Archives of Photogrammetry, Cartography and Remote Sensing, Vol. 22, 2011, pp. 215-226
ISSN 2083-2214

DETECTION OF WINDOWS IN BUILDING TEXTURES FROM AIRBORNE AND TERRESTRIAL INFRARED IMAGE SEQUENCES

Dorota Iwaszczuk [1], Ludwig Hoegner [2], Uwe Stilla [3]

[1, 2, 3] Photogrammetry & Remote Sensing, Technische Universitaet Muenchen (TUM), Munich, Germany - (iwaszczuk, hoegner, stilla)@bv.tum.de

KEY WORDS: Infrared, Image Sequences, Texture Mapping, Structure Detection

ABSTRACT: Infrared (IR) images depict thermal radiation of physical objects. Imaging the building façades and the roofs with an IR camera, thermal inspections of the buildings can be carried out. In such inspections a spatial correspondence between IR-images and the existing 3D building models can be helpful. Texturing 3D building models with IR images this spatial correspondence can be created. Furthermore in textures heat leakages can be detected and the heat bridges can be stored together with 3D building data. However, before extracting leakages, the windows should be located. In IR images glass reflects the surrounding and shows false results for the temperature measurements. Consequently, the windows should be detected in IR images and excluded for the inspection. The most common algorithms for window detection were developed for the images in the visual band. In this paper, an algorithm for window detection in textures extracted from terrestrial IR images is proposed. In the first step, small objects have to be removed by scaling down the image (texture). Then in the scaled image, regions are detected using a local dynamic threshold. Morphological operations are used to remove false detections and unify substructures of the windows. For every extracted region, which is a candidate for a window, the center of gravity is calculated. It is assumed that windows on façades are ordered in regular rows and columns. First the points are grouped into rows using histogram of height created from extracted gravity centers. Then masked correlation is used to detect the position and size of the windows. Finally, the gaps in the grid of windows are completed. For the first experiments we use a dataset from densely build urban area captured in Munich, Germany. The IR image sequences were taken from a vehicle driving on the street around the test area. Camera was directed to the building in oblique view. According to the acquisition geometry, no façade could be completely seen in one frame. Therefore, we combine the textures from many frames. To these textures we applied our algorithm for window detection. First results are promising. Applying the method for our test dataset, 79% completeness and 80% correctness could be achieved.

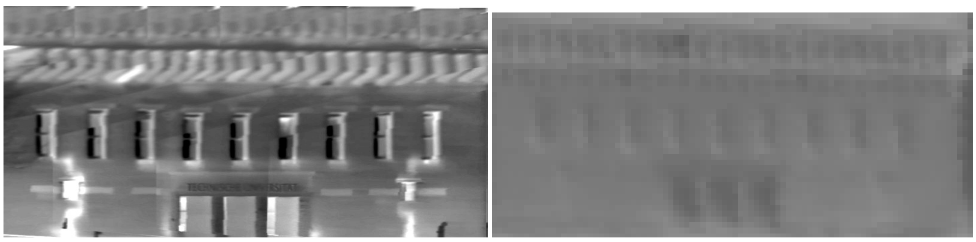
1. INTRODUCTION

Mobile mapping systems are used mainly for a fast acquisition of special data. A mobile mapping system usually consists of a moving platform (vehicle or aircraft), a navigation sensor and a mapping sensor (Li, 1997). Typically for navigation Global Positioning System (GPS) combined with Inertial Navigation System (INS) as well as with other navigation sensors (e.g. vehicle wheel sensors) are used. On one hand GPS provides rough knowledge of position and orientation, on the other hand, an INS yields good short-term accuracy, but in a longer time a systematic drift occurs. However, the combination of GPS and INS allows to avoid the INS drift and to bridge any loss of satellite signal by GPS (Yastikli & Jacobsen, 2005).

