POINT-BASED MATCHING OF OBLIQUE IMAGES ACQUIRED FROM AIRPLANE AND UAV PLATFORMS

UÇAK VE İHA PLATFORMLARINDAN ELDE EDİLEN EĞİK GÖRÜNTÜLERİN NOKTA TABANLI EŞLEŞTİRİLMESİ

SILA BAŞ

ASSOC. PROF. DR. ALİ ÖZGÜN OK

Supervisor

Submitted to

Graduate School of Science and Engineering of Hacettepe University

as a Partial Fulfilment to the Requirements

for the Award of the Degree of Master of Science

in Geomatics Engineering

2020

To My Family…

ABSTRACT

POINT-BASED MATCHING OF IMAGES ACQUIRED FROM AIRPLANE AND UAV PLATFORMS

Sıla BAŞ

Master of Science, Department of Geomatics Engineering

Supervisor: Assoc. Prof. Dr. Ali Özgün OK

March 2020, 73 pages

Images acquired from airborne oblique sensors and from Unmanned Aerial Vehicles (UAV) are widely used. On one hand, oblique imaging allows better visualization of side views of objects in three dimensional spaces. On the other hand, UAV imaging closes the gap between aerial and terrestrial photogrammetry. Therefore, combining airborne oblique and UAV images with a precise feature matching strategy can provide reliable information that can be useful for applications such as dense point cloud extraction for 3D city modeling, visualization, textured 3D city models, and so forth.

In this study, a novel framework for the point-based feature matching of the airborne oblique and UAV imagery is presented. The proposed framework makes use of the powerful A-KAZE descriptor for feature extraction in both oblique and UAV images. Feature extraction with an iterative scheme is developed to construct tentative matches as many as possible. During the iterations, Brute Force matching is utilized for the initial matching of the corresponding features and left-right consistency check together with Lowe's nearest-next distance ratio test are forced to filtering erroneous matches. In order to extract putative matches from the tentative matches, three different strategies are implemented. Each strategy employed outlines a different robust method

for the selection of matching points along with the epipolar constraint enforced between the two datasets.

The developed framework is tested for image pairs acquired over the Dortmund-Centre, Germany, from the International Society for Photogrammetry and Remote Sensing (ISPRS) image orientation benchmark dataset. The proposed framework yields successful results in terms of matching precision and provides a nice balance between the true-positive and false-positive matches. Besides, the results of the proposed framework for two different test pairs outperformed the results of the previously developed approaches in the literature.

ÖZET

UÇAK VE İHA PLATFORMLARINDAN ELDE EDİLEN EĞİK GÖRÜNTÜLERİN NOKTA TABANLI EŞLEŞTİRİLMESİ

Yüksek Lisans, Geomatik Mühendisliği Bölümü
Tez Danışmanı: Doç. Dr. Ali Özgün OK
Mart 2020, 73 sayfa

Eğik hava algılayıcılarından ve İnsansız Hava Araçları'ndan (İHA) alınan görüntüler yaygın olarak kullanılmaktadır. Eğik görüntüleme, üç boyutlu uzaydaki nesnelerin yanal görünümlerinin daha iyi görselleştirilmesini sağlar. Öte yandan İHA ile görüntüleme, hava ve yersel fotogrametri arasındaki boşluğu doldurur. Bu yüzden, eğik hava görüntüleri ve İHA görüntülerinin hassas özellik eşlemesinde birlikte kullanılması 3 boyutlu şehir modelleri için yoğun nokta çıkarımı, görselleştirme, 3 boyutlu şehir modellerine doku giydirilmesi vb. gibi uygulamalarda faydalı olabilecek güvenilir bilgiyi sağlayabilir.

Bu çalışmada eğik hava ve İHA görüntülerinin nokta-tabanlı özellik eşlemesi için yeni bir yaklaşım sunulmuştur. Önerilen yaklaşım, hem eğik hem de İHA görüntülerinde özellik çıkarımı için güçlü A-KAZE tanımlayıcısını kullanmaktadır. Mümkün olduğunca fazla sayıda potansiyel eşleme oluşturmak için yinelemeli bir şekilde özellik çıkarımı geliştirilmiştir. Yinelemeler sırasında karşılık gelen özelliklerin eşlenmesi için Brute Force, sol-sağ tutarlılık kontrolü ve hatalı eşlemelerin filtrelenmesi içinse Lowe'nin en yakının bir sonraki en yakına mesafe oranı testi kullanılmıştır. Potansiyel eşlemelerden başarılı eşlemelerin elde edilmesi için 3 farklı strateji uygulanmıştır. Uygulanan her strateji, iki veri seti arasında uygulanan epipolar kısıtı ile birlikte eşleme noktalarının seçimi için sağlam birer yöntem sunmaktadır.

Geliştirilen yaklaşım, Almanya Dortmund merkezi üzerinde çekilen ve Uluslararası Fotogrametri ve Uzaktan Algılama Birliği (ISPRS) görüntü yöneltme değerlendirmesi

çerçevesinde sağlanan görüntü çiftleri üzerinde test edilmiştir. Geliştirilen yaklaşım eşleme hassasiyeti açısından başarılı sonuçlar verirken doğru-pozitif ve yanlış-pozitif eşlemeler arasında güzel bir denge sağlamaktadır. Ayrıca, önerilen yaklaşımın iki farklı test çifti üzerindeki sonuçları daha önce literatürde geliştirilen yaklaşımların sonuçlarından daha iyi olarak bulunmuştur.

