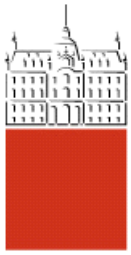
Univerza
v Ljubljani
Fakulteta
za gradbeništvo
in geodezijo

Jamova 2
1000 Ljubljana, Slovenija
telefon (01) 47 68 500
faks (01) 42 50 681
fgg@fgg.uni-lj.si



Univerzitetni študij geodezije,
Geodezija

Kandidatka:
Janja Avbelj

Koregistracija trirazsežnih modelov stavb z grafičnimi gradniki zaznanimi z infrardečih aero videoposnetkov

Diplomska naloga št.: 843

Mentor:
izr. prof. dr. Krištof Oštir

Somentor:
prof. dr. Uwe Stilla

Ljubljana, 2010

Izjava o avtorstvu

Skladno s 27. členom Pravilnika o diplomskem delu UL Fakultete za gradbeništvo in geodezijo, ki ga je sprejel Senat Fakultete za gradbeništvo in geodezijo Univerze v Ljubljani na svoji 10. seji dne 21. 4. 2010.

Podpisana **Janja Avbelj**, izjavljam, da sem avtorica diplomske naloge z naslovom: *Koregistracija trirazsežnih modelov stavb z grafičnimi gradniki zaznanimi z infrardečih aero videoposnetkov.* Noben del te diplomske naloge ni bil uporabljen za pridobitev strokovnega naziva ali druge strokovne kvalifikacije na tej ali na drugi univerzi ali izobraževalni inštituciji.

Izjavljam, da prenašam vse materialne avtorske pravice v zvezi z diplomsko nalogo na UL, Fakulteto za gradbeništvo in geodezijo. UL FGG ima ob pridobitvi izrecnega pisnega soglasja študenta in mentorja pravico do javne objave diplomske naloge.

I, the undersigned **Janja Avbelj**, hereby declare that I am the author of the diploma thesis titled: *Co-registration of three-dimensional building models with image features from infrared video sequences.*

Ljubljana, 27. september 2010

……

(podpis / signature)

Avbelj, J. 2010. Co-registration of three-dimensional building models with image features from infrared video sequences. Graduation Thesis – University studies. Ljubljana, UL FGG, Dep. of Geodetic Engineering, Geodesy.

Errata

Page	Line	Error	Correction

Bibliographic-documental information and abstract

UDC: 528.7/.8(043.2)

Author: Janja Avbelj

Supervisor: Associate Prof. Ph.D. Krištof Oštir

Co-supervisor: Prof. Ph.D. Uwe Stilla

Title: Co-registration of three-dimensional building models with image features from infrared video sequences

Notes: 127 p., 20 tab., 45 fig., 34 eq.

Key words: 3D building model, infrared radiation, image processing, image registration, feature extraction, photogrammetry, image sequences, matching

Abstract

In the European Union (EU) countries buildings consume 40% of the energy and cause 36% of CO_2 emissions. The thermal information of facades and roofs are important for building inspection and energy saving. Texturing the existing three-dimensional (3D) building models with infrared (IR) images enriches the model database and enables analysis of energy loss of buildings.

The main purpose of the presented thesis is to investigate methods for automatic extraction of the IR textures for roofs and facades of the existing building model. The correction of the exterior orientation parameters of the IR camera mounted on mobile platform is studied. The developed method bases on a point-to-point matching of the features extracted from IR images with a wire frame building model.

Firstly, extraction of different feature types is studied on a sample IR image; Förstner and intersection points are chosen for representation of the image features. Secondly, the 3D building model is projected into each frame of the IR video sequence using orientation parameters; only coarse exterior orientation parameters are known. Then the automatic co-registration of a 3D building model projection into the image sequence with image features is carried out. The matching of a model and extracted features is applied iteratively and exterior orientation parameters are adjusted with least square adjustment. The method is tested on a dataset of dense urban area. Finally, an evaluation of developed method is presented with five quality parameters, i.e. efficiency of the method, completeness and correctness of matching and extraction.

Bibliografsko-dokumentacijska stran in izvleček

UDK: 528.7/.8(043.2)

Avtorica: Janja Avbelj

Mentor: izr. prof. dr. Krištof Oštir

Somentor: Prof. Dr. –Ing. Uwe Stilla

Naslov: Koregistracija trirazsežnih modelov stavb z grafičnimi gradniki zaznanimi z infrardečih aero videoposnetkov

Obseg in oprema: 127 str., 20 pregl., 45 sl., 34 en.

Ključne besede: 3D modeli stavb, infrardeč spekter, digitalna obdelava podob, registracija podob, videoposnetki, fotogrametrija

Izvleček

V državah Evropske unije (EU) porabijo zgradbe 40 % energije in povzročijo 36 % vseh emisij CO_2. Podatki o temperaturi fasad in streh so pomembni pri določitvi temperaturne učinkovitosti zgradb, s čimer omogočajo prihranek energije. Teksturiranje obstoječega trirazsežnega (3D) modela stavb z infrardečimi (IR) posnetki dopolni podatkovno bazo modela zgradb in omogoči analize njihovih energetskih izgub.

Namen raziskovalne naloge je samodejna določitev tekstur streh in fasad stavb obstoječega trirazsežnega (3D) modela z IR posnetkov. Za to je potrebno izboljšanje natančnosti parametrov zunanje orientacije IR kamere pritrjene na mobilno platformo. Razvita metoda temelji na ujemanju točk samodejno (avtomatsko) zaznanih grafičnih gradnikov z IR videoposnetka in žičnega modela stavb.

V nalogi smo najprej proučili zaznavo različnih tipov grafičnih gradnikov na testnem IR posnetku. Förstnerjeve in presečiščne točke smo izbrali kot primerne grafične gradnike za predstavitev obravnavanih značilnosti stavb na IR posnetku. 3D model stavb je projiciran na vsak posamezen posnetek videosekvence ob upoštevanju orientacijskih parametrov, od katerih so parametri zunanje orientacije podani le s približnimi vrednostmi. Nato smo izvedli samodejno koregistracijo 3D modela stavb projiciranega na videoposnetek in grafičnih gradnikov zaznanih z istega IR videoposnetka. V iterativnem postopku samodejnega ujemanja 3D modela stavb in zaznanih grafičnih gradnikov smo parametre zunanje orientacije izravnali z metodo najmanjših kvadratov. Razvito metodologijo za koregistracijo in izravnavo zunanjih orientacijskih parametrov smo preizkusili na strnjenem poseljenem območju. Kvaliteto metodologije smo ocenili s petimi parametri: učinkovitostjo metodologije, ter popolnostjo in pravilnostjo algoritmov za ujemanje in zaznavo grafičnih gradnikov.

Acknowledgement

First of all, I would like to thank my mentor, prof. Krištof Oštir and co-mentor Prof. Dr. –Ing Uwe Stilla for all the support in the time I was writing my diploma thesis. The employees and PhD students of the department of Photogrammetry and Remote Sensing, Technische Universität München which made me feel welcome during my staying there. Especially many thanks to Dorota Iwaszczuk for all the constructive remarks and the time we spent together solving problems. To my parents - thank you for believing in me even in moments when I doubted myself. All of the abovementioned people have one thing in common; often they did not give me answers, they were patient and willing to teach me how to search for them. Last but not least I would like to thank my sisters, Eva and Mojca, friends and study colleagues.

Zahvala

Zahvaljujem se mentorju, izr. prof. Krištofu Oštirju in somentorju, Prof. Dr. –Ing. Uwe Stilla za vso podporo pri moji diplomski nalogi. V času mojega študija na Tehniški univerzi München so me zaposleni in doktorski študentje oddelka »Photogrammetry and Remote sensing« toplo sprejeli. Še posebej se zahvaljujem Doroti Iwaszczuk za konstruktivne kritike in najine ure iskanja rešitev. Mojima staršema – hvala, da verjameta vame tudi takrat, ko podvomim vase. Vsi, ki sem jih omenila imajo nekaj skupnega, s potrpežljivostjo so me učili iskati odgovore, namesto da bi mi na vprašanja odgovarjali. Ne nazadnje, hvala sestrama, Mojci in Evi, prijateljem, ter študijskim kolegom.

Generation of the diploma thesis

This diploma thesis was written under the supervision and with support of my mentor, Prof. Dr. Krištof Oštir, University of Ljubljana and co-mentor Prof. Dr. -Ing Uwe Stilla, Technical University of Munich (TUM). The data and technical support were provided by TUM, Department of Photogrammetry and Remote Sensing; the images of the flight campaign were provided by FGAN-FOM, Ettlinger. I was studying as exchange student from October 2009 to Mai 2010 as foundationer of *»Dr.-Ing. Otto und Karla Likar Stiftung«* at TUM.

Janja Avbelj

O nastanku diplomske naloge

Diplomsko nalogo sem izdelala pod mentorstvom izr. prof. dr. Krištofa Oštirja, Univerza v Ljubljani in somentorstvom Prof. Dr. -Ing. Uwe Stilla, Tehniška univerza München (TUM). Vse uporabljene podatke in programsko opremo mi je omogočila Katedra za fotogrametrijo in daljinsko zaznavanje TUM (Department of Photogrammetry and Remote Sensing). Uporabljene IR posnetke so priskrbeli v podjetju FGAN-FOM, Ettlinger. Od oktobra 2009 do konca maja 2010 sem bila štipendistka *»Sklada dr. Otta Likarja, dipl. inž., in Karle Likar«* na TUM.

Janja Avbelj

Table of contents

List of tables

List of figures

List of Graphs

Abbreviations

γ-rays	Gamma Rays
2D	Two-dimension(al)
3D	Three-dimension(al)
CIO	International Commission on Illuminance
CityGML	City Geography Markup Language
DGPS	Differential Global Positioning System
DTM	Digital Terrain Model
EM	Electromagnetic (spectrum)
EU	European Union
ExtOri	Exterior/External Orientation (parameters)
GIS	Geographic Information System
GML	Geography Markup Language
GPS	Global Positioning System
IEEE	Institute of Electrical and Electronics Engineers
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
IntOri	Interior/Inner Orientation (parameters)
IR	Infrared (spectrum, domain)
ISO	International Organization for Standardization
LIDAR	Light Detection and Ranging
LOD	Level of Detail
LS	Least Square (adjustment)
LWIR	Long Wavelength Infrared (spectrum)
MSE	Mean Square Error
MWIR	Medium Wavelength Infrared(spectrum)
NIR	Near Infrared (spectrum)
OGC	Open Geospatial Consortium
OpenGIS	Open Geographic Information System

px	Pixel(s), picture element(s) in digital image
RANSAC	Random Sample Consensus
RMSE	Root Mean Square Error
SWIR	Short Wavelength Infrared (spectrum)
TC	Technical Committee
TIR	Thermal Radiation
TUM	Technische Universität München (ger.) / Technical University Munich (eng.)
UV	Ultraviolet (spectrum)
VIS	Visible (domain, spectrum)
VLWIR	Very Long Wavelength Infrared (spectrum)
also FIR	Far Infrared (spectrum)
THz	Terahertz Radiation
FM	Frequency Modulation (radio)
AM	Amplitude Modulation (radio)
DC	Direct Current

Slovar izrazov

Slovenska terminologija na področjih računalniškega vida, digitalne obdelave in analize slik, fotogrametrije in daljinskega zaznavanja ni v celoti uveljavljena. Na področju računalniškega vida se slovenska terminologija izraziteje razlikuje glede na aplikacijo. Ta slovarček smo pripravili z namenom, da bralcu olajšamo branje in povezavo tako s slovensko kot angleško literaturo s tega področja.

Angleškim izrazom smo poiskali slovenske ustreznice in razlage priredili in povzeli po več virih, ki jih navajamo po abecednem vrstnem redu priimka prvega avtorja, ter v celoti v poglavju šest, Literatura. Slovar izrazov je izdelan na podlagi naslednje literature: Albert et al. (2003), Brown (1992), Campbell (1996), Catmull (1974), Gonzáles et al. (2002), Gonzáles et al. (2008), Guid (2010), Herakovič (2007), Kajfež-Bogataj (2005), Kosmatin Fras et al. (2008), Kraus (1993), CityGML (2008), Oštir (2006), Poženel (1999). Sonka et al. (2008), Stadler et al. (2007), Šumrada (2005a), Šumrada (2005b) in Zitová et al. (2003).

3D model mesta (*3D city model)* je digitalna predstavitev Zemljinega površja in z njimi povezanih objektov, ki pripadajo urbanim območjem.

3D model stavb *(3D building model)* je model, ki vsebuje podatke o stavbah. Omogoča prikazovanje tematskih in prostorskih aspektov stavb in delov stavb v različnih stopnjah podrobnosti. 3D model stavb je sestavni del 3D modela mesta.

Aeroposnetek ali **zračni posnetek** *(aerial image)* je posnetki zajet s senzorja na platformi v zraku oziroma atmosferi, na primer z letala ali helikopterja. Imenovan lahko tudi letalski posnetek.

Atmosfersko okno (*atmospheric window)* je območje valovnih dolžin, ki ga atmosfera prepušča in le malo absorbira. Atmosferska okna določajo dele spektra v katerih opazujejo senzorji daljinskega zaznavanja.

Atribut *(atribute)* je detajl s katerim je izbran objektni tip opredeljen, opisan, klasificiran ali je izraženo njegovo stanje in razmerje do drugih objektnih tipov. Atribut opisuje lastnost objekta v naravi.

City Geography Markup Language je standardni jezik za modeliranje mest, ki ga razvija OGC.

Digitalna analiza podob ali slik *(digital image analysis)* je področje, ki obravnava analizo in interpretacijo vsebine posnetkov, ter je povezano z digitalno obdelavo podob.

Digitalna obdelava podob ali slik *(digital image processing)* je področje digitalnega procesiranje signalov in obravnava postopke digitalne obdelave podob (slik) z uporabo računalniških algoritmov.

Filter *(filter)* je pri avtomatski obdelavi podob običajno kvadratna matrika. Z uporabo filtra na digitalni podobi se poudari ali zakrije nekatere elemente. Pri filtriranju digitalne podobe s filtrom se uporablja matematična funkcija konvolucija.

Geografski informacijski sistem *(Geographic Information System, GIS)* je sistem za zajemanje shranjevanje, vzdrževanje, obdelavo, povezovanje, analiziranje in predstavitev prostorskih geokodiranih podatkov.

Geography Markup Language (GML) je standardni jezik za označevanje geografskih podatkov, ki ga razvija OGC. Omogoča shranjevanje, zapis in prenos prostorskih podatkov.

Georeferenciranje *(georeferencing)* je postopek določitve prostorskih koordinat objektom in pojavom v prostoru. Na področju fotogrametrije in daljinskega zaznavanje je to vpenjanja satelitskih ali aeroposnetkov v obstoječ koordinatni sistem.

Global Positioning System (GPS) je sistem za globalno določanje položaja, ki temelji na tehnologiji vesoljske radijske navigacije. Vzpostavilo ga je ministrstvo Združenih držav Amerike. GPS je satelitski navigacijski sistem kot tudi GLONASS, Galileo in Beidu.

Grafični gradnik *(feature)* predstavlja geografski pojav v vektorskem 2D modelu. Osnovni grafični gradniki so točke, linije (segmenti ali vektorji) in območja (poligoni), njihova vozlišča in oznake. Grafičnim gradnikom so lahko pripeti atributni podatki.

Grobo pogrešeno opazovanje *(outlier)* je tisto opazovanje, ki vsebuje grobi pogrešek in ga je potrebno izločiti iz obravnave.

Homologna točka je ista točka v naravi, ki se pojavlja na več različnih posnetkih.

Inercialna merilna enota ali naprava *(inertial measurement unit)* omogoča meritev kotnih premikov. Sestavljena je iz pospeškomerov in/ ali žiroskopov. V povezavi z GPS meritvami se uporablja za določanje parametrov zunanje orientacije.

Inercialni navigacijski sistem *(inertial navigation system)* omogoča določitev položaja zračnega plovila, torej parametrov zunanje orientacije.

Infrardeč spekter *(infrared spectrum)* je del spektra elektromagnetnega valovanja valovnih dolžin med 0,74 in 1000 μm. Segreta telesa velik del energije sevajo v IR spektru, ki je človeškemu očesu neviden.

Izravnava *(adjustment)* je matematični postopek določitve neznank v predoločenem sistemu s statistično oceno njihovih natančnosti.

Kot gledanja *(look angle)* je smer gledanja senzorja. Na primer: kot gledanja kamere je lahko navpično navzdol, usmerjen v nadir, poševno *(oblique)*, naprej v smeri leta *(forward looking)*.

Ločljivost *(resolution)* je sposobnost sistema za ločevanje pojavov ali lastnosti. V fotogrametriji in daljinskem zaznavanju poznamo prostorsko, radiometrično, spektralno in časovno ločljivost.

Mestni toplotni otok *(urban heat island)* je pojav temperaturne razlike med mestom in okoliško naravno pokrajino. Pri nas, v Evropi, je ta pojav izrazitejši v poletnih mesecih.

Model *(model)* je poenostavljena podoba stvarnosti. Je abstrakcija in posplošitev tistega dela stvarnosti, ki je pomemben za določen namen in zajema njegove bistvene značilnosti.

Modeliranje *(modelling)* je postopek načrtovanja in ustvarjanja modela.

Nadir *(nadir)* je točka na zemeljskem površju neposredno pod letalom ali satelitom.

Nagib *(roll)* je naklon letala v smeri krila.

Naklon *(pitch)* je premik letala v smeri kljuna in repa.

Optični spekter *(optical spectrum)* je opredeljen z valovnimi dolžinami, ki jih je mogoče odbiti in lomiti s pomočjo leč in ogledal (zajema območje valovnih dolžin med 0,3 in 15 μm). Vsebuje viden spektra in del IR spektra.

Parametri ali elementi zunanje orientacije *(exterior orientation parameters)* so koordinate položaja kamere oziroma projekcijskega centra v prostoru in trije rotacijski koti, na primer: nagib, naklon in zasuk.

Parametri ali elementi notranje orientacije *(interior orientation parameters)* so: koordinati glavne točke v slikovnem koordinatnem sistemu in konstanta kamere.

Piksel *(pixel)* ali slikovni element je najmanjši del podobe. Podaja informacijo o karakteristiki točke, na primer: radiometrično ali sivinsko vrednost, barvo, intenziteto.

Platforma *(platform)* je nosilec na katerem so instrumenti za daljinsko zaznavanje. Lahko so na tleh (terestrične), v zraku (aero) ali satelitih in drugih vesoljskih plovilih.

Podpikselska natančnost *(subpixel accuracy)* je določitev položaja pojava na podobi, ki obsega več pikslov, z natančnostjo višjo od radiometrične ločljivosti podobe.

Poravnava ali registracija *(registration)* je postopek geometrijskega prileganja dveh ali več podob pri katerem se določa najboljša transformacijska funkcija (in parametri transformacije) med posnetkoma.

Posnetek videosekvence *(frame)* je posamezen posnetek videoposnetka.

Poševen posnetek ali fotografija *(oblique photography or image)* je posnetek zajet s kamere s poševnim kotom gledanja proti površju.

Potratnost izračuna *(computational cost)* je čas, ki je potreben za izračun pri uporabi operatorja ali metode. V aplikacijah računalniškega vida, digitalne obdelave in analize podob je potratnost operatorjev in algoritmov za izračun pogosto omejitveni dejavnik.

Prostorska ločljivost *(spatial resolution)* je velikost najmanjšega predmeta na podobi, ki ga zaznamo. Pri digitalnih aeroposnetkih je prostorska ločljivost velikost piksla v prostoru *(ground sample resolution).*

Računalniški vid *(computer vision)* je področje znanosti in tehnologije, ki se ukvarja z interpretacijo 2D podob v povezavi in z namenom razumevanja 3D okolja.

Radiometrična ločljivost *(radiometric resolution)* je sposobnost senzorja za ločevanje vrednosti na podobi, na primer 8-bitna ločljivost pomeni 256 različnih sivih vrednosti.

Slikovno ujemanje *(image matching)* je postopek iskanja homolognih vzorcev ali točk na dveh ali več posnetkih, ter iskanje določenega vzorca na sliki.

Stopnja podrobnosti ali detajla *(Level of Detail)* se navezuje na vsebino, položajno natančnost, stopnjo generalizacije in podrobnosti 3D modela mesta. Termin je definiran v standardu CityGML, ki ga razvija OCG.

Šum posnetka *(image noise)* se pojavi zaradi slučajnih pogreškov pri zajemu posnetka.

Tekstura *(texture)* predstavlja lastnosti površine zgradbe ali objekta. Tekstura je slika ali podoba, ki jo lahko projiciramo v večrazsežen prostor oziroma na površino modela.

Teksturiranje *(texturing, texture mapping)* je postopek, v katerem je 2D tekstura kartirana na površino v 3D prostoru.

Termični spekter *(thermal spectrum)* je del spektra elektromagnetnega valovanja valovnih dolžin med 0,1 in 1000 μm, kjer črno telo s temperaturo nad 0 K seva energijo v okolico. Pri daljinskem zaznavanju se termin uporablja ožje, in sicer za valovne dolžine okoli 3(8)-15 μm.

Topologija je veda o medsebojnih odnosih med objekti.

Ujemanje grafičnih gradnikov *(feature matching)* je iskanje povezav med grafičnimi gradniki na več posnetkih, ki predstavljajo isti pojav v prostoru. Je del postopka registracije (prileganja) podob.

Vidni spekter *(visible spectrum)* je del spektra elektromagnetnega valovanja valovnih dolžin med 380 in 740 nm. To je zvezen spekter barv od vijolične do rdeče, ki jih zaznava človeško oko.

Zasuk *(yaw)* je premik letala v smeri navpične osi.

Zaznava ali ekstrakcija grafičnih gradnikov*(feature extraction)* je postopek, pri katerem z metodami digitalne obdelave podob določimo grafične gradnike na podobi.

Žični model *(wire-frame model)* je način predstavitve in modeliranja 3D modela, sestavljen iz poligonov ali daljic.

1 INTRODUCTION

In the introduction the motivation for this work with the hypothesis (section 1.1) and a state of art in the field of research (section 1.2) is given. Problem overview and objectives of the thesis are presented in the section 1.3 and finally in the section 1.4 organisation of the diploma thesis is listed.

1.1 Motivation

In the last decades the world energy consumption has been increasing, with the exception of the year 2009. »World primary energy consumption - including oil, natural gas, coal, nuclear and hydro power - fell by 1.1% in 2009, the first decline since 1982« (Statistical Review of World Energy 2009, 2010). The energy consumption decrease is assumed to be connected with economic contraction in the year 2009. In European Union (EU) countries buildings consume 40% of the energy and cause 36% of CO_2 emissions (Directive 2010/31/EU, 2010). For improving the energy performance of buildings, it is necessary to decrease the energy consumption and greenhouse gas emissions. Importance of the energy consumption in buildings and its influence on the environment are commonly addressed topics in science, research as well as politics. These topics are motivation for numerous regulations, e.g.: recast of EU's The Directive on energy performance of buildings (Directive 2010/31/EU, 2010) adopted in May 2010 and standards. Technical committee (TC) 163: Thermal performance and energy use in the built environment of International Organization for Standardization (ISO) has published 83 standards in the field of building and civil engineering works, additional 27 standards are under development. These standards include energy performance of buildings, test and measurement methods, calculations and also thermal insulation products (ISO/TC 163, 2010).

Inspection of buildings can contribute to develop strategies for energy savings. The building data should be collected and regularly updated for building analysis. Numerous buildings lead to a large amount of data which must be acquired, processed and analysed. Therefore sensors mounted on mobile platforms, i.e. spaceborne, airborne or mobile terrestrial platforms, are appropriate to assure time and cost efficient data acquisition.

High population density is distinctive of urban areas and is often connected to high building density. Focusing on urban areas, many (virtual) city models exist. »Virtual 3D city models are digital representations of the Earth's surface and related objects belonging to urban areas« (Stadler, 2007). Three-dimensional (3D) building models contain buildings and building data and are a part of 3D city models. They are helpful for interpretation of the image contents (Stilla, 2000). The 3D building models can contain textures from images taken by sensors mounted on variety of platforms. Spaceborne and/or airborne images with top-down view provide data for roofs whereas terrestrial images provide data for facades. Additionally, oblique airborne images can provide data for roofs and facades (Frueh, 2004 and Stilla, 2009). Combination and coverage of data types is desired for enhancing existing 3D model.

The developed Geography Markup Language (GML) standard City Geography Markup Language (CityGML) allows integrating different Level of Detail (LOD) into one Geographic Information System (GIS) database (Hoegner, 2009). Existing city models have different LOD and are often represented with textures in visual (VIS) spectrum. Some structures appear in VIS and infrared (IR) spectrum, whereas some, e.g. heating systems, thermal leakages, can only be captured by IR data. For instance, urban heat islands can be observed in the IR spectrum in small scale, whereas inspection of buildings requires larger scale (Weng, 2009). Airborne IR images have higher resolution in comparison to space-borne images; therefore they have higher accuracy for locating thermal losses on buildings in larger scale. The thermal information of facades and roofs are important for building inspection and energy saving. Texturing existing 3D building models with IR images enriches the model database and enables the analysis of energy loss of buildings. »Thermal images of different parts of buildings acquired in various scales are analysed, however conducted analysis concern mainly the radiometric characteristics, avoiding dealing with the geometry of the images«

(Stilla, 2009). Locating e.g. thermal leakages on roofs or facades in larger scale also requires controlling the accuracy of texturing.

1.2 Related work

Preconditions for accurate texturing of building models are calibrated camera with known Interior orientation (IntOri) parameters, a geo-referenced building model and known position of the acquisition device, i.e. Exterior orientation (ExtOri) parameters (Stilla, 2009). are Besides cameras, there are usually also devices for tracking the trajectory of flight on airborne platforms, most commonly Global Positioning System (GPS) receivers and inertial measurement unit (IMU) to obtain coarse ExtOri parameters. Next to abovementioned preconditions, relative positions between all measuring devices must be known or calibrated. Projecting the 3D building model into the image using all these parameters should be made to extract sub matrix of the image corresponding to a specific building face. ExtOri parameters gained from GPS/IMU often do not have sufficient accuracy for high quality texture mapping. To refine the accuracy of position of the acquisition device, many researchers use image processing methods, specifically methods for automatic matching of a 3D model with images were developed. Defining the position of acquisition or measuring device is often addressed in literature as pose estimation (problem). In case of two non-calibrated images this problem is also referred to as simultaneous pose and correspondence problem.

Hsu et al. (2000) are using projection of the 3D model line segments into image and extracted line features to refine the pose of acquisition device. Firstly, an algorithm to extract and track features in video sequence is applied and tracked features are used to predict poses between frames. Secondly, line segments are projected with predicted pose in an image and pose is refined by aligning projected lines to oriented image gradient energy pyramids. The algorithm is tested on aerial and terrestrial video sequence of a large scale urban scene. Frueh et al. (2004) also use extracted lines for pose estimation. Canny edge detector is applied to detect image edges in high resolution oblique images in VIS spectrum, which are further divided into line segments by a recursive endpoint subdivision algorithm. To find a camera pose, a building model is projected into the image around an initial pose obtained from integration of

GPS and inertial navigation system (INS) and building lines are matched to extracted lines. Each pose is rated using weighted correlation function based on line-to-line matching algorithm. The pose with the highest correlation value is used for extracting the texture. A weakness of this method is very high computational cost. Problems with edge extraction and matching are also addressed in the article. Naming a few edges belonging to the same entity in object space are often extracted incompletely and inaccurately in a single image, non-single edge response, endpoints of extracted edges are not reliable (Tian, 2008, Frueh, 2004, Lee, 2002). Tian et al. (2008) propose edge matching across a video image sequence using geometric constrains based on reliable points. Reliable points are calculated by analyzing the endpoints of extracted edges and are qualitative evaluated. Using reliable points significantly reduces the search space for matching step and therefore also the computational cost of the method.

A possible approach for camera pose estimation is using vanishing point(s). In Lee et al. (2002) two or three vanishing points and 3D to 2D (two-dimensional) line matching to estimate ExtOri parameters of terrestrial camera are used. Ding et al. (2008) propose a two step process for refinement of camera position. Firstly, coarse ExtOri parameters are estimated using vanishing points and data obtained from GPS/INS measuring system. Second step refines the coarse ExtOri parameters estimated in first step by matching of orthogonal corners from oblique images and lidar model (LIDAR, Light detection and ranging).

Hoegner and Stilla (2008) used IR image sequences acquired by mobile terrestrial system for automated texturing of 3D building models. They discuss the special characteristics of IR spectrum compared to VIS spectrum which can cause problems in automatic texturing and also feature detection. Stilla et al., 2009 and Kolecki el al., 2010 describe direct geo-referencing with GPS/INS data and extended system calibration. They address problems given by vibrating platforms, such as helicopters, which are observable and may cause misalignment of projected model in an image in subsequent frames. In this diploma thesis an approach for refinement of the coarse ExtOri is presented using point-to-point matching. A correspondence between 3D wire-frame building model and image features extracted from video sequence are automatically searched.

1.3 Problem overview and objective of the thesis

If the accuracy of the GPS/INS data is not high enough to gain the needed accuracy of orientation parameters of the mobile platform, one possible approach is to model and estimate the influences on the GPS/INS measurements and correct them. A reverse approach is that image data, acquired with the camera mounted on the mobile platform, are used for the refinement of orientation parameters. This thesis applies the reverse approach, using image processing methods to detect features in the IR images acquired with the mobile platform. A goal is to propose an automatic algorithm for refinement of ExtOri parameters for every image in the video sequence.

Firstly, a special attention is devoted to choosing the appropriate features to be automatically extracted from the IR image sequence. The resolution of the IR images is lower in comparison to the capability of cameras acquiring in VIS spectrum. What is more, the properties of the IR spectrum differs from VIS, therefore the choice for feature type is studied on a few cases. Secondly, the 3D building model is projected into the images, using corrected orientation parameters calculated in the extended system calibration proposed by Kolecki et al. (2010). The 3D building model is projected into the 2D image under consideration of all the observed orientation parameters For example, distortions of the image are present due to the lens distortions of the camera, and when the 3D building model is projected into the image, it is distorted for values of the lens distortions. For projection of the 3D building model into each of the images of a sequence, different ExtOri parameters must be used. This thesis focuses next on developing a method for automatic co-registration of a 3D building model projected into the image sequence with image features. (Figure 1).

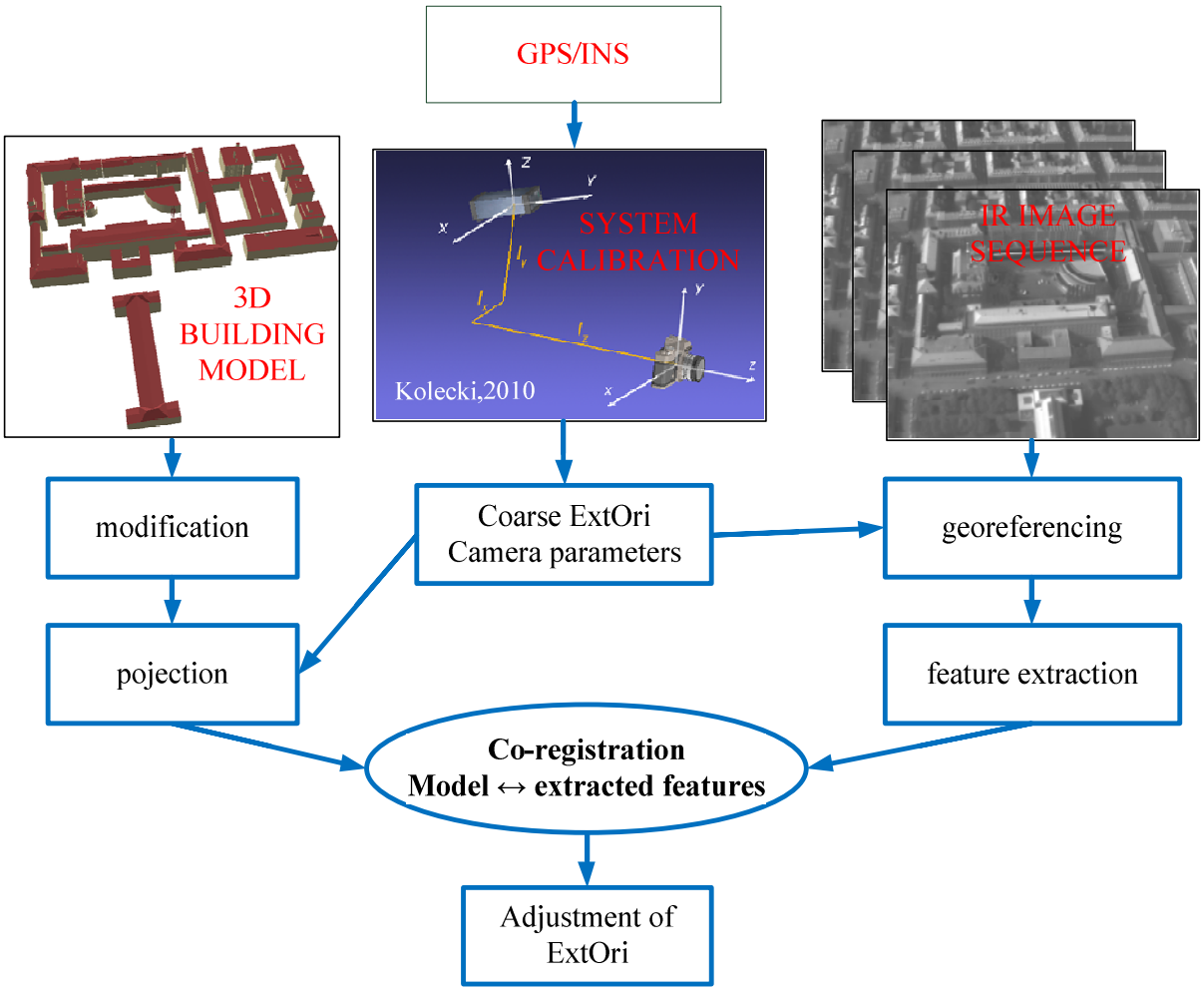


Figure 1: Flowchart of the method for co-registration of model and extracted features with input and output parameters.

Hypotheses:

- *Features extracted from the IR image registered with the 3D building model can improve the ExtOri parameters so that the projected building model is better aligned to the observed buildings in image.*
- *GPS/INS data and system calibration provide coarse ExtOri parameters, accurate enough for the first projection of the 3D building model into the image.*

1.4 Organisation of the diploma thesis

This diploma thesis is divided into eight main chapters. In this chapter the introduction and the motivation for this diploma thesis are given, then an overview of related work is presented and finally the approach to solve the given problem is described. Chapter 2 includes theoretical background of the topics linked to the given problem, datasets and developed method. A method developed for the solution of described problem is presented in chapter 3, firstly the different feature types extracted from IR images are discussed and according to the extracted features, a method is developed and described. A method is tested on the sample dataset, described in detail in chapter 4. Additionally, a comparison between the chosen feature types is presented, evaluation of the method, and the problems of the 3D building model are addressed. All the results are discussed. Finally, in chapter 5 the conclusions are presented and ideas for future work are proposed. Chapter 6 contains an extended abstract of this thesis in Slovene language. Chapters 7 and 8 are the lists of references and appendixes, respectively.

2 THEORETICAL BACKGROUNDS

In this work a method of registering a building model in Level of Detail 2 and extracted features detected in the infrared image sequences taken by a helicopter is presented. The Organisation of the chapter follows the structure of the developed methodology, used data and procedures. Firstly, three disciplines studying images are described, then the electromagnetic spectrum is presented, emphasising the infrared spectrum and the IR acquisition devices. Next we focus on used methodology, summarizing image registration steps, specifically the feature extraction and matching. Later a description of building models, which can be projected into the image using space central projection, also under consideration of the IntOri and ExtOri parameters, is given. A short description of the least square adjustment with observation equations is given. Finally, an overview on texture mapping of building models is presented.

2.1 Computer Vision, Digital Image Processing and Analysis

The Input data used in this work are images taken from the IR video sequence. Three related disciplines studying images are described in this section; computer vision, digital image processing and image analysis. Computer vision interprets images representing a 3D environment whereas digital image processing and analysis are dealing with 2D images. This section gives a short overview of abovementioned disciplines, which are connected to the photogrammetry and remote sensing, thus are not a part of them.

2.1.1 Computer vision

Human vision and understanding of its functioning from the psychological and physical point of view are analogue to research of computer vision. »Vision allows humans to perceive and understand the world surrounding them while computer vision aims to duplicate the effect of

human vision by electronically perceiving and understanding an image« (Sonka, 2008). Programming software to imitate human understanding of 3D world is a difficult task. A pair of eyes allows humans to see in 3D; in addition, humans have natural ability to recognize three-dimensional objects from 2D images. Computer vision tasks and applications mean giving computers ability to reconstruct and understand 3D scene from usually 2D sources, i.e. images, video streams, which requires detailed understanding of human congenital ability for pictorial recognition (González, 2008). Hochberg and Brooks (1962) described an experiment supporting hypothesis that pictorial recognition is an unlearned ability and therefore is not culturally conditioned. In the experiment a child was raised without being shown any pictures and avoided to be taught with help of 2D drawings of any kind. After 19 months, the child was shown black and white line drawings and photographs and was instantly able to identify objects on them (Hochberg, 1962).