Using the information from GPS/INS the images can be geo-referenced directly and matched with existing 3D building models (Frueh et al., 2004; Eugster & Nebiker, 2009; Avbelj et al., 2010). Images combined with a 3D city model can be used in variety applications, e.g. visualization, urban planning, navigation.

1.1 Motivation

In this paper we present an application of mobile mapping system for thermal inspection of buildings. Due to increasing energy costs and climate changes, energy efficiency of buildings became in recent years an important topic. Much effort is required for reducing the energy loss. Sustainable building as well as thermal inspection and monitoring of old buildings contribute to further development saving energy. Thermal radiation of the building hull can be captured by infrared (IR) thermal cameras. In these images weak spots can be detected. In such inspections a spatial correspondence between IR images and existing 3D building models can be helpful. Usage of the 3D building models enables processing of image sequences taken by a mobile platform capturing whole districts or cities. The correspondence can be created by geo-referencing of the images. This can be done directly using GPS/INS data from the devise mounted with the camera on the mobile platform. After system calibration the 3D model can be projected into the image and for each surface of the model a region of the image can be selected for texture. Using terrestrial image sequences from a camera mounted in a vehicle frontal faces can be captured, while airborne image sequences can be taken for roofs and inner yards. The IR cameras have smaller focal plane array compared to cameras working in visible (VIS) domain. To achieve geometric resolution comparable with VIS cameras, a smaller viewing angle has to be used. Therefore, on one hand, in terrestrial images it is not possible to record a complete building in dense urban areas with narrow streets in one image. Because of this, textures have to be combined from multiple images. On the other hand, in airborne images the resolution is low and not all details can be seen. A comparison of a terrestrial and an airborne texture of an exemplary face is presented in Fig.1.



a) b)

Fig .1 Comparison of a terrestrial (a) and an airborne (b)

In these 3D spatial referenced textures, certain structures (e.g. heat leakages, windows and other objects in walls and roofs) can be detected. These automatically get associated with the building. Glass in IR images reflects the temperature of surrounding. Hence paned, windows depict not the real temperature of the façade, but rather the temperature of the neighboring buildings or sky. Accordingly, windows cannot be taken into consideration for building inspection and should be detected in textures.

Infrared images are usually blurred, so that the windows edges cannot be easily detected. Furthermore, due to the reflections, windows can appear differently. For window detection in IR textures an algorithm is needed which can match different appearances of the window and can deal with reflections inside the window frame.

1.2 Related work

Existing algorithms for façade reconstruction were developed for images in visible domain (VIS) and point clouds. Dick et al. (2004) introduce a texture prior generated from training data to identify image areas as supposable façade elements. A group of histograms is generated for every texture using horizontal and vertical wavelet transforms at for different scale levels. A likelihood functions evaluates possible structures with their priors. Ripperda (2008) and Mayer and Reznik (2006) propose reversible jump Markov chain Monte Carlo (rjMCMC) (Green 1995) for the estimation of optimal parameters for windows. Ripperda (2008) is using a formal grammar to describe the behavior of windows. Becker (2009) uses cell decomposition to extract geometry from images as well as from LiDAR point clouds. Formal grammars are used to predict invisible parts and to enhance the window detection. A problem of grammars is the complexity of big scenes and the limitations of the rules defined for the grammars. Reznik and Mayer (2007) are using implicit shape models (Leibe and Schiele 2004) to define a set of windows to check with the given image. Werner and Zisserman (2002) use regular structure primitives like vanishing points or symmetry.

The low optical resolution and smooth edges in infrared images limit the usability of most approaches. The description of both implicit shape models and explicit geometry are limited to well-known optical behaviors whereas windows in infrared show a big variety of optical behavior. Edge points are hard to determine exactly. Window frames can appear lighter or darker than the façade. The façade often contains neither homogeneous nor regular structure due to temperature differences. Approaches dealing with more general geometric primitives should be preferred. In Sirmacek et al. (2011) L-shape structures are searched in textures and combined in a bottom up approach. In Hoegner and Stilla (2009) morphological operators are used to mask window areas and distinguish them from other objects like heat leakages.