Anahtar Kelimeler: Nokta-tabanlı Özellik Eşleme, Özellik Tespiti, Eğik Hava Görüntüsü, İHA Görüntüsü, A-KAZE Tanımlayıcısı, ISPRS Görüntü Yöneltme Karşılaştırmalı Değerlendirme Testi

ACKNOWLEDGEMENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

1. INTRODUCTION

1.1. Purpose and Scope

Airborne imaging is a classical technique to obtaining consistent and up-to-date information. Photogrammetric methods are applied to measuring and interpreting the collected information from state-of-the-art airborne sensors; thus, such (timely) information can be a critical input for various applications, e.g. mapping, 3D modelling, environment monitoring and disaster management and so forth.

At the present time, images collected from oblique airborne sensors and from Unmanned Aerial Vehicle (UAV) platforms become very popular. Oblique imaging allows better visualization of objects in three dimensional (3-D) spaces with a vertical extent. The actual interest in oblique imaging for mapping purposes arises from its primary characteristic: the exposure of the objects' side views. Many applications such as 3D city modelling, building detection and reconstruction, urban and infrastructural planning, tax assessment, and structural damage identification etc. are based on oblique aerial images.

UAVs (a.k.a Drones) are airborne platforms gaining huge interest in the field of remote sensing recently since they are closing the gap between aerial and terrestrial photogrammetry [1]. UAVs were originally developed for military purposes, but they have started to exploit in numerous civilian tasks in the past decade, thanks to their significant advantages in data acquisition, portability, and cost-efficiency. These small platforms are controlled by ground stations and can reach inaccessible areas. UAVs are particularly appealing for surveillance and emergency applications since a pilot onboard is not required. Moreover, a UAV can be used whenever it is needed and allows the acquisition of extremely-high-resolution (EHR) data (up to 1 mm [2]) with respect to the traditional imaging systems since it can fly very close to the objects under investigation. By giving opportunity to many users to repeatedly acquire low-cost imagery with a very high level of detail, UAV remote sensing techniques became revolutionizing in agricultural, environmental, archaeological and surveillance fields [3].

In addition to acquiring high level of detailed imagery from UAVs, accurate and reliable information about the surface objects becomes crucial in mapping tasks. Even though

UAVs make capturing up-to-date aerial imagery practical and with a low-cost, they have in general poor georeferencing information due to low-quality GNSS and IMU onboard. Therefore, precise georeferencing of UAV images is a requirement. At this point, feature matching with a pre-georeferenced oblique aerial images can be utilized to reach better UAV geo-registration performance beyond hardware limits [1].

Besides, oblique aerial images include facade details detected and recorded with high accuracy and completeness level. Hence, combining the use of UAV and oblique aerial images with feature matching can provide reliable tie points that can be useful for further applications such as dense point cloud extraction for 3D city modeling, visualization, textured 3D city models, etc. [4]. Up till now, a small number of research studies have been conducted for the feature matching of the UAV and oblique aerial imagery. However, the outputs of the approaches developed are still limited due to large differences in terms of scale, rotation, viewpoint, and illumination, between the two datasets. A crucial part in attempting to alleviate these problems involves a feature extraction step to identify a sufficient number of keypoints being input of the image matching step. Besides, the numbers of true matching correspondences remain at an insufficient level for most of the previously developed approaches.

