AUTOMATED GENERATION OF BUILDING TEXTURES FROM INFRARED IMAGE SEQUENCES

Ludwig Hoegner, Uwe Stilla
Institute of Photogrammetry and Cartography, Technische Universitaet Muenchen, Munich, Germany
Ludwig.Hoegner@bv.tum.de; Stilla@tum.de

KEY WORDS: Image Analysis, Image Sequences, Texture Extraction, 3d Building Model, Thermal Infrared, Terrestrial, Close Range, Urban

ABSTRACT:

In this paper the application of computer graphics and computer vision for texture extraction from infrared image sequences is described. These techniques normally are used in computer graphics to project a virtual scene onto the image plane. For the matching between the images and the given 3D model a strategy is presented based on the estimation of planes in the image sequence using homography. Textures are extracted using an algorithm based on the principles of ray casting to generate partial textures for every visible surface in every single image of the sequence. The textures generated from different images of the sequence belonging to the same façade are combined. By combining the intersection points of several images generated during the ray casting, a dense structure of points can be used to texture and analyse even big building complexes.

1 INTRODUCTION

The paper intends to investigate strategies for a detailed texture mapping of building façades which allow a thermal inspection. Typically, thermal inspections of building façades are carried out in single images from the observed objects. Larger building parts require several images to be analysed. An integral way of viewing buildings recorded from different images is difficult without combining those images. This problem is getting worse, when images from different cameras or views need to be combined and stored for further processing without any geometric reference.

The discussion of global warming and climate has focused on thermal inspection of single buildings on the one hand and urban environment on the other hand. Ground cameras are recording the irradiation of building facades (Klingert, 2006), for the specification of its thermal behavior. Satellite images are used for fire detection (Siegert, 2004), vegetation monitoring (Quattrochi, 1999) or the analysis of urban heat islands (Lo, 2003). Airborne IR-systems are applied for vehicle detection (Hinz, 2006, Stilla, 2002) or exploration of leakages in district heating systems (Koskeleinen, 1992). Typically, the analysis of the urban environment is done on the ground but without 3d buildings. A relatively new approach is developed by Janet Nichol (Nichol, 2005). For urban environmental quality studies, satellite IR data were combined with 3d city models. In IR images are projected onto the terrain and the building surfaces involved in the energy exchange were matched with the IR images, including vertical walls and the horizontal surfaces of the buildings seen from the satellite. This textured model permit 3d visualization for a better understanding of the factors controlling urban environment. But there is no detailed analysis of the single buildings.

In difference from areal 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. 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 (Triggs, 1999). For 6 and more correspondence points the Direct Linear Transformation (DLT) can be applied (Triggs, 1999). For homogene 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). 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. So this paper is focused on the homography based surface estimation. The extraction of surface textures from ir images is done by a reversed variant of ray casting. Instead of calculating the pixel color from rays that collect color values in the 3d scene, the rays project the known pixel values on the intersection points of the rays and surfaces in the 3d model. The 3d coordinates of the intersection points can be directly used to generate texture coordinates. This texture coordinates are used to generate 2d textures for the surfaces. Textures generated from different images of an image sequence can be combined to achieve a higher resolution and completeness of the façade texture.

2 DATA ACQUISITION

As the wavelength of infrared light is much longer than the visible spectrum, other optics and sensors are necessary to record infrared radiance. Current IR cameras cannot reach the optical resolution of video cameras or even digital cameras. The cameras that were used for the acquisition of the test sequences offer an optical resolution of 320x240 (FLIR SC3000) and 320x256 (MerlinMID) pixel with a field of view (FOV) of only 20° (Fig. 1). The SC3000 is recording in the thermal infrared (8 - 12 µm), whereas the MerlinMID is a midwave infrared camera (3 - 5 µm). For the methods used in this paper, both cameras were mounted on a rotatable and shiftable platform on the top of a van. Images in the mid-wave infrared are directly affected by the sunlight 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.

Figure 1: Used camera system: midwave (3-5µm) and the longwave (8-12µm) infrared camera (see the left and middle camera) and video camera right

To inspect the temperature loss of a building, the environment should be cold, where as the heating of the building should be running to observe a relevant temperature flux. Optimal time is after sunset and before dawn in spring. To minimize occlusion caused by vegetation a date before foliation is to be preferred.
Caused by the small field of view and the low optical resolution it is 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. That is why the optical resolution of the images is not constant. Image parts that show the complete building have a lower resolution than image parts that show only the first floor of the building. When several images are combined, the optical texture quality of the final surface texture is high in the first floor and decreases to the top of the façade. Due to this, the image sequences are 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 is recorded with a view forward looking and a view backward looking. The viewing angle related to the along track axis of the van must be constant. The position of the cameras is to be recorded with GPS.