There are two main reasons why research in computer vision is important. Firstly, to equip machines with knowledge and understanding of surroundings that humans have. This has a crucial meaning for automation of processes in many applications. Computer vision is closely connected with acquisition devices, i.e. cameras, video devices; data can be acquired in visible spectrum of electromagnetic radiation or any other and can be enhanced with data from different types of sensors. Fusion of different data types contains more informational value, therefore efficient algorithms – such as machine vision methodology provides - to process and evaluate large amount of input data are needed. Secondly, research is important to understand computational understanding of human vision. There are several attempts to connect computational results to the relevant areas of psychology. (González, 2008, Sonka, 2008).

2.1.2 Digital image processing and analysis

Digital image processing is a subfield of digital signal processing. It refers to the processing of digital images using computer algorithms. Digital images are 2D functions $f(x, y)$, where x, y are coordinates and the amplitude f at position (x, y) is called grey level or intensity (González, 2002). Digital image processing is rather independent of an application and does not include interpretation of the processed image. In contrast to digital image processing, the

crucial step of image analysis is image interpretation which often depends on the application. Inputs and outputs of digital image processing are images, extracted attributes from images and individual objects. It must be pointed out that there is no general agreement among authors regarding where digital image processing stops and other areas, such as computer vision and image analysis start. (González, 2008).

In this work, a 3D model is projected into the IR image on which features are detected. Feature detection is a result of a method of digital image processing. Interpretation of extracted features, for example analyzing which extracted point is the edge point of a building, is a task of image analysis. The understanding and interpretation of the 3D environment is a domain of computer vision. Thus connection of a 3D building model and extracted and interpreted features is an area of computer vision.

2.2 Infrared spectrum and infrared remote sensing

In this section, a representation of electromagnetic (EM) spectrum emphasising the infrared spectrum and IR acquisition devices is given. In this thesis IR image sequences are used and processed, therefore knowledge about used spectral band and its characteristics is important for understanding and interpretation of results, as well as for developing the methodology.

2.2.1 Electromagnetic spectrum

EM spectrum is illustrated in image in Figure 2 and divided into spectral bands, see Table 1. Human eyes are capable of detecting the visible spectrum (VIS) that corresponds to wavelengths in range of around 380–740 nm and corresponding continuous spectrum of colours from violet to red. The EM spectrum covers a broad range: from radio waves with long wavelengths (low energy) to gamma rays (γ-rays) with short wavelengths (high energy). Different divisions of the EM spectrum can be found in literature, thus the boundaries are set according to discipline. Additionally, neighbouring types of electromagnetic energy also overlap in some cases. (Campbell, 1996).

The EM spectrum can be expressed in terms of energy E, wavelength λ or frequency ν:

$$\lambda = \frac{c}{\nu} \qquad \textit{(Eq. 1)}$$

and

$$E = h \cdot \nu, \qquad \textit{(Eq. 2)}$$

where:

λ wavelength [m],

ν frequency [Hz],

c speed of light [m/s],

E energy [J] and

h= 6.62606896·10^{-34} m^2 kg/s, Planck's constant.

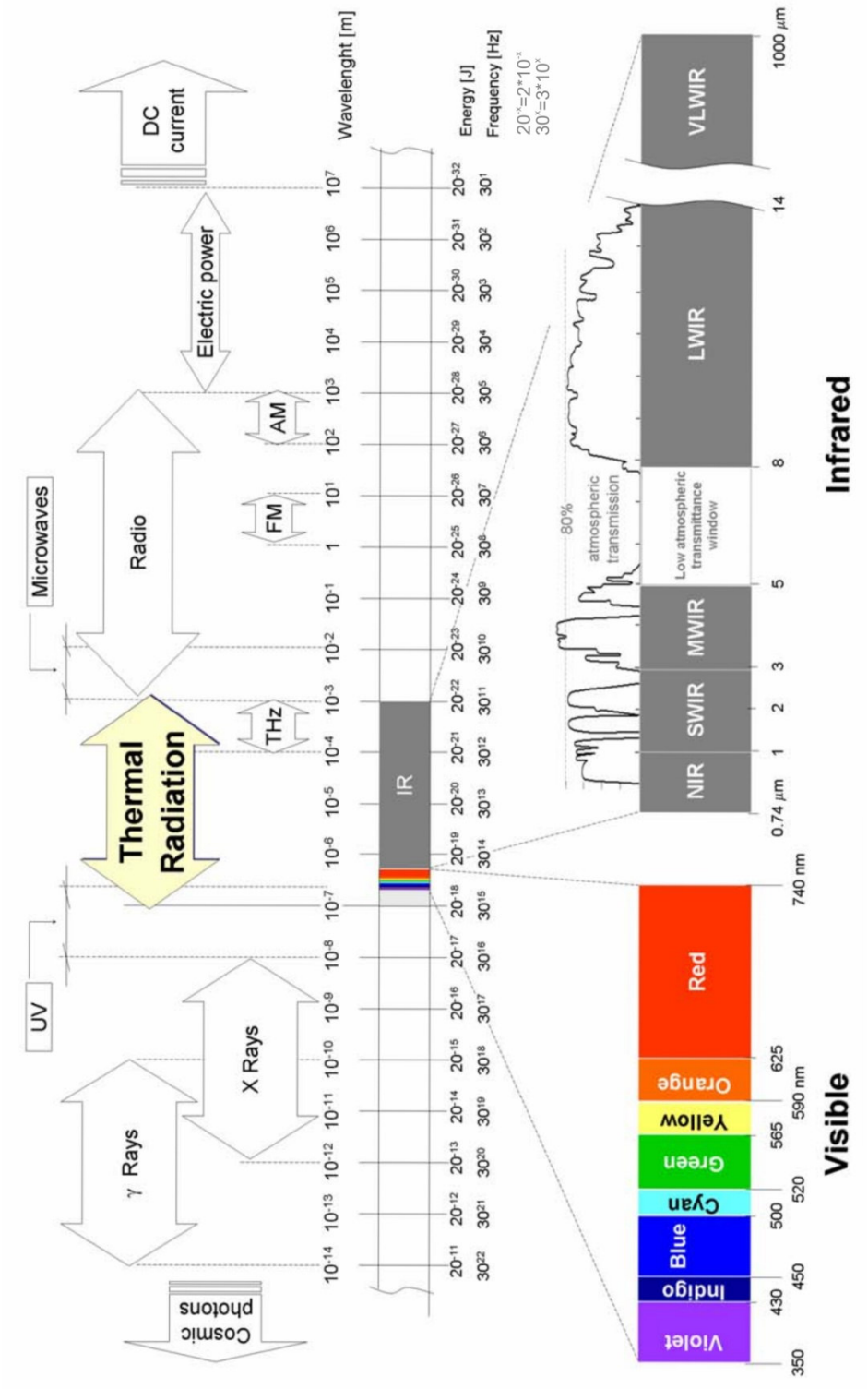


Figure 2: The electromagnetic spectrum emphasising infrared spectrum. (Source: Ibarra-Castanedo, 2005, p.128).

2.2.2 Infrared spectrum

The IR spectrum has wavelengths longer than red portion of the VIS domain and it extends to the microwaves with wavelengths of about 1 mm. The IR spectrum (0.74-1000 μm) can be divided into five spectral bands (CIO, 2010, Ibarra-Castanedo, 2005). »Although this subdivision is somehow arbitrary and varies from one source to another, it is based on the atmosphere high transmissivity windows, i.e. the wavelength bands where atmosphere interferes the less with the incoming IR radiation, and on the detector's spectral sensitivities« (Ibarra-Castanedo, 2005). The division of the EM spectrum outside the IR spectrum is adapted from Campbell (1996).

Table 1: Spectral bands with detail division of the infrared spectrum (adapted from Campbell, 1996 and Ibarra-Castanedo, 2005). The image sequence used for validating the developed methodology in this thesis was acquired with IR camera in MWIR spectral band 3-5 μm (highlighted in the table).

	Spectral band /Type of EM radiation	**Wavelength**
	Gamma Rays (γ-rays)	**< 0.03 nm**
	X-Rays	**0.03-300 nm**
	Ultraviolet (UV)	**0.30-0.38** μm
	Visible (VIS)	**0.38-0.74** μm
Infra red (IR) 0.74 - 1000μm	Near Infrared (NIR)	0.74-1 μm
	Short Wavelength Infrared (SWIR)	1-3 μm
	Medium Wavelength Infrared (MWIR)	3-5 (8) μm
	Low atmospheric transmittance window	
	Long Wavelength Infrared (LWIR)	8-14 μm
	Very long Wavelength Infrared (VLWIR, also FIR)	14-1000 μm
	Micro waves	1 mm-30 cm (1 m)
	Radio waves	> 30 cm (1 m)

Two important categories of the EM spectrum are not shown in Table 1, these are optical spectrum and thermal spectrum.

The optical spectrum (0.3–15 μm) is defined by wavelengths that can be reflected and refracted with lenses and mirrors. It includes the VIS and part of the IR spectrum (Campbell, 1996). Optical properties of the VIS spectrum are similar to optical properties of a

part of the IR spectrum; therefore similar cameras can be used for detecting the radiation in the optical domain.

An object with a temperature above 0 K is in an ideal case a black body, radiating energy, i.e. heat, to the environment. *The thermal radiation (TIR) band or thermal spectrum* (0.1-1000 μm) includes complete VIS spectrum, and parts of IR and UV (highlighted in Figure 2). Especially in remote sensing the term thermal IR radiation is referring to the wavelengths of MWIR and up to (14) 18 μm. (Campbell, 1996, Ibarra-Castanedo, 2005).

As mentioned in the previous paragraph, all objects with an absolute temperature above 0 K radiate IR energy. A background radiation from the Earth's surface is at about 300 K and causes thermal noise when observing IR radiation with wavelengths 3 μm or longer. To reduce background thermal noise, cold filters, cold shields or cooling systems of detectors are necessary. (Characteristics and use of infrared detectors, 2004.)

The IR radiation is invisible for the human eye, therefore it needs to be detected and transformed into a visible image with the aid of specialised equipment. Infrared imaging is used in military and civilian applications, naming a few: target acquisition, night vision, tracking, thermal efficiency analysis, temperature detection, vehicles detection and tracking, glacier monitoring, spectroscopy, astronomy, industrial and machine control. (Campbell, 1996, Ibarra-Castanedo, 2005, Laval University MiViM, 2010).

Important criteria for band selection:

- emissivity of the object of interest,
- temperature,
- indoor or outdoor application and
- operating distance (close range, aerial image acquisition, etc.).

Detection of high temperature objects, having peak emission at short wavelengths a MWIR band is appropriate, if focusing on emissivity of the object of interest as main criteria. On the contrary, for detection of room temperature object, LWIR band is more appropriated.

Additionally, the LWIR is more appropriate for outdoor applications, because they are less affected by the sun radiation. (Ibarra-Castanedo, 2005, Laval University MiViM, 2010).

2.2.3 Infrared cameras

The human visual system cannot evaluate detected light in absolute terms in contrast to the acquisition devices, e.g. cameras. Data acquired with cameras can be qualitatively evaluated and are not limited only to the visible EM spectrum as humans are. Sensors are capable of detecting energy radiated by a selected band of the EM spectrum and visualised so that humans can detect the object of interest. A significant limitation of sensors is that wavelength, in which we are observing must be the same or possibly smaller than the size of the observed object. There are also other technical limitations in materials, construction and physical properties of sensors. (González, 2008, Sonka, 2008).

An IR sensor is a measuring device that converts radiated IR energy into electrical current or any other measurable form. IR sensors can be divided into two classes, thermal and photonic or quantum detectors. A *thermal detector* responds to incident radiation by raising its temperature when it rises above the equilibrium state of the sensor. The excess of the temperature is removed by conduction. The thermal detector is at equilibrium when there is no conduction of energy; i.e. the sensor radiates and absorbs energy at the same rate. These types of detectors are (theoretically) independent of observed wavelengths and their peak of sensitivity is at room temperature. Main principle of *photonic or quantum detectors* is that the incident radiation excites electrons from the valence to the conduction atomic bands. They are made from semiconductor materials and often require cooling to reduce noise-to-signal ratio. The thermal detectors have lower sensibility and response speed than photonic sensors, which need cooling to obtain relative high sensitivity. However, thermal detectors are less expensive. Both types of IR sensors are being developed with the aim to increase sensitivity, response time and lower the price. (Ibarra-Castanedo, 2005, Laval University MiViM, 2010).

2.3 Image registration

The main steps of image registration are presented and shortly described in this section, stressing out feature extraction and feature-based methods for matching. A focus of the thesis is re-calculation of orientation parameters of acquisition device using feature extraction and feature matching.

Image registration is a process of geometrically alignment of two or more images (reference and sensed images) by estimating optimal transformation between them. For points in reference image, the corresponding points in sensed image are defined and related with the transformation function. Image registration is widely used in applications of remote sensing, photogrammetry, medical imaging, computer vision, etc. The registration methods can be divided into four main groups with regard to image acquisition: from different viewpoints, at different time, with different sensors or scene to model registration. (Zitová, 2003, Brown, 1992). Scene to model registration is in literature also addressed as prototype or template registration in which the alignment between image and a template is searched (Brown, 1992). In this work scene to model registration is used where model (template) is a projected 3D building model into the scene, i.e. each frame of video sequence.

Zitová and Flusser (2003) present a review of recent and classic registration methods. They suggest division of registration methods in four steps:

- feature detection or feature extraction,
- feature matching,
- transform model estimation is defining and estimating parameters of mapping function and
- image re-sampling and transformation is transforming the sensed image with regards to mapping function. In this step interpolation methods are often used.

All the steps are not required for every application and/or can be jointed together. A different division of registration methods and steps are presented in Brown (1992) ensuing from the different type of variations in images and registration techniques. Main stress is given to the

choice of the transformation class, search for optimal transformation parameters and the evaluation of chosen parameters.

2.3.1 Feature extraction

Features are denoting pieces of information in an image relevant for solving certain problem. Therefore definition of a feature often depends on the application and given task. According to Brown (1992), the first step in image registration is the decision on the feature space which extracts the information in the images that will be used for matching. In this work, chosen feature spaces are addressed as features, more precisely points, contours, line segments and area borders that represent building edges and corners. Feature extraction is an image processing method of which results are extracted features in an image; automatic feature extraction implies on using computers to detect them (Zitová, 2003, Sonka, 2008). Two major categories of feature extraction, area-based and feature-based methods are described below, adapted from Zitová and Flusser (2003).

Area-based methods are emphasising the matching step which is merged together with the extraction. Using area-based methods, no salient image objects are detected in contrast to feature-based methods. For area-based methods, a window of predefined size (»template«) is used to search for correspondence in an image. This window in not a feature, nevertheless, there is a high probability for wrong matching, if the extracted window does include smooth area without any prominent details.

Feature-based methods; this approach is based on extracting features in an image. Commonly used features in computer vision tasks are region, line and point features. More complex extracted features can also be used, e.g. statistical features, higher-lever features (Brown, 1992).

2.3.2 Feature matching

Feature matching is a registration step in which correspondence between detected features in sensed and reference image is established or correspondence between a »template« cut out of a reference image and sensed image is searched for feature-based or area-based methods, respectively. The section 2.3.2 is adapted from Zitová and Flusser (2003).

In literature, *area-based methods* are also named correlation like methods or template matching. Brown (1992) mentions these techniques also to refine or to estimate the parameters of chosen transformation.

Feature-based methods. Two sets of features are extracted in reference and sensed images and a pair wise correspondence between them is searched. In regard to extracted feature type, point-to-point algorithm or line-to-line matching algorithm can be applied. More complex extracted features require more extensive matching algorithm.

Correspondence between features can be searched using:
Spatial relations. In certain distance from a feature in the reference image a correspondence to a feature in the sensed image is searched.
Invariant descriptors. Correspondence is searched using the description of features which should be invariant to expected image deformation.
Relaxation methods. Each feature from the sensed image is labelled with an identificator of a feature from the reference image. The method is iterative and is recalculating pairs until a stabile solution is found.
Pyramids and wavelets. Aim of these methods is mainly to reduce computational cost of matching algorithm.

Next two steps of the image registration, transform model estimation and image re-sampling and transformation are not described in detail. The definition and estimation of the mapping parameters for used mapping function in this thesis is adapted from Kolecki et al. (2010). Image re-sampling is in this case model re-sampling with regards to the mapping function, without applying any interpolation methods.

2.4 Building models and Level of Detail

Building models can be projected into the 2D images for better interpretation of image's content (Stilla, 2000). The LOD of a building model is defining the amount of details included in a model as well as the accuracy. This section is adapted from the Open Geographic Information System (OpenGIS) CityGML, Encoding Standard, Open Geospatial Consortium (OGC) (CityGML, 2008).

»The building model allows for the representation of thematic and spatial aspects of buildings, building parts and installations in four levels of detail, LOD1 to LOD4«. (CityGML, 2008, p. 56). CityGML supports different LOD, that is, a same object can be represented in different LOD simultaneously. Additionally, generalisation of the relations between objects in different LODs is supported. This allows analysis and visualisation of the same object with regard to different degrees of resolution.

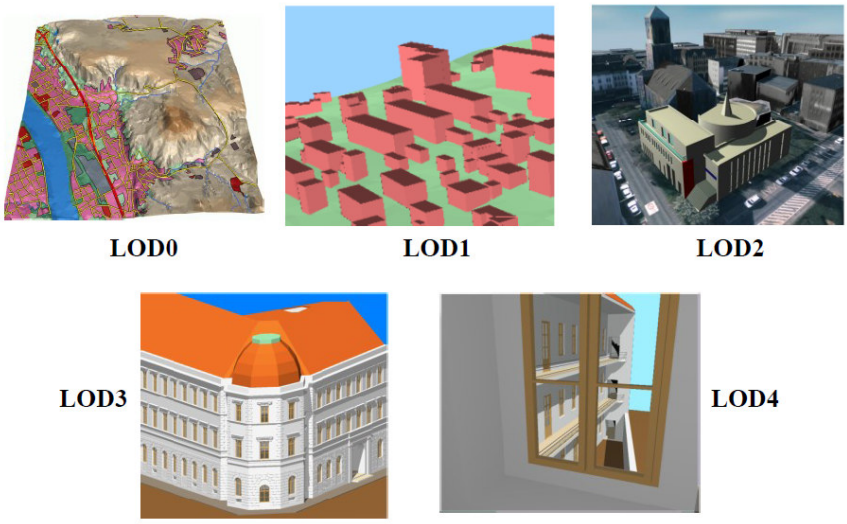


Figure 3: The five levels of detail defined by CityGML (source: CityGML, 2008 p. 9).

Table 2: Level of detail in city models with description of details according to CityGML standard (Adapted from CityGML, OGC 2008, p.9-10 and p. 56).

Level of Detail	Description of details of a building model
LOD0	Two and a half dimensional (2.5) Digital Terrain Model (DTM), over which an aerial image or a map may be draped.
LOD1	Block model with prismatic buildings and flat roofs.
LOD2	Differentiated roof structures Additionally, vegetation objects may be presented and high-resolution textures can be mapped onto represented structures.
LOD3	Detailed vegetation and transportation objects
LOD4	Interior structures for 3D objects For example, buildings are composed of rooms, interior doors, stairs, and furniture

Building models in certain LOD are differing by the contents (Table 2) as well as by accuracies and minimal dimensions of the objects that are included. Accuracy is a standard deviation of the absolute 3D position of a point; relative position is not yet defined in CityGML (2008) standard, but it should be higher in comparison to the absolute position. For a model in LOD2, the positional and height accuracy must be 2 m or better, and all objects with a footprint of at least 4 m × 4 m have to be considered (Albert, 2003). In the Appendix A is a table with LOD 0-4 of CityGML with its accuracy requirements.

2.5 Central projection in space

A central projection as a geometrical model is commonly used in photogrammetry and remote sensing to reconstruct position and shape of objects from photographs (images) or to obtain position of the acquisition device, i.e. calculation of ExtOri parameters on the ground of measured image points. For determination of the position of the acquisition device, the IntOri parameters should be known, and the ExtOri parameters calculated. In this thesis, a position of an IR camera is re-calculated for each frame of the video sequence.

The relation between the 2D image coordinates *(x, y)* of the image point of the object P and the *(X, Y, Z)* coordinates of the object P is given with co-linearity equations 3 and 4 (Kraus, 1993).

$$x = x_0 - ck\frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{22}(Z - Z_0)}, \quad \textit{(Eq. 3)}$$

$$y = y_0 - ck\frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{22}(Z - Z_0)}, \quad \textit{(Eq. 4)}$$

where:

x_0, y_0 image coordinates of principal point,

x, y 2D image coordinates of the object P,

ck camera constant or principal distance,

X_0, Y_0, Z_0 object coordinates of the camera position,

X, Y, Z coordinates in 3D object coordinates system of the object P; frequently in the national system of ground coordinates and

R spatial rotation matrix of size 3 × 3 with elements r_{ij}, i, j = 1, 2, 3.

Spatial rotation matrix R is given with

$$R = R_{yaw} \cdot R_{pitch} \cdot R_{roll}, \quad \textit{(Eq. 5)}$$

where:

$$R_{yaw} = \begin{bmatrix} cos\,(yaw) & -sin\,(yaw) & 0 \\ sin\,(yaw) & cos\,(yaw) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \textit{(Eq. 6)}$$

$$R_{pitch} = \begin{bmatrix} cos(pitch) & 0 & sin(pitch) \\ 0 & 1 & 0 \\ -sin(pitch) & 0 & cos(pitch) \end{bmatrix}, \quad \textit{(Eq. 7)}$$

$$R_{roll} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\,(roll) & -sin\,(roll) \\ 0 & sin\,(roll) & cos\,(roll) \end{bmatrix} \quad \textit{(Eq. 8)}$$

Yaw, pitch, roll are special kind of Euler rotation angles named also Tait-Bryan angles (Figure 4 and Figure 5). Order of multiplications of rotation matrixes (Eq. 5) can be changed in some cases, as well as signs in front of sinus or cosines function. The change of sign in Eq. 6, 7 and 8 changes the orientation of the axes of rotations.

Interior orientation parameters (IntOri) are x_0, y_0 and ck as well as additional parameters. Additional parameters are a consequence of inaccuracies in construction or materials of

camera, i.e. radial and tangential distortion parameters. The camera constant, the position of a principal point and the distortions are periodically re-evaluated in a calibration process. If the higher accuracy is to be achieved, more IntOri parameters must be set in calibration.

Exterior orientation parameters (ExtOri) are X_0, Y_0, Z_0 and *roll, pitch, yaw* angles. X_0, Y_0, Z_0 are coordinates in the global coordinate system, which determinate the position of an aircraft in a certain moment and are usually acquired with GPS measurements. To control the orientation of the flying aircraft (helicopter or any other aerial vehicle), yaw, pitch and roll motion (Figure 4) should be observed. Yaw, pitch and roll motion are rotations about yaw, pitch, roll axes, respectively (Figure 5).

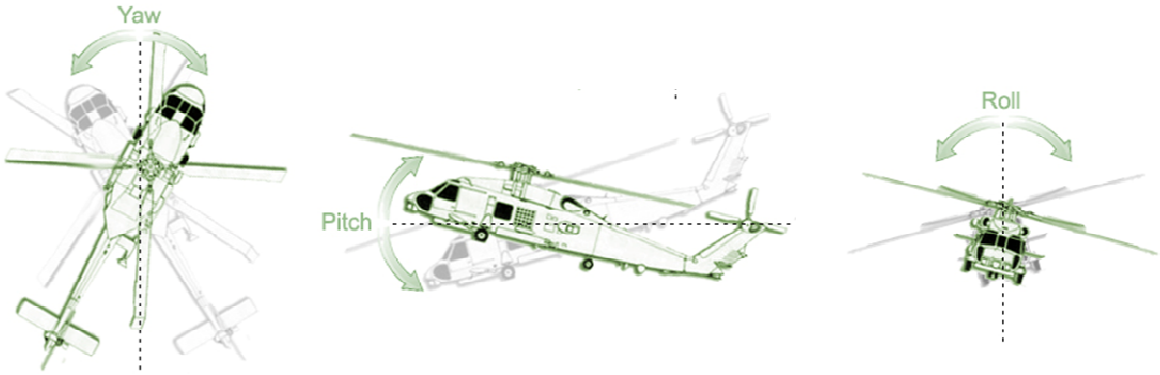


Figure 4: Yaw, pitch, roll motions of a helicopter (Source: ACME, 2010).

Rotation angles define a rotation between a reference axis system and a vehicle-fixed axis system which is defined with:

Origin is centre of gravity of aircraft,

yaw axis is perpendicular to the plane of the wings with its origin in the centre of gravity and directed towards the bottom of the aircraft,

pitch axis is perpendicular to the yaw axis and is parallel to the plane of the wings with its origin at the centre of gravity and directed towards the right wing tip and

roll axis is perpendicular to the yaw and pitch axes, with origin at the centre of gravity, directed towards the nose of the aircraft. (Aircraft Rotations, NASA, 2010).

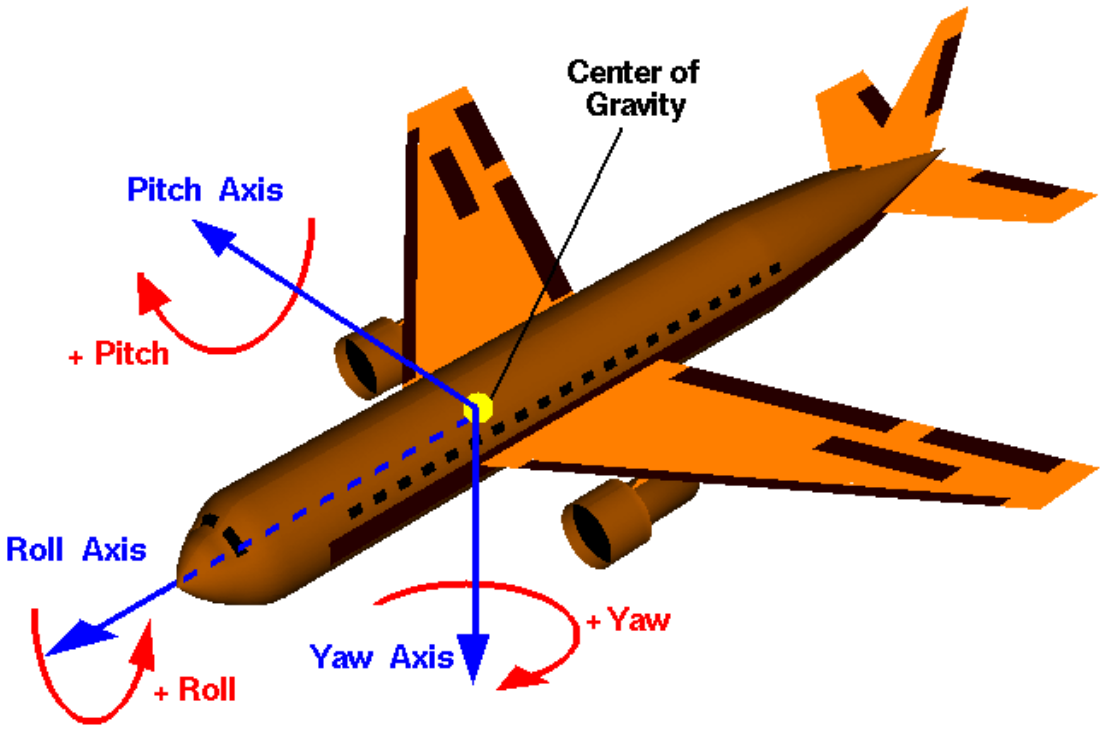


Figure 5: Vehicle fixed coordinate system and aircraft rotations (Source: Aircraft Rotations, NASA, 2010).

Rotation angles are measured with inertial measurements unit (IMU) which uses accelerometers and/or gyroscopes. Inertial navigation system (INS) is similar measurement device to IMU, thus compared to IMU does not need integration with GPS. Both IMU and INS measure besides orientation angles also the velocity of an aircraft. Relative position between all measurements devices mounted on a platform must be known or calibrated.

Luhmann (2010, p. 48-67) describes standard model for camera calibration with many additional IntOri parameters as well as simultaneous estimation of the IntOri and ExtOri parameters.

2.6 Least square adjustment with observation equations

Least square (LS) adjustment is an adjustment commonly used in numerous of tasks. The LS is easy to apply, since normal equations (A^TA, see Eq. 11) are linear, it gives unique solutions and it allows statistical testing of estimated parameters. In this work it was used to adjust the measurements of the image points to re-calculate the position of the helicopter at the time of acquisition, i.e. the ExtOri parameters of the helicopter (see 2.5 Central projection in space).

Adjustment is required in case of redundant measurements. Mathematically formulated, the over determinate system of equations is solved using least square adjustment.

Fundamental condition of least square adjustment

$$\sum_i^n v_i^T p_i v_i = min, \qquad \textit{(Eq. 9)}$$

where:

v_i residual of ith observation,

p weight of ith observation. Weights are expressing relative values of observations and

n number of independent observations, i = 1,..., n.

Mathematical model consist of functional and stochastic model.

Functional model is set by n independent observation equations $\hat{b} = f(\hat{x})$ with n adjusted observations and u adjusted unknowns $\hat{x}$.

Stochastic model is set by covariance matrix of observations $Q_{bb} = P_{bb}^{-1}$, where P_{bb} is weight matrix with diagonal elements equal to variance of each observation. If the observations are dependant, non-diagonal elements of P_{bb} are co-variances of belonging observations.

If observation equations are not linear, they are first linearized by approximation to a first-order Taylor series expansion. Coarse values x_0 for unknowns are used in first iteration and least square adjustment is applied iteratively.

$$\hat{x} = x_0 + \Delta\hat{x}, \qquad \textit{(Eq. 10)}$$

$$\hat{v} = A \cdot \Delta\hat{x} - w, \qquad \textit{(Eq. 11)}$$

where:

x_0 coarse value of unknowns,

v residuals of observations,

A design matrix (matrix of coefficient by the linear observation equations),

w observations,

^ least squares estimate of the parameter beneath and

Δ increments to the assumed values of unknowns.

(Kraus, 1993, p. 384-386, Stopar, 2006, 2007).

2.7 Texture mapping

Texturing an existing city model enhances its visualisation and also enables extraction of information from textures. Sufficiently accurate texture mapping is a preliminary condition for improving the city model with information gained from textures. In this thesis, a possible method for improving the detection of the edges of surfaces of a building is researched. These edges of surfaces should be further used to cut out the part of image and project it into the 3D model space as a texture. Furthermore, textures in the model can be evaluated.

Definition of texture is not precise due to its wide variability and it differs according to application. The following two quoted definitions of the texture are sufficient for the application in this thesis. Texture refers to the properties that represent the surface or structure of an object (Sonka, 2008, p.718) or more generally, »a multidimensional image that is mapped to a multidimensional space« (Heckbert, 1986).

Texture mapping is a technique in which a texture or a raster images are mapped onto a surface in a 3D scene, as wall paper is applied on wall (Heckbert, 1986, Catmull, 1974). »Texture mapping can be used to define many surface parameters besides colour. These include the perturbation of the surface normal vectors to simulate bumpy surface (bump mapping), transparency mapping to modulate the opacity of a translucent surface, specularity

mapping to vary the glossiness of a surface, and illumination mapping to model the distribution of incoming light in all directions« (Heckbert, 1984) and also many others.

The advantage of texture mapping is that details are added to the scene or model and it appears more realistic. Example of a realistic texture mapping is draping an ortophoto that is a geometrically corrected aerial photography over a DTM. According to the CityGML standard (2008) this is LOD0. A 3D building model in the LOD2 can have high-resolution textures mapped onto the surfaces (CityGML, 2008); the textures can be digitally designed or cut out of photographs (images) and clipped to the edges of model (Figure 6).

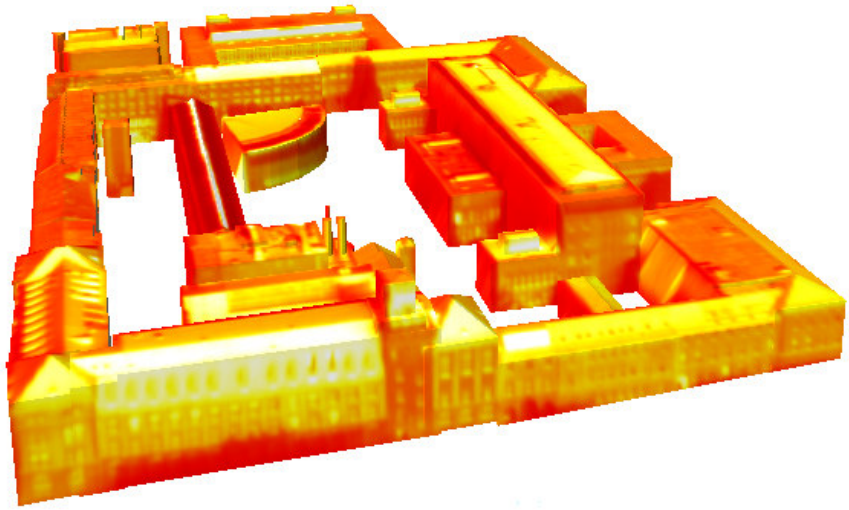


Figure 6: Textured 3D building model of TUM in LOD2 with IR textures (Author: D. Iwaszczuk).

Urban information can be extracted from images (Hoegner, 2009), thus texturing an existing 3D city model with textures extracted from images can enhance the model. Further, detailed analysis of textures in 3D city model requires considering the geometry of the captured scene and accuracy of data acquisition (Kolecki, 2010) in order to apply sufficiently accurate geometrical position to detected detail in an image.

3 METHODOLOGY

This chapter is divided into two parts; the first part describes the survey that was made about the feature extraction from infrared images and the second part presents the developed methodology which is tested on a dataset, described in chapter 4.

3.1 Overview of feature extraction from infrared images

To find an appropriate feature extraction method, a survey was made on a sample IR image sequence. Choosing an appropriate feature extraction method is of great importance for developing a method for refining orientation parameters of a sensor. Aim of this survey about feature extraction is to find appropriate features, which are good representation of buildings and will enable automatic matching with a building model. In this section a few examples of the feature extraction are shown. This survey was done in program Halcon MVTec[1]. Examples are shown on a sample image number 13200, 4th stripe; the main campus of the Technische Universität München (TUM) is seen in the scene.

Choosing the appropriate feature type and parameters for feature extraction procedure is not a trivial task and it requires detailed study, which was performed within the thesis in this section. Changing one parameter of feature extraction operator/procedure significantly influences the amount and appearance of extracted features. The feature type and parameters should be chosen in regards to given data and should be the same for one task. However, this problem is addressed later in developing a methodology, but it is not the main focus of this thesis. Images in the IR domain have lower accuracy compared to images in the VIS domain.

[1] Halcon is a commercial software product specialised for tasks for machine vision, produced by MVTec. (Halcon, 2010).

3.1.1 Subpixel edge extraction

Subpixel edges extraction operation detects step edges by linking the edge points into the edges (contours). A Variety of edge filters can be used for edge detection, naming a few: Sobel, Canny, Derichle, Shen, etc. Filtering an image is a local image operation that uses small neighbourhood of a pixel and its radiometric values as an input values to calculate new radiometric value in output image in order to enhance chosen characteristic (Sonka, 2008). In digital image processing filter is represented by usually a square matrix of a size n × n, also called filter matrix. The values of a filter matrix depend on a desired effect of filtering, i.e. for relatively small filter widths, ca. up to 11 × 11 pixels (px), different filters result in similar edge detection. Edge operators can be applied in recursive way or conventionally using filter masks. Computational cost of recursive filters is independent of filter width, whereas it significantly increases for non-recursive filters for increasing filter width.

Both represented examples of subpixel edge extraction are realized through filter masks (non-recursive). Three parameters of subpixel edge extraction can be set:

- alpha is width of a filter matrix (e.g. when alpha = 3 px, the size of a filter matrix is 3 × 3 px),
- threshold value low and
- threshold value high.

Both threshold values are set for automatic acceptance or rejection of candidate points for contours.

Gaussian smoothing kernel of size alpha is used to determinate quadratic polynomial in x and y direction for each image point. Both are direction of axes in pictorial coordinate system that are directions of rows and columns of digital image. The parameters of the polynomial are used to calculate the line direction for each pixel. Candidates for edge points are pixels, which have local minimum in the second directional derivatives perpendicular to the line direction. Candidates for edge points that have the second derivative smaller than low value are automatically rejected, and points with the second derivative higher than high value are automatically accepted to be a part of contours (extracted edges). Other points can be also

accepted as a part of the edge if they are connected to secure points. Thus, secure points influence their surroundings.

