of detail of the given building model, only few point correspondences between IR image and 3D model can be identified. Strategies based on the orientation of the image sequence itself like homography and Nistér are more useful for the given scenario.

This link between 3d building models and infrared image sequences allows dealing with the analysis big building complexes that cannot be observed in one single image. By integrating the infrared image data and the 3d model data, it becomes possible to assign infrared information to a building and store them together in a GIS database. Images taken with different aspect ratio, from different IR bands or taken at different time can be combined for analysis. Several effects of warming and cooling of façades can be described including 3d model information, i.e. shadows caused by occlusion.

This paper focuses on the usability of the 5-point algorithm for image sequences with constant viewing direction, low contrast and low resolution and the integration of a given building model. Surface hypotheses are not used to create a building model from the images and image sequences, but are used to match the relative oriented scene generated by the 5-point algorithm with a given building model from a GIS database. We will concentrate on the evaluation of the quality of the relative estimated orientation of cameras and estimated extracted building façades according the given building model of the GIS database and the measured path of the recording infrared camera.

In chapter 2, the 5-point position estimation is briefly described and the special behaviour and problems with recording building façades in the infrared spectrum are introduced together with the a short look at the used camera as well as the given building

model. Chapter 3 presents the application of the 5-point algorithm with infrared image sequences recorded as described in chapter 2 and the strategy for quality comparison of the given building model's façades and the estimated surfaces of the 5-point algorithm as well as the comparison of the measured camera path and the estimated camera path of the algorithm. In chapter 4 there are given some experimental results and quality measurements using Nistér's position estimation and chapter 5 finishes up with a conclusion.

2. NISTÉR'S 5-POINT ALGORITHM, THE SPECIAL BEHAVIOUR OF INFRARED LIGHT AND CAMERAS

2.1 A Short introduction to Nistér's 5-point algorithm

Nistér's 5-point algorithm was developed as an efficient solution to the relative pose estimation problem of a camera between two calibrated views using 5 corresponding image points. From images only, it is possible to reconstruct only the relative orientation of the image pair and thus the relative position of the corresponding points and the cameras of the views can be determined. The scale of the scene cannot be reconstructed as well as of course the absolute positions. This limitation to a relative and unscaled orientation is one of the main problems for the integration in a given building model. The algorithm uses a hypothesis generator within a random samples consensus scheme (RANSAC) (Fischler and Bolles, 1981). The precondition of intrinsic calibration of the camera given an improvement of the accuracy and robustness, especially for the special case, the algorithm is used for in this paper. The calibration of the camera minimizes problems with planar scenes and building façades normally appear planar. Without calibration the methods fails in coplanar scene points as there remain many correct solutions. Using not only image pairs but image triplets, the RANSAC scheme with the 5-point algorithm resolves all ambiguities. One precondition is a sufficient change in the observed scene between the images which is normally achieved by changing the camera position and viewing direction. A detailed mathematical description of the recovering of the translation and rotation of the second and third view corresponding to the first view, can be found in Nistér (2004).

2.2 Recorded infrared image sequences

Current IR cameras cannot reach the optical resolution of video cameras or even digital cameras. Like in the visible spectrum, the sun affects infrared records. Images in the mid-wave infrared are directly affected as in addition to the surface radiation caused by the building's temperature the radiation of the sun is reflected. 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 to get an acceptable texture resolution. The image sequences were recorded with a frequency of 50 frames per second. The viewing angle related to the along track axis of the van was constant. Figure 1 shows a set of images from the sequence. The position of the camera was recorded with GPS with an accuracy of 2-5 meters and, for quality measurements from tachymeter measurements from ground control points.

Fig. 1: Images of one test sequence showing the angular view and the camera movement.

2.3 Description of the given 3d building model

The information extracted from the infrared image sequences has be assigned to the corresponding building in a GIS database. To link extracted façade textures and GIS database, the given polygonal building model stored in the database is taken. This model is given in LOD 2 and represents the façades as one polygonal surface with the vertices in world coordinates.