1.3 Overview

In this paper we present a technique for window detection in textures extracted from IR images. First we discuss the appearance of windows in IR images and define the model we apply for detection (Section 2). Then we propose the workflow to detect windows (Section 3). For detection we use local dynamic threshold, masked correlation and geometric constrains The presented algorithm was applied for a test data set “Technische Universität München” (Section 4). First results of the experiment are presented in Section 5. The results and further work are discussed in Section .

2. APPEARANCE OF WINDOWS

The appearance of diffuse surfaces in visible domain (VIS) depends on material properties, illumination and viewing angle. In infrared (IR) domain, illumination plays no role, but the appearance of an object is influenced by the temperature distribution. The viewing angle influences the object's appearance in IR images less than in VIS images, however only for objects made from material with diffuse reflection. In case of specular surfaces the appearance in IR depends on the viewing angle as well as on temperature and material of reflected objects. Basing on this knowledge we define the window model in IR image.

2.1 Window model

First of all we expect the window to appear in the image on a homogenous background, it means on the wall. This assumption is usually fulfilled in the close surrounding of the window, because the wall is usually made from a homogeneous material around the window and the temperature of the wall does not change a lot in a small surrounding. Exceptions are the heat leakages caused by poor isolated pipes or damages. But in most cases they occur locally and have a longish form. However this irregularity has minor relevance and should not influence the model in this step.

Next, the appearance model of a window includes a window frame, which is usually made from the material of the wall, having mostly a different emissivity factor. Thus the window frame appears differently compared to the background, concerning the intensity. Typical materials used for window frames (e.g. wood, plastic) have diffuse properties, thus their appearance don't depend on reflections. Inside the frame panes are placed which usually made of glass (specular reflection). Besides inside the window frame other elements (e.g. sashes) can be recognized. These elements and the panes, according to the reflections, cannot be modeled in detail.

The proposed window model is presented in the Fig. 2. Basing on this model we will search for the boundary between the window frame and background. For this purpose we design a binary mask searching for intensity changes in the image. The mask has a predefined shape which corresponds to the expected window shape. However, before creating mask one more fact should be considered. In a textures created from image sequences in oblique view (Fig.1a), we can observe that the window opening partially appears in the texture. It is connected to the fact that windows do not lie in the wall plane. Consequently, it is possible that at one corner we have a change from light to dark, and at another corner of the same window we have change from dark to light. Also roller blinds mounted at the top of the opening can cause this effect. Hence the mask should be designed in the way which allows considering the corners separately.

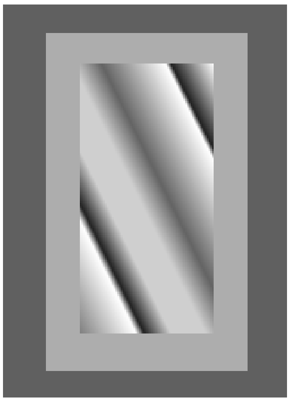


Fig .2 The window model: dark grey - homogenous background, light grey - window frame and inside - glass.

According to our assumptions we design the mask in the way presented in the Fig. 3. This binary mask consists of “on” and “off” fields and of “don’t care” areas. Since we search for intensity change between background and the window frame, the “on/off” fields are placed around expected corners. However, the corners can be moved independently during the correlation process, as long as they build a rectangle. So the window size can also be adapted. The rest of the mask is labeled as “don’t-care-area” and is not used for calculations. Consequently, the edges between the background and window frame and reflections inside the frame are masked out for the correlation process. Then, using equation (1) (Stilla 1993), the mask is correlated within a region of interest (ROI). Determination of the ROI is described in Section 2.3.