This thesis presents a novel feature matching framework (Fig. 1.1) and evaluates its performance for the airborne oblique and UAV imagery acquired over a test area in Germany. The motivation of the developed framework is to mitigate the problems identified above, and is to enhance the performances of the point-matching between the oblique and UAV images. The principal motivations of the proposed framework are (*i*) to increase the number of unique point-based features detected, (*ii*) to provide tentative matches as large as possible, and (*iii*) to extract putative matches from the tentative matches up to the best possible extent. In this respect, the framework proposed in this thesis consists of three essential steps; (*i*) feature detection with an iterative scheme to collect keypoints, (*ii*) initial matching of the collected keypoints between the oblique and UAV images, and (*iii*) final matching step including outlier
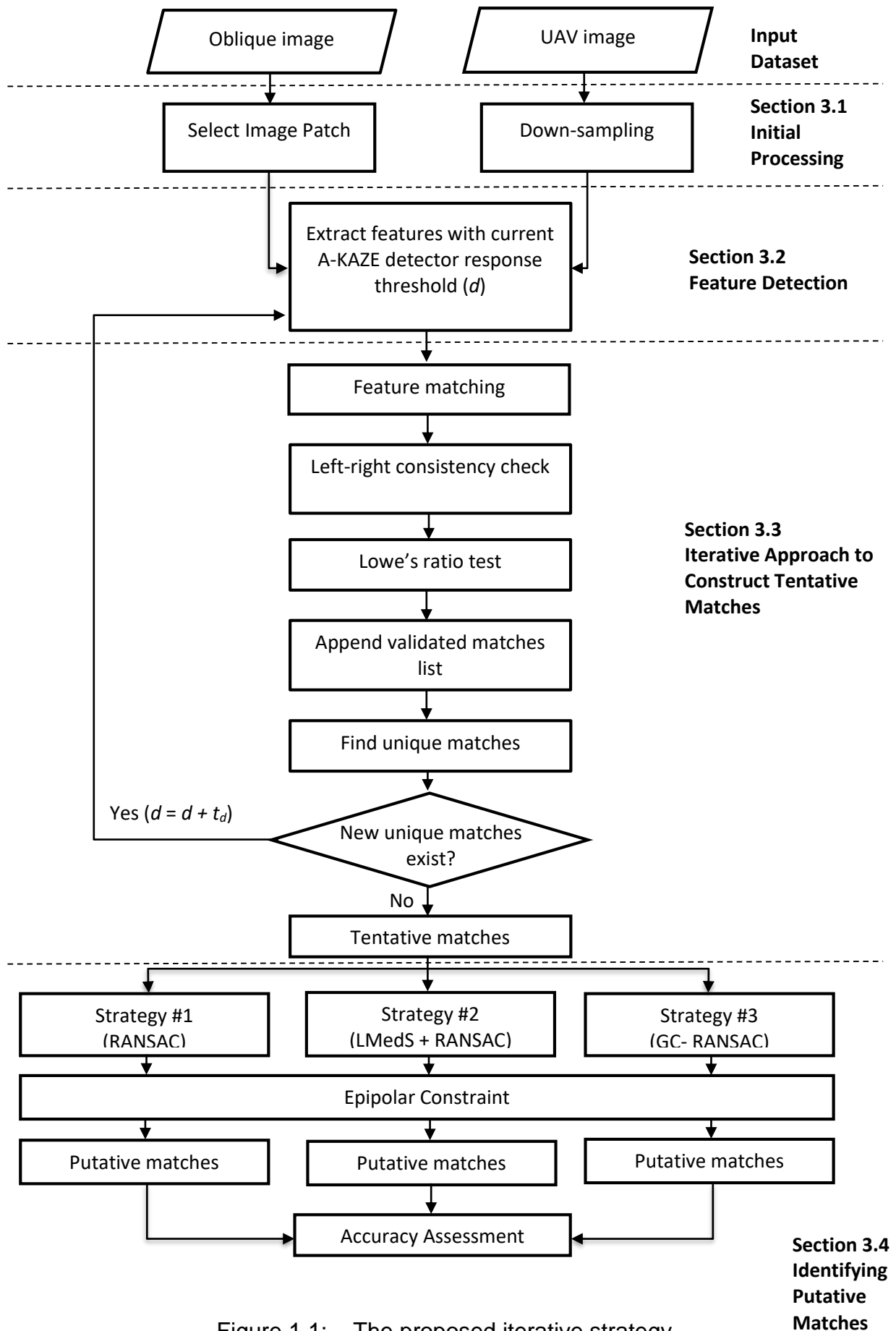
Figure 1.1:   The proposed iterative strategy

3

removal. Figure 1.1 depicts a flowchart showing briefly the proposed strategy to perform a point-based feature matching between the oblique an UAV images.

1.2. Contributions

The novelties of this thesis are as follows:

- In this thesis, a novel framework for the matching of oblique images acquired from airplane and UAV platforms is proposed. To the best of our knowledge, there are only a few studies conducted in the literature in this context.
- An iterative strategy is proposed to increase the number of tentative matches. Thus, a combination of the outputs of different detector response thresholds is manipulated to increase the number of initial matches.
- Three different strategies are implemented and evaluated to accurately extract putative matches from the tentative matching candidates. To the best of our knowledge, it is the first time in the literature that the performance of a graph-based RANSAC strategy is assessed and revealed in this context.
- A Python script to implement the proposed framework is developed.

1.3. Organization of the Thesis

This thesis consists of six chapters. Chapter 2 provides a brief summary of the airborne oblique and UAV imaging with a literature review related to previous studies in the field of point-based feature extraction and feature matching of the remote sensing images.

Chapter 3 presents the details of the developed matching approach. The chapter begins with explaining the initial processing applied. Next, feature detection method and the iterative approach to construction of the potential matching candidates are described. Finally, three different strategies for identifying putative matches are introduced.

The information related to the dataset and the evaluations of the experiments conducted are provided in Chapter 4. First, the study area and the dataset utilized are

described. Thereafter, the accuracy assessment strategy and the evaluation of the critical parameters are reported.

Chapter 5 involves the evaluation of the results of the proposed framework. The comparative assessment of the results of the developed approach with previously developed feature matching approaches is also provided.

Finally, in Chapter 6, the outcomes of this study and the potential future works are stated.

2. BASICS AND STATE-OF-THE-ART

The standard environment of photogrammetric imaging and processing has been well known and established for decades. Digital imaging sensor technologies and camera systems have significantly improved compared to the former designs. Since aerial oblique and UAV images are utilized in this thesis, a short overview of the airborne oblique cameras and UAV platforms are firstly presented in this section. Thereafter, a summary of literature about the previous studies regarding point-based feature extraction and feature matching from remotely sensed images are provided.