3 TRANSFORMATION OF 2D IMAGE DATA TO 3D POINTS AND TEXTURE COORDINATE

3.1 Position estimation and image matching

The camera parameters must been known in order to project the 3D model onto the image plane. The interior orientation of the camera is determined by a calibration. The position of the camera is recorded during the image acquisition. Pan, tilt and roll angles of the camera are estimated from ground control points given by the vertices of the 3D model. Caused by the high buildings the recorded GPS signal is inadequate and only allows a position accuracy of about 5 meters. This is to inaccurate to use the position directly for the projection and does not give the pan, tilt and roll angles.

Instead of calculating the position for individual key frames, the image sequence is used to orient the images. For homogene 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). Many buildings have an approximately planar façade. Assuming that façade structures lie in a plane, a correspondence between points of two subsequent images can be found using a homography matrix H. Points of interest are found searching eigenvalues and performing a non-maxima-suppression. The homography matrix is then calculated for corresponding points selected by RANSAC (Fischler and Bolles, 1981). From subsequent image pairs with frame distance d a set of homographies are calculated to allow the estimation of a trajectory of the camera relative to the façade plane which is averaged by combining the planes of the homographies. For the first image pair, the initial camera position is used to estimate a façade plane that is close to the corresponding 3d model surface. This strategy works well for homogeneous images with only one façade covering most of the image.

3.2 Selecting the visible objects

Ray casting searches for intersection points between the rays sent through the pixels of the image plane and all surfaces of the 3d scene. Using a typical resolution for an IR image of a video sequence, there are 320 x 240 pixels per image (values of camera FLIR SC3000 used for the data acquisition). This means at least 76800 pixels which have to be checked against all surfaces. The surfaces of the 3d models consist of triangles. A building façade consists of at least two triangles to form a rectangle. In result, for every surface there are at least 153600 intersection points to calculate. When using several rays per pixel to achieve a higher accuracy and avoid holes in the resulting textures, this number gets even higher. It seems reasonable to minimize the number of surfaces, which have to be checked for intersections. To achieve this, frustum culling is used for all polygons of the surfaces of the given 3d model using the camera parameters estimated in the homography. This way, only surfaces being partially within the field-of-view are processed in the projection and texture extraction steps.
A second way to minimize the number of surfaces for ray casting is the back face culling. That means, that surfaces within the field of view but invisible can be removed from processing. Back faces are the back sides of buildings completely occluded by the building's front side. In computer graphics the distinction between front and back faces is made by the order of the vertices defining a triangle. Unfortunately, the models that are used for the test in our project, have not

always correctly defined triangles to distinguish between front and back faces. That is why back face culling is not used in this project, but of course can be used for 3d models with correct back face definition.

3.3 Projection of image pixels into the 3d scene

Depending on aspect and position of the camera view, parts of the buildings are invisble due to self occlusion or occlusion from other objects. For every pixel of the image of the virtual camera the corresponding surface point is searched. First, for every surface a plane equation is calculated to receive the depth value and texture coordindates. This plane is used for a ray casting (Foley 1995), where every pixel is projected into the scene and assigned to an intersection point of the plane which has the smallest depth value along the ray from the camera through the pixel. For the intersection point of this plane texture, the texture coordinates are interpolated from the texture coordinates of the vertices of the plane. The texture coordinates count from (u,v) = (0,0) at the left lower vertex of the surface and are going up to (u,v) = (1,1) at the upper right vertex. The ID of the intersected surface and the texture coordinates of the intersection point are returned to the pixel of the infrared image. After this 3D to 2D transformation of the model surfaces into the image plane of the IR image a 2D transformation is carried out to transform the IR image pixels to texture coordinates of the model surfaces. After this process the pixel values of the IR image have been transformed to texture coordinates of points on the model surfaces.

A further 2D transformation is conducted to transfer the texture points to pixel coordinates for the surface textures. At first, the individual pixel coordinates of the texture are transferred into texture coordinates of the surface. Then, their interpolated values are calculated by means of a bilinear interpolation. If a pixel has only three surrounding points in u and v direction, the pixel value is interpolated using barycentric coordinates. If the pixel is outside the triangle defined by the three surrounding points, it is outside the visible part of the façade. Pixels with only two or less surrounding texture points also are outside the visible part of the façade and left in black.