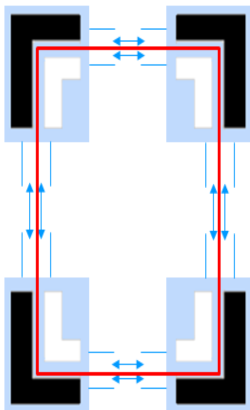


Fig .3 Binary mask: red – expected shape of the window; black & white – binary mask; blue – don’t-care-areas.

$$c = sgn(\rho_{\oplus} - \rho_{\ominus}) \cdot sgn(\overline{g_{\oplus}} - \overline{g_{\ominus}}) \cdot \frac{1}{\sqrt{\frac{m}{m_{\ominus}}\left(\frac{\sigma_{\oplus}}{\overline{g_{\oplus}}-\overline{g_{\ominus}}}\right)^2 + \frac{m}{m_{\oplus}}\left(\frac{\sigma_{\ominus}}{\overline{g_{\oplus}}-\overline{g_{\ominus}}}\right)^2 + 1}} \tag{1}$$

c – correlation coefficient
$\rho_{\oplus}$ – value of "*on*" mask
$\rho_{\ominus}$ – value of "*off*" mask
$\overline{g_{\oplus}}$ – mean value of intensity values in the image covered by "*on*" mask
$\overline{g_{\ominus}}$ – mean value of intensity values in the image covered by "*off*" mask
$m_{\oplus}$ – number of "*on*" pixels in the mask
$m_{\ominus}$ – number of "*off*" pixels in the mask
m – number of "*on*" and "*off*" pixels in the mask
$\sigma_{\oplus}$ – standard deviation of intensity values covered by "*on*" mask
$\sigma_{\ominus}$ – standard deviation of intensity values covered by "*off*" mask

2.2 Regular grid

Window on the façade usually build a regular grid: rows and columns. This regularity can be used to learn a machine to recognize them. In this research we assume that the windows in one row have the same size and the gaps between windows are the same size. Also existence of a window in one row implicates high probability of a window at this position in other rows. The number of rows can be determined using a high histogram, it means a function of the frequency of the window height. The number of maxima in this histogram is equal the number of points.

2.3 Detection of region of interest

Before applying the masked correlation the ROI in the texture should be defined. Usage of a global threshold is not appropriate for detection of window candidates. Therefore, in this paper we propose local dynamic threshold. To compute the local dynamic threshold a second image is used as a reference. In the reference image local threshold values are stored. The reference image is determined using a smoothing filter. The filter mask size has to be adapted to the window size, so that the windows are removed in the reference image. The advantage of the local dynamic threshold is the extraction of regions which differ from the background, and not which has intensity value lower or higher than a given value, what is a case using a global threshold.

3. WORKFLOW

3.1 Region detection

Terrestrial images of building facades captured from a vehicle usually are taken from small distance. Hence the resolution of the textures extracted from these images is relatively high. In these textures many details can be seen. For windows detection the small objects have to be removed by scaling down the image (texture). For the airborne images it is not necessary, because they are taken from a bigger distance and have lower geometric resolution. In the next step the texture is segmented using a local dynamic threshold.

After segmentation some windows can be fused into one object. It is related to the heat leakages, which sometimes are also included in the results of the local dynamic threshold. In such cases two or more windows together with the leakage build one object. To separate the windows morphological operations are used. These operations also help to remove false detections (small objects). Then, for every extracted region, which is a candidate for a window, a rectangular bounding box and its center of gravity is calculated.

3.2 Geometric constrains

For façade reconstruction a pre-knowledge about the regularity of windows can be used. In most buildings the windows build regular grid. The gaps between the columns and between rows are often the same size. However the gravity centers extracted from IR textures in the previous step usually don't build a regular grid of windows. It is related to the fact that the IR images depict distribution of thermal radiation which does not have sharp borders, as well as to the still included thermal leakages which could not be removed through the morphological operations. Hence, in the next step of our method the extracted gravity points are adjusted to regular grid. The gravity centers which are close to each other are merged.

