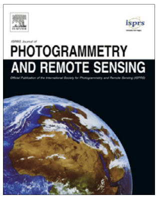
ISPRS Journal of Photogrammetry and Remote Sensing 132 (2017) 33–47

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs



Camera pose refinement by matching uncertain 3D building models with thermal infrared image sequences for high quality texture extraction

Dorota Iwaszczuk[a,b,*], Uwe Stilla[a]

[a] *Photogrammetry and Remote Sensing, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany*
[b] *Astronomical and Physical Geodesy, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany*

ARTICLE INFO

Article history:
Received 1 December 2016
Received in revised form 5 July 2017
Accepted 14 August 2017

Keywords:
3D
Building
Thermal
Sequences
Matching
Texturing

ABSTRACT

Thermal infrared (TIR) images are often used to picture damaged and weak spots in the insulation of the building hull, which is widely used in thermal inspections of buildings. Such inspection in large-scale areas can be carried out by combining TIR imagery and 3D building models. This combination can be achieved via texture mapping. Automation of texture mapping avoids time consuming imaging and manually analyzing each face independently. It also provides a spatial reference for façade structures extracted in the thermal textures. In order to capture all faces, including the roofs, façades, and façades in the inner courtyard, an oblique looking camera mounted on a flying platform is used. Direct georeferencing is usually not sufficient for precise texture extraction. In addition, 3D building models have also uncertain geometry. In this paper, therefore, methodology for co-registration of uncertain 3D building models with airborne oblique view images is presented. For this purpose, a line-based model-to-image matching is developed, in which the uncertainties of the 3D building model, as well as of the image features are considered. Matched linear features are used for the refinement of the exterior orientation parameters of the camera in order to ensure optimal co-registration. Moreover, this study investigates whether line tracking through the image sequence supports the matching. The accuracy of the extraction and the quality of the textures are assessed. For this purpose, appropriate quality measures are developed. The tests showed good results on co-registration, particularly in cases where tracking between the neighboring frames had been applied.

1. Introduction

Energy efficiency of buildings became a important topic in the recent years. In European countries, buildings consume 40% (Baden et al., 2006) of all produced energy and 47% of the energy consumed in buildings is used for heating (Hennigan, 2011). Due to climate change, increasing energy costs, and energy performance directives, the energy efficiency of buildings should be improved. Using new technologies, new buildings are typically constructed with higher energy performances. Regarding the older buildings, there is an increasing demand for inspections and improvement of their energy efficiency. For this purpose, thermal infrared (TIR) images can be used. Thermal cameras capture the heat radiation of the building hull with the radiometric resolution up to 0.01 [K] and record it as an intensity image. The intensity values correspond to the remotely measured temperature on a defined scale. For building inspections however, the exact temperature measurement is of minor interest, while the main focus is on the temperature differences, which enable us to detect heat leaks as well as damaged and weak spots in building structures (Balaras and Argiriou, 2002). The knowledge of the location of heat leaks is used for renovation planning or for quality control after the renovation.

In practice, the TIR images are often taken manually for a single building. However, in the recent years a trend can be observed to use mobile mapping systems and carry out large scale inspections for urban quarters (Hoegner et al., 2007) or entire cities (Chandler, 2011). For this purpose, the TIR cameras are mounted on a mobile terrestrial platform, usually on a vehicle. Terrestrial TIR images are used to document frontal faces (façades) visible from the street level, but they do not capture roofs and inner yards. Roofs can be seen from a flying platform. Using oblique view images, inner yards can be captured as well (Iwaszczuk et al., 2011). Combining TIR

---

[*] Corresponding author at: Astronomical and Physical Geodesy, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany.

E-mail addresses: dorota.iwaszczuk@tum.de (D. Iwaszczuk), stilla@tum.de (U. Stilla).

http://dx.doi.org/10.1016/j.isprsjprs.2017.08.006

images with three-dimensional (3D) geometries allows for the spatial reference of the thermal data and facilitates their interpretation. In literature, examples can be found for combination of thermal data with Building Information Models (BIM) (Mikeleit and Kraneis, 2010), with 3D building models via texture mapping (Hoegner et al., 2007; Iwaszczuk et al., 2011), with 3D point clouds via assignment and interpolation of the measured temperature to the points (Cabrelles et al., 2009; Borrmann et al., 2012; Vidas et al., 2013) or with aerial photographs combined with a point cloud (Boyd, 2013). Using a point cloud as spatial reference enables fast generation of results with a high level of detail and is appropriate for visual interpretation, while 3D building models deliver more generalized and structured representations to support automatic analysis. Embedding thermal data in a geo-database (Kumke et al., 2006) allows for spatial queries and analysis. Particularly, a combination of thermal data with other image and spatial data stored within the 3D building model is of high interest and can be used for wider understanding of buildings and urban scenes.

Thermal textures can be generated together with the 3D reconstruction from thermal images (Westfeld et al., 2015) or assigned to already existing 3D geometries (Iwaszczuk et al., 2012). 3D reconstruction from thermal images ensures the perfect fit between 3D geometries and thermal textures, but it is challenging, compared to visible domain, due to low number of tie points in thermal image pairs resulting from low resolution and contrast (Hoegner, 2014). In addition, thermal inspections should be conducted in time intervals and updated textures should be assigned to the same geometries. Therefore, the main challenge in generation of thermal textures is co-registration of thermal images and 3D building models.