3.4 Combination of the partial textures of one surface

For composing a complete texture for a building façade, it is necessary to combine several partial textures generated as described in 3.1 to 3.3. Because the camera parameters are containing a small error even after correction, like mentioned in 3.1, the surface textures generated from different images of the same sequence may not be congruent. But, as the input image sequence is recorded with several frames per second, in our case study with 50 frames per second, and the viewing angle related to the along track axis of the van is constant, the disparity of two adjacent images is very small in movement as well as in rotation and so is the possible mismatch between the extracted surface textures of those two IR images.
Due to this, corresponding points in both surface textures are searched within a small area of only several pixels for every point. After two surface textures are matched, the intensity values for the combined texture have to be calculated. For an image sequence, the resolution of the surface textures is not constant. The resolution decreases in the viewing direction of the camera caused by the perspective view. Texture points, which have a bigger distance to the camera, have a lower spatial resolution. When recording forward, the visible part of the façade texture of each following image has a higher spatial resolution than the image before, but does not show the complete part of the image before. Figure 2 illustrates the resolution distribution of a combined surface texture.

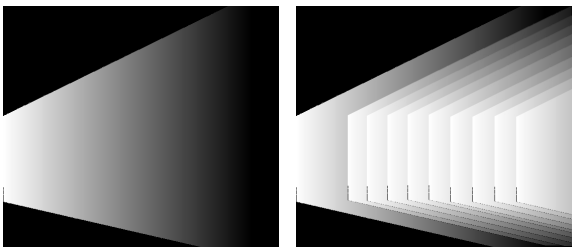

Figure 2: left: Resolution image of a single partial texture, right: Resolution image of a combination of several partial textures

For backwards view, every previous image as a higher spatial resolution, but does not show the complete part of the following image. This fact allows the use of a very simple texture combination method. For forward viewing, initially the first partial surface texture is copied to the combined texture. The second partial surface texture is then copied to the combined texture and overrides the pixel copied from the first texture, that are also present in the second one. But, as mentioned before, all pixels of the second texture have a higher resolution than the first texture. For the first texture, only the pixels that are not part of the second texture remain in the combined texture. This procedure is continued for all partial surface textures of the image sequence. In the end, the combined final texture has the highest possible resolution that can be achieved from the input image sequence. For the backward viewing image sequence, the last partial surface texture is initially copied to the combined texture and then the partial surface textures of the sequence are added in reverse order.
Generating the final texture in this way leads to a texture with the highest resolution for every texel that is extractable from all partial textures. Figure 3 shows a continuous distribution of the resolution of the complete texture of the façade assuming a high image rate and a low velocity.

Figure 3: Resolution image of the final texture

The white region in the middle corresponds to the first floor of the building, which lies in the view axis of the camera and can be seen at the left high resolution edge of all partial textures. Due to the oblique view the resolution decreases downwards and upwards. These texture parts cannot be seen at the left edge

of the images, but more to the right, where the resolution decreases. This effect depends on the angles of the camera view towards the camera path and towards the facades. The bigger the angle between the viewing direction of the camera and the surface normal of the façade, the bigger gets this effect.

Because the appearance of IR signatures is changing over time and depends for example on the weather conditions, the combination of textures from different records at different days or daytimes 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.

The final combined textures of all visible facades are assigned to building surfaces and stored together with the 3d model in a hierarchical object oriented database for further processing like feature extraction. The 3d intersection points are also stored as they can be used later to create geometry if the feature detection in the textures identifies geometry.

4 PROCESSING OF TEST DATA

For the experiments with the camera system mentioned in section 2 a building complex is chosen that shows long façades as well as discontinuous façade structures in narrow streets. The given 3d building model containing the surface polygons of the façades (Fig. 4) is combined with the longwave infrared image sequence of the SC3000 camera (Fig. 5)