3.3 Standardized masked correlation

To refine the location of windows a standardized masked correlation is applied. Locally around the point (gravity center) the correlation with a priori mask is applied to find the best position of the window and determine its size. The results of the masked correlation are again adjusted into rows. Then the gaps in the window grid are replaced by hypothetic windows. Again the masked correlation is used to check the hypotheses.

4. EXPERIMENTS

The algorithm presented in the Section 3 was tested with an exemplary dataset consisting of airborne and terrestrial image sequences of a test area "Technische Universitaet Muenchen".

4.1 Terrestrial image sequences

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 a focal plane array 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 Fig.1. The position of the camera was recorded with GPS and, for quality measurements from tachymeter measurements from ground control points.

4.2 Textures

The terrestrial textures used for the experiment where extracted from terrestrial image sequences applying the method which is described in detail by Hoegner et al. (2007). In first step the image sequences were relatively oriented using 5 point algorithm. Further, the relative path was matched with the GPS path registered during the measurement and 3D building model. Finally, the sequences where projected on the 3D building model and combine into the textures.

4.3 Airborne textures

We tested our method also with textures extracted from airborne IR images. Airborne images usually have lower resolution than textures from terrestrial images. Therefore scaling down is often not necessary. Using our test data the windows have size of few pixels. Thus, creating a proper mask and matching are difficult.
Our experience shows that better results can be achieved applying the segmentation and adjustment in rows only.

5. RESULTS

Our first results on segmentation and regions extraction are shown in the Fig. 4. Many extracted gravity centers correspond to the approximated position of windows, however in two top rows only. In the lowest row (ground floor) windows could not be extracted. It is related to the irregularity of the windows and entrances in this floor.

In the Fig. 5 results on window extraction after masked correlation are shown. In few cases the detection was false. These false detections were strongly related to striking leakages which connect two windows together. The windows from the lowest row were not detected according to low correlation coefficient. An example of window extraction from airborne textures is presented in Fig.6.

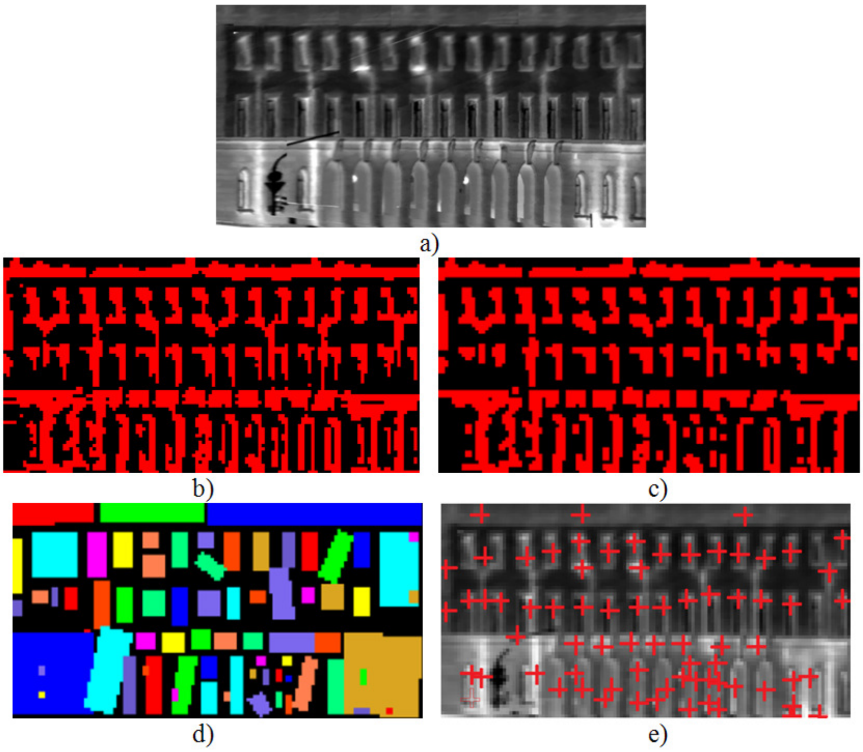


Fig .4 Region detection: a) input b) extraction of candidates for windows from IR texture using local dynamic threshold; c) separation of the regions and removing false detections using morphological operations; d) rectangular bounding boxes for the extracted regions; e) gravity centers