Input variables: input image, alpha, low, high.
Output variables: subpixel contours, edge direction, edge amplitude.

Sobel filter

Edge detection using the Sobel filter is convolving the image with the selected filter in horizontal and vertical direction. The Sobel filter matrix is a numerical approximation of the gradient of the image intensity function.

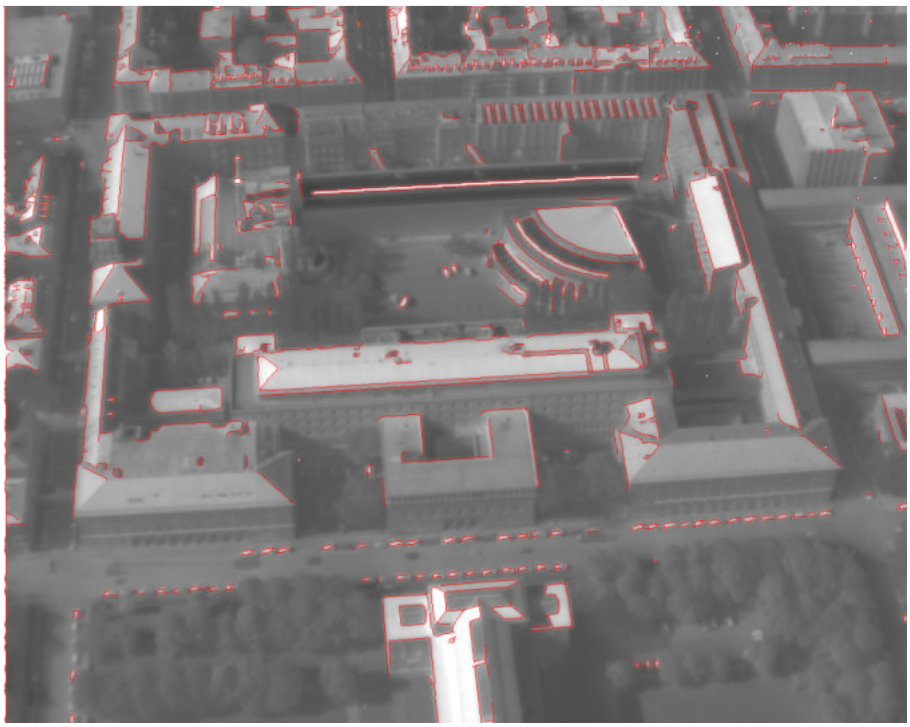


Figure 7: Extracted subpixel contours using Sobel filter (red). Input variables are: input image 13200, alpha: 3 px, low: 20, high: 40.

Canny Edge Detector

Subpixel edge extraction using the Canny filter (edge detector) is realized through convolution of the image with the Gaussian filter.

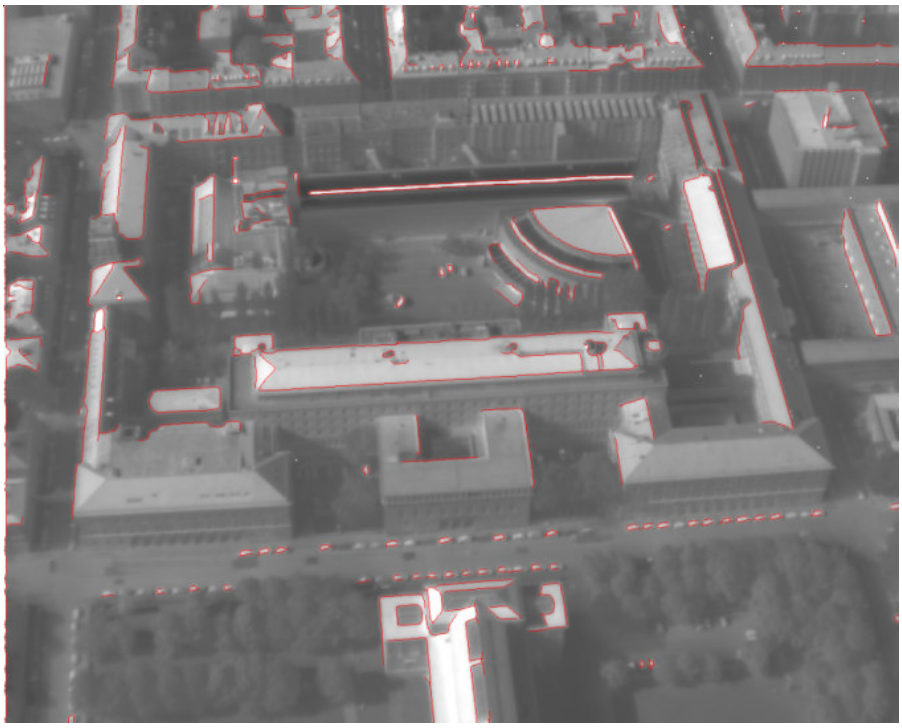


Figure 8: Extracted subpixel contours using Canny edge detector (red). Input variables are input image: 13200, alpha: 3 px, low: 20, high: 40.

3.1.2 Subpixel contours approximated by line segments

Buildings are usually rectangular, with straight roof ridges and building details, e.g. chimneys, windows, are often describable by simple geometric forms. For these reasons, buildings can be good represented by straight line segments, instead of arcs or other curves.
Both subpixel edge extraction operators mentioned in 3.1.1 result in contours that represent detected edges. Straight line segments are simpler features and can be easier processed in matching algorithm than contours. In this section, a possible algorithm to approximate contours with line segments is described. Moreover, contours can be approximated not only with line segments, but also with circular and elliptic arcs.

Firstly, contours are smoothed with the purpose to make the algorithm more robust and to avoid very short line segments (Figure 10). Secondly, the polygonal approximation is done by the Douglas Peucker algorithm (Douglas, 1973), in literature also referred to as the Ramer algorithm. The Douglas Peucker algorithm is applied in an analogue way as in Stilla et al. (1995). The article also includes stepwise image representation of the algorithm on p. 12 (Stilla, 1995, p. 9-15). A parameter maximum distance in pixel units must be set. The contours are segmented in a way that the Euclidian distance of the approximating polygon to the contour is at most the value set with parameter maximum distance. When segmenting contours with lines and arcs, a two-step algorithm is proposed to decrease the computational effort of the procedure. The result is segmented contours (polygons).

In the next step, these segmented contours are approximated by the longer line segments using, e.g. linear regression. Line fitting can also be calculated with other methods, using weighted mean square algorithm, etc. Some properties can be set to improve approximation of line fitting, for instance maximum number of contour points used for computation, number of points at the beginning and end of the contour to be ignored for fitting. A statistic method is used to estimate the standard deviation of the distances from the contour points. Special attention should be dedicated to detecting outliers that can significantly influence line fitting. Outliers are not taken into computations; however, the quantity of the outliers is controlled with the clipping factor, i.e. a scaling factor for the standard deviation. Outliers can be excluded or weighted with a small pound in the iterative line fitting algorithm.

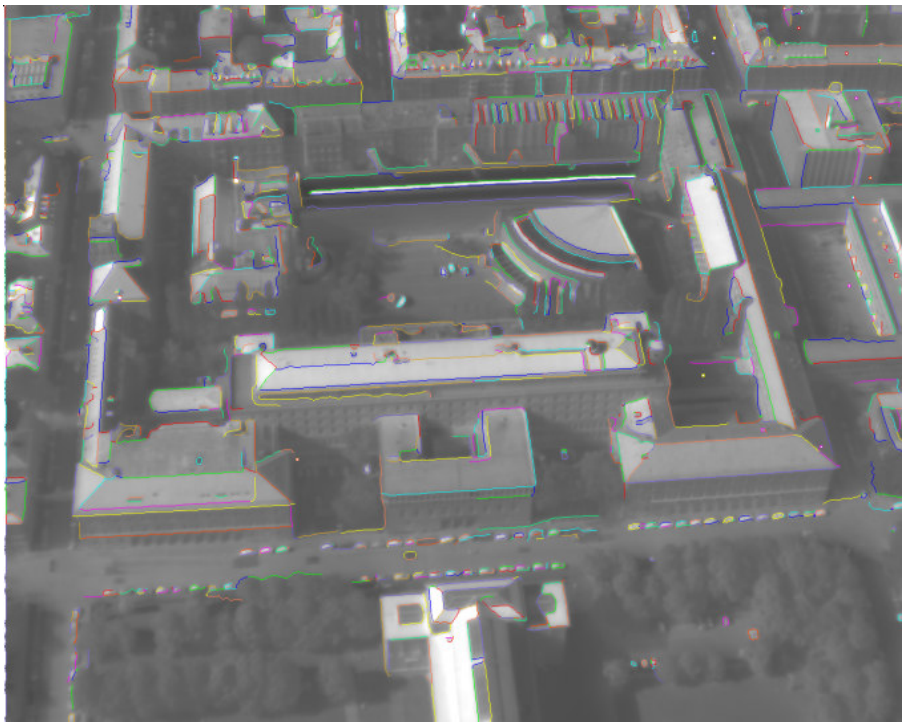


Figure 9: Smoothed extracted subpixel contours on image number 13200 (eight colour representation). These extracted contours are approximated by line segments shown in Figure 10.

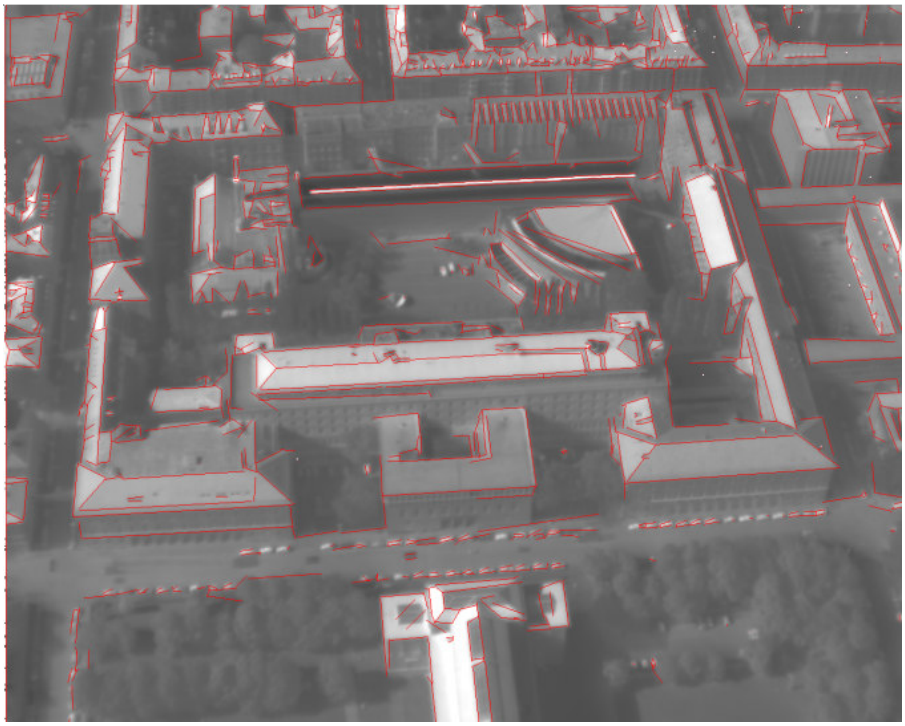


Figure 10: Extracted line segments (red) which are the result of approximating subpixel contours with line segments on image number 13200.

3.1.3 Region extraction with minimum covering rectangle or circle

Firstly, a threshold values low and high are applied on the image. All pixels with radiometric values in between low and high value are accepted. Instead of specified threshold values, automatic threshold values can be used, which are defined according to the histogram of the image and pre-used Gaussian smoothing filter. Secondly, the selected pixels are connected into regions using the neighbourhood. For extracting foreground, 8-neighbourhood is recommended (Halcon Reference Manual, 2010).

Regions with an area larger than the value areaLimMin and smaller than the areaLimMax are selected. Appropriate regions can also be selected according to different criteria, e.g. height, width, maximum diameter, orientation of region or more complex criterion. Regions extracted by this procedure can have gaps; therefore an algorithm to fill them up can be applied. This step of gap removal is important, if further method is processing region margins. At last, minimum covering circle and rectangles are calculated for each region and their centre of gravity.

Input variables: input image, low, high, neighbourhood, areaLimitMin, areaLimitMax.
Output variables: regions, minimum covering rectangle (circle), centre of gravity of minimum covering rectangle (circle).

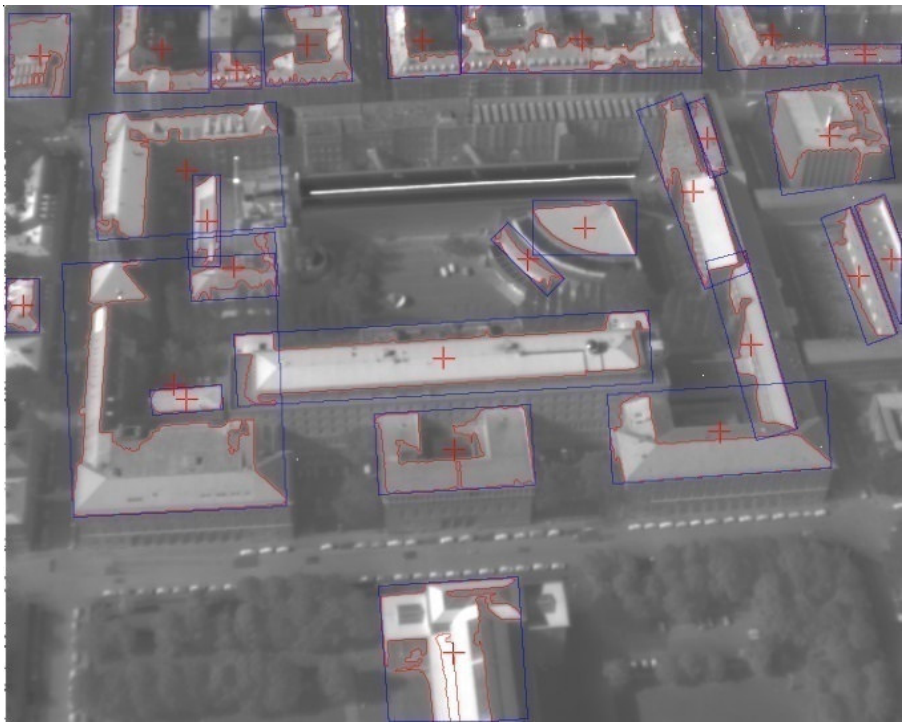


Figure 11: Extracted regions (red contours) with defined minimum covering rectangle (blue) and the centre of gravity of minimum covering rectangle (red cross). Input variables are: input image: 13200, low: 135 high: 225, neighbourhood: 8 px, areaLimitMin: 500 px, areaLimitMax: no max limit, algorithm to fill up holes is applied.

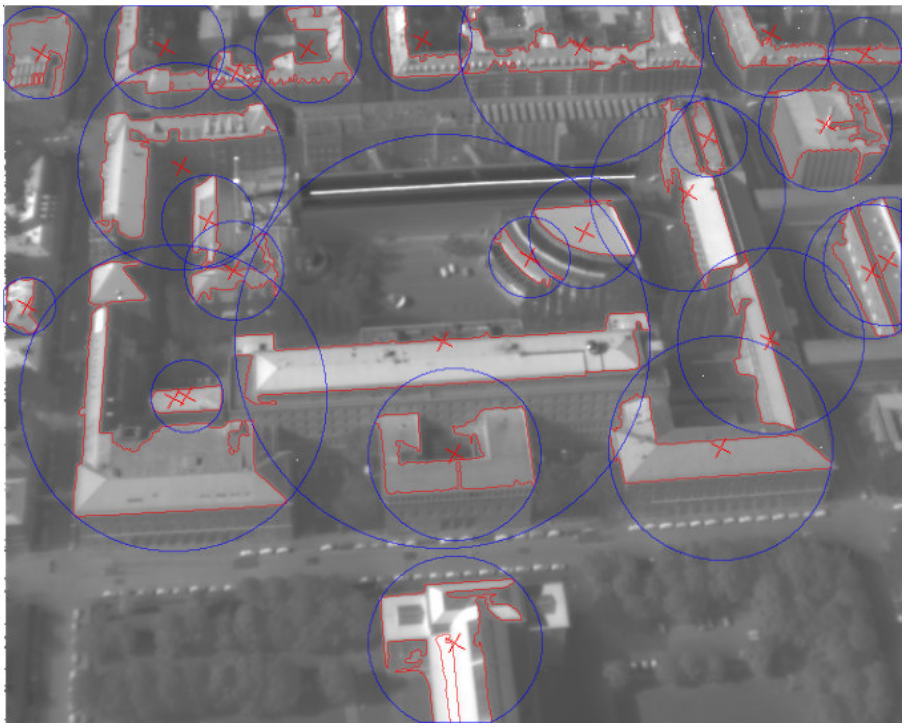


Figure 12: Extracted regions (red contours) with defined minimum covering circle (blue) and

the centre of gravity of minimum covering rectangle (red cross). Input variables are input image: 13200, low: 135 high: 225, neighbourhood: 8 px, areaLimitMin: 500 px, areaLimitMax: no max limit, algorithm to fill up holes is applied.

3.1.4 Förstner points extraction

The procedure of Förstner point detection described in this section is mainly adapted from the Halcon Reference Manual (Halcon Reference Manual, 2010). The Förstner operator was introduced in Förstner and Gülch (1987) as well as in Förstner (1993); the operator was later named after the author.

The Förstner operator detects significant points in an image that differ from the neighbourhood. There are two types of Förstner points; so called »junction points« which appear on the intersection of image edges and »area points«, where colour or brightness differs from the surrounding neighbourhood.

In the first step the point regions (inhomogeneous, isotropic regions) are extracted from the image.

Smoothed matrix *M* is calculated

$$M = S \cdot \begin{bmatrix} \sum_{c=1}^{n}(Ix,c)^2 & \sum_{c=1}^{n}(Ix,c \cdot Iy,c) \\ \sum_{c=1}^{n}(Ix,c \cdot Iy,c) & \sum_{c=1}^{n}(Iy,c)^2 \end{bmatrix}, \qquad \textit{(Eq. 12)}$$

where:

(lx, c), (ly, c) are first derivatives of each image channel *c* and

S is smoothing matrix, and can be realized with Gaussian or Sobel derivatives.

Gaussian smoothing: The derivatives are computed with Gaussian derivatives of size sigmaGrad and smoothing is performed by a Gaussian of size sigmaInt.

Mean smoothing: Derivatives are computed with the Sobel filter size 3×3, smoothing is performed by the mean filter size sigmaInt x sigmaInt.

Degree of inhomogeneity in the image is calculated

$$inhomogeneity = trace(M). \quad \textit{(Eq. 13)}$$

Degree of the isotropy of the texture in the image

$$isotropy = 4 \cdot \frac{det\,(M)}{trace(M)^2}. \quad \textit{(Eq. 14)}$$

Further examined points in the second step must fulfil the condition

$$further\ examined\ points = (inhomogeneity(pt) \geq ThreshInhom) AND\ (isotropy\ \geq ThreshShape), \quad \textit{(Eq. 15)}$$

where:

pt is image point, defined with row and column and

threshInhom, *threshShape* are threshold values for inhomogeneity and isotropy, respectively.

In the second step, two optimization functions for each resulting point from the first step are calculated. These optimization functions average for each point the distances to the edge directions (for »junction points«) and the gradient directions (for »area points«) within an observation window around the point. Averaging is performed with the same smoothing matrix set in the first step, but with different matrix size, that is a Gaussian of a size sigmaPoints or with the mean filter of size sigmaPoints × sigmaPoints. Local minimum of the optimization functions define the extracted Förstner points with subpixel precision. For each extracted point the corresponding covariance matrix is returned with the precision of the calculated position. However, to obtain the actual precision value, the amount of image noise should be estimated.

The nature of the Förstner point operator is that corners often results in pairs of extracted points; a »junction point«, where the edges of the corner actually meet, and one »area point« inside the corner. These doublets can be eliminated with regards to the precision of the point position. The point with higher precision of position is accepted.

Input variables: input image, sigmaGrad, sigmaInt, sigmaPoints, threshInhom, threshShape, smoothing.

Output variables: coordinates of »junction« and »area« Förstner points and covariance matrixes for each coordinate of extracted points.

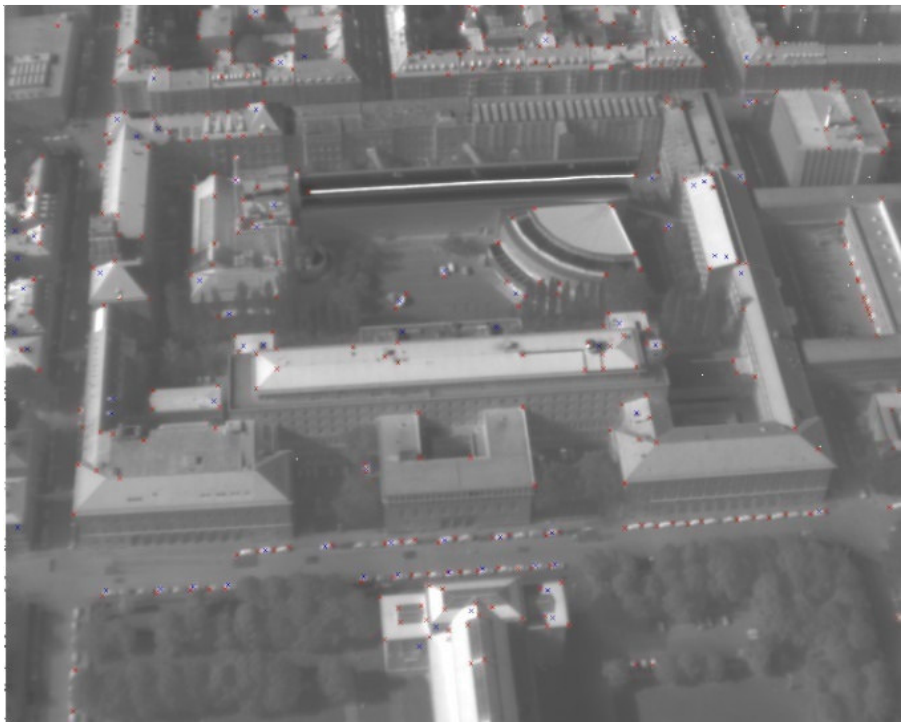


Figure 13: Extracted »junction« (red cross) and »area« (blue cross) Förstner points. Input variables are input image: 13200, sigmaGrad: 1.0, sigmaInt : 3, sigmaPoints: 4.0, threshInhom: 300, threshShape: 0.1, Gaussian smoothing is set, doublets are not eliminated.

Doublets are not eliminated, because »junction points« are points of interest in this work and »area points« do not represent significant, physically interpretable objects that could be connected with a building model.

3.1.5 Straight edge extraction

Extraction of straight edge segments results in dataset of coordinates of the start and end points of extracted line segments. For edge detection the Sobel filter is used, which is based on the matrixes A and B (Eq. 16 and 17), filter size can be set, for example 3 × 3, 5 × 5, etc.

The Sobel filter matrixes sizes 3 × 3 are

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad \textit{(Eq. 16)}$$

and

$$B = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}. \quad \textit{(Eq. 17)}$$

Convolving an image for one pixel with A and B results in a and b. The filter response, i.e. amplitude, of pixel is than calculated

$$filter\ response = \frac{|a|+|b|}{4}. \quad \textit{(Eq. 18)}$$

The candidates for edge points are pixels with a filter response larger than the specified threshold value for minimal amplitude minAmp. These edge point candidates are thinned and split into straight edge segments. Additional two parameters can be specified, the first one to control the maximum allowed distance between an edge point and its approximating line (MaxDist), and the second one to remove extracted line segments shorter than the set minimum length (minLength). This operation usually does not return closed object contours.

Input variables: image, filter size, minAmp, maxDist, minLength.
Output variables: coordinates of begin and end points of extracted straight edges.

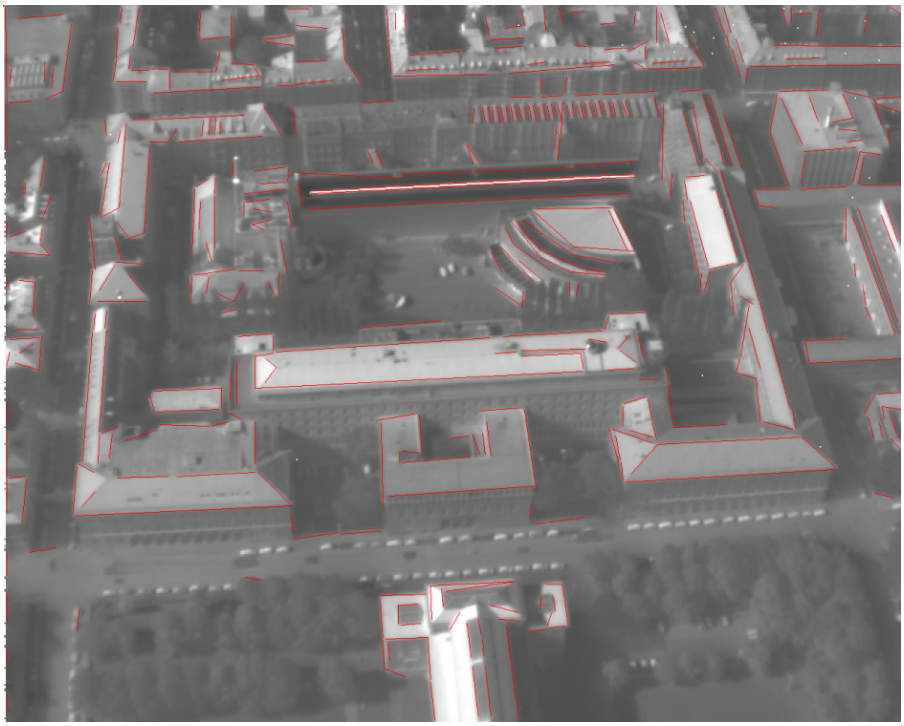


Figure 14: Extracted straight edges (red). Input variables are input image: 13200, filter size: 9 × 9 px, minAmp: 18 px, maxDist: 5 px, minLength: 10 px

In this section five different feature extraction algorithms in IR images are presented. In the Table 3 a comparison between them with regards to the desired properties is made. The suitability of each of the presented feature type and extraction is discussed in the next section 3.2.

Table 3: Comparison between feature extraction algorithms.

	Subpixel edge extraction		Contours approximated by line segments	Regions extractio	Förstner points	Straight edges
	Sobel filter	Canny edge detector				
Precision	Subpixel	Subpixel	Subpixel	Pixel	Subpixel	Pixel
Computational cost[2] [unit]	Ca.1.3	Ca. 1.2	Ca. 4	Ca. 1	Ca. 2	1
Type of features	Contours	Contours	Line segments	Regions	Points	Line segments
Stability over sequence[3]	High	High	High	Low-medium	High to medium	High

[2] Computational cost was defined for computational time in HALCON 9.0 MVTec software, all computations were made on same personal computer. Computational time may vary on different settings of procedures – especially with larger smoothing matrixes and with more complex procedures.

[3] Stability over sequence is defined by a visual check on a sample of image sequence consisting of 128 frames.

3.2 Developed method

An automatic texturing process requires geo-referenced data and a very precise co-registration between images and faces of the model. Before texturing, polygons of the building model should be projected to the images using coarse orientation parameters. By a matching process of projected polygons and features extracted from the image, the values of the projection matrix should be improved. The aim of this work is to investigate the automatic feature extraction in infrared images to improve orientation parameters of the sensor.

Image registration is a process of overlying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors (Zitová, 2003). Image registration is described in detail in the section 2.3. In this example, scene to model registration is carried out, i.e. the IR images (sensed images) of the scene and a 3D building model (reference image) are registered.

The main stress of this thesis is the developed method presented in this section. Firstly, the argumentation for choosing two feature types is given and secondly the method is described in six steps (Figure 15).

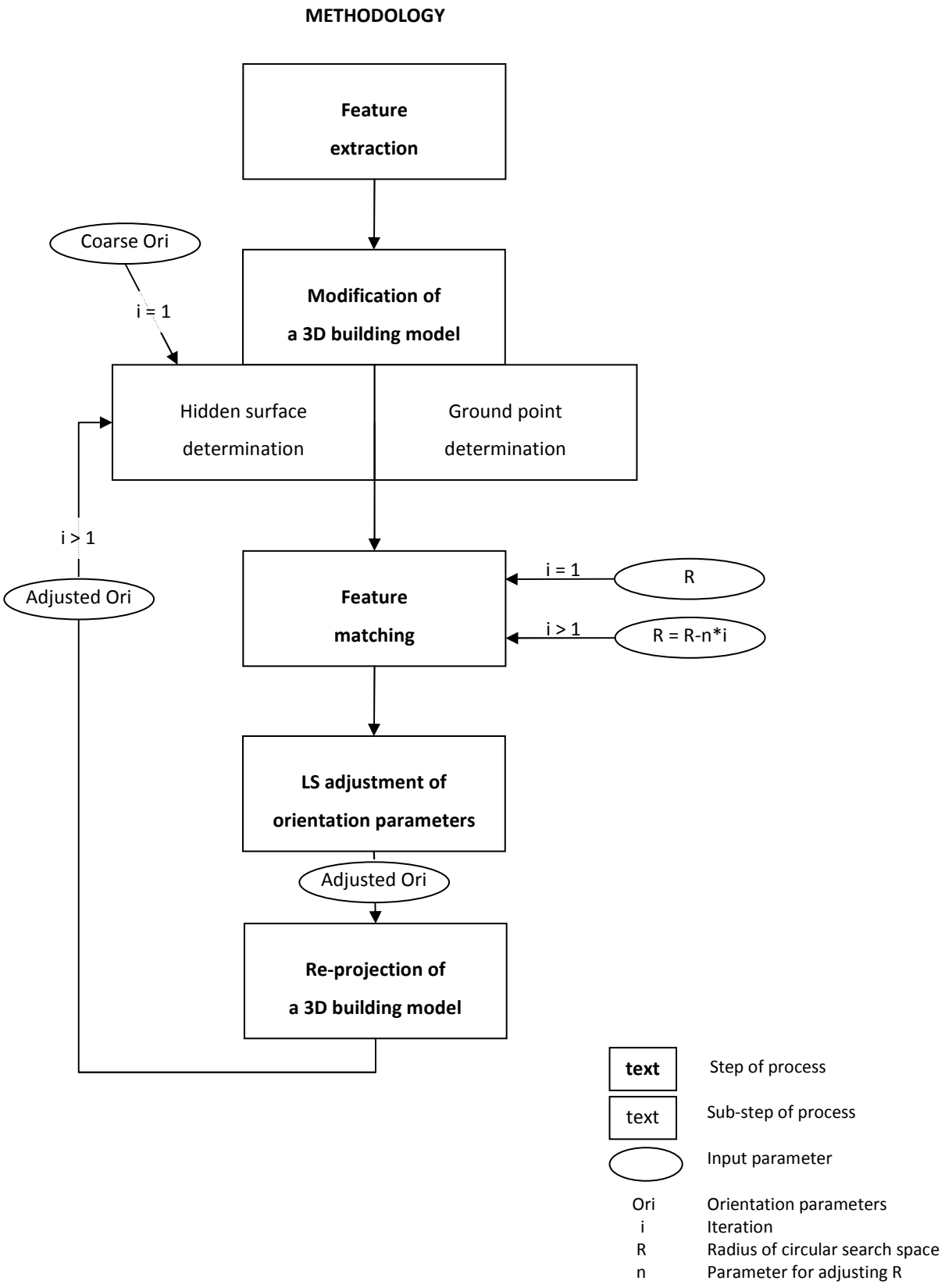


Figure 15: Diagram of methodology steps for feature extraction and matching.

3.2.1 Feature extraction from infrared image sequence

In this section a comparison and evaluation of the described feature detection algorithms presented in the section 3.1 is given. Secondly, the argumentation is given for choosing the Förstner point extraction and extraction of intersection points.

3.2.1.1 Discussion of different feature types

Feature extraction is widely used in computer vision, image processing and analysis tasks and its applications. Local features provide a compact description of the searched objects in an image based on their appearance. In the previous section 3.1 some feature extraction algorithms are described on a sample IR image. Deciding which type of feature(s) is appropriate for a given task is a first step of image registration. This section focuses on pros and cons of the features described in the section 3.1 with regards to the given problem.

Extracted features should be:

- highly distinctive, with low probability of mismatch,
- numerous and possible also regularly spread over image,
- easy to extract,
- physical interpretable and
- in certain range invariant to image deformation e.g., image noise, changes in illumination, uniform scaling, rotation and minor changes in viewing detection.

Dealing with image sequences in subsequent frames nearly the same scene is acquired multiple times from different viewing directions. Illumination has different effect on the same object in subsequent frames. When acquiring the same scene in different flight directions, the angle of illumination also changes. In images acquired in oblique view, the same object appears in significantly different size in different frames. If the acquisition device is forward looking, the same acquired object will appear relatively small when in top of the image and relative large when in lower part of the image. For all these reasons, stability of extracted features over a sequence should also be observed. A stabile feature in sequence is the one, which appears in several subsequent frames on same position. If the coverage of images is not

complete or occlusions are present, detected features in sensed image must have enough common elements with reference image to enable enough registered points for matching. The method of feature extraction should have sufficient localisation accuracy. Detected features should also be invariant to the certain degree to the image noise. (Zitová, 2003).

An ideal feature detection algorithm should detect all the features with high accuracy, independently of small changes in images mentioned in previous paragraph. In addition, the matching algorithm should ensure full completeness, i.e. the quotient between the number of matched points and all the possible matching points, and correctness of the matching algorithm. For refinement of the ExtOri parameters a connection between the same points in subsequent frames is not relevant, whereas for re-calculation of the IntOri parameters it is necessary for gaining higher accuracy.

When dealing with real datasets, a decision for a feature extraction is a compromise between the above mentioned desired properties and the computational effort in regard to given task. For instance, smoothing of image improves feature detection but is simultaneously worsening localisation accuracy. Feature extraction step cannot be entirely separated from the feature matching, because a number and type of extracted features directly influence the chosen matching algorithm and its computational effort. A higher number of extracted features results in higher computational cost of feature extraction and matching algorithm.

Buildings are in 3D building models often described with planes (surfaces) or wire-model (line segments or polygons). Models with a higher LOD are not required as resolution of the IR domain does not enable to extract very small and sophisticated details. This intuitively leads to the decision, that extracted features should be simple.

Regions

Extracted regions appear to be less stabile over a sequence than extracted contours, line segments or points and are therefore less appropriate for matching with model. However, the algorithm for the region extraction results in a stabile number of regions as well as stabile central point of minimum covering circle or rectangle over a sequence. These properties could

be further researched and potentially used for connection between frames or for estimation of coarse ExtOri parameters. Completeness of extracted regions that correspond to roofs is comparing to other represented feature extraction operators the highest, but has worse localisation accuracy. In addition interpretation problems can occur, because one region does not always represent one roof.

Subpixel edge extraction

Optimal edge detection algorithm should have:

- good detection - low probability of failure to detect an existing edge or detecting a non-existing edge,
- good localisation – a detected edge should be as close as possible to a true edge and
- single response - a true edge should be represented with only one detected edge.

(Sonka, 2008, p. 144).

Subpixel edge extraction results in contours with high geometrical position precision. A question arises, whether the subpixel feature extraction is required according to the resolution of acquired images (geometric, radiometric), accuracy of model, demanded accuracy of result and ability of applied edge extraction operator. There is no uniform answer to this question; a decision should be made in regards to the given task. Moreover, a precision of the extracted feature does not directly result in a better localisation of the building edges. For instance, a comparison between detecting features in image in the IR and the VIS domain is presented. A characteristic of the IR domain is that it appears blurred. In images acquired in the VIS domain, the building edges are sharper, with larger edge response and therefore easier detectable than in the blurred IR domain. Consequently, very high precision of feature extraction in images in the IR domain is not reasonable and does not improve localisation. What is more, geometric resolution of the IR cameras compared to geometric resolution of the VIS cameras is lower.

The computational cost of the subpixel edge extraction can quickly increase when using larger smoothing matrixes or when the difference between the low and high threshold values for immediate rejection and acceptance of pixels is larger. The number of extracted contours is

sufficient, thus many of the extracted contours are very short and they represent objects in space, which are not part of building models, e.g. cars. Even with comprehensive manipulation of the input parameters of the edge extraction, completeness of the edge detection is low. Short contours in majority represent non-existing or unwanted edges, and can be excluded from further computation with a simple length threshold value. A good single response of extracted edges can be achieved by adjusting parameters of the edge detector to the data type.

Contours approximated by line segments

Straight line segments are simpler feature as contours and they can satisfactorily represent buildings in building models. When approximating contours with straight line segments, different strategies can be applied (section 3.1.2). A computational time increases several times when calculating the approximation of contours with line segments. The line segment approximation requires additional input variables and possibly also detection of outliers. These additional variables make the procedure of line segments extraction complicated and difficult to control. Nevertheless, straight line segments appear to be the appropriate feature type for representing buildings in the IR image sequence.