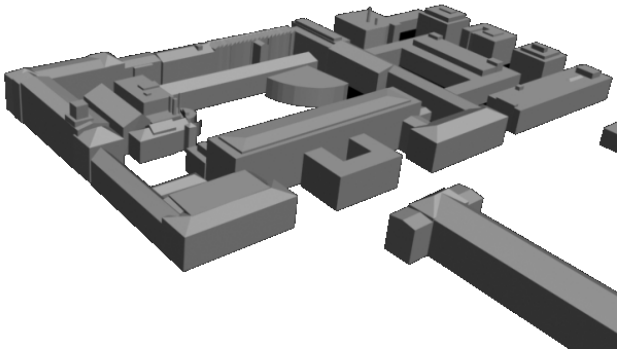

Figure 4: 3d building model with polygon surfaces

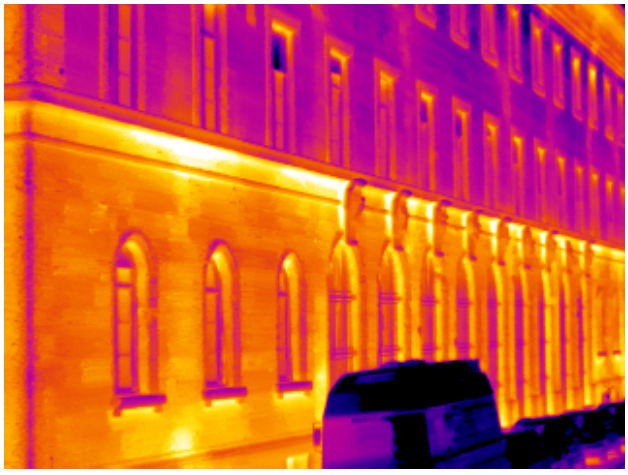

Figure 5: IR image, intensity values are coded as 256 color table

The homgraphy surface estimation leads to a relative camera path and estimated surface for the image sequence (Fig. 6). Using the given GPS coordinates, we can assign each of the estimated surfaces to a polygon surface of the model.

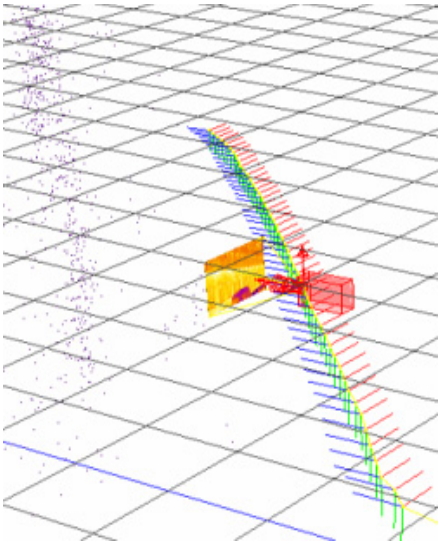

Figure 6: Camera path and point cloud of corresponding points of the image sequence

As mentioned, raycasting is used to determine the visible surface for every pixel and then calculate its texture coordinates. From these texels of one image a partial surface texture for every visible surface is generated. Figure 7 demonstrates the sampling rate and the resolution distribution. To the right, the resolution is decreasing due to the perspective view. Areas of the surface, which could not be seen in one image, are left in black. The black stripes are image pixels, where no texel is transformed to.

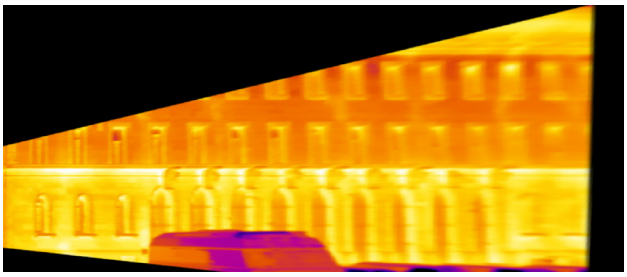

Figure 7: Partial surface texture for the image of Fig. 5, intensity values are coded as 256 color table.

For composing a complete texture for a building façade, for every pixel of the final texture its value is interpolated from the stored partial surface textures. Figure 8 shows a combined surface texture built of twenty partial textures from one sequence in forward view. The black areas from the partial textures are filled from other partial textures. In difference to the partial texture, where the resolution is decreasing from left to right, here the resolution is decreasing from the first floor to the roof and the first floor to the ground, caused by the viewing direction of the camera and the texture combination (see section 3.4 and Fig. 2 and 3).

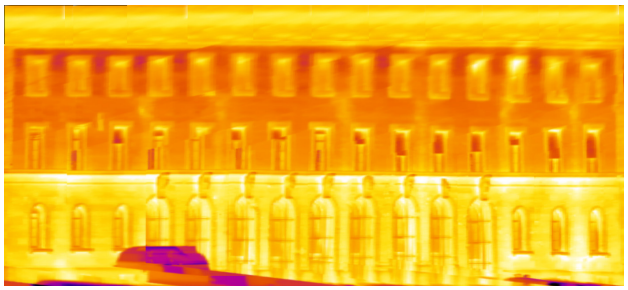

Figure 8: Surface texture generated from twenty partial textures.

For the texture in Fig. 8, no invisible parts remain sue to the fact that there is no occlusion for this façade. Other façades are not completely visible from the sequence so that there remain invisible parts left in black (Fig. 9).