3. AUTOMATED EXTRACTION OF SURFACES AND TEXTURES

3.1 Application of the 5-point algorithm on infrared image sequences

Nistér's 5-point pose estimation can be used to extract point clouds from image triplets and a relative camera path. Those point clouds can then be used to estimate surfaces in a scene observed from several images. Mayer (Mayer 2007) has introduced an approach for wide-baseline image sequences. In this approach, Förstner points (Förstner and Gülch, 1987) are matched via cross-correlation. RANSAC is used with the RANSAC scheme of Chum et al. (2003) for the estimation of the fundamental matrix F and trifocal tensor T of the image triplet. The found inliers are used for a robust bundle adjustment (Hartley and Zisserman 2003). To orient the whole image sequence, the triplets are linked based on homographies and already known 3d points of the already oriented sequence part. For the reconstruction of planes from the point clouds, vanishing points are detected for groups of images. Because building façade are often vertical, the medians in x- and y-direction can be takes as the vertical direction. Planes are searched defining a maximum distance of a point to a plane. The best plane is the plane with the smallest distance to a hypothesized plane. From the plane parameters and projection matrices homographies are computed between the planes and the images.

All images are recorded from a moving vehicle with the same viewing direction. This means, that the camera is only moving along a path. In a first glance, this seems to be a simplification. But, although the angle between the camera and the moving direction is constant, there are changes in the viewing direction caused by the movement of the vehicle. So the viewing direction cannot be seen as fixed. For the reconstruction of the

3d point cloud from the image sequence three different cases of façades have to be investigated.

1. The first façade type is standing parallel to the street and the camera is moving along this façade. These façades fulfil the conditions for the 3d point extraction, because the points are moving through the image from the right to the left and so their relative 3d position can be determined from their movement.
2. The second class on façades is the class of occluded invisible façades. They, of course remain invisible as holes in the generated point cloud.
3. The third class of façades is standing perpendicular to the street. Those façades are not moving along the camera, but are changing their scale as they are moving towards the camera. For those façades, the movement of points is very low and thus it is very difficult to estimate the correct coordinates.
4.

In addition to the limitations in the viewing position of the camera the low resolution and low grade of details cause a small number of points of interest.

3.2 Description of the quality measurements

It is necessary to make a quality investigation including the estimated camera path, the position and completeness of façades and the extracted textures of the façades in comparison to the given 3d building model of the GIS database and the recorded GPS camera path. For all images and GPS positions a synchronized time-code is stored. The camera path of the GPS is corrected using a Kalmann filter and interpolated for every time-code corresponding to an image of the sequence. Now, the first and the last position of the estimated camera from the 5-point pose estimation are moved to the interpolated positions of the GPS path with the corresponding time codes. This leads to a scale, rotation and translation of the estimated model onto the given 3d model and the GPS path.

The estimated planes are now transformed to the world coordinate system. The scale factor is calculating comparing the length of the given camera path and the length of the estimated camera path for the first and last camera position. Afterwards, a 3d rotation and translation is calculated to rotated and move the estimated camera path onto the given measured one. Because of the assumption that the GPS path is not afflicted with an error, the corresponding surfaces of the given building model and the generated planes are given by the smallest error in their orientation and the smallest translation vector of their barycenters. To avoid a systematic error caused by the GPS position, the translation vectors of the barycenters of all generated planes to their corresponding model surface are used to calculate a mean translation vector. The remaining translation vectors of the planes to the surfaces are the remaining positioning errors of the planes, the generated mean translation vector is used to move the generated camera path of the image sequence. The distance between this path an the GPS path can be caused by the 5-point algorithm, an incorrect camera calibration or the inaccuracy of the GPS positions.

In addition to the positioning error of the planes, the completeness is a criterion of the quality of the surface reconstruction. For every estimated plane its length and height are determined calculating its bounding rectangle and compared to the length and height of the given corresponding façade. Caused by the reconstruction from points of interest, the generated planes are estimated to be a little bit smaller than the original façade.

4. EXPERIMENTS

The camera that was used for the acquisition of the test sequences offers an optical resolution auf 320x240 (FLIR SC3000) pixel with a field of view (FOV) of only 20°. The SC3000 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. Different scenarios of image sequences were acquired. The first scenario (Figure 2a) deals with several small façades belonging to one building block. The second (Figure 2b) shows a long façade with regular structure and a specific entry with overlap. The third scene (Figure 2c) shows a long façade with different structures. The forth scenario (Figure 2d) deals with bridges between buildings crossing a street.