2.1. Overview of Airborne Oblique and UAV Imaging

2.1.1. Oblique Imaging

Oblique aerial images, unlike vertical nadir-viewing counterparts, better expose the side views of the objects on the Earth's surface. Therefore, observing an oblique image is more natural and intuitive compared to viewing a vertical image, and even uninformed users can identify and interpret features on the ground [5]. Oblique images are exposed with the camera axis intentionally tilted between the vertical and horizontal. The total area captured with oblique images is much larger than that of vertical images. Generally, oblique imaging is divided into two categories, low and high oblique imagery, according to angular position of the camera utilized:
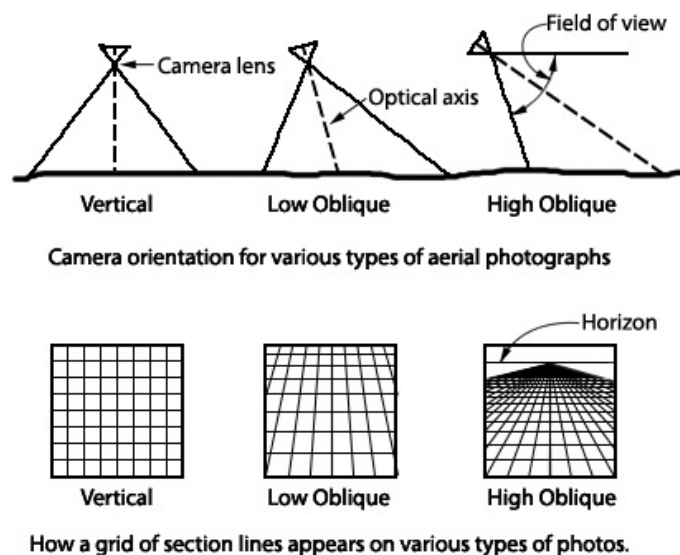
Figure 2.1: Differences between vertical and oblique imaging [6].

6

- **Low Oblique Aerial Imagery:** A low oblique image is taken with the camera inclined about 30° from the vertical. This type of aerial image does not include horizon (Fig. 2.1). The ground area covered is a trapezoid, but the image is square or rectangular. Distance cannot be measured due to scale changes in the entire image. Parallel lines located on the ground are not parallel on image; thus, direction (azimuth) cannot be measured. Relief information is detectable but distorted.

- **High Oblique Aerial Imagery:** A high oblique image is taken with the camera inclined about 60° from the vertical. In this type of aerial image, horizon is visible and it captures a large area (Fig. 2.1). The ground area covered is a trapezoid, while the image is square or rectangular. Distances and directions cannot be measured on image due to the same reasons stated for the low aerial oblique imaging. Relief may be quite detectable but distorted as in any oblique view [7].

Oblique images are usually acquired by multi-camera systems as introduced by many companies such as IGI, Leica, Midas, Vexcel/Microsoft and VisionMap and can be classified into three categories according to their configurations [8, 9] (Table 2.1). These systems comprise of more than one lenses mounted in the same camera body or more than one cameras mounted in an array in order to keep the angles between their optical axes fixed (Fig. 2.2). The shutters are generally synchronized for obtaining exposures at the same time. Eventually, the output accuracy mostly depends on the calibration of the optical axes [5].

2.1.2. UAV Imaging

In the past, UAV systems were mainly developed for military purposes and tasks such as surveillance, reconnaissance, unmanned inspection or recording of inimical regions. But recently for mapping purposes, UAV imaging has opened a variety of new application fields in the close-range domain, providing a cost effective alternative to the traditional airborne photogrammetry for large-scale topographic mapping or three dimensional capturing of terrain information, and therefore, being a fundamental complementing solution to terrestrial photogrammetry [10].

Table 2.1: Summary of oblique multi-camera systems [9]. MC: Maltese-Cross, F: Fan, B: Block.

System	Type	Number of Sensors	Geometric Resolution (px)	Pixel Size (μm)	Spectral Bands	Focal Length (mm)
Vexcel Osprey 2	MC	4+1	11674 x 7514 8900 x 6650	6.0	RGB, NIR	80-120
MIDAS 5	MC	4+1	5616 x 3744 (Canon EOS-1D) 7360 x 4912 (Nikon D800E)	6.4 4.8	RGB	27-90
IGI DigiCAM Penta	MC	4+1	7304 x 5487 8176 x 6132 8956 x 6708 (Hasselblad)	6.8 6.0 6.0	RGB, CIR	50-80
Pictometry	MC	4+1	2672 x 4008	9.0	RGB	65-80
Optron/Trimble AIC	B	4	7228 x 5428 (RolleiMetric)	6.8	RGB, CIR	60-100
VisionMap A3 Edge	F	2	4864 x 3232 (KODAK)	7.4	RGB, CIR	300

(a)   (b)
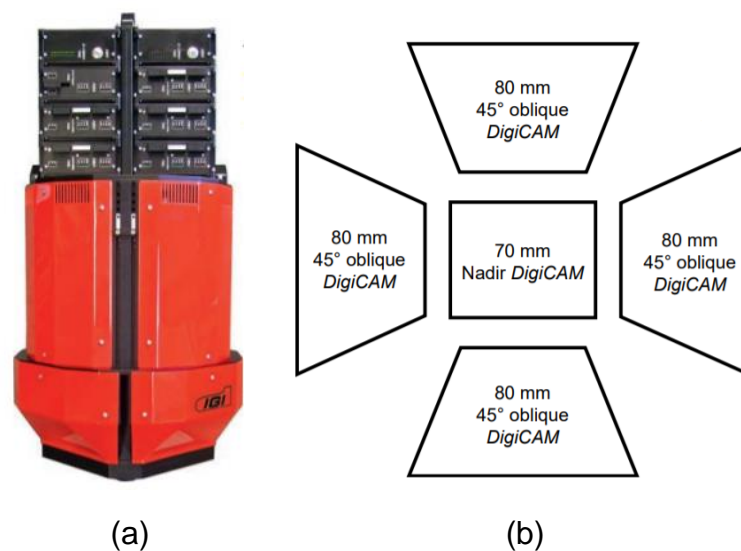
Figure 2.2: (a) a view of the IGI DigiCam Penta [5], (b) the configuration of the camera system [11].