Automated co-registration and texture mapping are of high interest for images in all spectral domains. In the following, therefore, we will focus on the general purpose to use any images and 3D building models and return to the thermal data in the experimental part.

1.1. Related work

Texture mapping on 3D models is a widely used technique and results in adding an image to the existing 3D geometry. Such operations became of high interest also in urban scenes and enrich 3D building models with their visual properties (Weinhaus and Devarajan, 1997; Allen et al., 2001). Textures are typically assigned to the 3D building models during the 3D reconstruction process (Debevec et al., 1996; Bornik et al., 2001; Kuschk, 2013), or added from another data set (Hsu et al., 2000; Früh et al., 2004; Hoegner and Stilla, 2007).

Texture extraction from geo-referenced images has already been implemented in several commercial software and imaging systems (Groneman, 2004; Grenzdörffer et al., 2008) as well as in scientific works (Früh et al., 2004; Klinec, 2004; Karbo and Schroth, 2009). Textures can be extracted from the images taken in different spectral bands, such as visible (VIS) images (Hsu et al., 2000; Früh et al., 2004; Abdelhafiz and Niemeier, 2009), multi-spectral images (Pelagotti et al., 2009), or infrared images (Hoegner and Stilla, 2007; Iwaszczuk et al., 2012). In the recent years, some imaging systems have appeared on the market, which are dedicated to photorealistic urban scene modeling. Particularly, oblique view cameras are useful for capturing whole scene including the building façades (Grenzdörffer et al., 2008; Wang et al., 2008), however such acquisition geometry requires special treatment for flight planning (Grenzdörffer et al., 2008) and image measurements (Höhle, 2008).

Co-registration can be carried out by direct geo-referencing, However direct geo-referencing is often not sufficient to accurately co-register the data (Früh et al., 2004; Kada et al., 2005; Grenzdörffer et al., 2008) and is used only as approximate alignment. In airborne photogrammetry, geo-referencing can be carried out in the frame of aerotriangulation. However, aerotriangulation can require manual selection of control points and does not yield good results for stripe-wise acquired oblique images with one perspective angle (Grenzdörffer et al., 2008). Manual selection of control points can be particularly time consuming for image sequences with a high frequency rate. Results of aerotriangulation for oblique images carried out with four perspective angles is also insufficient for high accuracy model-to-image co-registration (Kolecki et al., 2010). The mismatch can be due to inaccurately estimated exterior and interior orientations of the camera or inaccuracies in the 3D building model. Therefore, some authors propose model-to-image matching in order to improve the co-registration (Früh et al., 2004; Ding and Zakhor, 2008), which can be supported by tracking (Hsu et al., 2000).

1.1.1. Model-to-image matching

Model-to-image matching is a widely discussed topic and various methods for implementation have been developed. One of the strategies for the matching is to fit 3D models directly to the contours to simultaneously determine projection and model parameters by minimizing the perpendicular distance from the points on the image edge to the projected model curve. This can be successfully applied to parametric 3D models to images for recognition and tracking purposes (Lowe, 1991) as well as for model-driven semi-automatic 3D reconstruction from nadir (Vosselman, 1998) and oblique-view images (Panday and Gerke, 2011). Image gradients can also be used to match edges, faces, or whole buildings with the 3D models using mutual information technique (Nyaruhuma et al., 2012). Mutual information can also be used for co-registration of the image with the depth image generated from the 3D model (Pelagotti et al., 2009). Another strategy for co-registration is to consider not only the agreement between image features and model features, but also takes the relationship between features into account in a relational matching (Vosselman, 1992; Eugster and Nebiker, 2009). In some urban scenes, it is helpful to assume so called *Manhattan* or *Legoland scenes* which consist of piece-wise planar surfaces with dominant directions (Lee et al., 2002; Ding and Zakhor, 2008; Förstner, 2010b). In such scenes, calculating vanishing points of the vertical and horizontal lines in combination with global position system (GPS) data can be used for the computation of exterior orientation parameters. These methods require a calibrated camera system and the extraction of many vertical and horizontal lines in the image. Because of the lack of vertical and horizontal lines, which can be unambiguously extracted, these methods can fail in residential areas (Ding and Zakhor, 2008).

The methods for model-to-image matching can also be differentiated based on the image features used for matching. For model-to-image matching in urban areas, some authors propose points (Ding and Zakhor, 2008; Avbelj et al., 2010), but most consider lines more natural for building structures used them for co-registration (Debevec et al., 1996; Lee et al., 2002; Klinec, 2004; Eugster and Nebiker, 2009). Hybrid methods employing points and lines at the same time have been also discussed (Zhang et al., 2005; Tian et al., 2008). Finding line-to-line correspondences between the image lines and model edges can be solved with different techniques, such as clustering (Sester and Förstner, 1989), relational matching (Vosselman, 1992; Eugster and Nebiker, 2009), similarity measures (Ok et al., 2012), energy minimizing (Hsu et al., 2000; Sawhney et al., 2002; Früh et al., 2004), which are often supported by RANSAC approach (Lee et al., 2002; Früh et al., 2004; Cheng et al., 2013). Unlike points, lines cannot be directly used for estimation of the exterior orientation of the camera with collinearity equations. The collinearity equations have to