Straight edges

Detection of straight edges has, in comparison to the previous described contour extractions and the contours approximated by the line segments extraction, lower computational cost. The localisation is not significantly different, whereas in comparison to the contours, the completeness is approximated by the line segments extraction a little worsen. On the other hand, there are fewer detected lines which do not correspond to the model lines. Weakness of all described edge detectors is that the contours or line segments often do not finish at the edge points of buildings. That can require more complicated matching methods.

Förstner point extraction

The Förstner point extraction algorithm results in many extracted points frequently spread over image. The extraction method has more input parameters than the extraction of straight edges. However, the Förstner operator is an uniform standard procedure and not a sequence of

different standard procedures, so the influence of the input variables can be easier controlled. »Junction Förstner points« representing edges of buildings are stabile over a sequence, thus the extracted »area Förstner points« and »junction points« which represent smaller object in scene are less stabile. The localisation accuracy is in general lower in comparison to the straight edge extraction. A Benefit of this operator is the number of extracted points and the simplicity of extracted feature type (point). A disadvantage is the low completeness in detection – the Förstner operator fails to detect some existing edge points and results in detecting many other objects, e.g. cars. The problem with the completeness of feature detection is similar to detection of contours. For both operators, manipulating the input parameters can results in higher number of distinctive extracted features, but simultaneously results in detection of more non-relevant objects in scene.

3.2.1.2 Points, the chosen feature type

The choice for the feature types was made in regards to the sections 3.1 and 3.2.1.1. With the intention to simplify, the matching algorithm extracted points and the straight line segments are suitable for searching the correspondences. For the matching extracted lines or line segments and building model lines projected into images, a correlation function is needed. Computational effort of line-to-line matching method is high and additionally, extracted line segments often do not finish at the edge points. Problems of multiple responses of edges and detection of non-desired edges (extraction of objects that do not represent building edges) could aggravate the matching algorithm. The point-to-point matching algorithm has significantly less computational cost. The Förstner points can be used for feature extraction. However, the Förstner point extraction has shown smaller localisation inaccuracy on the example image which might noticeably influence the ability of the matching algorithm.

The straight edges extraction is efficient with regard to the computational cost, the representativeness of buildings as it was in detail explained in the section 3.2.1. To overcome the previously mentioned problem of the line-to-line matching computational effort and the problem of extracted line segments which do not end on building edges, intersection points of extracted straight segments can be computed. Two different point types are extracted and used

for matching: the Förstner points and the intersection points of extracted straight edge segments.

Intersection points of the extracted straight edge segments (intersection points) are calculated using intermediate step of extracting straight edge segments (described in section 3.1.5). the distances *d* between all the end points of all extracted straight edge segments are calculated. The maximum distance limit between the edge points d_{max} in pixels is set. In other words, around each endpoint of extracted straight edge segments, a circular search space with radius size d_{max} is established. A large d_{max} would result in a higher number of extracted intersection points, though most of them could not be physically interpretable. Furthermore, this would cause more wrong matches between model and extracted points. For all pairs of end points, which fulfil the condition $d \leq d_{max}$, an intersection angle α of the belonging edge segments is calculated. The intersection angle should be between the limits α_{min} and $\alpha_{max} = \pi - \alpha_{min}$ where $\alpha \in [0, \pi]$. The intersection angles between the nearby buildings edges are seldom very acute or very obtuse. Angle limits α_{min} and α_{max} prevent the algorithm to result in points defined by two almost parallel lines. Almost parallel lines should be avoided, because they are most likely a result of their multiple edge response, i.e. bad feature detection, not very accurate detection of two parallel edges or in parallel edges of the same building on different heights, which appear very near in images acquired in oblique view. For all pairs of edge points that fulfil both conditions: $d \leq d_{max}$ and $\alpha_{min} \leq \alpha \leq \alpha_{max}$ the intersection points are calculated. Both, angular and distance limits simultaneously restrict the intersection point algorithm to result in many non-reliable points. In other words, around each endpoint of the extracted straight edge segments a circular search space with the radius size d_{max} may also be defined as an ellipse or rectangle with the intention to weight the influence of the distance and angle limits.

Extraction of the intersection points and Förstner points is chosen as the appropriate feature extraction to assure good extraction. Both extraction algorithms are evaluated and compared in chapter 4.

3.2.2 Modification of 3D building models

The term modifying a 3D building model is used for changing the model in a way to be more appropriate for given task but without influencing the geometry. First part of this section addresses the problem of hidden surface determination as a consequence of the viewing angle. The second part explains the meaning of excluding the ground points of the building model before the matching algorithm is applied.

Hidden surface determination in 3D models

Representation of 3D models requires dealing with invisible planes and parts of planes. Depending on the viewing point, different surfaces of the 3D model are seen. In order to ensure the best possible matching between the projected model and the extracted features, occluded parts of the projected model should be switched off. Thus, some false matches can be avoided.

Planes or part of planes are not seen due to:

- viewing angle (self-occluding planes in the model, orientation of a model plane according to the position of acquisition device) and
- other objects between the acquisition device and plane.

Hidden surface determination algorithms are implemented in almost all commercial accessible software. The main problem occurs with partly occluded surfaces. In computer graphic this is called a clipping problem and is non-trivial task. However, several clipping algorithms have been introduced. Special attention should be paid to choosing the appropriate software for hidden surface determination as some of them are using rasterization, ray-geometry, etc. For instance, when projecting a building model into a 2D image, some polygons are occluding and they create fictive intersection points. Some algorithms only hide the part of polygon that is not seen, which is sufficient for representation, but might not be satisfying for all analysis of data. In this work, the vector data of fictive intersection points are required.

Ground points

In visual (VIS) domain the boundary lines between objects in an image are easily recognized. The precision, with which a certain boundary line can be detected, depends on spatial and radiometric resolution. When analyzing the IR images characteristic of the IR spectrum should be considered, e.g. in IR the images appear to be blurred. Two neighbouring points acquired in the IR domain have similar radiometric value and the boundaries are not as exact as the VIS domain. Blurred appearance of objects in the IR images depends more on materials of the observed objects and their physical characteristics and not only on the colour. In images in the IR domain, differences in radiometric values between low parts of facades and pedestrian areas are small, therefore only roof edges are considered. Ground points can be switched off using DTM or by a threshold value in lowlands. Alternatively, roof surfaces can be extracted from a model based on the direction of a normal vector of a plane.

3.2.3 Feature matching

Feature matching can be performed using area-based or feature-based methods. Area-based or correlation-like methods merge the step of feature extraction and matching. The matching process is based on the pre-defined size matrix cut out from a reference image, which is searched in sensed image using e.g. correlation, Fourier method. Area-based methods are usually not dealing with detection of salient objects. (Zitová, 2003). In this method, feature extraction is performed separately and is described in the section 3.2.1.

Feature based methods are searching correspondence pairs of extracted features from sensed and reference image. 3D building models are a computer representation of the scene and are projected into a 2D IR image with coarse orientation parameters. In an IR image point extraction is used to define significant points (sensed, extracted points). Establishing correspondences between model points projected into the IR image and extracted points is based on spatial relation among them. Point-to-point matching algorithm is applied. For each model point a circular search space with the radius R is defined in pixels. Different cases of relations among extracted points and model points occur. If an extracted point from a sensed image is in the search space of a model point, they are connected, a correspondence is found.

In case that none of the extracted points is in the search space of a model point, there is no correspondence. In case that two or more extracted points are in the search space of a model point, both or all of them are registered to the model point. In case that one extracted point lies in two or more search spaces of model points, it is connected only to the nearest model point. Found correspondences are stored in a correspondence matrix using point identifiers.

3.2.4 Least square adjustment of orientation parameters

The co-registrated points are adjusted with the LS algorithm and corrected orientation parameters are calculated. Input data for the LS adjustment are observations, i.e. picture coordinates of extracted points which correspond to the model points, model points in world coordinate system and coarse orientation parameters. Orientation parameters are unknowns in adjustment. The equations for the estimation ExtOri and/or IntOri are non-linear, thus the parameters are improved iteratively. It is assumed that the camera is calibrated (principal point, focal length and additional parameters), and that lever arm and boresight corrections, shift of the GPS measurement unit according to the centre of IMU are known. A time correction is calculated and applied on the data, thus IMU and GPS measurements refer to the centre line of the acquired image. With the LS adjustment ExtOri and IntOri parameters can be estimated if there are enough correspondence points found and the measurements are sufficiently accurate. The IntOri parameters can be assumed constant for one flight, therefore for estimating the IntOri parameters, correspondence between frames should be considered.

Statistical testing of the estimated orientation parameters (unknowns) with the LS adjustment is given with: standard deviations of each orientation parameter, calculated in variance-covariance matrix and mean square error (MSE) $\sigma_0^2 = s_0$ or root mean square error (RMSE).

3.2.5 Re-projection of 3D buildings model

With corrected orientation parameters a model should be projected into the image with the same algorithm used for first projection of a model.

3.2.6 Iteration of algorithm

Corrected orientation parameters can be calculated iteratively. Feature extraction is separated from the matching extracted features with model points. Therefore, projection of a model with corrected orientation parameters into the image can be used as coarse orientation parameters in second iteration.

The radius R of a circular search space can be set as fixed for all iterations, or adaptive according to the iteration. If the coarse orientation parameters have high accuracy, a fixed radius is sufficient. With less precise coarse orientation parameters in the first iteration, some mismatches might be present than influence the adjustment. In the second iteration, some of the mismatches can be avoided using smaller search space for correspondence between model points and extracted points.

4 EXPERIMENT

This chapter is divided in seven main sections. Firstly, the used data are described. Secondly, the test of the applied method on the dataset is given and thirdly, the results are presented. Then, the comparison of two point extraction algorithms is presented and the problem of the used 3D building model is addressed. Finally the results are discussed. Application of the method is realized with two commercial programming software packages, i.e. Halcon MVTec and Matlab[4].

4.1 Data description

In this section, an overview of used data and acquisition of them is given. Firstly, IR image sequence acquisition is presented with characteristic of the IR camera. Secondly, the GPS/INS measurement required for calculations of coarse ExtOri parameters are described. Last, the modelling of the 3D building model is paraphrased.

4.1.1 The infrared image sequence and geometry of acquisition

The thermal images were taken in a dense urban area with a high resolution IR camera AIM 640 QLW FLIR with the frame rate of 25 images per second, mounted on a platform carried by helicopter. The flight height was approximately 400 m above ground level and the camera was forward looking with oblique view of approximately 45° (pitch angle). The resolution of images is 640×512 pixels, the acquisition of one image lasts for 0.04 s, thus a rolling shutter effect is noticeable. Geometry of acquisition is shown in Figure 16.

[4] Matlab is a numerical computing environment and high-level technical computing language developed by MathWorks.

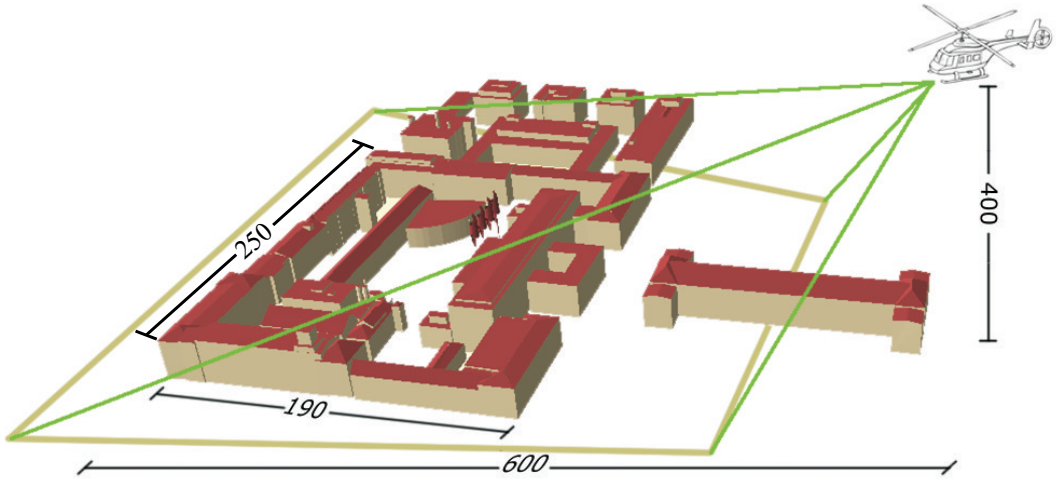


Figure 16: The geometry of the acquisition of IR image sequence. (Source: Stilla, 2009).

A helicopter flew over the test area, i.e. main campus of the TUM, four times recording a sequence of IR images. From a sequence four stripes were cut out, each with approximately 125 images (frames). The trajectory of the helicopter flight and the direction of flight for each stripe are shown in Figure 17.

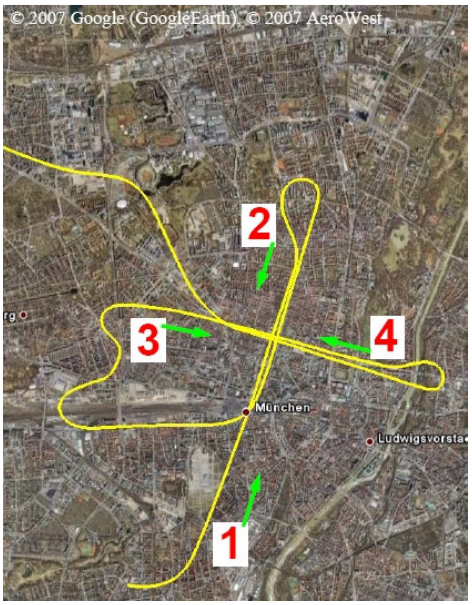


Figure 17: Test area and flight trajectory of the helicopter. (Source: Hebel, 2007).

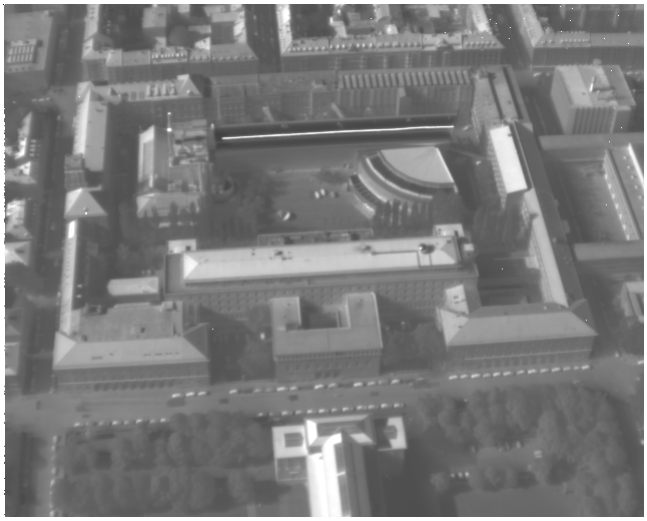


Figure 18: An IR image number 13200 from the 4th stripe. (Source: Stilla, 2009).

4.1.2 GPS and Inertial data

ExtOri parameters used in calculations in this thesis are the result of the extended system calibration proposed by Kolecki et al. (2010). First, the acquisition of raw GPS and INS data and then the extended system calibration are presented.

Raw data

The helicopter was equipped with the GPS/INS Applanix POS AV 510 system that measures position, GPS time of the measurements, altitude/heading and current speed of the helicopter (approximately 160 km/h). GPS antenna is mounted on the helicopter cockpit and registers position with the frequency of 1 Hz. Differential global positioning system (DGPS) corrections were not available for the time of the flight campaign, thus GPS/INS data have lower accuracy. The frequency of the inertial measurement unit (IMU) output is 200 Hz measuring roll-pitch-yaw angles. The GPS measurements are used to correct the IMU drift with Kalman filtering (Kolecki, 2010). Measured WGS 84 coordinates refer to the IMU centre, which is close to the IR camera. For each frame a time of acquisition is known; this enables the assignment of the ExtOri parameters to an appropriate image.

Initial projection of a building model in the IR image using GPS/INS ExtOri parameters have shown about 8° misalignment of the camera and IMU coordinate system in pitch rotation. Roll and yaw misalignments are smaller (Kolecki, 2010). Therefore a direct texture mapping cannot be utilized and the ExtOri parameters were corrected.

Extended system calibration

Only approximate IntOri parameters and distortion values of the IR camera were known. A system calibration was done solving extended bundle adjustment with camera self calibration. Approximately 140 control points were measured on an accurate reference stereo model (Stilla, 2009). Camera optics parameters that were estimated and/or corrected are: camera constant, principal point coordinates, radial and tangential distortions coefficients, y axis scaling and skewness in x and y direction.

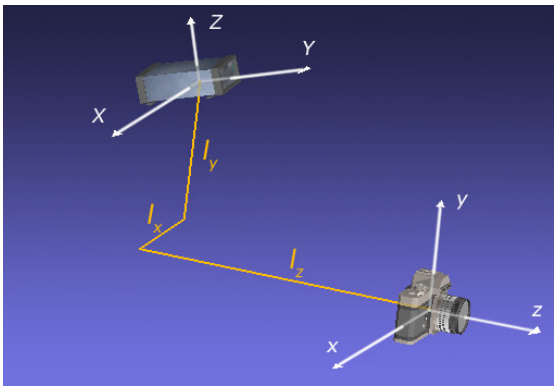


Figure 19: Relative position of IR camera and IMU presented with leverarm vector (l) and IMU misalignment (boresight). (Source: Kolecki, 2010).

Kolecki et al. (2010) and Stilla et al. (2009) suggested extended system calibration. The estimated values are IntOri, camera optic parameters, leverarm and boresight corrections, position systematic error and its change in time. Leverarm parameters (l) i.e. the distance between the camera centre and the IMU centre are assumed to be constant during the flight (Figure 19), whereas boresight parameters (angular misalignment) are corrected with a linear model. The GPS measurements are used to correct the IMU data every second; within a second, the position of the IMU is calculated based on gyros and accelerometers. Particularly due to the lack of DGPS data, the GPS measurements have a certain level of inaccuracy which appears to be nearly systematic and therefore the GPS shift was modelled. The time of acquisition assigned to each frame does not exactly correspond to the IMU measurement which also results in systematic time error that was estimated.

Applied corrections improve the projection of a building model into IR images and the remaining misalignment is only a few pixels. Kolecki et al. (2010) address an effect of platform vibration, which could be observed in angular measurements of the IMU unit over time. The frequency of vibrations for yaw and roll angles is estimated to 10 Hz and the amplitude exceeds 0.5°. The effect of the platform vibrations causes smaller jumping of the building model projected into the IR image sequence which is observable in subsequent frames.

4.1.3 The 3D building model

A 3D wire model in LOD2 was produced with semi-automated method in photogrammetric software INJECT[5]. The model was produced from aerial images and is showing generalized view of the main campus of the TUM. Due to generalization, some inaccuracies in the position of the buildings edges occur. (Frey, 2006).

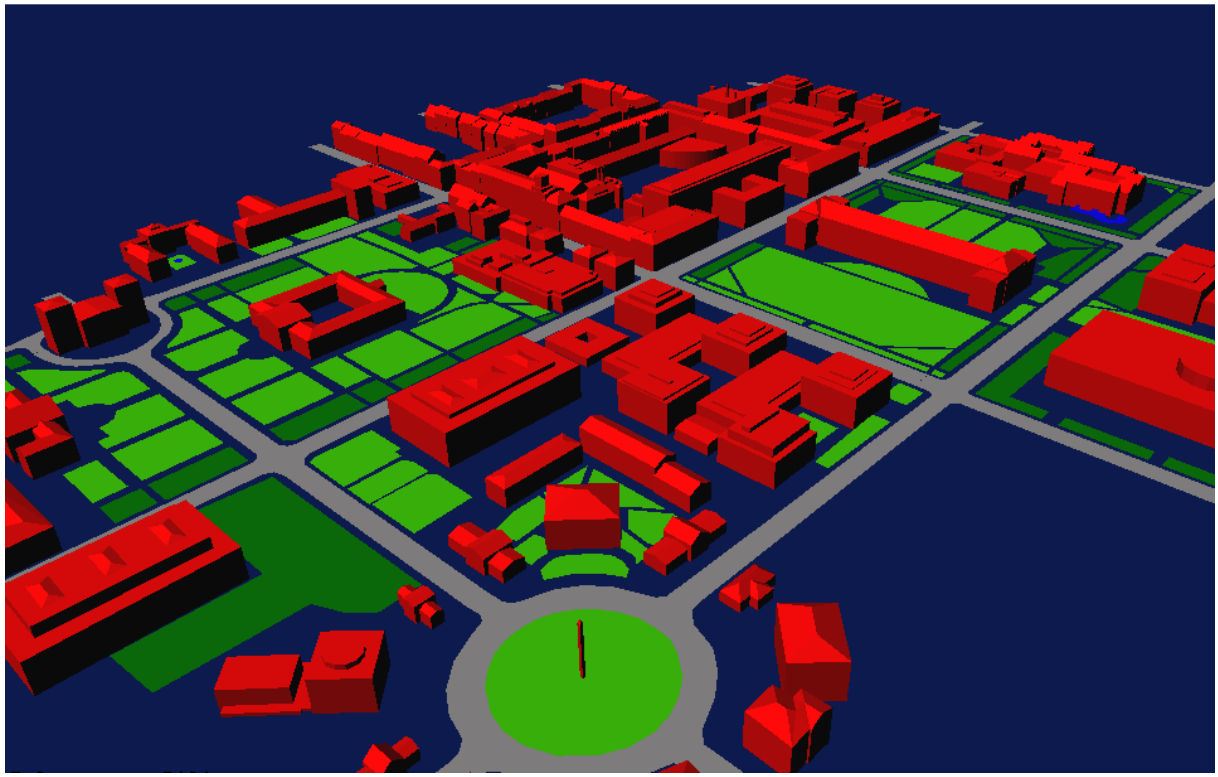


Figure 20: 3D building model of the TUM and surroundings produced with semi-automatic method. (Source: Frey, 2006, p. 13).

Semi-automatic method consists of manual selection of a building type (primitive), choice of two or three significant building points and a distinctive ground point. With this input data INJECT models a 3D building, adjusting primitive to the given data. The number of

[5] INJECT is a software for semi-automatic extraction of 3D structures from digital aerial images, emphasising the building extraction. The software also allows extraction of vegetation, water and roads. An overlapping satellite or aerial images are required with known ExtOri and IntOri parameters, as well as height data.

parameters that are calculated for each building depends on the selected building type (primitive) and is estimated using Random Sample Consensus (RANSAC). Primitives can be created manually and added to the program. Overall, this model required 18% of manual modelling; the rest was done with the semi-automatic method. The accuracy of the model was estimated to 0.5-1.0 m which is sufficient for urban planning, but it is unclear if it is sufficient for combination and registration with IR images. (Frey, 2006).

4.2 Test

The method described in the section 3.2 is applied to the data set presented in the section 4.1. From each stripe (#1-#4) three images are taken, one from the beginning, one from the middle and one from the end (Table 4). Two point extraction algorithms are used, i.e. Förstner points and intersection points of extracted straight edge segments (intersection points).The tests are made projecting the whole 3D building model into the image and also on two sub models.

Table 4: Selected IR images from all four stripes.

	Image number from			
Stripe	Begin	Middle	End	Stripe limits
#1	2011	2061	2120	[02003, 02128]
#2	4982	5032	5091	[04974, 05099]
#3	10140	10190	10249	[10132, 10257]
#4	13142	13200	13259	[13141, 13268]

Firstly, a methodology is tested on a sub model of one and two buildings (Figure 21) and secondly applied to the whole model.

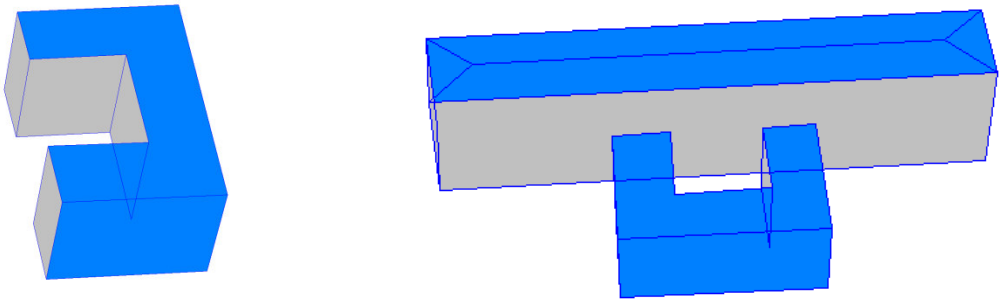


Figure 21: Two sub models: Building 1 (left) and two buildings (right).

For building 1, a subset of images from stripe #4 is taken and an extended evaluation of the adjusted ExtOri parameters is carried out. Coarse ExtOri parameters are changed and the efficiency of method is observed.

Table 5: Selected 95 IR images from stripe #4. On this subset of images an extended evaluation of the adjusted ExtOri parameters was made.

	Image number		
Stripe	from	to	Number of images
#4	13141	13235	95

4.2.1 Modification of the 3D building model

A 3D building model was produced in specialised photogrammetric software and the format of the output is not directly compatible with the import data formats of other available software. Therefore determination of hidden surfaces was not completely carried out.

An algorithm for back-face culling is applied. For each plane in the building model a normal vector is calculated. Then angles between the normal vectors of planes and the vector of acquisition device are determined. The planes with larger or equal angle than 90° to the normal vector of a photo plane are not considered (Figure 22 a, b).

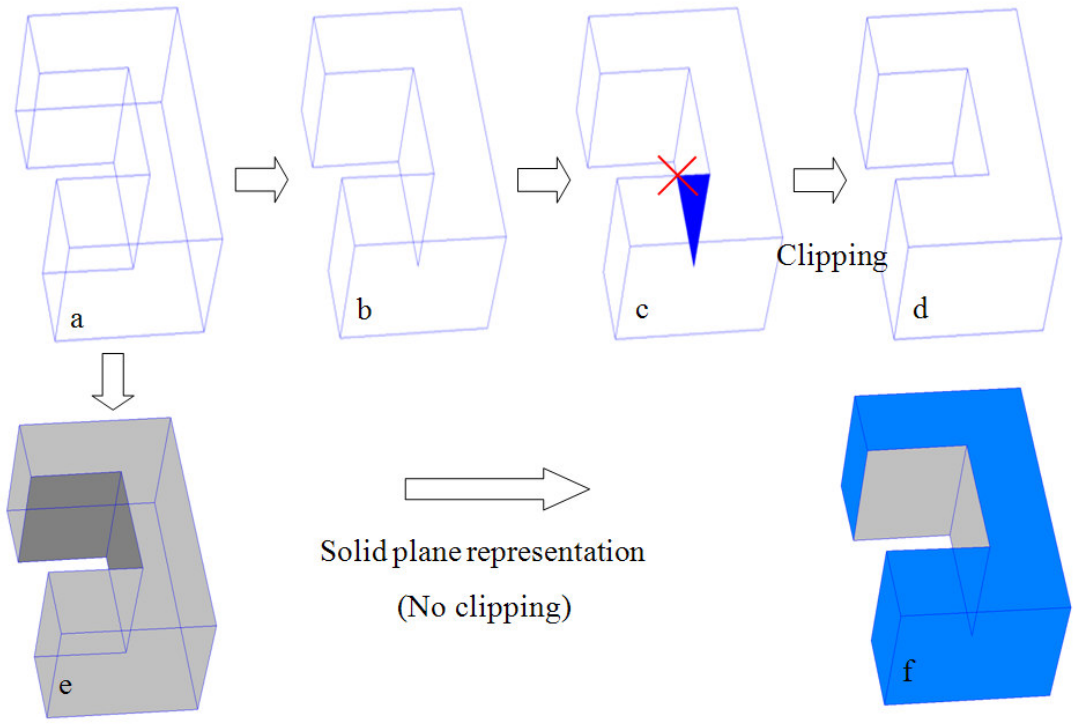


Figure 22: Projection of a building from a 3D space to a 2D (a). Invisible planes are removed by the algorithm described above (b). Self-occluding planes and part of planes of the model are not detected. A full algorithm should detect occluded planes (c, marked blue) and determinate and calculate the coordinates of the intersection points (c, red cross). The result of such an algorithm is presented in d. Visualisation of a building with solid planes appears to give the same result, but no clipping points are calculated. Oblique view of the acquisition device also causes apparent overlapping of buildings in images.

Ground points of the model were switched off and are not taken into further calculations. Roof edges and corners were determinate from the 3D coordinates of planes. A unit right circular cone, around z-axes, with vertex in origin of the coordinate system and aperture 30° is determined. If a normalized normal vector of the plane lies in space determinate with the cone the plane belongs to roof, else the plane belongs to the ground or wall of a building. In Figure 23 and Figure 24 the 3D building model in LandXplorer with solid plane representation and a wire-model representation is visualised, respectively.

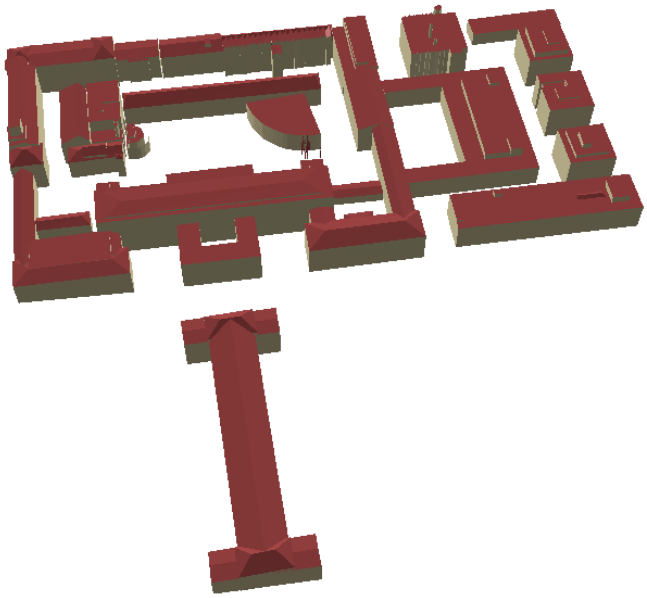


Figure 23: Visualisation of the 3D building model in LandXplorer CityGML Viewer. Roofs are in dark red and other buildings surfaces in brown.

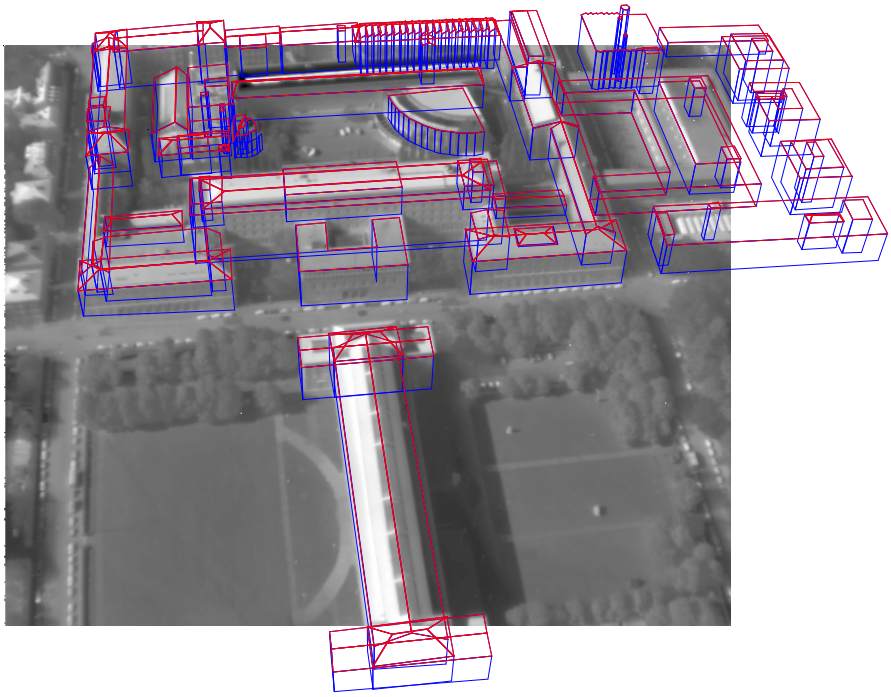


Figure 24: The 3D building model in wire-model representation (LOD2); roofs are in red and other building surfaces in blue. Model is projected into the image number 13142, stripe #4.

4.2.2 Feature extraction and matching algorithm

Feature extraction with the Förtsner point algorithm and straight edges extraction is described in section 3.1. Argumentation for the choice of the feature type – points and in the procedure for extraction of the intersection points of extracted straight edge is given in section 3.2.1. Extracted (sub) model is projected with the coarse ExtOri parameters into the image with extracted features. Furthermore, the correspondence between extracted points and model points is searched. On example of straight edges extraction and intersection points the feature extraction and matching algorithm are presented in this section; Building 1 is used as a sub model and is projected in the image number 13200, stripe #4. In the Appendix B extraction and matching parameters for Förstner points and straight edges which were used in this test can be found.

Straight edge extraction and classification

Extracted features – lines are represented in an image together with the projected 3D sub model. A threshold value for a minimum length of the extracted lines is applied (minLength), described in section 3.1.5. The extracted line segments are grouped into three groups on the basis of length in pixels, two threshold values are set. Lines with a length below the threshold 1 are represented in red, lines with a length between values threshold 1 and threshold 2 are represented in yellow and the longest lines are coloured green and have a length equal or larger than threshold 2. Lines with a length equal or larger than threshold 1 value are taken into further computation. Threshold values for length of lines should be chosen regarding to the relative size of building edges in an image. To each extracted line, a quality parameter is assigned

$$\textit{Quality of line} = \begin{Bmatrix} Good: length(red\ lines) < threshold\ 1 \\ Middle: threshold\ 1 \leq lenght(yellow\ lines) < threshold\ 2 \\ Bad: length(green\ lines) \geq threshold\ 2 \end{Bmatrix}. \qquad \textit{(Eq. 19)}$$

Longer lines, in these case green and yellow lines, are assumed to be more reliable. A larger weight value can be set for green lines in LS adjustment. Extracted straight line segments and classification is shown below in Figure 25.

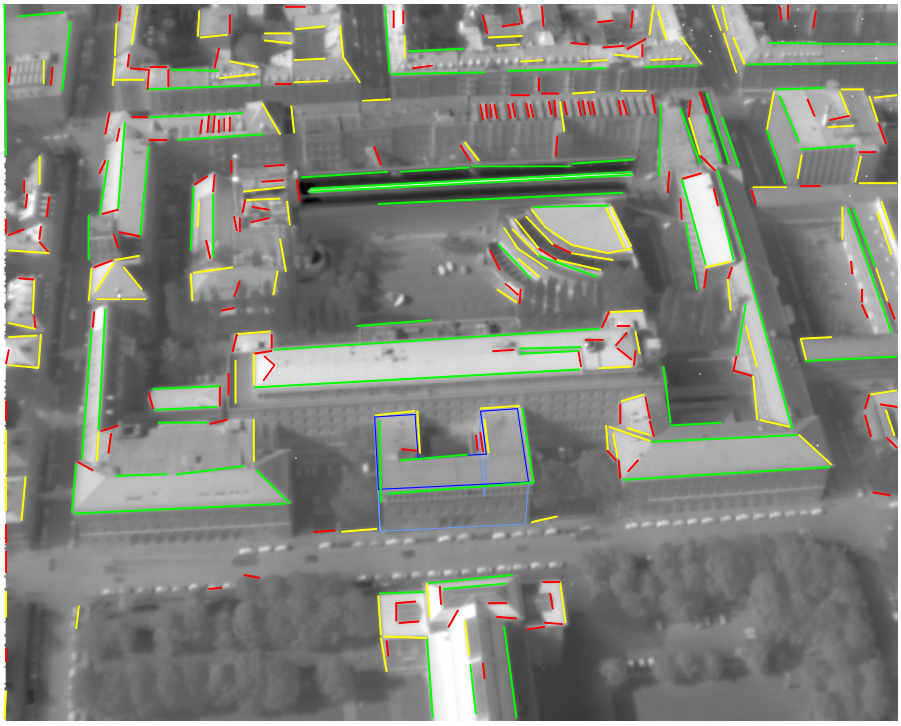


Figure 25: Projected sub model building 1 and extracted straight edge segments with traffic light colour coding. In red: short, not reliable lines (10-16 px); in yellow: middle length lines (16-32 px); in green: the longest, most reliable (>32 px); In dark blue: roof of sub model; In light blue: other surfaces of sub model.

Intersection points

Intersection points between lines with »good« and »middle« quality are calculated. Additional two parameters are set according to the description in section 3.2.1.2, these are:

d_{max} maximum distance limit between edge points in pixels and

α_{min} minimum intersection angle between two line segments.