Fig. 2: a) several façades with occlusion, b) façade with regular structure, c) façade with irregular structure, d) façade with building bridge

For scenario 2 and 3, the situation is quite different. Both scenarios consist of only one long façade. This façade can be estimated quite easy because there are no occlusions of the façade and the complete façade is almost in one plane. Scenario 4 has to deal with two building bridges crossing the street. Those bridges are separating the façade into segments. The bridges cannot be detected.

5. RESULTS

5.1 Extracted surface planes and camera path

In general, the surface estimation works well for façades going parallel to the street. Façades standing perpendicular to the street are much more difficult to extract correctly. For the scene in figure 2a, three parallel surfaces can be extracted. For these surfaces, the relative camera path containing relative camera positions for all images of the sequence is generated. The result of the automated extraction is shown in figure 3.

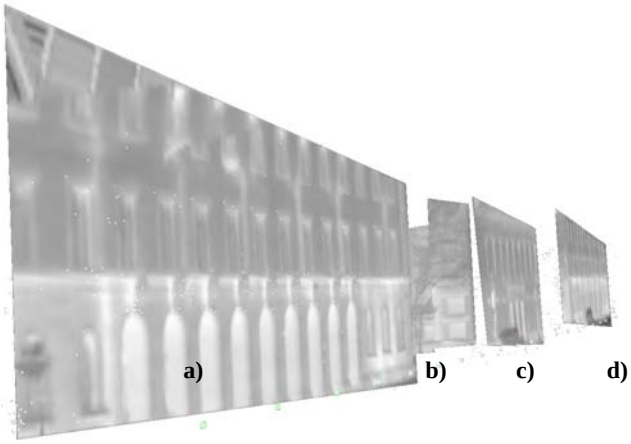

Fig. 3: Extracted surfaces of the sequence shown in figure 1 and 2a) with extracted textures. The small dots show the estimated camera positions.

It can be seen that the textures on the surfaces are looking tilted in movement direction. This is caused by the angular view including a little upwards looking angle that is not correctly estimated and leads in fact to a tilt of the surfaces. The textures are then extracted correctly from the image sequence and pasted on the tilted surfaces. This error is proportional to the up looking angle of the camera. Another interesting fact is the perpendicular façade (Fig. 3b). Although a tree is standing before the façade, the algorithm could estimate a surface. In this case, even two surfaces were detected. The right part of the façade shows two non occluded windows and was estimated as façade. The position error is caused by the nadir view of the camera to the façade causing only rough positioning information. The left part was detected at the position of the tree. Most of the texture of this part shows the trunk and branches of the tree. Between the second (Fig 3c) and the third (Fig 3d) big façade at the right, no façade can be seen. In the original image sequence (Fig. 1 and 2a) one can see a tree and in addition some flag staffs. This produces too many wrong pairs of points of interest to detect the surface plane.

5.2 Comparison with the given 3d model and measured camera path

The façades are not extracted in their correct size. In figure 3, at the left corner the façade textures shows only two windows left of the bright entrance gallery. In fact, there must be three. The façade does not contain its roof and floor borders. This is on the one hand caused by the tilt; on the other hand, the surface is estimated as to small. This is caused by the point extraction density, which leaves a small band around the façade unexplored because the points of interest are found at the façade and not at the border. Only the right border of the façade is found correctly because of the high number of points of interest. These points are extracted from the strong image border between the façade and the tree. In the up direction, the façades do not contain a tilt. But this is clear, because the vertical direction is set fix in the algorithm and so the estimated planes are forced to be almost vertical. Figure 4 shows a front view of the building with both the measured and estimated model.

The image sequence starts at the left. The planes are shifted with about 3 meters to the right. The size of the left façade plane is almost the correct size of given façade except for the tilt in the moving direction of the camera (see Fig. 3). The central façade is also detected with almost the correct size, but shifted. The small rectangular plane at the left edge of the central façade is the wrong detected surface belonging to the tree (see Fig. 3). For the right façade, the height is estimated almost correctly but the plane is stretched in moving direction. This is caused because the perpendicular small façade (Fig. 3 at the right) is not detected and so some points of this small façade are added to the front façade. Except this, this façade shows only the shift. This shift is caused by the errors of the given GPS position of the camera. This error can be seen in figure 5.