UAV-based images can be obtained by operating flying unmanned platforms with different sizes and usually with small weights. The costs of such platforms are significantly lower than the operations conducted by conventional aircrafts. UAV can be classified into variety of types and classes. The UAVs can be classified as fixed and rotary wings based on their main airframe types (Fig. 2.3). A runway area is not needed for the Rotary wing UAV platforms which have high flexibility for the take-off and landing in a vertical manner; however, fixed wing UAVs enable capturing larger survey areas in a relatively short amount of time [12].

(a)

(b)
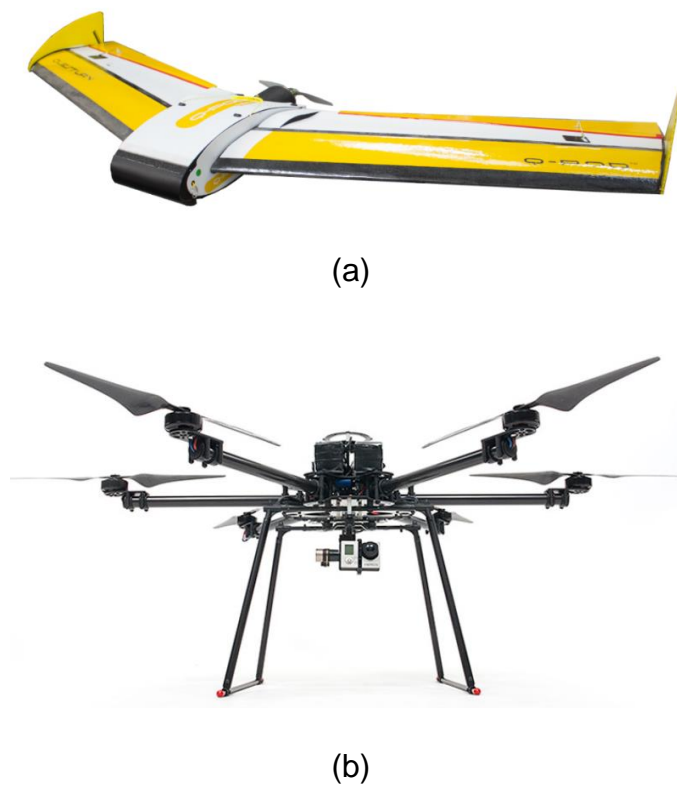
Figure 2.3:   Fundamental airframe types of UAVs [13]; (a) a fixed wing UAV, (b) a rotary wing UAV.

UAV based data collection can replace the terrestrial acquisitions performed in small scale applications. The derived high-resolution products can be used in 3D modelling, dense point cloud generation, texture mapping on existing 3D data, map production or orthophoto generation. In comparison to long-established airborne platforms, UAVs may provide better accuracies and significantly reduce the risk of recovering information in brutal areas. However, small or medium format cameras especially on

low-cost and small payload systems, enforce the acquisition of a higher number of images in order to obtain the same image coverage at a comparable level of detail [12].

2.2. Related Work

Image matching is defined as establishing the correspondence between the image data, while feature extraction can be thought of as a pre-processing step for the image matching tasks [14]. Generally, image matching techniques are classified into two major groups, dense (energy-based) and sparse (feature-based), according to their principal ideas [15].

The main objective of a dense correspondence algorithm is to establish the corresponding pixels between images. In this context, the corresponding pixels may look quite different because of different illumination conditions and different image acquisition angles. Also, in some cases such as uniform surfaces, it is not easy to compute a discriminative feature that can describe a pixel. Therefore, the dense correspondence problem can be seen as an optimization problem to find the best correspondence between pixels of two images. The taxonomy proposed by [16] defines four stages (matching cost computation, cost aggregation, disparity computation and disparity refinement) as building blocks of a dense correspondence algorithm. Early dense matching approaches rely on area-based methods, also termed as correlation-like methods or template matching, and there is no pre-processingf stage for the extraction of salient features from the images [17]. These matching approaches use algorithms that searching for the matches between images based on areas within the images, by using predefined matching windows. The matching cost computation block of a modern dense correspondence algorithm defines a way of assessing the resemblance of image locations. An extensive evaluation of different matching cost computation methods can be found in [18]. Their evaluation shows that, BilSub [19] performs best for low radiometric difference; Mutual Information [20] is more stable for image noise; and as overall performance, Census performs best. The cost aggregation step accumulates the matching costs within a certain neighborhood. Once the cost calculations are done, the disparity estimation can be done locally, globally or semi-globally. A comprehensive study on different techniques can be found in [21] and [22]. More recently, deep learning also became popular for generating dense correspondences [23].

In sparse image matching, the first step is to detect image features such as corners, blobs, edges, etc. in given input images. Once image feature descriptors are computed by feature extraction methods, they are compared, e.g. in stereo geometry, to reveal a relationship between the input images to realize image matching.

Each of the matching methods has specific characteristics and hence they are suitable for solving different type of problems. Dense correspondence is mostly required when dense 3D surface is necessary, e.g. DSM generation, orthophoto production, optical flow etc., whereas sparse correspondence is mostly used to relate two or more images, e.g. image registration, robotics applications etc. The literature on both methods is vast. In accordance with the topic of this thesis, we only discuss point-based (sparse) methods dealing with optical remote sensing images.