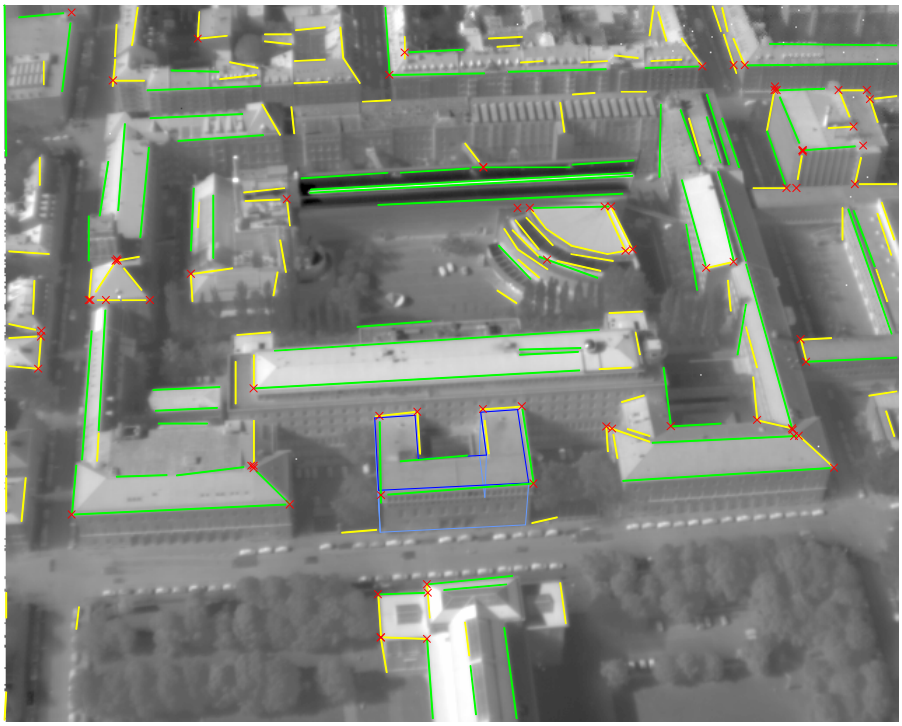


Figure 26: Extracted intersection points (red cross) calculated from extracted straight edges with a length larger or equal than 16 px (in green and yellow). Parameters for intersection points calculations: d_{max}=10, α_{min}=30°.

Matching of the building model and extracted points

Points are connected to the model (Figure 27, a). A circular search space R is defined for searching for correspondence between projected model lines and extracted points (section 3.2.3).

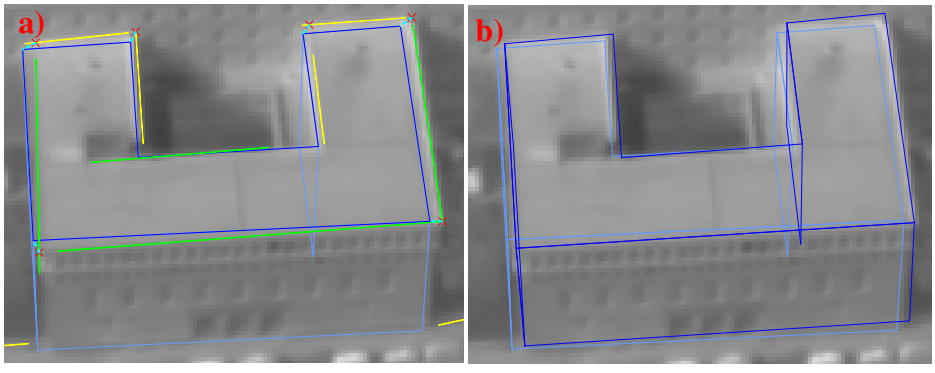


Figure 27: Matching of the building model and extracted points (left) and the re-projection of model (b) with refined ExtOri parameters. Left (a): Extracted lines (in green and yellow) and intersection points (red cross) connected with the projected model (blue); The connection is represented by cyan dashed line, R=5. Right (b): Projection of sub model with initial ExtOri

parameters (light blue); Projection of sub model with adjusted ExtOri parameters, after matching algorithm was applied (dark blue).

LS adjustment of ExtOri parameters and re-projection of the 3D building model

LS adjustment of the ExtOri parameters and a re-projection of the 3D building model are described in sections 3.2.4 and 3.2.5. The results of these two method steps are seen on an example in Figure 27, right.

Registration between the projected building model and the extracted Förstner points is equal to the one described above from »Matching of the building model and extracted points« on. There was no quality parameter set for the extracted Förstner points. However, the results of the Förstner point extraction are also covariance matrixes with the precision of the calculated position for each Förstner point (see section 3.1.4).

4.3 Results

The results are presented on the sample dataset of 12 images, three from each of the four stripes; first, two sub models are projected into this sample images, and secondly, the whole model. Additionally, more tests were made on the 95 subsequent images of the 4th stripe, projecting the sub model building 1 into them. In this section, the results of these extended tests are presented, including manual movement of the projected model before applying the matching algorithm.

4.3.1 The sub models: building 1 and two buildings

Sub models building 1 and two buildings presented in Figure 28are projected into the 12 selected images from four stripes. As initial ExtOri parameters the corrected values are used. Out of 12 sample images, on which the method was tested, the sub model building 1 (and two buildings) is seen in ten images (Figure 28 and Table 6).

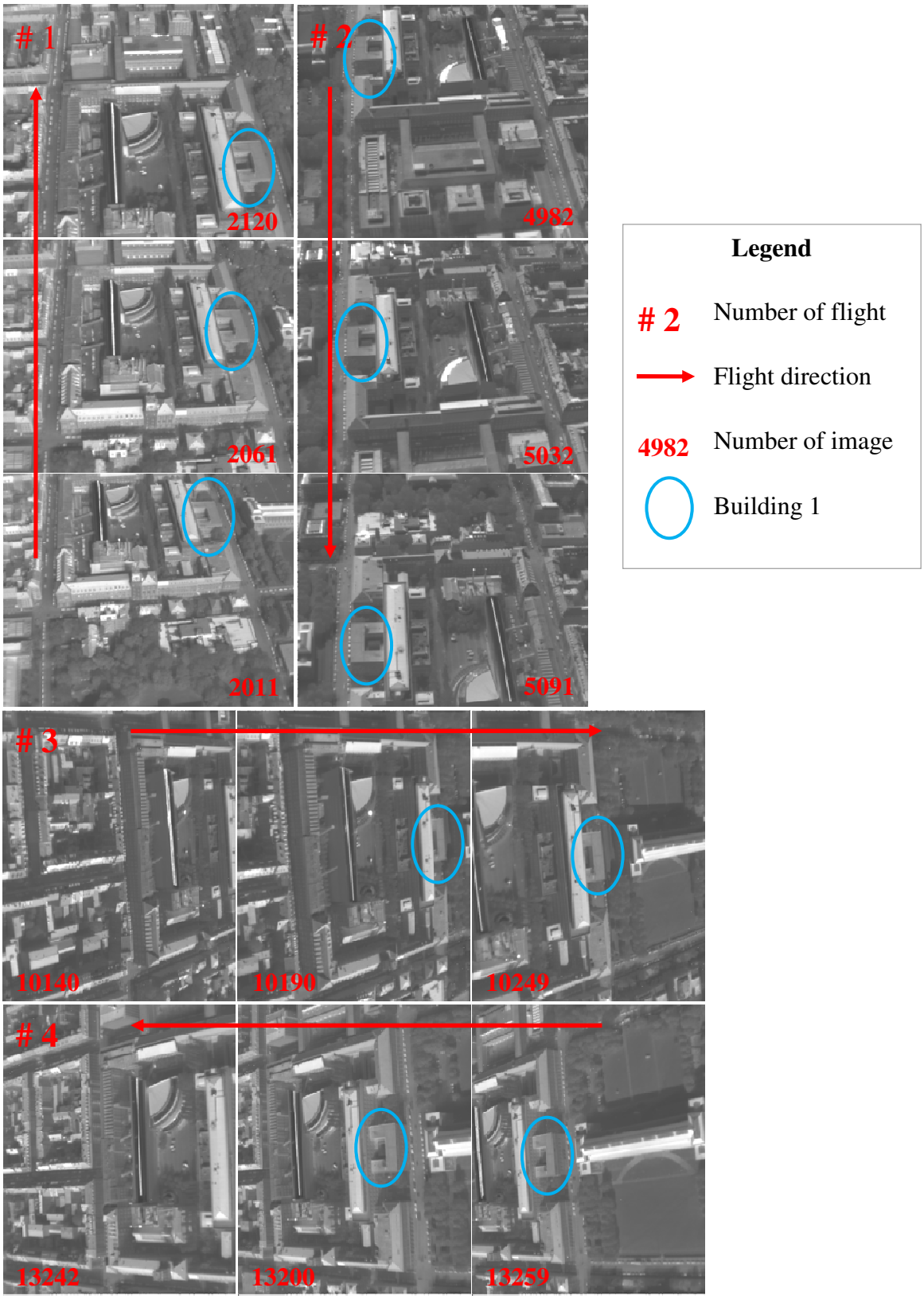


Figure 28: Sample images for testing the method with marked sub model building 1 (blue ellipse).

Table 6: Description of the position of the sub model building 1 on the sample images.

	Position of building 1 is sample images from all four stripes			
Stripe	**#1**	**#2**	**#3**	**#4**
Begin	Upper-right part of image.	Upper left corner	Not in scene.	Upper middle part
Middle	Middle–right part of image.	Left part of image in middle	Upper middle part of image, party occluded by other building	Middle of image.
End	Lower right corner	Lower left corner	Middle of image, but partly occluded by other building.	Not in scene.

Applying method on the sub model of building 1 gives poor result, using Förstner point extraction as well as intersection points (Table 7 and Table 8, respectively). Texts corresponding to images on which the method results in improving the ExtOri parameters are highlighted in grey. We consider that the re-adjusted ExtOri parameters are improved, if the projected building model is better aligned to the observed buildings.

Table 7: Results of applying the method on 12 sample images; A sub model building 1 and Förstner point extraction are used. Highlighted: Text corresponding to sample images where ExtOri parameters are improved.

Förstner point (building 1)				
Stripe	#1	#2	#3	#4
Begin	Two correspondences found.	Worse ExtOri parameters (smaller translation of projected model, shrunken).	Not in scene.	Corrected ExtOri.
Middle	Worse ExtOri parameters (moved model).	A bit shrunken model, relative OK.	Worse ExtOri parameters due to wrong matching.	Corrected ExtOri
End	Worse ExtOri parameters (smaller translation and rotation of model).	A bit shrunken model, relative OK.	Rotated, poor matching due to occlusion.	Not in scene.

Table 8: Results of applying the method on 12 sample images; A sub model building 1 and extracted Intersection points are used. Highlighted: Text corresponding to sample images where ExtOri parameters are improved.

Intersection points (building 1)				
Stripe	#1	#2	#3	#4
Begin	Corrected ExtOri parameters.	Worse ExtOri parameters (rotation, due to almost singular normal matrix in LS adjustment).	Not in scene.	A bit worse ExtOri parameters, due to small number of corresponding points.
Middle	OK No significant improvement or worsening of ExtOri parameters.	Smaller worsening of ExtOri parameters.	Not enough corresponding points found.	Corrected ExtOri parameters.
End		Rotation due to small number of corresponding points.	No significant improvement or worsening of ExtOri parameters.	Not in scene.

Testing the methodology on the same image set and sub model two buildings has shown some additional problems with wrong correspondence between model and extracted points. Matching of the sub model two buildings and the extracted Förstner points is unsuccessful, only two examples out of 12 sample images result in improved ExtOri parameters, i.e. the efficiency of method is 2/12. Furthermore, the efficiency of the method tested on the sample dataset with sub models can also be written with the denominator 10, e.g. 2/10; since out of 12 sample images on ten the sub model two buildings (and building 1) are in scene. Matching based on extracted intersection points is slightly better, but as well results in 3/10 better, 2/10 not significantly improved or worse and other 5/10 worse ExtOri parameter, of which one is evidently a mismatch.

The position of sub models in image significantly influences the possibility of a successful matching algorithm; the inspected building should not be occluded and should lie in the bottom middle to the middle of the image. Radial and tangential distortions have bigger influence on border of the image; the upper part of the image is more deformed and buildings appear relatively smaller to the ones on the bottom of the image due to oblique view

(geometry of the acquisition). Other evident problems when searching for a correspondence between extracted points and sub model of two buildings are model incorrectness and poor feature detection due to small edge response and other extraction problems.

Model incorrectness and inappropriate building modelling lead to some mismatches. Building 1 is simple and correctly modelled, whereas the quality of the sub model of two buildings is questionable. The larger building (of the two buildings sub model) has a flat roof with a massive fence around. There is another floor with smaller area compared to the main building on the top of the building. In the 2D building model this is modelled as building in a building. Closeness of edges of the building in a building causes mismatching. A different matching strategy would be more appropriate in this case, i.e. line to line matching. In stripe #3, the building 1 is occluded and because the intersection points are not calculated, this disables the correction of the ExtOri parameters.

In stripes #1 and #2, the building 1 has a low edge response, therefore less features are extracted on the edges of the building 1. This leads to a smaller number of corresponding points and less reliable adjustment. Good localization of extracted features is required, so smaller inaccuracies in position consequently have a large influence on re-adjustment of the ExtOri parameters.

4.3.2 The whole model

A method is applied on the 12 sample images into which the whole building model is projected. For projecting the model into the image, corrected ExtOri parameters are used; the extracted features are Förstner points and intersection points. The method is not applied iteratively.

Firstly, the number of extracted, corresponding and unique corresponding points in the sample dataset is inspected for extraction of Förstner and intersection points. The results are presented in the Table 9 and Table 10.

Table 9: Number of extracted, corresponding and unique corresponding points for extracted Förstner points on a sample dataset.

	Förstner point extraction R = 5 px		
Image number	Number of extracted points	Number of corresponding points	Number of unique corresponding points
2011	491	78	78
2061	559	102	101
2120	475	113	109
4982	307	92	91
5032	355	79	76
5091	399	59	56
10140	456	46	42
10190	305	49	47
10249	170	45	45
13142	232	68	66
13200	329	76	75
13259	403	51	51

Number of extracted points is a total number of extracted points which are possible candidates for a correspondence with the projected model. For the Förstner point extraction, this number is equal to number of the junction Förstner points and for the intersection points, this number is the number of intersection points between the extracted straight lines under condition of the angle limit and distance limit (see section 3.1.4).

Number of corresponding points is a number of extracted points which correspond to the edge points of the projected model found in the circular search space of the radius R. (For example: in Figure 27 points which have drawn cyan dashed line between model point and extracted point).

R is the radius of the circular search space for searching correspondence between the projected model into the image and the extracted features [px].

Number of unique corresponding points equals the number of model points which have at least one correspondent extracted point.

Table 10: Number of extracted, corresponding and unique corresponding points for extracted intersection points on a sample dataset. Highlighted text corresponds to the images on which the applied methodology is efficient.

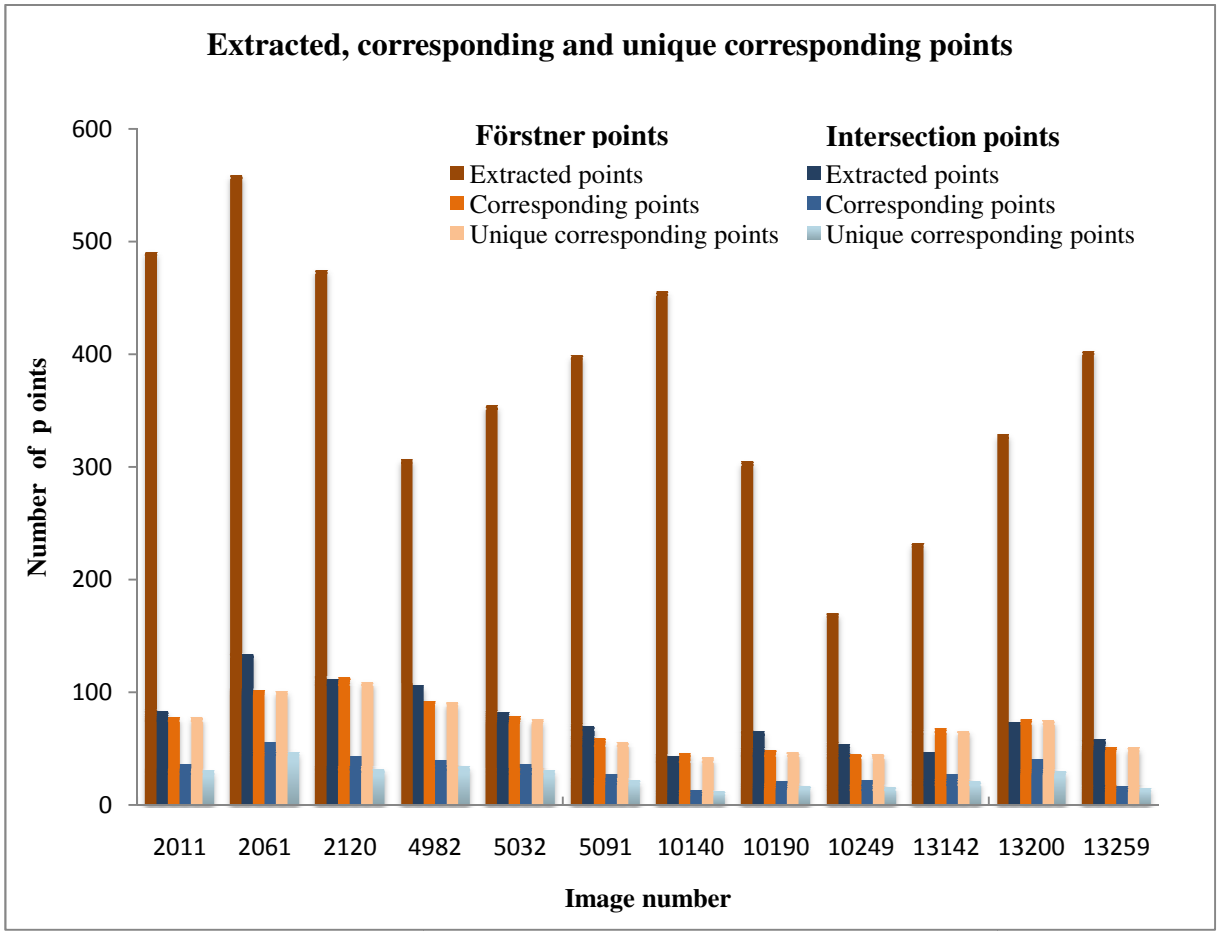
Intersection points Distance Limit = 10 px, Angle Limit = 30 °, threshold 1 = 16 px, threshold 2 = 32 px, R = 5 px					
Image number	Number of extracted lines		No of extracted points	Number of corresponding points	Number of unique corresponding points
	»Good+middle«	»Bad«			
2011	244	231	83	36	31
2061	314	238	134	56	47
2120	288	229	112	43	32
4982	257	150	106	40	34
5032	210	168	82	36	31
5091	227	182	70	27	22
10140	209	276	43	13	12
10190	217	166	65	21	17
10249	147	99	54	22	16
13142	162	118	47	27	21
13200	206	172	74	41	30
13259	274	236	58	17	15

Number of extracted lines is a total number of extracted lines from an IR image.

»Good + middle« is a number of extracted lines with a length higher or equal to the threshold 1 value in pixels.

»Bad« is a number of extracted lines with a length lower than the threshold 1 value in pixels.

The comparison of the numbers of extracted, corresponding and unique corresponding using Förstner and intersection point extraction is presented in Graph 1, that is comparing data from Table 9 and Table 10.



Graph 1: Comparison of number of extracted, corresponding and unique corresponding points for Förstner and intersection points extraction.

Förstner point extraction on the given dataset results in a relatively high number of extracted points, but they usually do not lie on the corners of the buildings. Furthermore, only few of the extracted Förstner points correspond to the model (Graph 1). Extraction of the intersection points gives small number of extracted points. However, in contrast to the Förstner point extraction, a relatively high number of extracted points correctly correspond to the model. Small roof structures represented by short line segments in the 3D building model often cause false matches, on the contrary, the matching based on the long edges gives better result. For this reason the minimum length of extracted straight edges, from which the intersection points are calculated, must be carefully chosen.

The number of unique corresponding points is with Förstner point extraction algorithm the same or just a bit smaller than the number of all corresponding points. On the other hand, this difference is bigger with extraction of straight edge segments. Non-unique corresponding points can be caused as a result of a non-single edge or corner response of extracted feature, closeness of two model points and extracted edge points as well as due to the projection – in some viewing directions, roof edges can appear very close to the ground points, or to the points of the nearby buildings. Model incorrectness causes some mismatches (see section 4.5).

Matching over the sequence with the described methodology applied on the whole building model is not efficient with the given data. Further investigation on a different 3D building model would be required for a better estimation of the quality of the method.

4.3.3 Subsequent images of stripe #4 (sub model)

Building 1 is projected into the images of stripe #4, 13141-13235. Images with lower numbers were acquired earlier. The image number 13141 is the beginning of the stripe #4 and the image number 13235 is the last image of this stripe where the whole sub model building 1 is in scene (Figure 29).

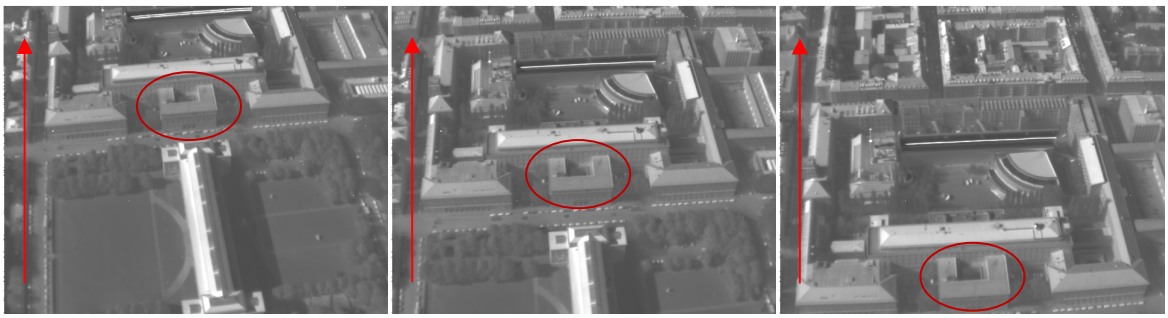
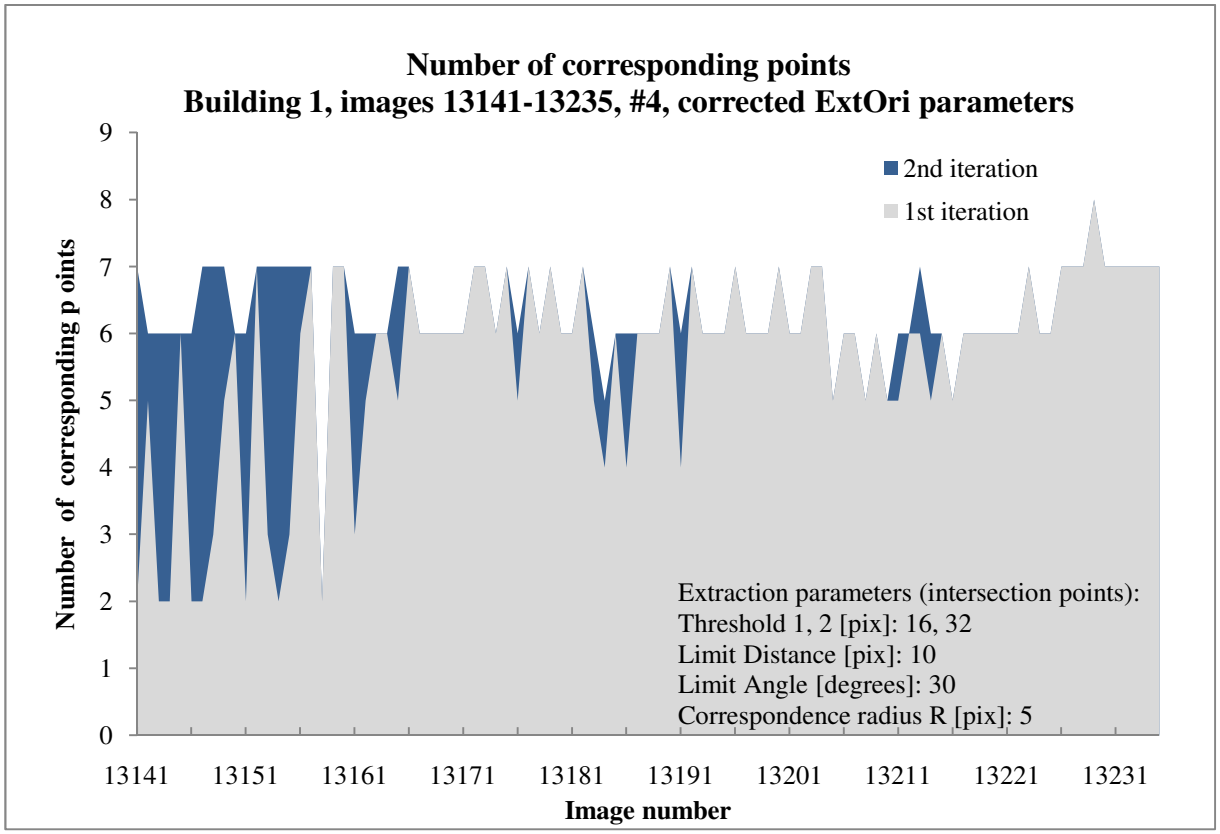


Figure 29: Position of sub model building 1 (red ellipse) used for testing the method. Left: Image number 13141; Middle: 13188; Right: 13235. Red arrow is flight direction.



Graph 2: Comparison of number of corresponding points in first and second iteration of matching algorithm. (Corrected ExtOri parameters are used, a method is applied on a sub model building 1 on images 13141-13235, flight #4; extracted features are intersection points).

Extraction of intersection points results in high completeness of the extraction and correspondence (section 4.6.1). The number of model points, i.e. roof points of the building 1, is 8. The number of correspondence points is shown in Graph 2. However, the Graph 2 represents the number of corresponding points, but does not exclude the possibility that two extracted points can correspond to one model point, that are so called not-unique corresponding points. This can be a consequence of a non-single response of an edge or detection of ground edges, which appears close to roof edges in a 2D projection.

This paragraph is an explanation to the Graph 2, where in some images from the beginning of stripe #4 only two extracted points correspond to the model, but the algorithm does not return an error. For exact solution and calculation of six ExtOri parameters, three extracted and

registered points are needed. Moreover, an overdetermined system is a consequence of more than three registered points, which leads to a LS adjustment. Used mathematical models are the co-linearity equations in which the ExtOri parameters (X0, Y0, Z0, Roll, Pitch, Yaw) are six unknowns. Each image point (extracted and registered point) has (x, y) coordinates in the picture coordinate system. Solving the system of linear equations with Matlab, a build-in operator, called matrix left division is used. This operator allows non unique solutions. Using this operator, a minimum of two correspondence points is required, however the output solution should be additionally evaluated.

Matrix left division »\«. A\B solves the symbolic linear equations A * X = B for X.

$$\begin{Bmatrix} if\ m = n\ \rightarrow System\ is\ square.\ Exact\ solution\ is\ found. \\ if\ m > n\ \rightarrow Overdetermined\ system.\ LMS\ solution\ is\ found \\ if\ m < n\ \rightarrow underdetermined\ system.\ A\ basic\ solution\ withat\ most\ m\ nonzero\ components\ is\ found. \end{Bmatrix}, \qquad (Eq.\ 20)$$

where:

A is coefficient matrix size m × n

(Solving Linear Systems of Equations, Matlab, 2007).

Results for manually moved sub model

In order to evaluate the efficiency of the method, initial, e.g. corrected values, of the ExtOri parameters are manually changed. Two examples are represented, first example results in successfully re-calculated the ExtOri parameters (Figure 30) and in the second example the problems and possible causes for unsuccessfulness of the algorithm are presented (Figure 31).

A sub model building 1 is projected into the image number 13170, stripe #4; in Figure 30 cut-out of the image is seen. Corrected coordinates are manually moved for 5 m in position (X, Y, Z direction), initial rotation angles are equal to the corrected ExtOri parameters. In all three iterations six points are extracted and registered with the model. However, in the 1st iteration, there is one mismatch, which does not appear in the further iterations (Figure 30, a, red arrow). In first column, extracted lines (yellow, green) and projection of the sub model building 1 with the initial ExtOri parameters in the IR image is seen. The initial parameters

for the 1st iterations are corrected ExtOri parameters, which are moved; in the 2nd iteration, the initial parameters are adjusted parameters from the 1st iteration, etc. In dark blue: roof of the sub model; In light blue: other surfaces of the sub model; Red cross: extracted intersection points. In second column, movement of the model is presented: in the ith iteration in dark blue and ith-1 iteration in light blue colour, where iteration i =1, 2, 3 and i-1 is projection of the sub model with moved ExtOri parameters. R = 15 px. The result is the corrected position of the acquisition device, that are corrected ExtOri parameters.

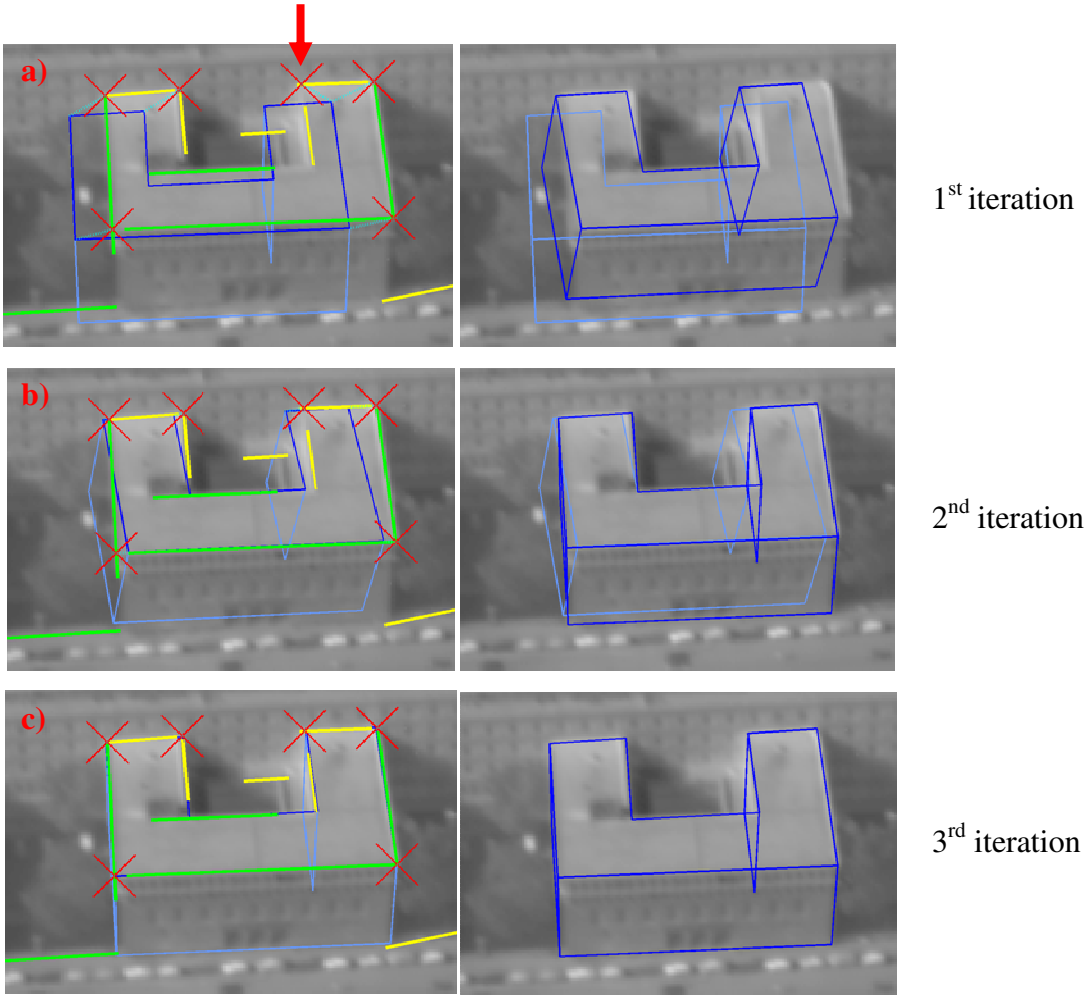


Figure 30: Example of successful stepwise extraction and matching algorithm.

Table 11: Comparison of ExtOri parameters, raw, corrected and manually moved (for 5 m in position) initial values, and after applying the matching algorithm iteratively.

Data Type/ iteration	Image number	X0 [m]	Y0 [m]	Z0 [m]	Roll [degrees]	Pitch [degrees]	Yaw [degrees]
RAW	13170	4468328.9	5334433.9	928.6	-2.439	34.160	297.268
Corrected	13170	4468322.1	5334430.6	920.9	-1.868	42.727	297.270
1st iteration	13170	4468292.5	5334439.8	937.9	-2.485	39.020	297.274
2nd iteration	13170	4468292.5	5334439.8	937.9	-2.485	39.020	297.274
Moved	13170	4468327.1	5334435.6	925.9	-1.869	42.727	297.270
1st iteration	13170	4468239.0	5334355.6	972.3	-12.668	35.981	297.290
2nd iteration	13170	4468292.5	5334439.8	937.9	-2.485	39.020	297.274
3rd iteration	13170	4468292.5	5334439.8	937.9	-2.485	39.020	297.274

For the example presented in Figure 30, two iterations are sufficient to achieve the same result applying LS adjustment for the ExtOri parameters when the corrected coordinates are moved for 5 m, (initial angles remains corrected values of ExtOri parameters) (Table 11). In the presented example, the iteration of the matching algorithm is efficient. However, the relative large residuals of the unknowns in the LS adjustment implies on the presence of systematic errors, possibly caused by i.e. the localisation inaccuracies of the extracted features, insufficient resolution of the IR image sequence or characteristics of the observed spectrum.

In Figure 31 an example of wrong matching is shown. The sub model building 1 is projected into the image number 13164; the coordinates are manually moved 5 m in position, angular values are the same as the corrected ExtOri parameters.

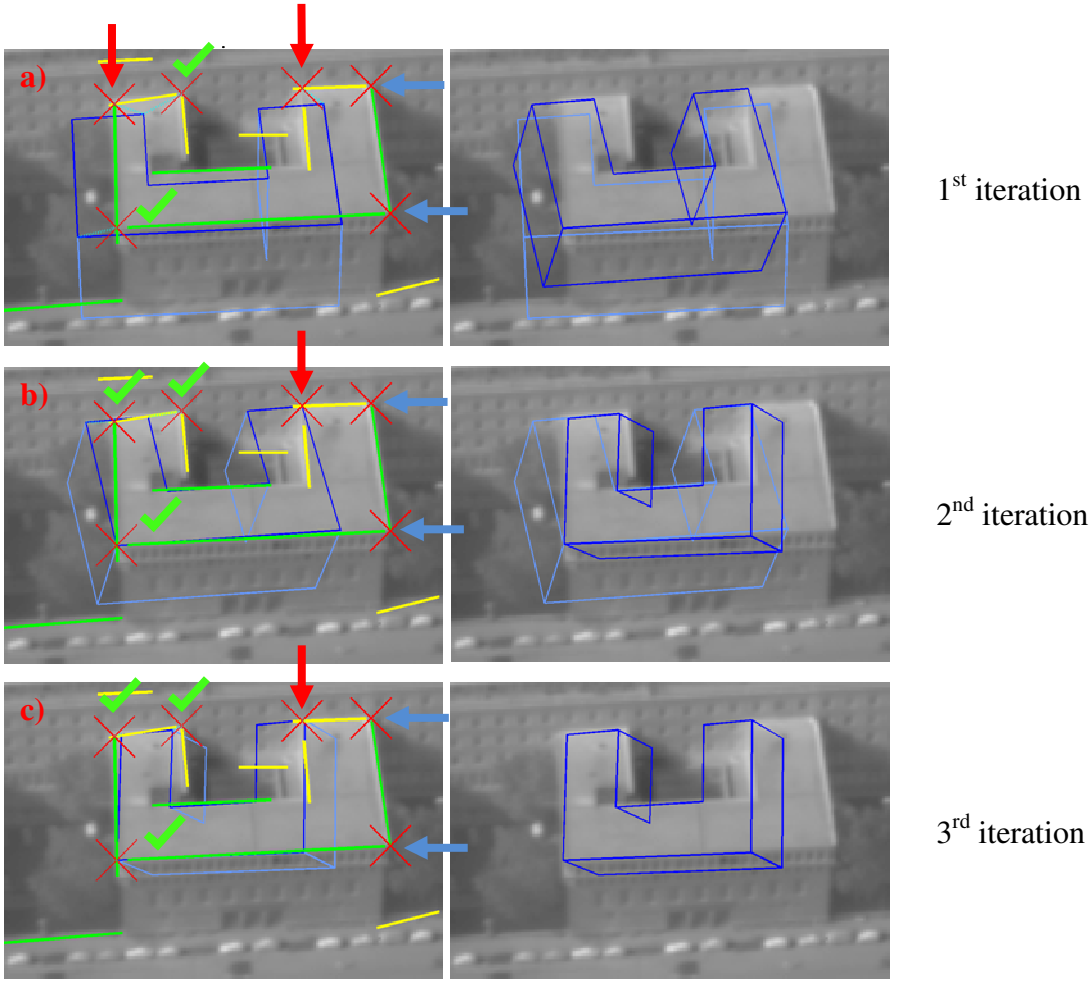


Figure 31: Example of an unsuccessful stepwise extraction and matching algorithm.