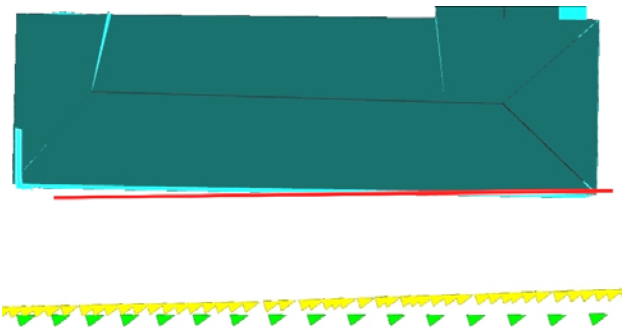

Fig. 5: Top down view of a building façade with measured GPS camera path (yellow) and estimated camera path (green). The estimated surface plane is red.

It can be seen that the measured camera path from the GPS is leading towards the façade. In fact, the camera path must be parallel to the façade, so this is an error in the GPS positioning. When the estimated camera path is moved to the GPS path, there remains the shift error as seen in figure 4 and this angular error. When including the estimated surface planes in the positioning, the results are looking like in figure 5. The estimated camera path is almost parallel to the façade and the estimated surface plane is almost parallel to the model façade. Another aspect is the distance between the camera and the building which is estimated very good with an error of about half a meter for this scenario. The angular error of the GPS

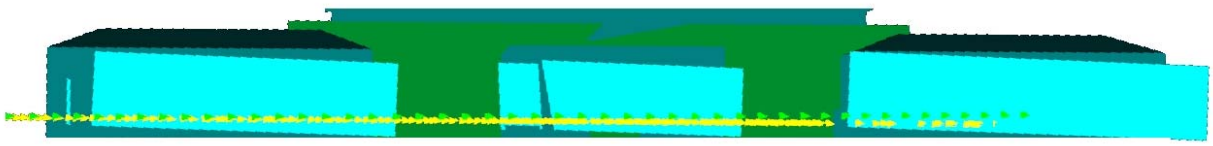

Fig. 4: Front view of the building. Dark blue and green: Measured model, Light blue: estimated façade planes, Yellow: measured camera path, Light green: estimated camera path

oriented scene is about 4 degree and can be reduced to about 1.5 degree. In addition to a global adjustment, every estimated surface plane can be corrected locally. So, for this scenario with several façades, only the façade at the front of the street can be extracted. Other façades are at least partially occluded and can not be estimated correctly.

5.3 Comparison to other scenarios

In scenario 2 and 3, the façade is tilted in moving direction caused by the small upwards viewing angle of the camera. The distance to the building is adequate with about 1 meter and the shift is in moving direction is about 4 meters. Both values are somehow higher because the GPS coordinates are afflicted with a bigger error. This is caused by the record situation. Whereas in scenario one, you can find an open park landscape on the opposite side of the building, in scenarios two and three, there are buildings on the opposite side occluding the satellite signal and so leading to a bigger positioning error. For scenario two, there is a special case (see Figure 6). The first part of the façade has a very regular structure and thus the algorithm finds many wrong pairs of points of interest (Fig. 6a). This causes long computation steps and wrong planes estimations. This effect depends on the points chosen by RANSAC and produces changing hypotheses for every computation. A second problem of this façade is the gateway at the beginning (Fig. 6b). This gateway is standing back from the façade generating points of interest behind the surface plane. Together with pillars in front of the gateway, which are standing in the surface plane, a point cloud is computed with points jumping between the surface plane and the gateway plane. In this area, no surface plane can be estimated.

Fig. 6: Scenario two, a) regular façade structure causing wrong point pairs, b) gateway behind the façade plane with pillars in front.

In scenario 4 only parts of the façade can be extracted due to the occlusion caused by the bridges. The façade parts before and after the bridges can be estimated to about 70%. The algorithm aborts when the bridges cover more than 40% of the image. The positioning errors for the extracted planes are 5 meters along the façade and 2 meters to the correct façade plane.

6. DISCUSSION

Over all scenarios, some facts can be integrated. For all scenarios, the upward angle of the camera view cuased a constant tilt of 6 ° in the found surface planes. Façades parallel to the moving direction that are not occluded in parts can be extracted with about 85 to 90 % when the façade has an almost plane structure. The positioning errors are in the range of the accuracy of the GPS. The scenario with the bridges can not be extracted with appropriate quality, but this is a very special case.