In this section, the former studies are examined in an article by article manner, in which each article is investigated and summarized regarding their certain key aspects such as the methodology proposed, the data utilized, and (if available) the results reported.

2.2.1. Point-based Feature Extraction

Extracting features from image data is an essential step for variety of research fields, especially in computer vision domain. Image registration, object recognition, image classification, 3D scene reconstruction, motion tracking and robotics applications are based on defining a set of stable features [24-28]. In addition, many feature extraction methods have been developed in the literature to provide successful image matching.

An **image feature** is an image element that has a specific structure in the image, such as corners and edges (Fig. 2.4), which is possible to discriminate from surrounding parts of the image while providing rich information about the image content. **Feature extraction** can be defined as detecting and isolating desired features from the image or pattern for identifying or interpreting meaningful information from the image data. Feature extraction for local features can be divided into two different stages; feature detection and feature description [29]. Feature detector is a function which finds a set of stable distinctive features in an image, while the feature descriptor encodes information in spatial neighborhoods of the determined features.
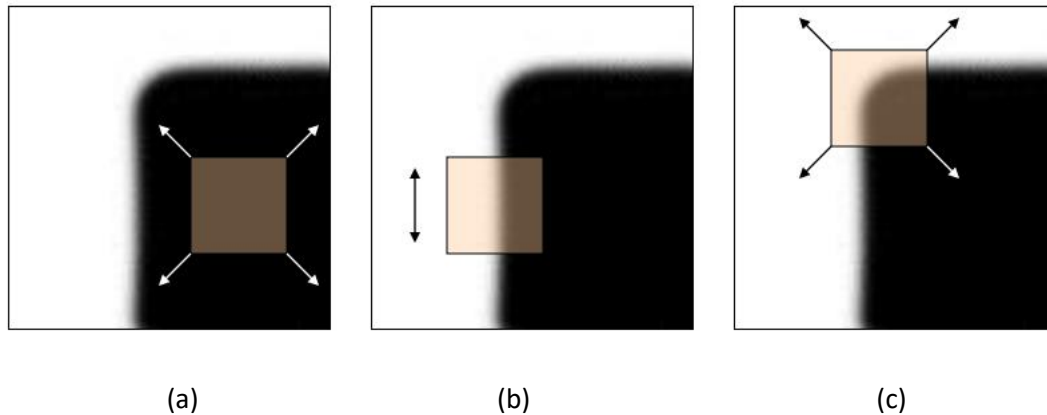
Figure 2.4: For (a) a "flat" region, shifting a window in any direction gives no change, (b) an "edge", shifting a window gives no change along the edge direction, (c) a "corner", shifting a window gives significant intensity change in all directions [30].

The most popular feature descriptor for detecting interest features from images was Scale Invariant Feature Transform (SIFT) proposed by D. Lowe [31] in 2004. SIFT descriptor aims to find features that are invariant to image scale and rotation, affine distortion, change in illumination and 3D perspective, and noise. SIFT algorithm includes four major steps for producing a set of image features. Initial step begins with computing a scale space extrema using the Difference-of-Gaussian (DoG) filters to ensure scale invariant keypoint set. Thereafter, the keypoint candidates are localized and refined by eliminating the low contrast keypoints and edge keypoints. Next, an orientation value is assigned to each keypoint based on local image gradient to achieve invariance to image rotation. For each keypoint, the local image descriptors are computed based on the image gradient magnitude and orientation in the final step. Although SIFT has been found successful various object recognition applications, the major drawback; however, is the relatively large computational cost.

The Speeded-up Robust Features Method (SURF) [32] which was an approximation of the SIFT, operated in a scale space with using Laplacian of Gaussian (LoG) approximation with box filters and utilized a Blob detector based on the determinant of Hessian matrix. Convolution with box filters can be faster if calculation done with the integral images and with this advantage, SURF performs faster than SIFT. However, although SURF features are rotation and scale invariant, they have almost no invariance to affine type distortions.

12

Yu et al. [33] introduced a fully affine invariant method called ASIFT (Affine-SIFT) to improve the SIFT method which is fully invariant concerning only translation, rotation, and scale. Unlike SIFT, ASIFT method also covers two camera axis orientation parameters, latitude and longitude angles. It performs this by simulating many views of the initial two images with changing rotation angles and projection. ASIFT method can detect many feature points but it requires large computational time both for feature extraction and matching.

Binary Robust Independent Elementary Features (BRIEF) [34] was developed as an alternative local binary descriptor for SIFT and SURF with almost similar matching performance unless there is large rotation difference between the input images. When feature extraction is performed on low-power devices, the selection of computationally efficient algorithms is crucial [35]. Therefore, BRIEF descriptor compares feature points relatively faster and stores in memory more efficiently. It dramatically reduces the matching time by means of the Hamming distance utilized in the approach.

ORB [36] is an acronym for Oriented FAST [37] (Features from Accelerated Segment Test) and Rotated BRIEF, and is another effective alternative for SIFT and SURF. ORB, as the name suggests, was based on the combination of FAST and BRIEF descriptors with certain modifications. In ORB, feature detection relies on the FAST detector, which is an effective corner detector suitable for real-time applications due to its computational advantages. It expands FAST through adding orientation component while overcoming BRIEF descriptor's lack of rotational invariance weakness.