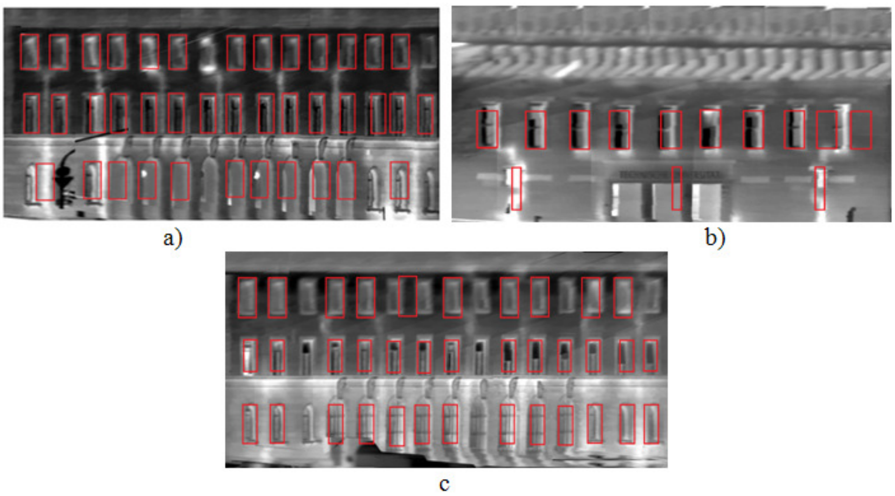


Fig .5 Windows extracted from terrestrial textures

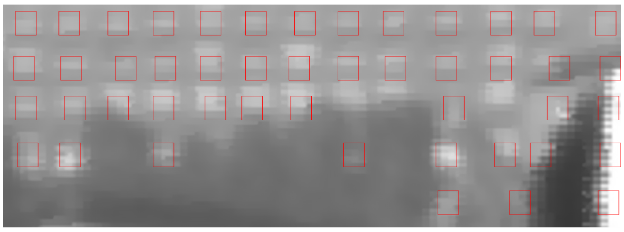


Fig .6 Windows detected in an airborne texture

To evaluate presented results completeness and correctness were calculated for our dataset. For terrestrial textures completeness, defined as correctly detected windows divided by all windows in reality, is 79%. Correctness, defined as correctly detected windows divided by all detections, is 80%. For airborne data completeness is 37% and correctness 49% (Tab. 1). The number of windows for ground truth was determined by a human operator.

Tab. 1. Completeness and correctness

	completeness	correctness
terrestrial	79%	80%
airborne	37%	49%

6. DISCUSSION AND FUTURE WORK

Our first results on region extraction show that usage of the local dynamic threshold allows extracting appropriate candidates for windows. This segmentation technique helps to bridge the different appearance caused by various objects which are reflected in panes of windows. Masked correlation and geometric constrains allow refining the position of each window. Usage of the mask, which switches off irrelevant parts of the window and search for intensity change in the image around the window corners, delivers promising results. However, the mask applied in this research does not match all possible window shapes (the lower rows in Fig. 5a and 5c). Hence, the window model should be extended, so that more shapes are allowed, for example arcs in the upper part of the window.

Further problem in window detection are occlusions by trees, traffic signs and lights as well as by other buildings (e.g. Fig.5a 3rd row 2nd window). Thus, more effort is needed on occlusion free texture extraction.