An overview over the errors in the different scenarios is given in table 1.

Scene	**1**	**2**	**3**	**4**
Pose error model				
X	3 m	4 m	4 m	5 m
Y				
Z	½ m	1 m	1 m	2 m
Tilt error	6°	6°	6°	6°
Pose error GPS	3 m	4 m	4 m	5 m
Angle x-z-plane error	4 ½ °	4 °	4 ½ °	6 °
Completeness façade	90 %	85 %	90 %	40 %
Completeness model	55 %	80 %	90 %	40 %

Table 1: Errors in the different scenarios: Pose error model: Error of the estimated position to the given model in X, Y and Z axis, Tilt error: Tilt of the planes caused by upward viewing angle of the camera, Pose error GPS: mean error of GPS position to corrected estimated position, Angle X-Z plane error: Error in the estimated camera path direction, Completeness façade: Percentage of completeness of a detected façade, Completeness model: Percentage of reconstructed façade surface of the hole scene including non-detected surfaces

7. CONCLUSION

The used algorithm was originally designed for high resolution images that show an object from different positions and viewing angles. The results show that the 5-point-algorithm can also be used for image sequences with constant viewing direction, if there are façades parallel to the moving direction. Façades observed in nadir view do not show enough movement of points of interest to extract surface planes. The images taken for the extraction do not need to be given in a high resolution. A small resolution as given in the scenarios of this paper often contains enough information to extract sufficient points of interest. Problems occur if the façade contains only few structure and if the structure is regular. When a façade is standing along the street with only little occlusion, the façade can be reconstructed with up to 90 %. For special façades like mentioned in chapter 4 the algorithm is not robust enough and reconstructs only small parts of the façades.

The algorithm does not only provide estimated surfaces but also an estimated camera path. Results show, that it is possible to reduce the position error caused by the GPS by combining the GPS and the given 3d model with the estimated surface planes and the estimated camera path. The estimated camera path is afflicted by a much smaller error than the GPS measurement. The GPS position are only necessary for the scaling, rotation and translation of the estimated model to fit the given 3d model. The position refinement is then done without the measured camera path comparing the positions of the façades and the surface planes.

8. OUTLOOK

Two further steps are now possible: The corrected camera path can be used to extract the infrared textures from the image sequence. Or, the 3d model surfaces are included in the plane estimation process to precise the estimated planes and thus

extract the correct infrared textures for the given building model. So, two main conclusions can be made.

First: The estimated model consisting of relative surface planes and a relative camera path can be used to refine the measured camera path.

Second: The estimated model can be used to link the given 3d model to the point cloud extracted from the image sequence to refine the surface plane estimation.

Both strategies allow improving the reconstruction of building parts, which can not be reconstructed directly. A better camera position leads to a better projection and a better surface plane estimation leads to a better assignment of points of the point cloud. Further investigation will concentrate on those aspects.

ACKNOWLEDGEMENTS

This 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" (STI 545/1-2) 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, Ettlingen, for their assistance during the recording campaign. And special thanks to Prof. Helmut Mayer for providing his software of the 5-point algorithm and plane estimation.

REFERENCES

Chum, O., Matas, J., Obdrzalek, S., 2003. Epipolar geometry from three correspondences. In: Drpohlav, O. (ed.): Computer vision – CVWW´03, Czech Pattern Recognition Society, Prague, pp. 83-88.

Fischler, M.A., Bolles, R.C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*, vol24(6), June 1981, pp. 381-395

Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In: ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305.

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

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

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

Hinz, S., Stilla, U, 2006. Car detection in aerial thermal images by local and global evidence accumulation, *Pattern Recognition Letter*, vol. 27, pp. 308-315

Klingert, M., 2006. 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

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

Lo, C.P., Quattrochi, D.A., 2003. 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

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

Nichol, J., Wong, M.S., 2005. Modeling urban environmental quality in a tropical city, *Landscape and urban Planning,* vol.73, pp.49-58

Nistèr, D., 2004. An efficient solution to the five-point relative pose problem, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 6, pp. 756-777, Jun. 2004

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

Stilla, U., Michaelsen, E., 2002. Estimating vehicle activity using thermal image sequences and maps, Symposium on geospatial theory, processing and applications. *International Archives of Photogrammetry and Remote Sensing*, vol.34,Part 4

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