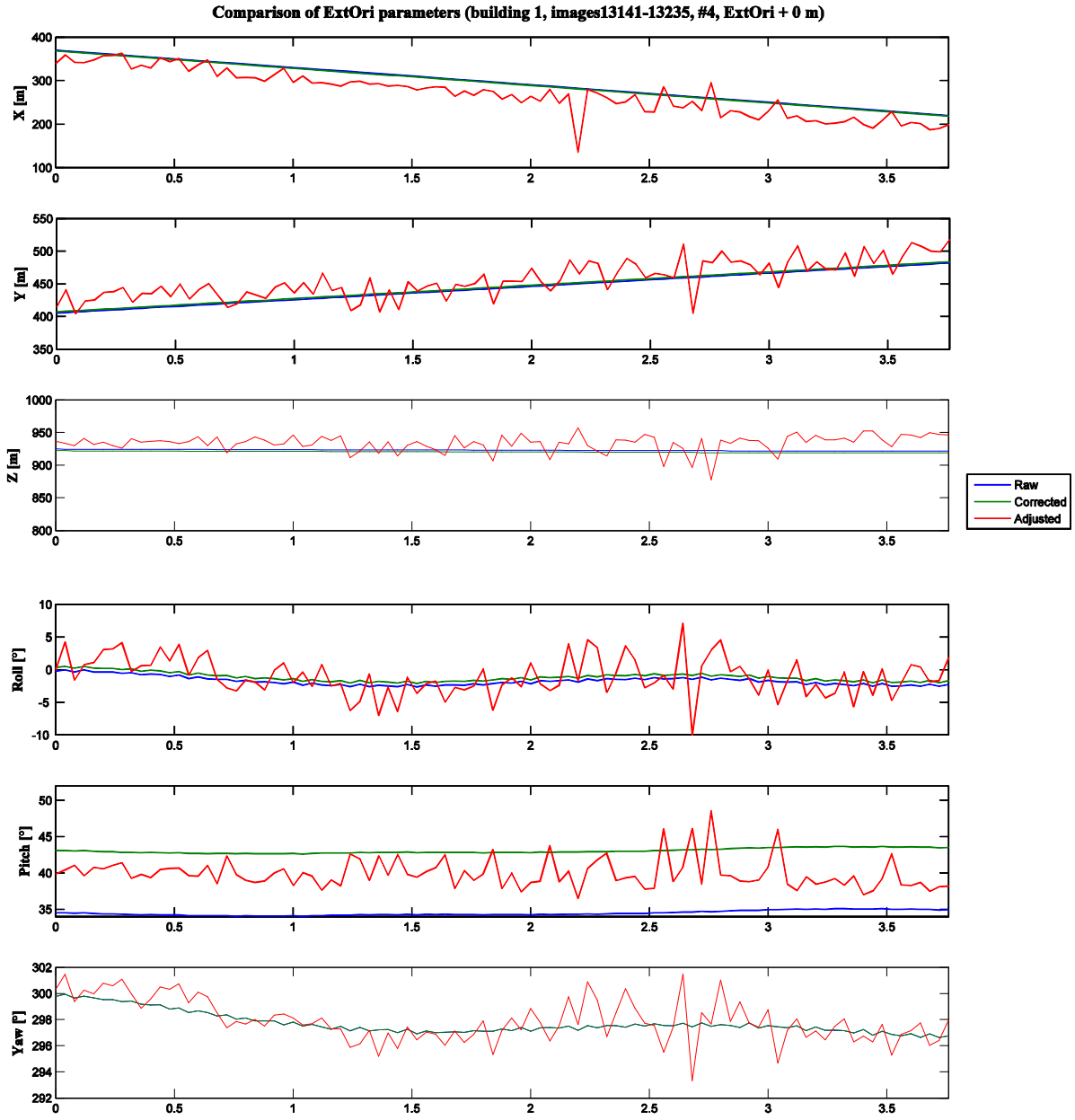
Figure 31: Red arrow points to the extracted points which were connected to the wrong building corner. Blue arrow points to the correctly extracted points, but due to the large misalignment they are not connected to any model points. Points marked with green are correctly extracted and matched to the model. Nevertheless, only four points are registered, of which one is mismatched. In this case the method results in worsening the coarse ExtOri parameters. This is a logic consequence of the LS adjustment, because the LS adjustment is efficient when all the measurements are free of gross errors; measurements should be corrected for all systematic errors. Denotations in Figure 31 are similar to the ones in Figure 30 (see above).

Efficiency as a quality parameter of the developed method is more detailed described in section 4.6 Evaluation.

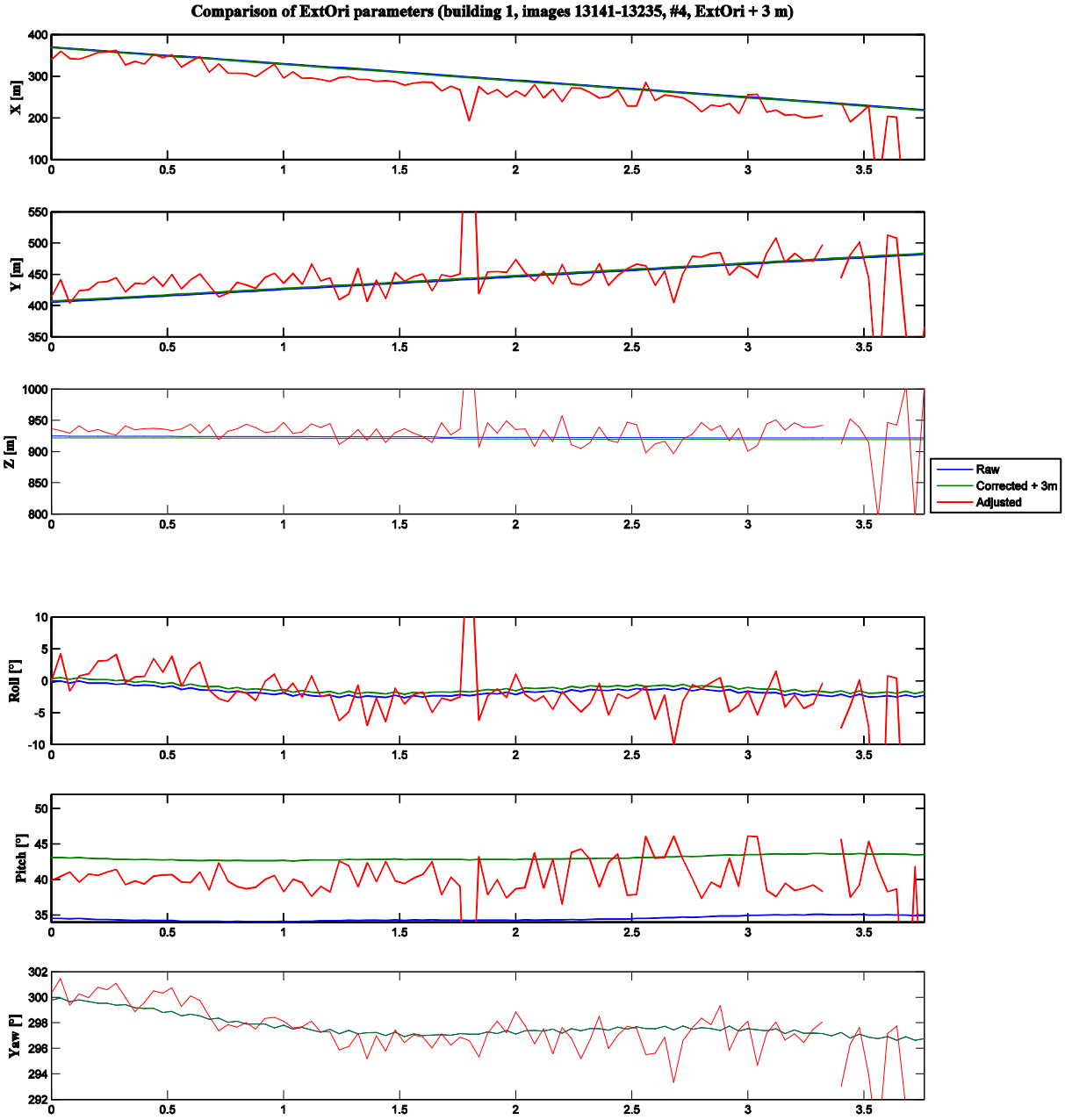
Comparison of the adjusted ExtOri parameters using manually moved and corrected ExtOri parameters as initial values

In Graph 3 and Graph 4 each ExtOri parameter – raw, corrected and adjusted (X0, Y0, Z0, roll, pitch, yaw) – is shown for the sequence of the images 13141-13235 from the stripe #4. Red line represents adjusted ExtOri parameters, cut line means no data, which is a consequence of not sufficient number of corresponding points or singularity of a normal matrix in LS adjustment. Raw and corrected data are presented with blue and green line, respectively. All the plots have as independent variable on horizontal axes normed time in seconds. Time is normalised to the acquisition of the frame 13141, stripe #4, time = 0.

ExtOri parameters were re-calculated according to the described methodology. Three iterations were performed – result of the 3rd iteration is presented in graphs - and for given accuracy and dataset 2 are sufficient. Iteration is not considered as an iteration of the LS adjustment, but as one iteration of the matching algorithm. In the 1st iteration initial projection of the building model is made using coarse ExtOri Parameters, in the 2nd iteration adjusted ExtOri parameters are used for projection of the model, etc. Iterations can improve the number of corresponding points, which is more obvious when using less accurate coarse ExtOri parameters for initial values (see also Table 11).



Graph 3: Comparison of raw, corrected and adjusted ExtOri parameters using corrected ExtOri parameters as initial values. (Sub model building 1, images 13141-13235, stripe #4).



Graph 4: Comparison of raw, corrected and adjusted ExtOri parameters using manually moved ExtOri parameters for 3 m as initial values. (Sub model building 1, images 13141-13235, stripe #4).

Comparing graphs, when for initial ExtOri parameters corrected ExtOri are used (Graph 3) and graphs when for initial value for ExtOri parameters 3 m in position is added (Graph 4), the algorithm is less efficient when the building is more on the edge. This can be observed in the graphs - the red line has higher amplitude, more outliers are present and the line is cut. For

example, Graph 4, around time 1.8 s in plots for X, Y, Z, roll and pitch over time it is clearly present an outlier. It can be assumed that the correction of ExtOri parameters for this frame is wrong.

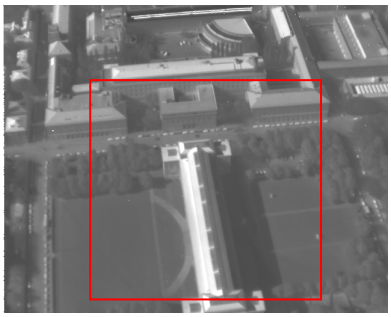


Figure 32: Area of the oblique image (red rectangle) where the inspected single building should lie that applied method gives good results.

It can be concluded, that a correction of ExtOri parameters by the described method is relatively efficient, if the inspected building lies in the centre and centre bottom area of the image (Figure 32). Preliminary condition is that the inspected building is not occluded by other objects. The same conclusion was made while testing the sample 12 images, three from each of the four stripes. If the building lies in the lower centre part of the image, it is relatively bigger compared to other object in the upper part of image, due to oblique view of the camera. Therefore, better extraction is possible in this image area and consequently more efficient matching. Apart from oblique view of the acquisition device, other reason, why the procedure is not so efficient on the borders of the image, could be the influence of not sufficiently calibrated and/or corrected IntOri parameters, i.e. lens distortions.

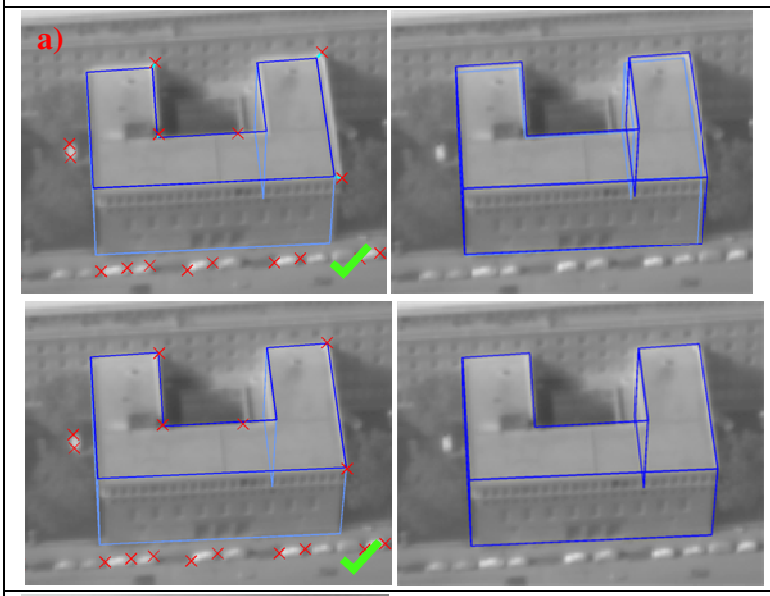
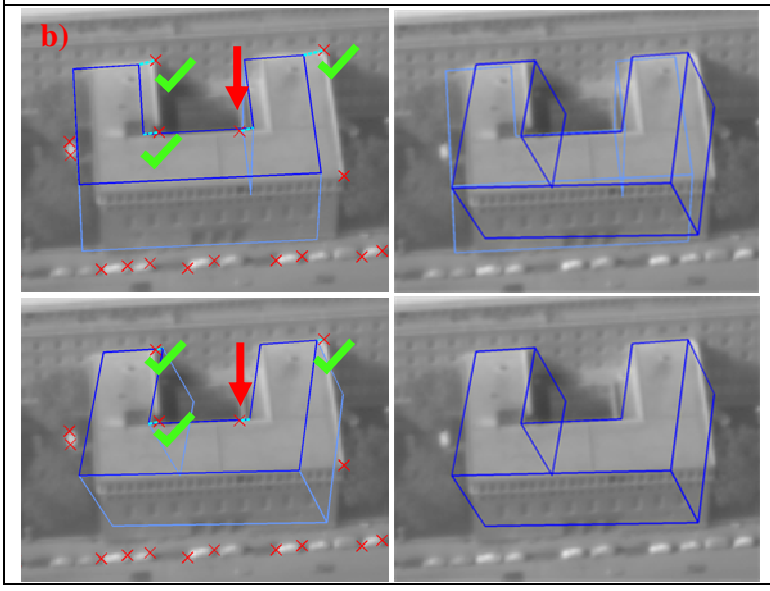
4.4 Comparison between extraction of Förstner points and intersection points

Two used point feature extraction algorithms are compared and discussed in this section. It is demonstrated, that the intersection points are in this example more robust compared to the extracted Förstner points.

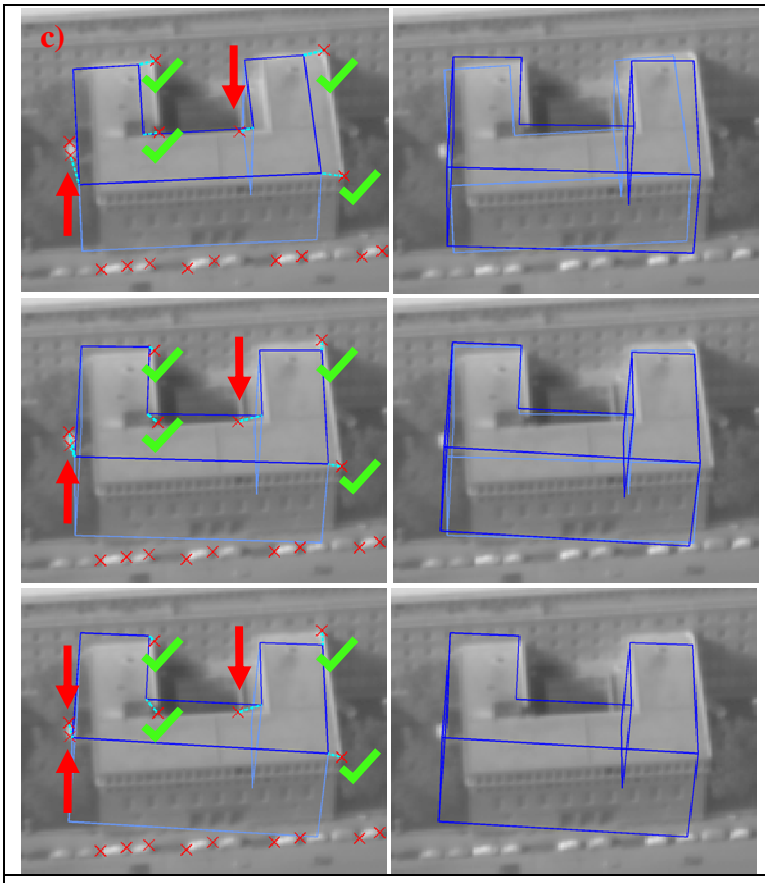
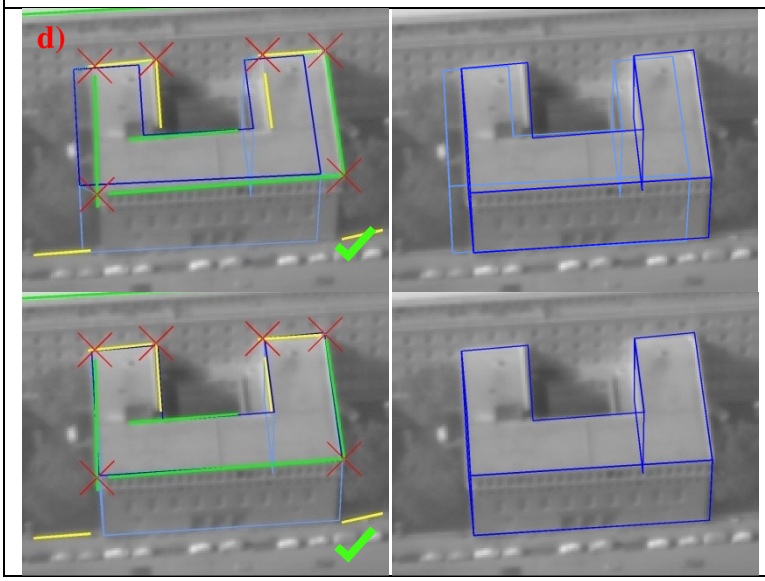
The Förstner point extraction results in a higher number of extracted points in comparison to the points of extracted straight edge segments (intersection points), but most of them cannot be connected to the model. If the initial projection of a 3D building model is very good, the matching and the LS adjustment Förstner point algorithm usually ensues in equal adjusted ExtOri parameters as the straight edge detection. However, in some cases due to the small, highly distinguishing object on the roofs or nearby building corners in the projection, false matching is more common when using Förstner points than extracted features. For instance, in Table 12, c, extracted Förstner points marked with red array all belong to the ground points, but appear on a near roof edge due to the projection in 2D. False corresponding points significantly influence the LS adjustment and the adjusted ExtOri parameters are worse than initial coarse ExtOri parameters (Table 12, b, c).

In the Table 12, a correct matching based on extracted Förstner points, which leads to corrected ExtOri parameters, is presented. When the model is manually moved (Table 12, b, c), i.e. when less exact ExtOri parameters are used, a small number of corresponding points and wrong matching results in not corrected ExtOri parameters. Table 12, a, b and d show only 2 iterations, more were calculated, but the result is the same within the accuracy of the measurements. Using extracted intersection points for matching results in the same adjusted ExtOri parameters – for manually moved (Table 12, d) or corrected initial values of the ExtOri parameters. Therefore, in the Table 12, d only the example with manually moved model, on which the matching algorithm is applied, is shown. What is more, the radius R of the circle, in which correspondence between extracted points and model edges is searched, has greater influence with the Förstner point extraction. An enlargement of the search space, i.e. a bigger radius R results in more corresponding points as well as more not correctly corresponding points (comparing Table 12, b and c). The algorithm using intersection points results in the same adjusted ExtOri parameters, if the $R \geq 10$ px. Low limit is crucial and needs to be adopted according to the expected accuracy of the coarse ExtOri parameters.

Table 12: Comparison between Förstner point extraction and intersection points. Left column: extracted point (a-d) is presented with a red cross; Green and yellow lines are extracted straight edges (d); Connection between extracted point and model point is in cyan dashed line (a-d); Green tick is correctly extracted and registered point (b, c) if in bottom right corner than all corresponding points are correct (a, d); Red arrow points to the wrong registered points (b, c). Middle column: In light blue: Projection of the sub model with adjusted (or manually moved) ExtOri parameters; In dark blue: sub model after the matching algorithm was applied. Right column: Parameters of the feature extraction and matching algorithm and explanation of images in middle and left column.

Förstner point extraction	
	Corrected ExtOri parameters X, Y, Z, roll, pitch, yaw. R = 10 px Extracted points: 16. Correctly corresponding points: 4.
	Moved ExtOri parameters X+3 m, Y+1.5 m, H+2 m, roll, pitch yaw are same as in corrected parameters. R = 10 px Extracted points: 16. Correctly corresponding points: 3. 1 wrong detected point and connected to the model.

Avbelj, J. 2010. Co-registration of three-dimensional building models with image features from infrared video sequences. Graduation Thesis – University studies. Ljubljana, UL FGG, Dep. of Geodetic Engineering, Geodesy.

	Moved ExtOri parameters X+3 m, Y+1.5 m, H+2 m, roll, pitch yaw are same as in corrected parameters. R = 15 px Number of extracted points: 16 1st and 2nd iteration: 4correctly corresponding points, 2 wrong detected points and connected to the model. 3rd iteration: 4 correctly corresponding points, 3 wrong detected points and connected to the model.
Intersection points	
	Moved ExtOri parameters X+3 m, Y+1.5 m, H+2 m, roll, pitch yaw are same as in corrected parameters. R = 15 px Extracted points: 6. Correctly corresponding points: 6.

4.5 The 3D building model problems

The automation of a whole process from acquisition to the textured 3D building model is a complex task. In this thesis, only a part of it was researched. The 3D building model is an input data, however it is not faultless. The inaccuracies, faults and method of modelling influence the efficiency of the developed method. This is why this section presents the problems of the used 3D building model.

The model was inspected in a wire model representation and in LandXplorer with solid surface representations from several viewing angles. However, some of the deficiencies still cannot be seen in both representations. The 3D building model was produced with the semi-automatic method in program INJECT. Due to export problems there are several severe deficiencies, of which some are marked in the Figure 33 and shown in detail on images in Figures 34-38.

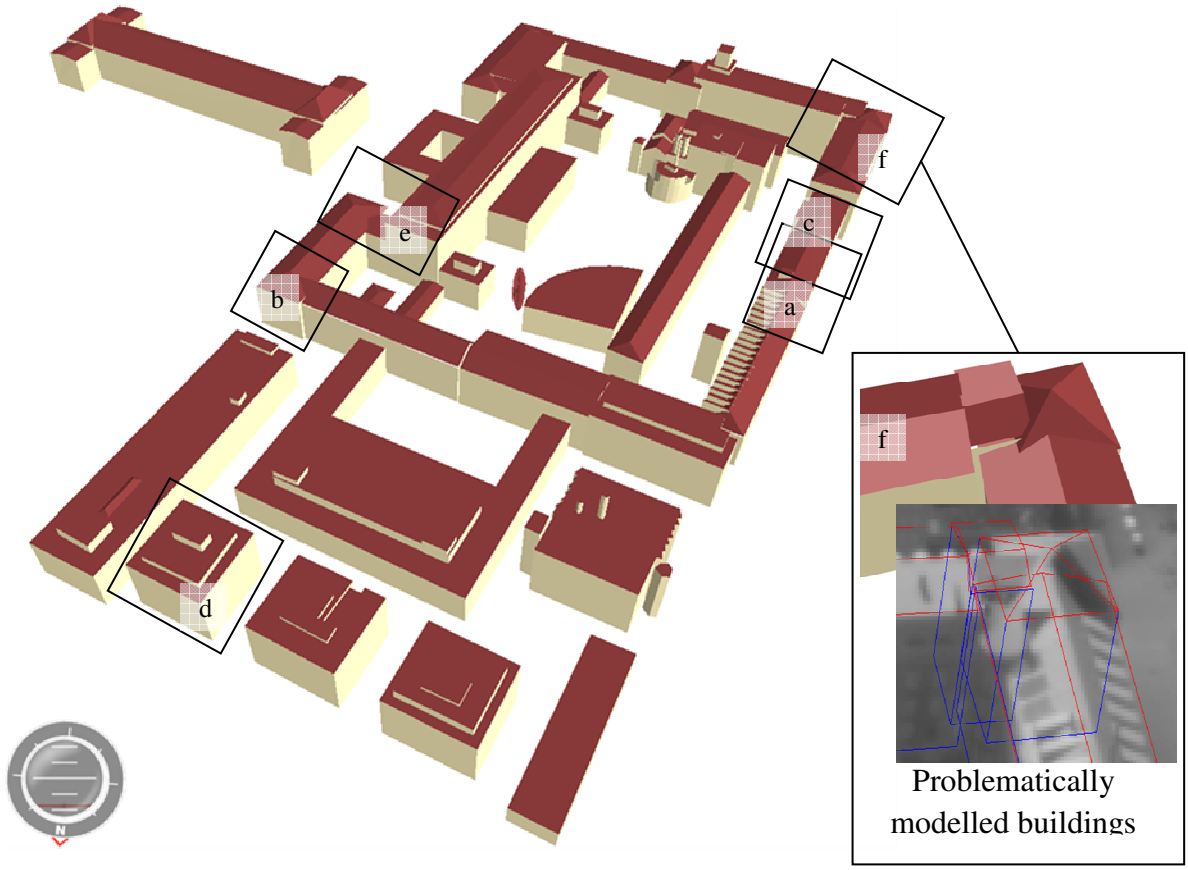


Figure 33: Problem areas of the 3D building model of the TUM visualised in LandXplorer

CityGML Viewer. The examples of problematic areas are detailed shown and explained in Figures 34-38.

In reality, many buildings are touching, but in a model the modelled buildings often appear to have gaps or are overlapping. As a consequence there are more model points, which do not exist, but can be connected to the extracted points. If these points are near the existing building edges, false matching cannot be avoided. There are several cases of gaps between roof and the belonging building. The buildings should be clipped together and connected with their roofs. Perpendicularity and parallelism of buildings is not maintained. Some buildings in the model are not on the correct height and/or are not placed on the ground. An additional semantics and attributes of surfaces in model could improve matching algorithm. For instance, small roof structures could be excluded from matching algorithm, etc.

The detail visualisation and explanation of problematic areas of the 3D building model of the TUM presented in Figures 34-38. The marked areas of Figure 33 are shown in larger scale.

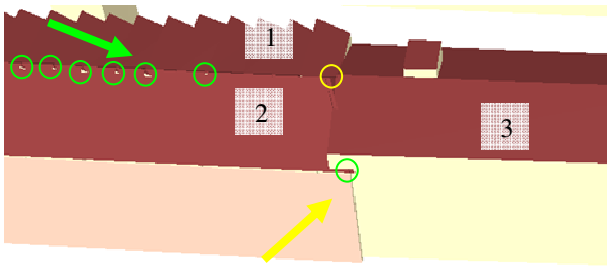


Figure 34: Overlapping of buildings and roofs (Figure 33, a).

Additional edge points that do not exist (encircled) are shown in Figure 34. Green arrow points to an overlapping of building 1 and 2. Building 1 is modelled as many small building segments, which are not touching in the model and are overlapping with the building 2. Consequently there are additional roof points (green circles) very near the ridge. Buildings 2 and 3 are overlapping and there is another additional point in the model. With yellow circle the area where the model has three different points with different coordinates, but should only be one point or three points with same coordinates, is marked.

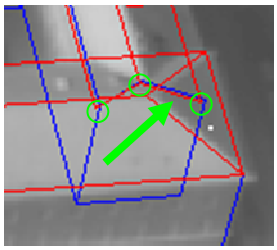


Figure 35: Lengthen buildings (Figure 33, b).

In the model are presented lengthened buildings (Figure 35), i.e. a roof of a building does not end at the wall or roof of a touching building. There are additional points (green circles) as a consequence of poor modelling.

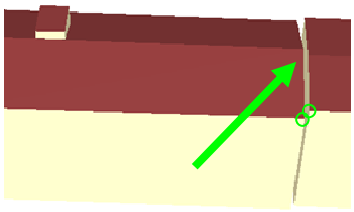


Figure 36: Gaps between buildings (Figure 33, c).

An opposite problem of overlapping of the buildings are the gaps between them (Figure 36). This influences the matching algorithm in the same way as overlapping does. Due to automatic modelling roof ridges are in some cases not oriented in the same direction.

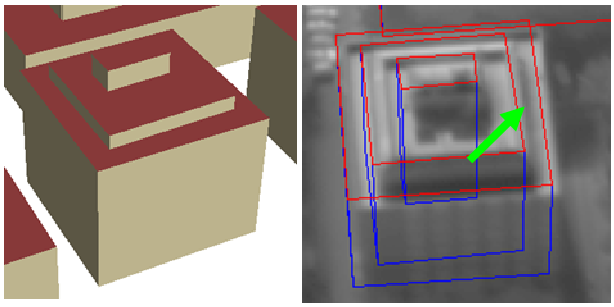


Figure 37: Building in building (Figure 33, d).

Roof structures, e.g. chimneys, fence on flat roofs, are not modelled as a structure attached to the roof but as separate buildings that extend to the ground level (Figure 37). In most of such cases it does not influence the matching algorithm. Thus, closeness of the roof structures and roof edges can result in false matching and in inaccuracies of the corrected ExtOri parameters,

that are not easy detectable. A problem of building-in-building modelling significantly aggravates automatic texturing.

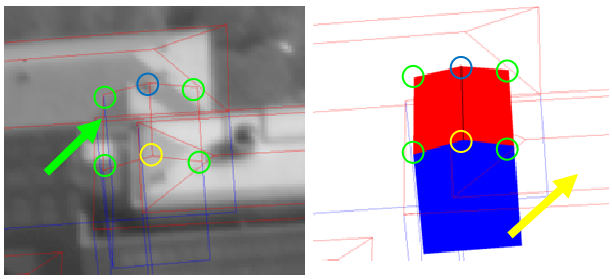


Figure 38: Typical multi-problem area in the building model (Figure 33, e).

In the Figure 38 overlapping of buildings (green arrow), building within a building (yellow arrow), and non-parallel buildings (green arrow) can be seen. Coloured building is wrongly modelled – all four green encircled roof edges cannot be seen from any viewing direction, because they are overlapped by two other buildings, this is wrongly modelled. Blue encircled roof edge can be detected from this viewing direction, whereas yellow encircled roof point can be detected from some other viewing directions.

4.6 Evaluation

On a dataset of the sub model building 1, stripe #4, images 13141-13235 (95 images) an evaluation of the method was made. First, the efficiency of the algorithm, then the completeness and correctness of the extraction and matching algorithm were observed. The evaluations of the method and the quality parameters were checked manually.

4.6.1 Efficiency of the method

The method is efficient, if the model re-projected with the adjusted ExtOri parameters is in a better position than with not adjusted ExtOri parameters. The visual check was made. Two examples of the algorithm efficiency are shown in the section 4.3.3 (Figure 30 and Figure 31).

$$Efficiency\ of\ algorithm = \frac{Number\ of\ correctly\ moved\ and\ rotated\ model}{Number\ of\ images}, \qquad (Eq.\ 21)$$

where:

Number of images is 95 for the given sample dataset.

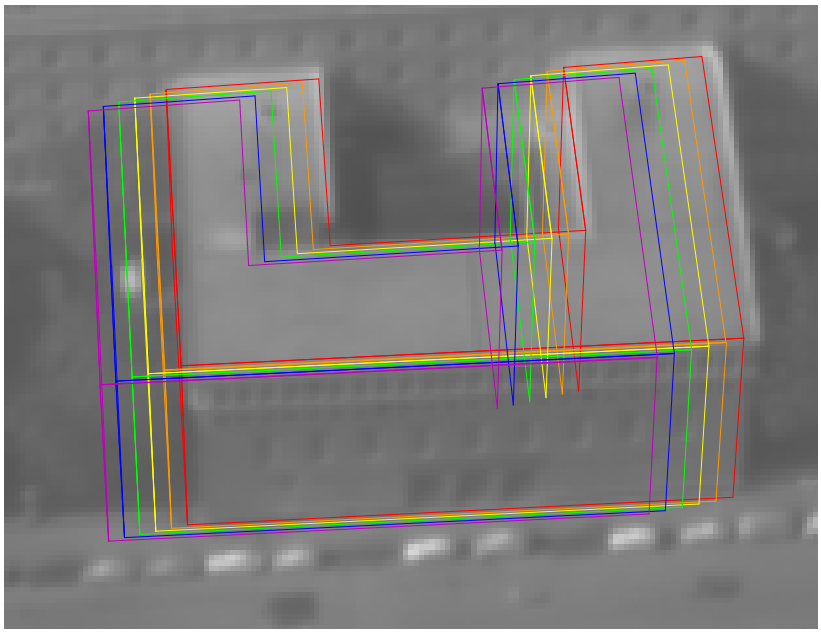


Figure 39: Projected building 1 into the image number 13200, stripe #4 with the corrected and manually moved ExtOri parameters. In red: the corrected ExtOri parameters are use for projection; In orange, yellow, green, blue and violet, the corrected ExtOri parameters are moved in position for 1-5 m, respectively.

Firstly, the initial, i.e. the corrected coordinates were manually moved for 1 to 5 m with 1 m increment in position in X, Y and Z direction (Figure 39, Eq. 22-24). Then the roll, pitch yaw initial values were moved for 10' to 50' with increment of 10' (Eq. 25-27).

$$X = X_{initial} + dX, where\ dX = 1, 2, \dots, 5;\ dX[m], \qquad (Eq.\ 22)$$

$$Y = Y_{initial} + dY, where\ dY = 1, 2, \dots, 5;\ dY[m], \qquad (Eq.\ 23)$$

$$Z = Z_{initial} + dZ, where\ dZ = 1, 2, \dots, 5;\ dZ[m], \qquad (Eq.\ 24)$$

$$roll = roll_{initial} + droll, where\ droll = 10, 20, \dots, 50;\ droll['], \qquad (Eq.\ 25)$$

$$pitch = pitch_{initial} + dpitch, where\ dpitch = 10, 20, \dots, 50;\ dpitch['], \qquad (Eq.\ 26)$$

$$yaw = yaw_{initial} + dyaw, where\ dyaw = 10, 20, \dots, 50;\ dyaw[']. \qquad (Eq.\ 27)$$

Table 13: Average efficiency of the method, using manually moved ExtOri parameters.

Shift	**dX=dY=dZ [m]**				
	+1	+2	+3	+4	+5
Efficiency	1.00	1.00	0.92	0.65	0.16
Shift	**droll=dpitch=dyaw[‘]**				
	+10	+20	+30	+40	+50
Efficiency	1.00	1.00	0.97	0.93	0.61

For instance, droll = dpitch = dyaw = 30’ and dX = dY = dZ = 3 m results in 1.0 efficiency. Efficiency does not depend only on the accuracy of the ExtOri parameters, as well as on the quality of extraction, and relative position of building in an image.

The matching algorithm is very sensitive to angular changes. The coarse ExtOri parameters should be known with high accuracy, better than 4 m in position and 0.5º in orientation. With DGPS measurements better than 4 m accuracy of position is expected. However, the problem remains with roll, pitch and yaw angles.

Secondly, the efficiency of the algorithm was checked on corrected coordinates as the initial values for the ExtOri parameters. The corrected ExtOri parameters are very close to the real value; therefore, the efficiency of the algorithm must be checked more closely. The correctness of the point extraction is 74%, but the position accuracy, i.e. the localisation accuracy of extracted feature is a more complex value to estimate. The precision of the used features is pixel precise. Due to oblique view of the camera, the pixel in upper part of the image does not cover the same area as the pixel in the lower part of same image. The problem of the position accuracy of the extracted features remains although the correctness of matching is over 90%. An adjustment of the ExtOri parameters cannot result in higher accuracy of the ExtOri parameters as the accuracy of extracted features is.

Position inaccuracy in the feature detection results in a smaller rotation and/or translation of the projected sub model with the adjusted ExtOri parameters in comparison to the projection

with initial corrected ExtOri parameters. Therefore, the efficiency of the algorithm using corrected ExtOri parameters was checked using three values:

- »better«, algorithm is efficient,
- »minor change«, re-projected sub model with adjusted ExtOri parameters appears to be a bit rotated or moved according to the initial position of the projected sub model and image data and
- »worse«, sub model projected with the initial ExtOri parameters is in a better position as with the adjusted ExtOri parameters, larger rotation and/or translation is present according to the initial projection of the sub model and image data.