Alcantarilla et al. [38] introduced KAZE features in 2012. The name of the algorithm arises from the Japanese word *kaze* which denotes wind and referring the flow of air on a large scale ruled by nonlinear processes. Their method was based on utilizing nonlinear scale spaces to overcome the drawbacks of Gaussian blurring employed by previous approaches, such as SIFT, to operate a scale space with DoG approximation. Non-linear diffusion filtering together with the AOS (Additive Operator Splitting) technique has been used to build the non-linear scale spaces. This allowed locally adaptive blurring of image data to extract feature points, thereby, reducing the noise level while preserving the boundaries of regions. KAZE detector calculates the response of the (scale normalized) determinant of the Hessian matrix at various scale levels. The maxima of the responses are labelled as the potential keypoints. KAZE features are invariant to scale, rotation and have more distinctiveness at varying image

scales. Although KAZE features require more computational cost than SURF, its' results are found to be comparable with SIFT.

One year later in 2013, Alcantarilla et al. [39] introduced Accelerated-KAZE (A-KAZE) to overcome the main drawback of KAZE; the high computational cost involved. A-KAZE utilizes a faster method, Fast Explicit Diffusion (FED) scheme embedded in a pyramidal framework in order to create the non-linear scale space. In addition to FEDs, A-KAZE exploits a modified version of the Local Difference Binary (LDB) descriptor that utilizes gradient and intensity information from the nonlinear scale spaces. Although the LDB descriptor follows the same basis as BRIEF algorithm, it computes binary tests in the influence region surrounding each keypoint.

2.2.2. Point-based Matching of Optical Remote Sensing Images

Up to now, many feature-based methods are developed and commonly used in a variety of application areas. Zhuo et al. [1] investigated the performance of SIFT-based image matching to match the low altitude nearly nadir UAV images with the high altitude nadir aerial images. The authors stated that SIFT and ASIFT methods often performed badly or even fail, because of the large differences between the images regarding scale and rotation, hence they proposed an approach [1] to overcoming these problems. Their method was comprised of the SIFT descriptor, a one-to-many strategy and a geometric verification of the putative matches with the help of pixel-distance histograms. Although their method can found thousands of correct matches even at image regions that have repetitive patterns, their approach requires pre-information about position and orientation for alignment of the UAV image.

Barath et al. [40] described a novel method called Graph-Cut RANSAC (GC-RANSAC) for separating inlier and outlier matchings. The graph cut algorithm proposed a new local optimization step that globally optimal and simple to implement. The authors concluded that the GC-RANSAC algorithm is more geometrically accurate than conventional methods on a variety of problems e.g. line fitting, homography, affine transformation, fundamental and essential matrix estimation.

A two-view matching technique, named as Matching on Demand with Synthesis (MODS) was presented by Mishkin et al. [41] in 2015. MODS employed a combination

of different detectors in which an iterative scheme that powerful still slower detectors and descriptors were applied. They also proposed correspondence selection strategy which generalized the standard strategy based on first to second closest distance ratio. They rigorously evaluated their approach in multiple datasets including a dataset comprising 30 pairs of airborne images.

Bansal et al. [42] proposed a method for facade matching of low altitude aerial images with the street-level images. Their approach was based on a novel descriptor called as Scale-Selective Self-Similarity (S^4) to perform matching of these images with huge viewpoint and illumination variations. In their study, street level queries matched with a database of pre-processed bird's-eye view aerial images for the rectification of the facades. The authors stated that the use of S^4 features instead of SIFT features provided a significant improvement in the final performance.

An approach for the purpose of building recognition was introduced by Chung et al. [43] in 2009. In this work, repetitive structures on buildings are found by the use of MSERs (Maximally Stable Extremal Regions) [44] and exploited for the construction of sketch-based representations of the buildings. Thereafter, the spectral graph theory is used for the matching of the images having large differences regarding scale and viewpoints.

Verykokou and Ioannidis [45] presented a novel algorithm for automatically estimating the exterior orientation parameters of the images including planar surfaces in the captured scenes. In their procedure, main stages composed of (i) the determination of overlapping images, (ii) image matching with using a template matching method, and feature tracking with a homography based technique, (iii) estimation of the exterior orientation parameters, (iv) and a Structure from Motion workflow combined with iterative adjustment methods for outlier removal. The proposed algorithm was tested on a subset of UAV oblique images of the multi-platform photogrammetry benchmark dataset provided.

A procedure for ultra-wide baseline image matching for urban environments was presented by Altwaijry et al. [46] in 2013. Their method benefited advantage of multiple ideas proposed in the literature, and principally based on creating synthetic affine views along with the self-similarity graph and a RANSAC-based scheme. 30 image pairs collected from Google Maps were utilized, and the authors argued that their

method demonstrated an improved performance than the alternate methods in the literature regarding the ultra-wide baseline matching.

Onyango et al. [47] proposed a strategy to register UAV and aerial oblique images for accurately orienting UAV images via already georeferenced oblique images. In this study, used image data is a sample of images in the framework of the multi-platform photogrammetry benchmark that available to researchers and provided by the ISPRS (Scientific Initiative) and EuroSDR. Their procedure implemented the A-KAZE descriptor since the A-KAZE was tested against SIFT, SURF, KAZE, BRISK, BRIEF and gave the best results. They used Brute force hamming distance to find putative matches between the images. To filter wrong matches, Lowe's ratio test was used and this followed by the computation of multiple homography to remove remaining outlier matches. The accuracy evaluation criteria of the study, based on the computation of the average residual error of the successfully matched points with using their corresponding epipolar lines and comparing this value with the average residual error computed from manually matched points. The authors stated that their procedure gave satisfying results in most of the challenging scenarios for the cases having scale, geometry and illumination differences.