The achieved results with completeness and correctness around 80% are satisfying for first results, but still can be improved. We expect an enhancement of completeness and correctness by extending the method with searching for window candidates using multiple levels of the image pyramid. Candidates, which are detected in many scales, can be weighted higher while adjusting. Another possibility to improve the results is to combine textures extracted from terrestrial and airborne IR image sequences. Also here, windows detected in both textures could get a higher weight. Moreover, also laser point clouds can be used to determined candidates for windows.

In further studies the presented method for window detection should be integrated into a stochastic process to determine possibilities of window positions. Then the presented geometric constrains could be described by probabilities, which would allow also exceptions from a typical case, which is most likely. Applying this solution also rows with different window size could be modeled.

REFERENCES

Avbelj J, Iwaszczuk D, Stilla U (2010) Matching of 3D wire-frame building models with image features from infrared video sequences taken by helicopters. PCV 2010 - Photogrammetric Computer Vision and Image Analysis. *International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, 38(3B): 149-154

Becker S (2009) Generation and application of rules for quality dependent facade reconstruction, *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 64, Issue 6, November 2009, Pages 640-653

Dick A, Torr P, Cipolla R, Ribarsky W (2004) Modelling and interpretation of architecture from several images. *International Journal of Computer Vision* 60(2), pp. 111–134

Eugster H., Nebiker S. (2009) Real-time georegistration of video streams from mini or micro UAS using digital 3D city models. Proceedings of 6th International Symposium on Mobile Mapping Technology, Presidente Prudente, São Paulo, Brazil, July 21-24, 2009

Frueh C, Sammon R, Zakhor A (2004) Automated Texture Mapping of 3D City Models With Oblique Aerial Imagery, Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04).

Green P (1995) Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* 82, pp. 711–732.

Hoegner L, Kumke H, Meng L, Stilla U (2007) 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

Hoegner L, Stilla U (2009) Thermal leakage detection on building facades using infrared textures generated by mobile mapping. Joint Urban Remote Sensing Event (JURSE 2009). IEEE

Leibe B, Schiele B (2004) Combined Object Categorization and Segmentation with an Implicit Shape Model. In: ECCV'04 Workshop on Statistical Learning in Computer Vision, pp. 1–15.

Li R (1997) Mobile Mapping - An Emerging Technology for Spatial Data Acquisition. *Journal of Photogrammetric Engineering and Remote Sensing*, Vol.63, No.9, pp.1085-1092.

Reznik S, Mayer H (2007) Implicit Shape Models, Model Selection, and Plane Sweeping for 3D Facade Interpretation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (3/W49A), S. 173-178

Ripperda N (2008) Grammar Based Facade Reconstruction using RjMCMC. PFG Photogrammetrie Fernerkundung Geoinformation. Stuttgart: Schweizerbartsche Verlagsbuchhandlung. 2008(2): 83-92

Sirmacek B, Hoegner L, Stilla U (2011) Detection of windows and doors from thermal images by grouping geometrical features. In: Stilla U, Gamba P, Juergens C, Maktav D (eds) JURSE 2011 - Joint Urban Remote Sensing Event, 133-136

Stilla U (1993) Verfahrensvergleich zur automatischen Erkennung in Metall geschlagener Zeichen. Karlsruhe: Universität, Fakultät für Elektrotechnik, Dissertation

Stilla U, Kolecki J, Hoegner L (2009) Texture mapping of 3D building models with oblique direct geo-referenced airborne IR image sequences. ISPRS Hannover Workshop 2009: High-resolution earth Imaging for geospatial information. 38(1-4-7/W5)

Werner T, Zisserman A (20029 New Techniques for Automated Architectural Reconstruction from Photographs. In: Seventh European Conference on Computer Vision, Vol. II, pp. 541–555.

Yastikli N, Jacobsen K (2005), Direct sensor orientation for large scale mapping – potentials, problems, solutions. *The Photogrammetric Record* 20(111), September 2005: 274-284