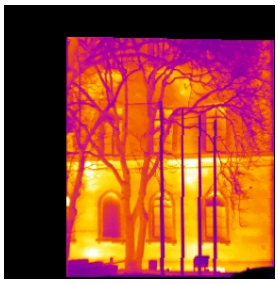

Figure 9: Texture of a partially occluded façade where the black part is not visible in any partial texture for this façade.

5 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. The appearance of a façade in the infrared depends on outer conditions like weather, daytime, sun, temperature and inner conditions like building structure or heating. Many things to be seen in one infrared image, can not be seen in another because the conditions have changed and many façade structures that can be seen in visible can not 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. The oblique view caused by the viewing angle and resolution of the cameras leads to self occlusion of buildings. This is partially solved by recording two different views per façade, one in oblique forward view and one in oblique backward view (Fig. 10). But, there are still remaining parts of buildings that are not visible. The given camera parameters for the projection have to be accurate, because the texture combination can only adjust small offsets between the partial textures. Errors in the rotation of the camera lead to completely wrong intersection points and thus wrong surface assignments and texture coordinates. A strategy based on the 5-point algorithm of Nistèr (Nistèr, 2004) will be investigated to deal with discontinuous façades.

In the strategy mentioned in this paper the resolution of the combined final texture is restricted to the resolution of the partial textures from the image sequence because one partial texture is simply overwriting another. In addition, this strategy is only feasible for the combination of textures of one sequence and with constant view direction. To reach a more general approach the texture coordinates generated during the ray casting from every single image should be combined as texutre coordinate 2d point cloud. With this hole set of texture coordinates the calculated combined texture is interpolated from a much more comprehensive basis and thus could reach a higher resolution than the input textures. A necessary precondition for this strategy is the improvement of the determination of the camera parameters to avoid blurred textures from incorrect texture coordinates.

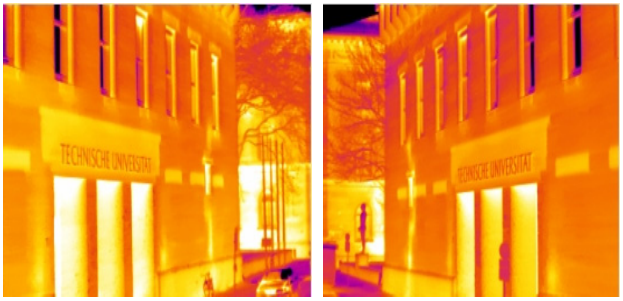

Figure 10: Forward and backward view from two sequences of the same building parts

As seen in figure 8, the windows of the façade can easy be seen and, because the whole façade is in one image, the rows and columns can be identified to search for further windows, which could be detected directly. In further processing steps, for structures like the windows, the intersection points corresponding to the pixels of the window corners can be used to create triangles in the 3d model to represent the window in geometry instead of texture.

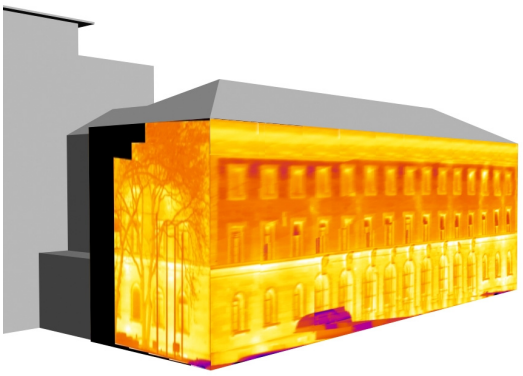

Figure 11: Textures of Fig. 8 and 9 mapped onto the 3d building model of Fig. 4

ACKNOWLEDGEMENTS

The authors thank Dr. Clement, Dr. Schwarz and Mr. Kremer of FGAN-FOM, Department Optronic Systems Analyses, Ettlingen, for equipment and their assistance during the recording campaign.

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“ (STI 545/1-2) as part of the bundle project “Interoperation of 3D Urban Geoinformation (3DUGI)”.

REFERENCES

Fischer, 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

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

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

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

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

Quattrochi, D.A., Luvall, J.C., 1999. Thermal infrared remote sensing for analysis of landscape ecological processes: methods and application, *Landscape Ecology* , vol. 14, pp. 577–598

Siegert, F., Zhukov, B., Oertel, D., Limin, S., Page, S.E., Rielay, O., 2004. Peat fires detected by the BIRD satellite, *International Journal of Remote Sensing*, vol. 25(16), pp. 3221-3230

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)