Table 14: Average efficiency of the method, using corrected ExtOri parameters.

Shift	**Efficiency**	
Non corrected ExtOri parameters	Better	0.64
	Minor change	0.22
	Worse	0.14
	Sum:	1.00

Using very accurate coarse ExtOri parameters, the efficiency is 64%, which is lower than having less accurate initial ExtOri parameters. 22% of the adjusted ExtOri parameters have minor changes, and they could still be used for an automatic texturing of a 3D building with images acquired in the IR domain, whereas in 14% the algorithm is inefficient.

4.6.2 Completeness and correctness of extracted features

Evaluating the extraction is given with two quality parameters, completeness and correctness of extracted intersection points. The extraction of the intersection points was evaluated. However, the Förstner point extraction was not evaluated, since it has shown less promising results (see section 4.4).

Completeness and correctness of the extraction are calculated

$$completeness\ of\ extraction = \frac{Number\ of\ correctly\ extracted\ points}{Number\ of\ model\ points} \qquad (Eq.\ 28)$$

and

$$corectness\ of\ extraction = \frac{Number\ of\ correctly\ extracted\ points}{Number\ of\ all\ extracted\ points}, \qquad (Eq.\ 29)$$

*w*here:

Number of model points is eight when the sub model building 1 in inspected. Only the roof points are taken into calculations.

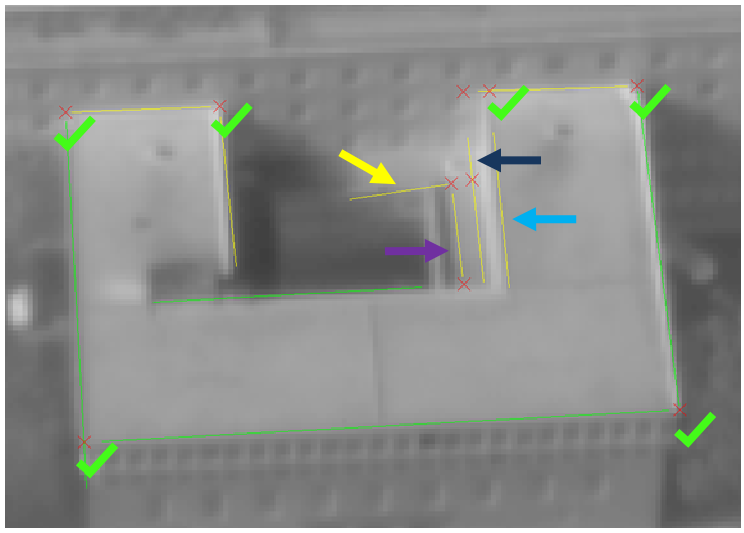


Figure 40: Correctness and completeness of intersection point (red cross) extraction presented on cut-out of image number 13229, stripe #4.

Number of correctly extracted points: 6 (green tick)
Number of all extracted points: 10 (red cross)
Number of wrong extracted points: 4
Completeness of extraction: 6/8 = 0.75
Correctness of extraction: 6/10 = 0.60
Green and yellow lines are extracted straight edges.

The reason for faulty extracted points in Figure 40 is a double response of an edge (dark and light blue arrow), detection of the ground line of the building (violet arrow) and the ground structure (yellow arrow). Double response of an edge should be avoided using the different parameters of the extraction, which is not always possible when processing a larger amount of images with same parameters of extraction.

The average completeness and correctness (Eq. 28 and 29) of the intersection point extraction for 95 images for parameters of the extraction in Appendix B are presented in the Table 15.

Table 15: Average completeness and correctness of extraction of intersection points. (Sample of 95 images, see Appendix B for extraction parameters).

Extraction	
Completeness	**Correctness**
0.88	0.74

4.6.3 Completeness and correctness of matching algorithm

The evaluation of the matching algorithm is made analogously to the evaluation of the feature extraction and on the same dataset (extracted intersection points with the extraction parameters shown in Appendix B).

The completeness and correctness of the matching algorithm are calculated

$$completeness\ of\ matching = \frac{Numberof\ correctly\ matched\ points}{Number\ of\ model\ points} \qquad (Eq.\ 30)$$

and

$$corectness\ of\ matching = \frac{Numberof\ correctly\ matched\ points}{Number\ of\ all\ maching\ points}. \qquad (Eq.\ 31)$$

The correctness and completeness were checked for the matching algorithm when the corrected ExtOri parameters were used as initial values and with the manually moved ExtOri parameters. The matching and adjustment was made for each image in three iterations.

Manually moved ExtOri parameters

$$dX = dY = dZ = 3\ m, \qquad (Eq.\ 32)$$

$$droll = dpitch = dyaw = 30'. \qquad (Eq.\ 33)$$

The corrected coordinates are very close to the real value, so one iteration is sufficient. However, it was calculated in three iterations to check the convergence of the algorithm.

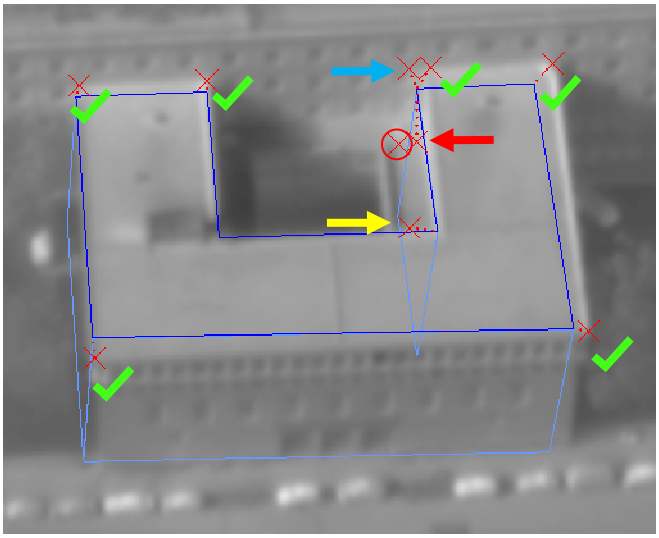


Figure 41: Correctness and completeness of matching algorithm presented on cut-out of image number 13229, stripe #4.

Number of all extracted points: 10 (red cross)

Number of correctly matched points: 6 (green tick)

Number of all matching points: 9 (red dashed lines between red cross and model points)

Number of model points is: 8

Number of points with no matching found: 1 (encircled with red)

Completeness of matching: 6/8 = 0.75

Correctness of matching: 6/9 = 0.67

In dark blue is the roof surface 1, in light blue are other surfaces of the sub model building 1.

For extracted points marked with arrows, wrong correspondence was found. The point marked with blue arrow is a result of double edge response, and therefore will not significantly influence the adjustment of the ExtOri parameters. That point is counted as an incorrect match, but is inaccurately detected, i.e. worse localisation. The point marked with red and yellow is a result of an incorrect extraction and found correspondence is severely influencing the adjustment. The encircled point is also a result of a wrong extraction, but it is correctly not matched to any model points.

The average completeness and correctness (Eq. 30 and 31) of the matching algorithm are presented in the Table 16; they are calculated for 95 images; for corrected and manually moved ExtOri parameters in three iterations.

Table 16: Average completeness and correctness of matching algorithm. (Sample of 95 images, three iterations, corrected and manually moved ExtOri parameters.)

Initial coordinates	**Matching**					
	Completeness			**Correctness**		
iteration	**1**	**2**	**3**	**1**	**2**	**3**
Corrected	0.75	0.75	0.75	0.90	0.90	0.90
Moved	0.73	0.74	0.74	0.87	0.87	0.90

The quality of the matching algorithm, presented with the correctness and completeness is high. No obvious change in correctness and completeness of the matching algorithm is seen in the iteration of a process when corrected coarse ExtOri parameters are used, because the corrected coarse ExtOri parameters are very close the correct value. However, using manually moved ExtOri parameters, iteration of the matching algorithm increases the values of the completeness and correctness of the matching algorithm (see also Figure 30).

4.7 Discussion

The tests of the developed method showed that the adjustment of the ExtOri parameters using the correspondence between the extracted Förstner »junction points« and the projected 3D building model does not improve them. This is a consequence of the Förstner point detector; many points extracted by this operator do not present roof edges, but some other smaller objects. Additionally, the number of extracted Förstner points is very high, but only few correspond to the model. On the contrarily extraction of intersection points gives more promising results when the inspected building is in the centre or lower part of an image. Although the number of extracted intersection points is much smaller compared to the extracted Förstner points, relatively high number of intersection points are co-registered with the projected model.

Due to the oblique view of the camera some ground objects appear to be close to the roof edges. This causes mismatching, especially in case of extracted Förstner points. A reason for that are abovementioned small objects detected by the Förstner operator; the procedure of intersection point extraction is less sensitive to detect small, spot-like objects. Oblique view of the acquisition also influences the appearance of the projected 3D building model. The model is not only occluded by other objects, but is also self-occluding. Apparent intersection points, caused by self-occluding of the model in projection were not calculated in the tests, which worsens the efficiency of the method.

What is more, the ground sample resolution in the lower part of the image is higher in comparison to the upper part of the image, what is the consequence of the geometry of the acquisition. The inspected building should lie in the bottom centre or centre part of an image. However, a building lying in the upper part of an image is relative small, and preferably, the textures should be acquired from another frame of a video sequence. If the discussing building is in the upper part of an image, the matching algorithm showed no improvement of the ExtOri parameters. The importance of the relative position of a sub model in the image can suggest that the distortions obtained in the system calibration are not accurate and the method is not robust against this inaccuracy. The lenses distortions of the camera can be clearly seen in the used IR images on all the image margins.

Used 3D building model has several inaccuracies and is in some cases unsuitable or even wrong modelled. For better estimation of the efficiency of the matching algorithm, the method should be applied on another building model.

The tests were made using projection of the whole model and sub models on the sample dataset. The projected model was manually moved before the matching algorithm was applied in order to evaluate the matching algorithm. Thus, the coarse ExtOri parameters should be known with high accuracy, better than 4 m in position and 0.5º in orientation. With DGPS measurements better than 4 m accuracy of position is expected (Grewal, 2007). However, the problem remains with measurements of roll, pitch and yaw angles.

The presented method is applicable for a refinement of the position and orientation obtained in the extended system calibration of the whole sequence under the condition that the inspected building does not lie on the very edges of the image. It can improve the texture mapping for a single building or building part caused by not modelled vibrations of the system addressed by Kolecki et al. (2010) or small inaccuracies in the model. However, on the given dataset the method does not improve the ExtOri significantly when applied on a whole model.

5 CONCLUSION AND FUTURE WORK

In this thesis the correction of the exterior orientation (ExtOri) parameters of the IR camera mounted on a mobile platform is studied: The main purpose of the research is to automatically extract IR textures for roofs and facades of the existing building model. The developed method bases on a point-to-point matching of the features extracted from IR images with a wire frame building model.

Firstly, different feature types were studied; Förstner points and intersection points are chosen as the suitable feature type for given task and compared. Secondly, extracted features and projected 3D building models are registered and the ExtOri parameters are re-calculated; matching of the features is applied iteratively. The 3D building model is modified before projection into a 2D image. The developed methodology was tested on the selected frames of the IR video sequence using the whole building model and two sub models.

Initial tests have shown promising results applying the methodology on IR images and the dataset of the sub model. The developed method for matching and extraction is evaluated on a sub model using five different quality parameters, i.e. efficiency of the method, completeness and correctness of extraction and matching. Efficiency of the method is estimated to over 90% for initial ExtOri parameters with accuracy of position 4 m and angular values 0.5° or better. The quality parameters for extracted intersection points are estimated to 88% and 75% for completeness and correctness of extraction of intersection points. The completeness of matching is estimated to over 73% and the correctness of the iterative matching is over 90%. However, the method is not applicable on the whole building model, so possible reasons for that are: localisation accuracy of the extracted features, resolution of the IR images, characteristics of the IR spectrum, lens distortions, accuracy of coarse ExtOri and influence of oblique view of image acquisition. The experiment also indicated some disadvantages of the input data, especially the 3D building model, and of the method. The procedures developed in

the frame of the thesis produce accurate results and are comparable to other studies. Several possibilities for future work and enhancement of the method are proposed in next paragraphs.

During the research many questions and ideas for different approaches or to enhance the developed method have appeared. Some of the ideas are addressed and presented in the experiment and also in this chapter. The discussed topic of the thesis is up-to-date with a vision to the fully-automated approach of texture mapping building models also with lower resolution images.

Extraction of intersection points from straight edge segments can be enhanced by using rectangular or ellipse search space for the line intersection instead of the circular search space. By changing the search space in this way, the direction and length of line are weighted. Furthermore, the intersection points extracted from long lines can be applied with the higher weight in weight matrix in the LS adjustment.

The matching algorithm can be divided in more steps, firstly using detected features in lower resolution images and secondly in iteration(s) in higher resolution of the same image. For this stepwise procedure, firstly a different approach for matching could be used, e.g. correlation methods. Although the correlation methods have great computational costs, they might still be efficient in lower resolution images due to a smaller amount of extracted features in the images. The processing of low resolution images could then provide coarse ExtOri parameters for the next iterations, so the less accurate initial GPS/INS should be sufficient.

The connection between subsequent frames could also provide better coarse ExtOri parameters. The coarse ExtOri parameters should be estimated for the first frame of the sequence and for the following ones, utilisation of the GPS/INS data could be sufficient. The connection between frames can be made for example by weight points of extracted regions. What is more, the registration of the extracted features between subsequent frames could provide the re-estimation of IntOri parameters.

Utilizing the dynamic search space for correspondence is one way to deal with some of the outliers and reducing the computational cost of the matching used in the method proposed in

this thesis. LS adjustment is sensitive to the outliers; the systematic influences on the measurements and outliers should be excluded from the data before the LS adjustment is applied. However, these conditions are difficult to fulfil in the automatic feature extraction and matching. On the other hand, »RANSAC is capable of interpreting smoothing data containing a significant percentage of gross errors, and is thus ideally suited for applications in automated image analysis where interpretation is based on the data provided by error-prone feature detectors« (Fischer, 1981). For this reason RANSAC is probably more suitable than LS adjustment in the presented case.

6 RAZŠIRJEN POVZETEK V SLOVENŠČINI

P.1 UVOD

P.1.1 Motivacija

V zadnjih desetletjih narašča svetovna poraba energije, izvzemši preteklo leto. (Statistical Review of World Energy 2009, 2010). V državah Evropske unije (EU) zgradbe porabijo 40 % energije in povzročijo 36 % emisij CO_2 (Directive 2010/31/EU, 2010). Izboljšanje energetske učinkovitosti zgradb je nujno potrebno za zmanjšanje porabe energije in toplogrednih plinov.
Številni obstoječi 3D modeli mest vsebujejo podatke o lastnostih stavb, ki jih lahko dopolnimo z analizami posnetkov zajetih z mobilnih platform. Za lažjo interpretacijo vsebine posnetkov si lahko pomagamo z obstoječimi 3D modeli (Stilla, 2000) in dele posnetkov uporabimo za teksturiranje modela. Posnetki zajeti iz zraka ali vesolja z navpičnim kotom gledanja so primerni za podatke o strehah, medtem kot so terestrični posnetki primerni za podatke o fasadah zgradb. Vendar lahko z aeroposnetki ustvarjenimi s kamerami usmerjenimi poševno glede na nadir, to so tako imenovani poševni posnetki (oblique photography), pridobimo podatke tako za strehe kot fasade stavb (Frueh, 2004 in Stilla, 2009). Nekateri pojavi in/ali strukture so vidni tako v VIS kot IR spektru, medtem ko so drugi, na primer sistemi za centralno ogrevanje, toplotni mostovi vidni le v posnetkih zajetih v IR delu spektra. Mestne toplotne otoke lahko opazujemo v IR spektru v večjih merilih, termalne lastnosti stavb pa zahtevajo zajem v večjih merilih (Weng, 2009, Kajfež-Bogataj, 2005). Teksturiranje obstoječega 3D modela stavb z IR posnetki povečuje informacijsko vrednost baze podatkov in omogoča analize termalnih izgub stavb.

P.1.2 Sorodna dela

Predpogoji za natančno teksturiranje 3D modelov stavb so: kalibrirana kamera z znanimi parametri notranje orientacije (NO), georeferenciran model stavb in znani parametri zunanje orientacije (ZO), to je položaj kamere v trenutku zajema (Stilla, 2009). Na platformah, ki

zajemajo posnetke iz zraka, so pogosto poleg kamer tudi naprave za zajem trajektorije leta, in sicer najpogosteje sistem za globalno določanje položaja (Global Positioning System, GPS) in inercialna merilna enota (Inertial Measurement Unit, IMU). GPS in IMU merita podatke za določitev približnih parametrov ZO. Poleg prej omenjenih predpogojev za natančno teksturiranje modela stavb mora biti znan ali s kalibracijo določen relativen položaj vseh merilnih naprav na platformi. Vsi ti parametri morajo biti upoštevani pri projekciji 3D modela stavbe v posnetek za določitev podmatrike digitalne slike za teksturo, ki ustreza izbrani ploskvi stavbe. Parametri ZO pridobljeni z GPS/IMU pogosto ne dosegajo natančnosti, zahtevane za visokokvalitetno teksturiranje. Številni avtorji v raziskavah za samodejno (avtomatsko) povezavo 3D modela s slikami uporabljajo metode digitalne obdelave podob (slik).

Hsu et al. (2000) projicirajo 3D daljice modela v posnetek in sočasno zaznajo linijske grafične gradnike, ter določijo položaj kamere v času zajema. Grafične gradnike najprej zaznajo in jim sledijo skozi zaporedne posnetke videosekvence, nato pa te grafične gradnike uporabijo za določitev približnih parametrov ZO med posameznimi posnetki. Daljice projicirajo s približnimi parametri ZO v posnetke. Položaj kamere ponovno preračunajo na podlagi ujemanja projiciranih daljic in usmerjenih gradientov energijskih piramid posnetka (oriented image gradient energy pyramids). Frueh et al. (2004) prav tako uporabljajo zaznane daljice za določitev položaja kamere. Na visokoločljivih poševnih posnetkih v VIS spektru zaznajo robove s Cannyjevim detektorjem. Zaznane robove nato razdelijo v daljice z rekurzivnim algoritmom delitve končnih točk (recursive endpoint subdivision algorithm). Položaj kamere določijo z registracijo projekcije modela stavb v posnetek z zaznanimi daljicami in ga ocenijo z vrednostjo korelacije, ki temelji na ujemanju »linije na linijo« (line-to-line matching). Tian et al. (2008) predlagajo ujemanje robov skozi videoposnetke z uporabo geometrijskih vezi določenih na podlagi zanesljivih točk. Zanesljive točke določijo z analizo koncev zaznanih robov, ki jih kvalitativno ocenijo. Z uporabo zanesljivih točk bistveno zmanjšajo prostor iskanja za ujemanje točk in s tem tudi zmanjšajo čas izračuna.

Možen pristop za določitev položaja kamere je uporaba enega ali več bežišč. Lee et al. (2002) uporabljajo dve ali tri bežišča in ujemanje 3D in 2D (dvorazsežnih) daljic za določitev parametrov ZO terestrične kamere. Ding et al. (2008) predlagajo dvostopenjski proces za

določitev položaja kamere. V prvem koraku določijo približne parametre ZO z uporabo bežišč in podatkov pridobljenih iz GPS/INS. V drugem koraku izboljšajo natančnost približnih parametrov ZO določenih v prvem koraku z ujemanjem pravokotnih vogalov zaznanih s poševnih posnetkov in modela stavb pridobljenega z lidarjem (LIDAR, Light detection and ranging).

Hoegner in Stilla (2008) uporabita IR videoposnetke zajete z mobilne terestrične platforme za samodejno teksturiranje 3D modelov stavb. Stilla et al. (2009) in Kolecki et al. (2010) opisujejo neposredno georeferenciranje s podatki GPS/INS in razširjeno kalibracijo merilnega sistema. Poudarijo problem vibrirajoče platforme, na primer helikopterja, ki lahko povzroči neskladje projiciranega modela z njegovo podobo na posnetku.

P.1.3 Opis problema in cilj raziskave

Če natančnost meritev GPS/INS ne zadostuje za določitev parametrov zunanje orientacije (ZO) mobilne platforme s potrebno natančnostjo, lahko vplive na meritve GPS/INS modeliramo in jih tako popravimo. Obraten pristop k rešitvi tega problema je uporaba vsebine posnetka za izboljšanje natančnosti parametrov ZO. V tej diplomski nalogi z metodami obdelave podob zaznamo grafične gradnike na IR posnetkih videosekvence in jih povežemo s 3D modelom stavb. Glavni namen naše raziskave je samodejna določitev tekstur za strehe in fasade stavb iz IR posnetkov za teksturiranje obstoječega 3D modela stavb. Razvijemo metodologijo za izboljšanje določitve parametrov ZO z metodo ujemanja »točke na točko« (point-to-point).

Hipotezi naloge sta:

- Ko-registracija zaznanih grafičnih gradnikov z IR videoposnetkov in 3D modela stavb izboljša parametre ZO tako, da se model stavb bolje prilega stavbam na posnetku.
- meritve GPS/INS in kalibracija celotnega sistema zagotavljajo dovolj natančne približne parametre ZO za prvo (izhodiščno) projekcijo 3D modela v posnetek.

P.2 METODOLOGIJA

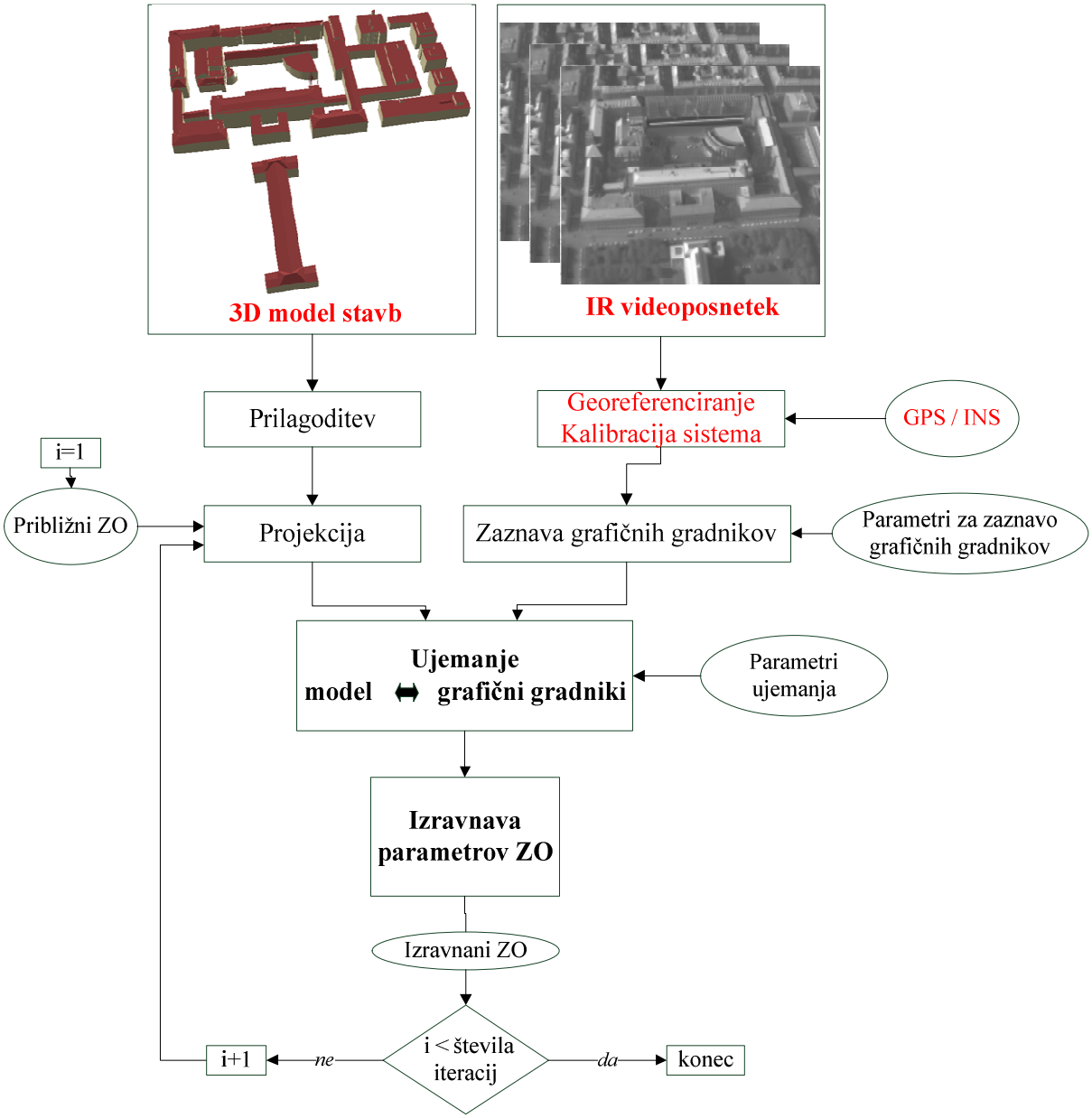
P.2.1 Pregled grafičnih gradnikov zaznanih z infrardečih posnetkov

Na testnem IR videoposnetku zaznamo različne vrste grafičnih gradnikov, kot na primer točke, linije, daljice. Uporabimo različne standardne operatorje za zaznavo grafičnih detektorjev: Förstnerjev operator, Cannyjev detektor robov in druge. Namen raziskave zaznave grafičnih gradnikov je izbira tistih, ki ustrezno predstavljajo robove stavb in omogočajo samodejno ujemanje s 3D modeli stavb. Izbira ustreznih grafičnih gradnikov in parametrov za njihovo zaznavo zahteva podrobno obravnavo, ki jo izvedemo in predstavimo v tem diplomskem delu. Posnetki v IR spektru imajo nižjo ločljivost in drugačne lastnosti od posnetkov v VIS spektru, kar vzamemo v obzir pri izbiri grafičnih gradnikov.

Zgradbe so najpogosteje pravokotnih oblik, z ravnimi slemeni in drugimi sestavnimi deli kot so na primer: dimniki, frčade in okna, zato jih je mogoče opisati z enostavnimi geometrijskimi oblikami. Pri obravnavi 3D modela stavb s stopnjo podrobnosti LOD2 so stavbe predstavljene dovolj dobro z daljicami namesto z bolj kompleksnimi gradniki kot so krivulje ali loki.

P.2.2 Razvita metodologija

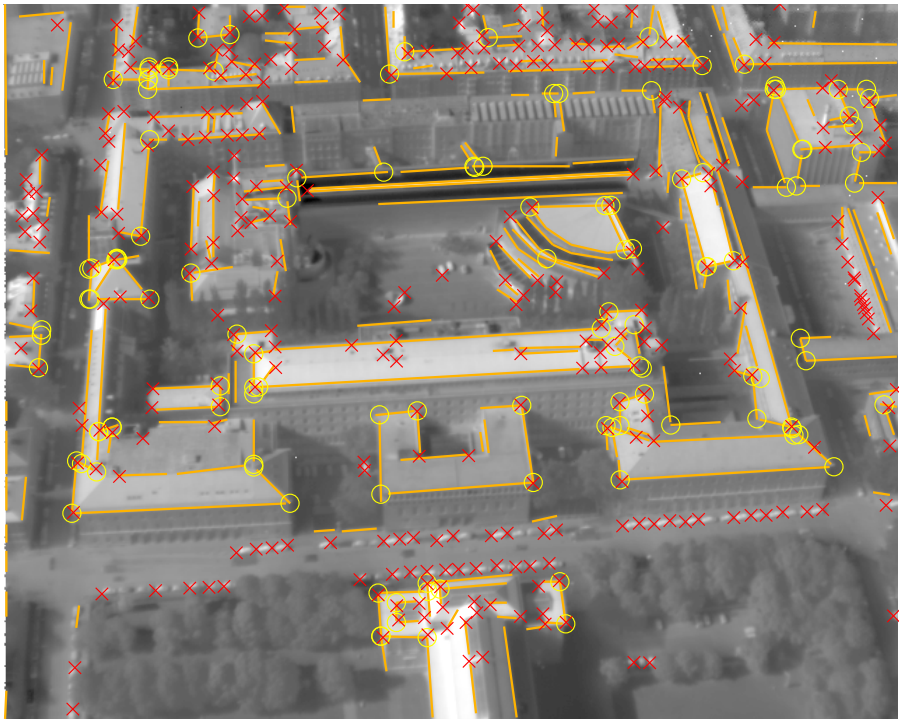
Na sliki P.1 je shematsko prikazana razvita metodologija. Na podlagi podatkov GPS/INS in razširjene kalibracije sistema projiciramo v vsak posnetek IR videosekvence prilagojen 3D model stavb s približnimi parametri ZO. Na IR posnetkih samodejno zaznamo grafične gradnike in nato iščemo ujemanje med projiciranim 3D modelom in zaznanimi grafičnimi gradniki. Na podlagi koregistriranih točk izravnamo po metodi najmanjših kvadratov parametre ZO snemalne naprave za trenutek zajema posnetka. Z izravnanimi parametri ZO 3D model ponovno projiciramo v isti posnetek. Ujemanje projiciranega modela in zaznanih grafičnih gradnikov iščemo iterativno. Cilj uporabe metodologije na IR posnetkih je, da se projiciran model zgradbe bolje prilega zgradbi na posnetku, takrat je metoda uspešna.



Slika P. 1: Shematski prikaz razvite metodologije. Z rdečo so obarvani vhodni podatki.

Izbira vrste grafičnih gradnikov in metode za njihovo zaznavo z IR posnetkov. Za ujemanje daljic in projiciranih poligonov 3D modela stavb na posnetek je potrebna korelacijska funkcija. Časovna potratnost izračuna ujemanja »linije na linijo« je visoka (Frueh, 2004) in zaznane daljice (linije) se pogosto ne začnejo in končajo v vogalih zgradb. Neenojen odziv zaznanih robov (non single edge response) lahko otežkoča postopek ujemanja. V primerjavi z ujemanjem »linije na linijo« je ujemanje »točke na točko« izrazito manj časovno potratno za

izračun. Zato se omejimo na točke, kot primerno vrsto grafičnih gradnikov, izberemo dva algoritma za njihovo zaznavo in uporabimo program Halcon MVTec[6] za zaznavo grafičnih gradnikov. Uporabljena sta Förstnerjev algoritem in algoritem za določitev presečiščnih točk, izračunanih iz zaznanih daljic s posnetka (slika P.2). Förstnerjev operator zazna na posnetku značilne točke, ki se razlikujejo od okolice. Obstajata dve vrsti Förstnerjevih točk, in sicer tako imenovane »stične točke«, ki se pojavijo na presečiščih robov na posnetku in »točke ploskve«, ki označujejo spremembo barve ali svetlosti glede na okolico. »Stične Förstnerjeve točke« uporabimo pri iskanju ujemanja med točkami in projiciranim modelom, saj »točke ploskev« ne zastopajo značilnih točk objektov, ki bi jih lahko povezali z modelom stavb.



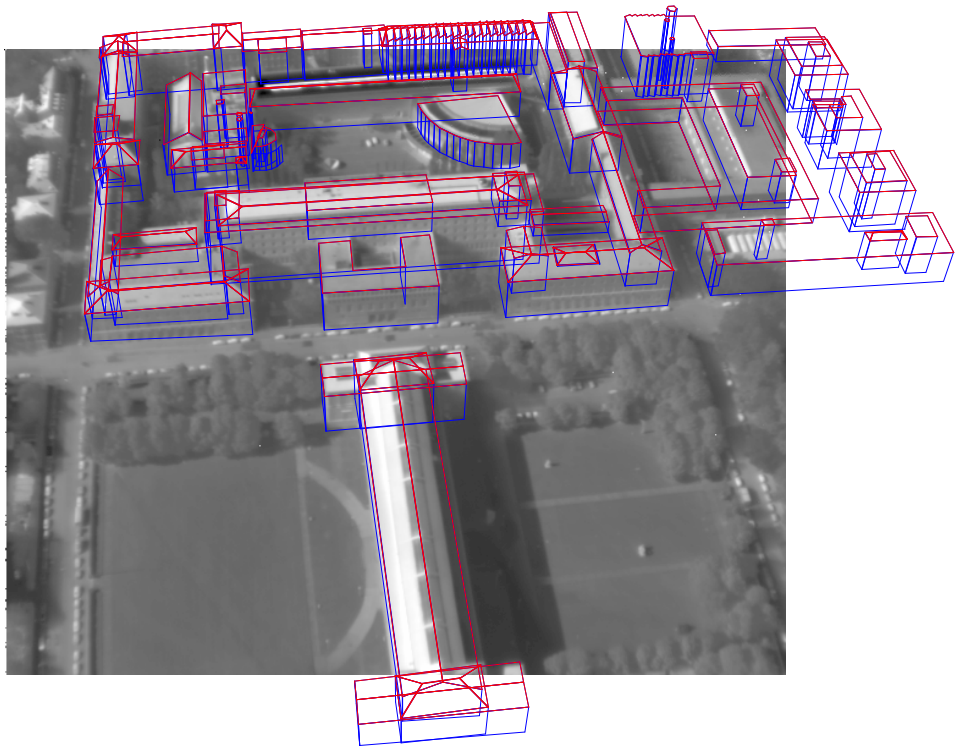
Slika P. 2: Zaznani grafični gradniki. »Stične Förstenrjeve točke« so označene z rdečimi križci, , središča rumenih krogov prestavljajo presečiščne točke določene na podlagi zaznanih ravnih robov, ki so predstavljeni z oranžnimi daljicami.

Presečiščne točke določene iz daljic izračunamo z vmesnim korakom zaznave ravnih robov (slika P.3). Izračunamo razdaljo *d* med končnimi točkami vseh zaznanih ravnih robov (daljic). Postavimo mejo, ki določa največjo dovoljeno oddaljenost med končnimi točkami d_{max} v

[6] Halcon je komercialna programska oprema za aplikacije računalniškega vida, ki ga razvija MVTec. (Halcon, 2010).

pikslih. Povedano drugače, okoli vsake končne točke daljice ustvarimo okroglo območje iskanja s premerom d_{max}. Za vse pare končnih točk daljic, ki izpolnjujejo pogoj $d \leq d_{max}$ izračunamo presečiščni kot α pripadajočih daljic. Presečiščni kot mora biti med vrednostmi α_{min} in $\alpha_{max} = \pi - \alpha_{min}$, kjer je $\alpha \in [0, \pi]$. Ti meji določimo, ker je presečiščni kot med bližnjimi zgradbami redko zelo oster ali zelo top. Skoraj vzporednim daljicam (linijam), ki ležijo blizu skupaj se izogibamo, saj so najverjetneje rezultat neenojnega odziva detektorja robov, to je »slabe« zaznave robu ali pa ne dovolj natančne zaznave dveh vzporednih robov. Za vse pare točk, ki zadostijo pogojema: $d \leq d_{max}$ in $\alpha_{min} \leq \alpha \leq \alpha_{max}$ izračunamo presečiščne točke. Z določitvijo obeh mej, kotne in dolžinske, algoritem za določitev presečiščnih točk omeji določitev manj zanesljivih točk.

Prilagoditev 3D modela stavb. 3D modeli stavb se med seboj razlikujejo po vsebini, načinu shranjenih podatkov o 3D modelu in drugem. Zaradi tega je potrebno 3D model prilagoditi, preden ga uporabimo in z njim preizkusimo razvito metodologijo. S prilagajanjem ne posegamo v geometrijo modela in njegovo natančnost.



Slika P. 3: Žični prikaz 3D modela stavb (LOD2); 3D model stavb je prilagojen, strehe, ki jih obravnavamo v raziskavi, so rdeče, ostali deli stavb modri.

Iz modela najprej odstranimo zakrite robove in ploskve, torej tiste ploskve in dele ploskev, ki so nevidne glede na položaj gledišča in smer gledanja kamere. Posnetki v IR spektru imajo majhne razlike (radiometričnih) vrednosti med nizkimi deli fasad, blizu tlom in pločnikom. Nato iz nadaljnje obravnave izločimo vse tiste točke 3D modela, ki ne pripadajo streham stavb (slika P.3).

Ujemanje grafičnih gradnikov z modelom. Vzpostavimo povezavo med točkami modela in zaznanimi točkami z IR posnetka na podlagi njihove medsebojne lege ter uporabimo ujemanje »točka na točko«. Za vsako točko modela določimo krožno območje iskanja z radiem R v pikslih, v katerem iščemo zaznate točke IR posnetka. Nastopijo različni primeri povezav točk modela in zaznanih točk. Če zaznana točka leži v območju iskanja jo povežemo z modelno točko, točki se ujemata (določimo ju za homologni točki). Če nobena zaznana točka ne leži v območju iskanja modelna točka nima povezave. V primeru, da več točk leži v območju iskanja modelne točke vse povežemo z njo. Če ena zaznana točka leži v več območjih iskanja modelnih točk je povezana le z najbližjo modelno točko.

Izravnava orientacijskih parametrov po metodi najmanjših kvadratov. Ko-registrirane oziroma ujemajoče točke izravnamo po metodi najmanjših kvadratov. Vhodni podatki izravnave po MNK so opazovanja, to so slikovne koordinate zaznanih točk, ki se ujemajo s točkami projiciranega 3D modela, 3D modelne koordinate ujemajočih točk v globalnem koordinatnem sistemu in približni orientacijski parametri. Parametri NO so konstante v izravnavi, parametri ZO so neznanke.

Ponovna projekcija 3D modela stavb. Z izravnanimi orientacijskimi parametri model ponovno projiciramo v posnetek z isto funkcijo (matematičnim predpisom), ki smo ga uporabili za prvo projekcijo modela.

Ponavljanje (iteracije) algoritma. Iterativno uporabimo metodo ujemanja zaznanih in projiciranih točk 3D modela. Radij R območja iskanja ujemanja grafičnega gradnika s točko projiciranega 3D modela stavb lahko določimo kot konstanto ali pa kot spremenljivo vrednost za vsako iteracijo ujemanja.

P.3 UPORABA METODOLOGIJE

P.3.1 Opis podatkov

Infrardeč videoposnetek in geometrija zajema podatkov. IR posnetki so zajeti na urbanem območju z visokoresolucijsko IR kamero AIM 640 QLW FLIR, ki zajema s frekvenco 25 posnetkov na sekundo. IR kamera je vgrajena na platformi na helikopterju. Višina leta je približno 400 m nad površjem. Kamera je usmerjena naprej v smeri leta (oblique forward looking) s kotom opazovanja 45° glede na nadir (pitch kot). Ločljivost posnetkov je 640×512 pikslov. Helikopter je štirikrat letel preko testnega območja, to je območja glavnega kampusa TUM. Z videoposnetka so izrezani štirje snemalni pasovi, vsak s približno 125 posnetki.

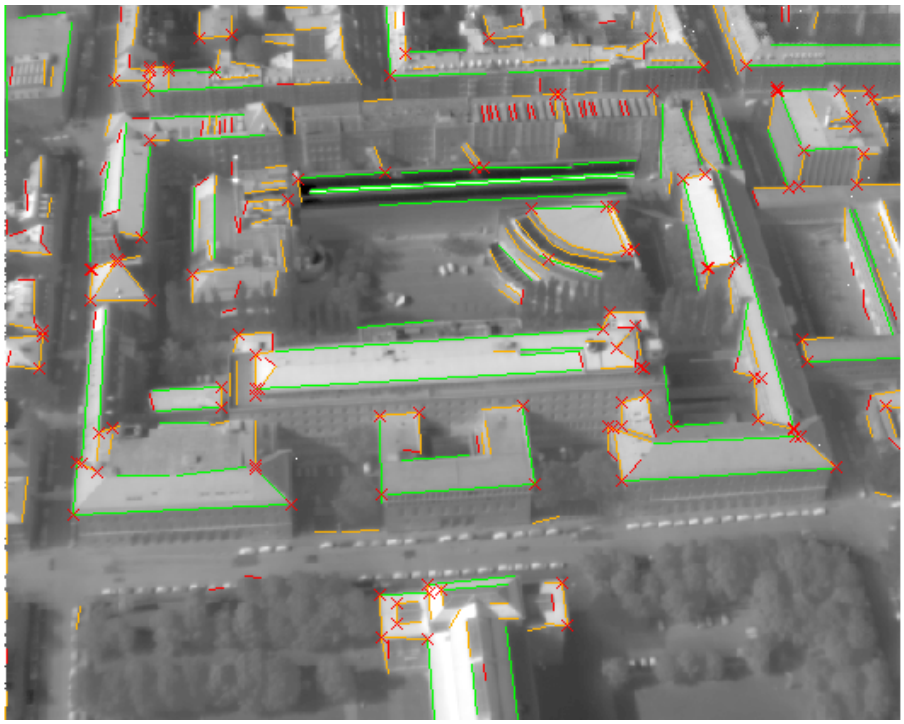
GPS/INS meritve. Na helikopterju je GPS/INS Applanix POS AV 510 enota, ki meri: položaj, GPS čas za trenutek zajema posameznega posnetka, višino, smer in trenutno hitrost leta (približno 160 km/h). GPS antena je pri pilotski kabini in zajema položaj s frekvenco 1 Hz (epoha). Popravki za diferencialni GPS (Differential GPS, DGPS) niso bili dostopni za čas snemanja, kar povzroča slabšo natančnost meritev GPS/INS enote. V tej diplomi uporabljamo popravljene parametre ZO, izračunane po postopku razširjene kalibracije merilnega sistema, ki ga predlagajo Kolecki et al. (2010).

3D model stavb. 3D žični model stavb s stopnjo podrobnosti LOD2 je izdelan s polavtomatsko metodo v programskem orodju INJECT[7]. Glavni kampus TUM je modeliran na podlagi stereoparov aeroposnetkov. Zaradi načina izdelave 3D modela in generalizacije so v modelu prisotne nekatere nepravilnosti in nenatančnosti. Položajna natančnost modela, ocenjena na testnih zgradbah modela, je ocenjena na 1 m. (Frey, 2006).

[7] INJECT je programsko orodje za polavtomatsko ekstrakcijo 3D objektov iz digitalnih aeroposnetkov. Namenjeno je predvsem zajemu stavb, omogoča pa tudi zajem vegetacije, cest in vodnih površin. Za ekstrakcijo 3D objekta je potrebno prekrivanje več aeroposnetkov (lahko tudi satelitskih), znani morajo biti parametri ZO in NO, ter podatki o višinah (DMR).

P.3.2 Preizkus razvite metodologije s podatki

Metodo preizkusimo na testnih podatkih, in sicer na dvanajstih izbranih posnetkih v katere projiciramo celoten model in dva podmodela. Iz vsakega snemalnega pasu (#1-#4) izberemo tri posnetke: iz začetka, sredine in konca vsakega pasu. Nato projiciramo podmodel na 95 izbranih zaporednih posnetkov videosekvence in razvito metodologijo kakovostno ovrednotimo ter uporabimo Forstnerjev algoritem in algoritem za določitev presečiščnih točk. Zaznane ravne robove klasificiramo glede na dolžino in šele nato iz njih določimo presečiščne točke (slika P.4).



Slika P. 4: Zaznane ravne robove klasificiramo in iz njih določimo presečiščne točke (rdeči križci). Rdeče daljice so kratki, manj zanesljivi robovi (8-12 pikslov); oranžni so robovi srednjih dolžin (12-32 pikslov); in zeleni so najdaljši robovi (več kot 32 pikslov).

dm_{ax} = 12 pikslov.

Število zaznanih Förstnerjevih točk z izbranih posnetkov je veliko, vendar relativno malo teh točk leži na vogalih zgradb, zato se tudi malo točk ujema s projiciranim modelom. V primerjavi s številom zaznanih Förstnerjevih točk je število presečiščnih točk majhno, vendar se jih relativno veliko pravilno ujema z modelom (slika P.2). Majhni objekti na strehah, na

primer: dimniki in frčade, so v 3D modelu stavb prikazani s kratkimi daljicami ali poligonskimi stranicami. Pri povezavi teh majhnih objektov z zaznanimi točkami pogosto prihaja do napačnega ujemanja. Napačno ujemanje z vogali majhnih objektov na strehah je bistveno zmanjšano, če so zaznane točke presečiščne točke izračunane na podlagi dolgih robov. Da zmanjšamo verjetnost napačnega ujemanja modela in presečiščnih točk je potrebno mejno vrednost za najmanjšo dovoljeno dolžino zaznanega ravnega robu skrbno izbrati (d_{max}).

P.3.3 Rezultati

V tem podpoglavju predstavljamo rezultate testa, ki je opravljen na dvanajst izbranih posnetkih. Najprej projiciramo v te posnetke oba podmodela, zgradba 1 in zgradba 2, ter nato še celoten model. Podmodela in celoten model projiciramo v posnetke s parametri ZO popravljenimi po postopku opisanem v Kolecki et. al (2010), ki jih v nadaljevanju imenujemo popravljeni parametri ZO. Dodaten test naredimo na 95 zaporednih posnetkih četrtega snemalnega pasu na katerih je v celoti vidna zgradba podmodela v katere projiciramo podmodel s to zgradbo. Nato projiciran podmodel ročno premaknemo, torej uporabimo manj natančne parametre ZO za projekcijo 3D modela na posnetek in ponovno preizkusimo algoritem ujemanja.

Podmodela. Uporaba metodologije na izbranih dvanajstih posnetkih in podmodelih, da nezadovoljiv rezultat, tako z uporabo zaznanih Förstnerjevih kot presečiščnih točk. Relativen položaj projiciranega podmodela v posnetku vpliva na uspešnost metode in algoritma za ujemanje. Obravnavana zgradba ne sme biti prekrita (occluded) z okoliškimi objekti ali sencami in mora ležati v sredini posnetka glede na levi in desni rob posnetka, ter v območju od spodnjega roba do sredine. Vpliv distorzij leč kamere ima večji vpliv na robovih posnetka. Zaradi poševnega kota gledanja kamere, torej zaradi geometrije zajema podatkov, ki smo jih uporabili, so objekti ob zgornjem robu bolj popačeni in so na posnetku relativno manjši v primerjavi z objekti ob spodnjem robu. Ostale težave pri aplikaciji metode na dane podatke so: (pre)majhno število zaznanih točk, ki se ujemajo z modelom, nepopolnosti in nepravilnosti pri zaznavi in ujemanju grafičnih gradnikov.

Celoten model. Ko opazujemo kvaliteto, to je pravilnost in popolnost ujemanja celotnega projiciranega modela z zaznanimi točkami v zaporednih posnetkih videosekvence zaključimo, da razvita metodologija ni učinkovita za dane podatke. Menimo, da bi morali metodo preizkusiti na drugih podatkih, predvsem z drugačnim 3D modelom stavb, da bi lahko bolj objektivno ocenili njeno učinkovitost.

Zaporedni posnetki četrtega snemalnega pasu (podmodel). Podmodel projiciramo v 95 zaporednih posnetkov na katerih zaznamo presečiščne točke (parametri zaznave so v prilogi B). Za začetne parametre ZO uporabimo popravljene parametre ZO in premaknjene parametre ZO, ki so glede na popravljene premaknjeni za 3 m po položaju. Rezultat je visoka popolnost zaznave presečiščnih točk in njihovega ujemanja z modelom. Zaključimo, da je način za izboljšanje parametrov ZO po predlagani metodologiji učinkovit, če obravnavamo eno stavbo, ki leži v sredini posnetka glede na levi in desni rob posnetka, ter v območju od spodnjega roba do sredine. Predpogoj za učinkovitost metode je, da obravnavana stavba ni prekrita s sosednjimi objekti ali drugimi ovirami. Do podobnih zaključkov smo prišli pri preizkusu metode na dvanajstih izbranih posnetkih.

P.3.4 Ovrednotenje metodologije

Metodo, ki smo jo razvili smo ovrednotili na podlagi projekcije podmodela zgradba1 v 95 posnetkov (s številkami 13141-13235) četrtega pasu snemanja. Najprej ocenimo učinkovitost metode, nato še popolnost in pravilnost algoritma za zaznavo in algoritma za ujemanje. Vrednotenje metode z določitvijo teh petih parametrov kvalitete naredimo z vizualno analizo vsakega posnetka.

Učinkovitost metode

Metoda je učinkovita, če se ponovno projiciran model z izravnanimi parametri ZO bolje prilega položaju stavbe na posnetku od projekcije z začetnimi vrednostmi parametrov ZO in je dana z izrazom:

$$učinkovitost\ metode = \frac{Število\ pravilno\ premaknjenih\ in\ rotiranih\ modelov}{Število\ posnetkov}. \qquad (Eq.\ P.1)$$

Tabela P. 1: Povprečna učinkovitost metode pri uporabljenih premaknjenih začetnih vrednostih ZO.

	dX = dY = dZ [m]				
Premik	+1	+2	+3	+4	+5
Učinkovitost	1.00	1.00	0.92	0.65	0.16
	dnaklon = dnagib = dzasuk[‘]				
Premik	+10	+20	+30	+40	+50
Učinkovitost	1.00	1.00	0.97	0.93	0.61

Če uporabimo za projekcijo začetne parametre, ki so glede na vrednosti popravljenih parametrov premaknjeni za 3 m po položaju, ocenjujemo učinkovitost metode na 92 %. Učinkovitost metode ni odvisna le od natančnosti začetnih vrednosti parametrov ZO, ampak tudi od kvalitete zaznave grafičnih gradnikov in relativnega položaja stavbe na posnetku.

V našem primeru je algoritem ujemanja zelo občutljiv na kotne spremembe (tabela P.1). Za začetne vrednosti parametrov ZO morajo biti znane koordinate z natančnostjo vsaj 4 m in vsaj 0,5º za kote vrednosti. Z DGPS meritvami, ki pa niso bile dostopne v času zajema uporabljenega IR videoposnetka, je takšno položajno natančnost mogoče doseči. Nerešen pa ostaja problem natančnosti določitve kotnih vrednosti, nagiba, naklona in zasuka.

Učinkovitost metode preverimo tudi za popravljene vrednosti orientacijskih parametrov, kot začetnih vrednosti za neznanke v izravnavi. Ker so popravljeni parametre ZO zelo blizu pravim vrednostim, za ta primer podrobneje obravnavamo učinkovitost metode. Uvedemo tri kategorije za vrednotenje učinkovitosti metodologije. Prileganje modela glede na položaj objekta na posnetku je lahko:

- »*Izboljšano*«, algoritem je učinkovit.
- Z »*majhno spremembo*«, projekcija modela je malo premaknjena in/ali rotirana glede na prvo projekcijo modela in položaj objekta na posnetku.
- »*Slabše*«, projekcija z začetnimi parametri ZO je boljša. Večja rotacija in/ali premik projiciranega modela z izravnanimi parametri ZO je opazna glede na prvo projekcijo modela in položaja objekta na posnetku.

Tabela P. 2: Povprečna učinkovitost metode pri uporabljenih popravljenih parametrih ZO za začetne vrednosti v izravnani.

Premik	**Učinkovitost**	
Popravljeni parametri ZO	Izboljšana	0,64
	Majhna sprememba	0,22
	Slabša	0,14
	Kontrolna vsota:	1,00

Za uporabljene popravljene parametre ZO ocenjujemo učinkovitost metode na 64 %, kar je manj od učinkovitosti metode pri uporabi manj natančnih začetnih vrednosti parametrov ZO (tabela P.2). 22 % izravnanih parametrov ZO je z majhnimi spremembami in jih še vedno lahko uporabimo za samodejno določitev tekstur iz IR posnetka, medtem ko je za 14 % metoda neučinkovita.

Popolnost in pravilnost algoritma za zaznavo grafičnih gradnikov

Za oceno kvalitete algoritma za samodejno zaznavo grafičnih gradnikov uporabimo dva pokazatelja, in sicer popolnost in pravilnost. Ocenjujemo kvaliteto zaznave presečiščnih točk. Parametri zaznave presečiščnih točk za obravnavani primer so dani v prilogi B.

Popolnost in pravilnost algoritma za zaznavo točk je kvocient med številom pravilno zaznanih točk in številom modelnih točk, oziroma številom pravilno zaznanih točk in številom vseh zaznanih točk. Povprečna popolnost in pravilnost algoritma za zaznavo presečiščnih točk za 95 posnetkov je dana v tabeli P.3.

Tabela P. 3: Povprečna popolnost in pravilnost zaznave presečiščnih točk (Vzorec 95 posnetkov, za parametre dane v prilogi B).

Zaznava	
Popolnost	**Pravilnost**
0.879	0.743

Popolnost in pravilnost algoritma za ujemanje

Algoritem za ujemanje projiciranega modela in zaznanih točk na IR posnetkih ovrednotimo analogno kot algoritem za zaznavo grafičnih gradnikov. Uporabimo isti niz podatkov in algoritem za zaznavo presečiščnih točk s parametri, danimi v prilogi B. Popolnost in pravilnost algoritma za ujemanje samodejno zaznanih točk s posnetka in projiciranega 3D modela stavb je dana s kvocientom števila pravilno ujemajočih točk in številom točk modela, oz. številom pravilno ujemajočih točk in številom vseh ujemajočih točk.

Parametra kvalitete algoritma za ujemanje ocenimo za projekcijo podmodela zgradba 1 v posnetke s popravljenimi in premaknjenimi parametri ZO. Metodologijo uporabimo iterativno, za vsak posnetek v treh iteracijah. Premaknjeni začetni parametri ZO so popravljeni parametri ZO ročno premaknjeni za 3 m po položaju (X, Y, Z) in 30’ za kotne vrednosti (nagib, naklon, zasuk).

Tabela P. 4: Povprečna popolnost in pravilnost algoritma za ujemanje. (Vzorec 95 posnetkov, tri iteracije ujemanja, za popravljene in premaknjene parametre ZO.)

Začetne vrednosti ZO	**Ujemanje**					
	Popolnost			**Pravilnost**		
iteracija	**1**	**2**	**3**	**1**	**2**	**3**
Popravljene	0,75	0,75	0,75	0,90	0,90	0,90
Premaknjene	0,73	0,74	0,74	0,87	0,87	0,90

Kvaliteta algoritma za ujemanje, podana s parametroma popolnosti in pravilnosti ujemanja, je visoka (tabela P.4). Za popravljene parametre ZO se popolnost in pravilnost algoritma za ujemanje ne zvišuje z iteracijami, saj so popravljeni parametri zelo blizu pravim vrednostim. Pri projekciji modela s premaknjenimi parametri ZO iterativno opazimo izboljšanje popolnosti in pravilnosti ujemanja zaznanih točk z modelom.

P.3.5 Razprava

S podrobno študijo smo pokazali, da je izboljšanje parametrov zunanje orientacije s postopkom ujemanja zaznanih presečiščnih točk in projiciranega 3D modela stavb učinkovito.

Če za zaznane grafične gradnike uporabimo »stične Förstnerjeve točke«, je metoda manj učinkovita. Zaradi poševnega kota gledanja kamere so nekateri objekti na tleh in spodnji robovi stavb navidezno blizu robovom streh. Posledica tega so nepravilna ujemanja modela z zaznanimi točkami, kar je bolj očitno pri Förstnerjevih točkah. Förstnerjev operator zazna številne majhne objekte, medtem ko je postopek za določitev presečiščnih točk manj občutljiv na značilne majhne točkaste objekte na posnetku.

Prostorska ločljivost posnetka ni enaka za vse piksle posnetka in je večja v spodnjem kot v zgornjem delu. To velja za vse poševne posnetke zajete s kamero, ki gleda poševno naprej v smeri leta. Ugotovili smo, da mora obravnavana stavba na posnetku ležati v sredini, od spodnjega roba posnetka, do sredine glede na smer leta. Metoda je neučinkovita, če obravnavana zgradba leži ob robovih. Iz vpliva relativnega položaja zgradbe v posnetku na učinkovitost metode sklepamo, da parametri NO in distorzije leč niso bili določeni z dovolj visoko natančnostjo v postopku kalibracije, ter da je razvita metoda občutljiva na te nenatančnosti.

Preizkus metode smo naredili s projekcijo celotnega 3D modela stavb in dvema podmodeloma v IR videoposnetek. Za oceno kvalitete metode ujemanja smo projiciran model premaknili in šele nato uporabili metodo ujemanja. Ugotovili smo, da morajo biti začetne vrednosti parametrov ZO znane z natančnostjo položajnih parametrov boljšo od 4 m in natančnostjo za kotne vrednosti boljšo od 0,5º. Z meritvami DGPS je mogoče doseči takšno natančnost merjenega položaja aero platforme (Grewal, 2007), vendar pa ostaja problem natančne določitve kotnih vrednosti parametrov ZO, nagiba (roll), naklona (pitch) in zasuka (yaw).

Z metodologijo, ki smo jo razvili je mogoče izboljšati položaj in orientacijo snemalne naprave na mobilni platformi za celoten videoposnetek, za vsak trenutek zajema posnetka, pod pogojem, da obravnavana stavba ne leži na robu vidnega polja kamere. Neskladje projiciranega modela z njegovo podobo na posnetku, ki ga povzroči vibriranje platforme (Kolecki, 2010), lahko zmanjšamo in tako izboljšamo tudi določitev položaja teksture IR posnetka za posamezno zgradbo ali del zgradbe. Vendar pa metoda ni učinkovita, če sočasno obravnavamo vse stavbe 3D modela stavb.

P.4 ZAKLJUČEK IN NADALJNJE RAZISKAVE

V diplomski nalogi obravnavamo popravo parametrov ZO IR kamere na mobilni platformi z metodami samodejne obdelave podob. Glavni namen opravljene raziskave je samodejna določitev tekstur za strehe in fasade stavb iz IR posnetkov za teksturiranje obstoječega 3D modela stavb. Razvita metoda temelji na ujemanju »točke na točko« (point-to-point matching) grafičnih gradnikov zaznanih z IR videoposnetka in 3D žičnega modela stavb.

Rezultati preizkusa metode za dane podatke in projiciran podmodel so obetajoči. Učinkovitost razvite metodologije je več kot 90 % za uporabljene začetne parametre ZO s položajno natančnostjo vsaj 4 m in natančnostjo kotnih vrednosti vsaj 0,5°. Parametra kvalitete, popolnost in pravilnost zaznanih presečiščnih točk sta 88 % in 75 %. Popolnost ujemanja modela s presečiščnimi točkami ocenjujemo na 73 % za iterativen postopek ujemanja, pravilnost ujemanja pa na preko 90 %. Vendar ugotavljamo, da dana metoda ni učinkovita, če obravnavamo celoten model hkrati. Možni vzroki za to so: zaznava grafičnih gradnikov s premajhno natančnostjo, ločljivost IR posnetkov in lastnosti IR spektra, distorzije leč, nezadostna natančnost parametrov ZO, ter vpliv poševnega pogleda kamere.

Pri preizkusu smo opazili pomanjkljivosti metodologije in uporabljenih vhodnih podatkov, zlasti 3D modela stavb. Možnosti za nadaljnje raziskavo in izboljšanje metodologije predlagamo v nadaljevanju. Izračun presečiščnih točk iz zaznanih ravnih robov, kot smo ga predlagali, lahko izboljšamo s spremembo oblike krožnega območja iskanja. Z območjem iskanja v obliki elipse ali pravokotnika utežimo vpliv določenih mejnih vrednosti za najmanjši in največji dovoljen presečiščni kot, ter oddaljenost med obravnavanima daljicama. Dodatno lahko zaznane daljice klasificiramo v razrede glede na dolžino in jim glede na razred, ki mu pripadajo določimo, višjo utež v matriki uteži za izravnavo po MNK. Ujemanje med zaznanimi grafičnimi gradniki in modelnimi točkami lahko naredimo v več korakih, in sicer najprej v posnetkih z zmanjšano ločljivostjo in nato s polno ločljivostjo, ter tako pridobimo približne parametre ZO. Le-te lahko izboljšamo s povezavo zaznanih grafičnih gradnikov med posnetki. Namesto izravnave po metodi najmanjših kvadratov lahko uporabimo metodo RANSAC (Fischler, 1981) in se s tem izognemo velikemu vplivu grobo pogrešenih opazovanj na izravnavo parametrov ZO.

7 REFERENCES

Aircraft Rotations, The National Aeronautics and Space Administration (NASA), Glenn Research Centre. http://www.grc.nasa.gov/WWW/K-12/airplane/rotations.html (09.06.2010).

Albert, J., Bachmann, M., Hellmeier, A. 2003. Zielgruppen und Anwendungen für Digitale Stadtmodelle und Digitale Geländemodelle. Erhebungen im Rahmen der SIG 3D der GDI NRW. Institut für Geodäsie und Geoinformation, Universität Bonn, Germany. http://www.ikg.uni-bonn.de/fileadmin/sig3d/pdf/Tabelle_Anwendungen_Zielgruppen.pdf. (09.06.2010).

Brown, L.G. 1992. A survey of image registration techniques. ACM computing Surveys, 24, 4: 326-376.

Campbell, J. 1996. Introduction to remote sensing. Second edition. London, Taylor and Francis: p. 22-52, 239-242.

Catmull, E. 1974. A Subdivision Algorithm for Computer Display of Curved Surfaces. Doctoral Dissertation, Salt Lake City, University of Utah.

Characteristics and use of infrared detectors. Technical information SD-12. Hamamatsu Photonics, Japan, November 2004. http://sales.hamamatsu.com/assets/applications/SSD/Characteristics_and_use_of_infrared_detectors.pdf (10. 06. 2010).

Department of Electrical and Computer Engineering, Multipolar Infrared Vision Chair: Theory. Quebec City, Canada, Laval University. http://mivim.gel.ulaval.ca/dynamique/index.php?idM=57 (10.06.10).

Ding, M., Lyngbaek, K., Zakhor, A. 2008. Automatic registration of aerial imagery with untextured 3D LiDAR models. In: Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR, Anchorage, Alaska, June 2008.

Douglas, D. H., Peucker, T. K. 1973. Algorithms For The Reduction Of The Number Of Points Required To Represent A Digitized Line Or Its Caricature. Cartographica: The International Journal for Geographic Information and Geovisualization 10, 2: 112-122.

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 24, 6: 381–395.

Förstner, W. 1994. A framework for low level feature extraction. In: Computer Vision - ECCV 94: 383-394.

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

Frey, D. 2006. Erstellung eines 3D–Modells für das Testgebiet TUM mit Hilfe des Softwarepaketes inJECT, Vertiefungsprojekt. Technische Universität München, November 2006. (Personal archives).

Frueh, C., Sammon, R., Zakhor, A. 2004. Automated Texture Mapping of 3D City Models With Oblique Aerial Imagery. In: Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission *3DPVT*, Thessaloniki, Greece, September 2004: 396–403.

González, R. C., Woods, R. E. 2008. Digital image processing. Third edition. New Jersey, Pearson Prentice Hall: p. 1-27, 72-104.

González, R. C., Woods, R. E., Eddins, S. L. 2002. Digital Image processing using MATLAB. Second Edition. New Jersey, Pearson Prentice Hall: 793.

Grewal, M. S., Weill, L. R, Andrews, A. P. 2007. Global Positioning Systems, Inertial Navigation, and Integration. Wiley-Interscience, A John Wiley & Sons, Inc., Publication: p. 382-424.

Guid, N. 2001. Računalniška grafika. Maribor, Fakulteta za elektrotehniko, računalništvo in informatiko: 440.

Halcon MVTec software. http://www.mvtec.com/halcon/ (23.3.2010).

Halcon Reference Manual, Version 9.0.2 MVTec Software GmbH. 2010. http://www.mvtec.com/download/documentation/reference-9.0/hdevelop/ (23.03.2010).

Hebel, M., Stilla, U. 2007. Automatic registration of laser point clouds of urban areas. In: PIA07-Photogrammetric Image Analysis 2007, International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences, 36 (3/W49A): p. 13-18.

Heckbert, P.S. 1986. Survey of texture mapping. IEEE Computer Graphics and Applications 6, 1: 56-67.

Herakovič, N. 2007.Računalniški in strojni vid v robotizirani montaži = Computer and Machine Vision in Robot-based Assembly. Strojniški vestnik 53, 12: 858–873.

Hochberg, J. E., Brooks, V. 1962. Pictorial recognition as an unlearned ability: A study of one child's performance. American Journal of Psychology, 75: 624-628.

Hoegner, L., Stilla, U. 2009. Thermal leakage detection on building facades using infrared textures generated by mobile mapping. IEEE In: 2009 Joint Urban Remote Sensing Event JURSE 2009, Shanghai, China: p. 1-6. http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F5076157%2F5137466%2F05137681.pdf%3Farnumber%3D5137681&authDecision=-203. (24.10.2009).

Hsu, S., Samarasekera, S., Kumar, R., Sawhney, H.S. 2000. Pose estimation, model refinement, and enhanced visualization using video. In: IEEE Proceedings of Computer Vision and Pattern Recognition CVPR00, Hilton Head, South Carolina: p. 488-495.

Ibarra-Castanedo, C. 2005. Quantitative subsurface defect evaluation by pulsed phase thermography: depth retrieval with the phase. Doctoral Dissertation, Québec City, Laval University: p. 127-140.

International Comission on Illuminance (CIO). http://www.cie.co.at/ (04.08.2010).

International Organization for Standardization, Technical committee 163: Thermal performance and energy use in the built environment (ISO/TC 163). http://isotc.iso.org/livelink/livelink/open/tc163 (07. 06. 2010).

Kajfež-Bogataj, L. 2005. Climate change impacts on quality of human live = Podnebne spremembe in njihovi vplivi na kakovost življenja ljudi. Acta agriculturae Slovenica (Slovenia), 85: 41-54.

Kolecki, J., Iwaszczuk, D., Stilla, U. 2010. Calibration of an IR camera system for automatic texturing of 3D building models by direct geo-referenced images. In: Proceedings of Calibration and Orientation Workshop EuroCOW. Castelldefels, Spain February 2010: p. 10-12.

Kosmatin Fras, M., Vezočnik, R., Gvozdanović, T., Kogoj, D. 2008. Avtomatizacija celotnega postopka relativne orientacije stereopara = Complete automation of the relative orientation of a stereopair. Geodetski vestnik, 52, 2: 241-266.

Kraus, K. 1993. Photogrammetry. Vol. 1. Fourth edition, rev. and enl., Bonn, Dümmler: p. 4-41, 247-290, 397-286.

Lee, S. C., Jung, S. K., Nevatia, R. 2002. Automatic Pose Estimation of Complex 3D Building Models. In: Sixth IEEE Computer Society Workshop on Applications of Computer Vision WCAV, 3-4. December 2002, Orlando, Florida: 148-152.

Luhmann, T. 2010. Erweiterte Verfahren zur geometrischen Kamerakalibrierung in der Nahbereichsphotogrammetrie. Doctoral Dissertation, München, Technische Universitaet Dresden: 48-60.

OpenGIS City Geography Markup Language (CityGML). 2008. Encoding Standard, Open Geospatial Consortium (OGC). http://www.opengeospatial.org/standards/citygml (20.08.2008).

Oštir, K. 2006. Daljinsko zaznavanje. Ljubljana, Založba ZRC: 250.

Poženel, P. 1999. Trojezični elektrotehniški slovar: angleško-nemško-slovenski. Ljubljana, Tehniška založba Slovenije: 324.

Simulator Engineering Specialists - ACME Worldwide Enterprises (ACME). Image: Helicopter roll-pitch-yaw motion. http://www.acme-worldwide.com/dynamic_motion_seat_Rotary.htm (07.06.2019)

Solving Linear System of Equations, Matlab 7.0.5 2007b, MathWorks. www.mathworks.com/access/helpdesk/help/techdoc/ (15.05.2010).

Sonka, M., Hlavac, V., Boyle, R. 2008. Image Processing, Analysis, and Machine Vision, International student edition. Third edition. Thomson Learning: 829.

Stadler, A., Kolbe, T. H. 2007. Spatio-Semantic Coherence in the Integration of 3D City Models. In: Proceedings of the 5th International Symposium on Spatial Data Quality, Enschede.

Statistical Review of World Energy 2009, British Petroleum. June 2010. http://www.bp.com/sectiongenericarticle.do?categoryId=9023753&contentId=7044109 (07. 06. 2010).

Stilla U., Soergel, U., Jaeger, K. 2000. Generation of 3D-city models and their utilization in image sequences. In: International archives of photogrammetry and remote sensing, 33, part B2: 518-524.

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

Stilla, U., Quint, F., Sties, M. 1995. Analyse von Luft und Satellitenbildern zur automatischen Ermittlung der Bodenversiegelung städtischer Siedlungsbereiche:DFG-Bericht II. Ettlingen/Karlsruhe: FIM/IPF: 9-15.

Stopar, B. Izravnalni račun I in II, Zapiski predavanj = Adjustment theory I and II, Lecture Notes. 2006, 2007. University of Ljubljana, Faculty of Civil and Geodetic Engineering. (Personal archives).

Šumrada, R. 2005a. Tehnologija GIS. Ljubljana, Fakulteta za gradbeništvo in geodezijo: 330.

Šumrada, R. 2005b. Strukture podatkov in prostorske analize.Ljubljana, Fakulteta za gradbeništvo in geodezijo: 284.

The energy performance of buildings (recast). 2010. Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010. http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2010:153:SOM:EN:HTML (07. 06. 2010).

Tian, Y., Gerke, M., Vosselman, G., Zhu, Q. 2008. Automatic Edge Matching Across an Image Sequence Based on Reliable Points. In: Chen J., (ed.), Jiang J., (ed.), Förstner W., (ed.). The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Beijing: p. 657-662.

Weng, Q. 2009. Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. ISPRS Journal of Photogrammetry and Remote Sensing, 64, 4: 335-344.

Zitová, B., Flusser, J. 2003. Image registration methods: a survey. Image and Vision Computing, 21: 977–1000.

Other used references

Autodesk LandXplorer CityGML Viewer. http://usa.autodesk.com/adsk/servlet/index?siteID=123112&id=13133842&linkID=13111311 (08.05.2010).

Avbelj J., Iwaszczuk D., Stilla, U. 2010. Matching of 3D wire-frame building models with image features from infrared video sequences taken by helicopters or UAVs. In:

ISPRS Commission III symposium on Photogrammetric Computer Vision and Image Analysis, Sain-Mandre, France, 1.-3. September 2010, XXXVIII, part 3B: p.149-154.

Chang, Y., Aggarwal, J. K. 1997. Line Correspondences from Cooperating Spatial and Temporal Grouping Processes for a Sequence of Images. Computer Vision and Image Understanding, 67, 2: 186–201.

Eric, W., Grimson, L. 1991. Object Recognition by Computer. The MIT Press: 532.

European Comission for Energy: Energy Efficiency. http://ec.europa.eu/energy/efficiency/index_en.htm (07. 06. 2010).

Gorjup, Z. 2001. Temelji fotogrametrije in postopki izvrednotenja. Ljubljana, Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo: 142.

Hamamatsu Photonics, Japan. http://www.hamamatsu.com/ (10. 06. 2010).

Iwaszczuk, D., Avbelj, J., Stilla, U. 2010. Matching von 3D Gebäudemodellen mit Wärmebildern einer flugzeuggetragenen IR-Kamera. In: 3-Ländertagung DGPF - OVG - SGPBF, Technische Universität Wien, 1.-3. July 2010.

Koler – Povh, T. 2005. Navodila za oblikovanje zaključnih izdelkov študijev na FGG in navajanje virov. Ljubljana, Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo: 24.

Lowe, D. G. 1987. Three-dimensional object recognition from single two-dimensional images, Artificial Intelligence, 31.

Lowe, D. G. 1995. Perceptual Organization and Visual Recognition. Second printing. Kluwer Academic publishers, Boston/Dordrecht/Lancaster: 162.

Nalwa, V.S. 1993. A guided tour of Computer Vision. Addison-Wesley Publishing Company: 361.

Rosenfeld, A., Kak, A. C.1982. Digital Picture Processing. Computer science and applied mathematics series, London, Academic press, Inc: 435.

8 APPENDIX

Appendix A: Table with Level of Detail 0-4 of CityGML with accuracy requirements.

Appendix B: Extraction and matching parameters for Förstner points and straight edges.

Appendix A:

Table with Level of Detail 0-4 of CityGML with accuracy requirements

	LOD0	LOD1	LOD2	LOD3	LOD4
Model scale description	regional, landscape	city, region	city districts, projects	architectural models (out-side), landmark	architectural models (interior)
Class of accuracy	lowest	low	middle	high	very high
Absolute 3D point accuracy (position / height)	lower than LOD1	5/5m	2/2m	0.5/0.5m	0.2/0.2m
Generalisation	maximal generalisation (classification of land use)	object blocks as generalised features; > 6*6m/3m	objects as generalised features; > 4*4m/2m	object as real features; > 2*2m/1m	constructive elements and openings are represented
Building installations	-	-	-	representative exterior effects	real object form
Roof form/structure	no	flat	roof type and orientation	real object form	real object form
Roof overhanging parts	-	-	n.a.	n.a.	Yes
City Furniture	-	important objects	prototypes	real object form	real object form
Solitary Vegetation Object	-	important objects	prototypes, higher 6m	prototypes, higher 2m	prototypes, real object form

(Adapted from CityGML, OGC 2008, p. 10 and Albert, 2003)

Appendix B:

Extraction and matching parameters for Förstner points and straight edges

1. Förstner points

Extraction parameters – Förstner points:
sigmaGrad: 1.0
sigmaInt: 3.0
sigmaPoints: 4.0
threshInhom: 300
threshShape: 0.1
Gaussian smoothing is set.
Doublets are not eliminated.
Junction Förstner points are extracted (Area Förstner points are excluded).

Matching parameter – Förstner: Search space for correspondence is circular and size of 5 pixels.

2. Straight edge extraction

Extraction parameters - Straight edge extraction
filter size: 9x9 px
minimal amplitude: 18
maximal distance: 5 px
minimum length: 10 px

Matching parameter – Straight edge extraction: Search space for correspondence is circular and size of 5 pixels.