In the past years, deep learning has mainly utilized for computer vision tasks such as image classification, object recognition and image segmentation. But recently, the advances in deep learning also allow finding good feature representations for feature extraction tasks. The layers of Convolutional Neural Networks (CNN) manage to intercept complex image characteristics and perform better than traditional algorithms [48]. Therefore, many researchers have implemented these networks in the feature extraction step (Fig. 2.5) rather than SIFT or similar algorithms [49].

Fischer et al. [50] introduced a CNN trained on ImageNet, and a CNN only trained with unlabeled data. While supervised CNN training is advantageous for descriptor matching, the supervised trained network is superior when the features are used for further classification tasks. The authors concluded that these features outperformed SIFT on descriptor matching while requiring more computational cost than SIFT.

In 2018, Yang et al. [51] proposed a feature based image registration approach using a CNN based feature extraction method. In their procedure, they used a pre-trained VGG network to construct a feature descriptor that exploited high-level convolutional

information while keeping localization capabilities. Similar to the previous approach, these descriptors provide better accuracy compared to SIFT especially where SIFT includes a large portion of outliers or performed matching with an insufficient number of keypoints.

A vision-based framework for geolocalizing a UAV with using its on-board camera and satellite imagery was presented by Nassar et al. [52] in 2018. The proposed method allows navigation in GPS-denied regions and improving existing GPS modules. In few words, their strategy combined traditional computer vision techniques and deep learning networks to registering a reference satellite image with a UAV image to guide the UAV localization.
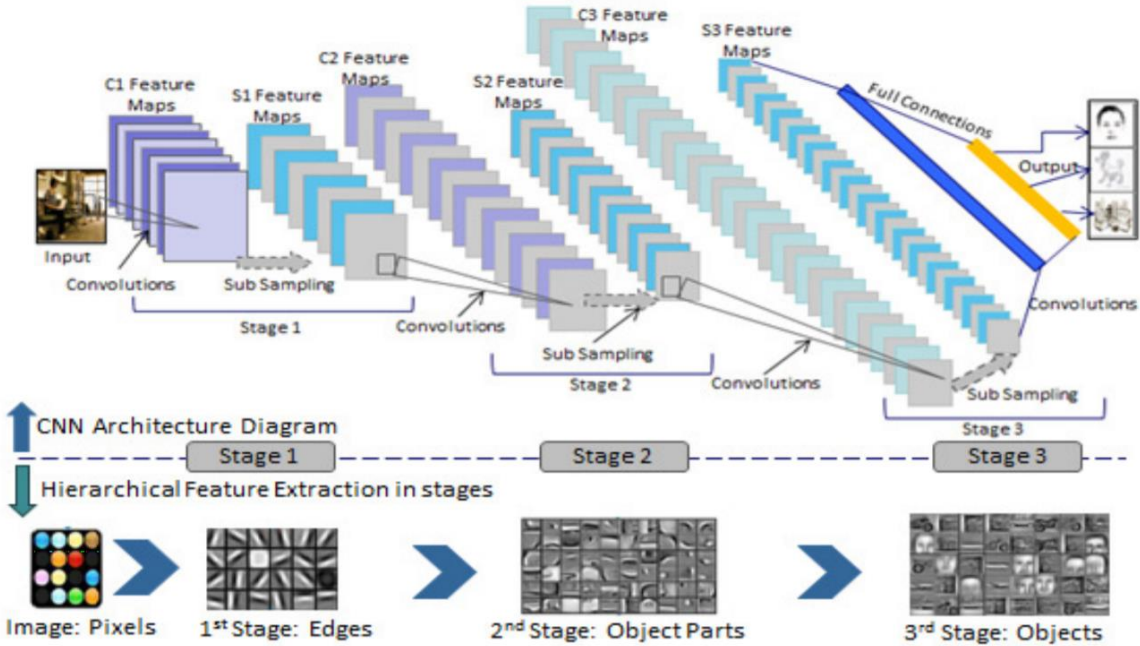
Figure 2.5: Hierarchical learning of features in CNN architecture [53].

3. PROPOSED FRAMEWORK

This chapter presents the details of the developed matching approach. The chapter begins with explaining the initial processing applied. Next, feature detection method and the iterative approach to construction of the potential matching candidates are described. Finally, three different strategies for identifying putative matches are introduced.

3.1. Initial Processing

In the first step, large scale differences between the oblique and UAV images are mitigated by applying a down-sampling process to the UAV image (Fig. 3.1). For that purpose, the "RasterResampler" function of the Feature Manipulation Engine (FME) is utilized. To do that, a down-sampling ratio (down-sampling ratio indicates the percentage of the original image resolution, see Chapter 4.4 for the details on this parameter) is defined and successively applied with using nearest neighbor interpolation to all UAV images available. In this way, scale discrepancies amongst the oblique and UAV images are reduced before any further processing.

(a) Total # of pixels: 6000x4000

(b) Total # of pixels: 1800x1200

(c) Total # of pixels: 1200x800

(d) Total # of pixels: 600x400
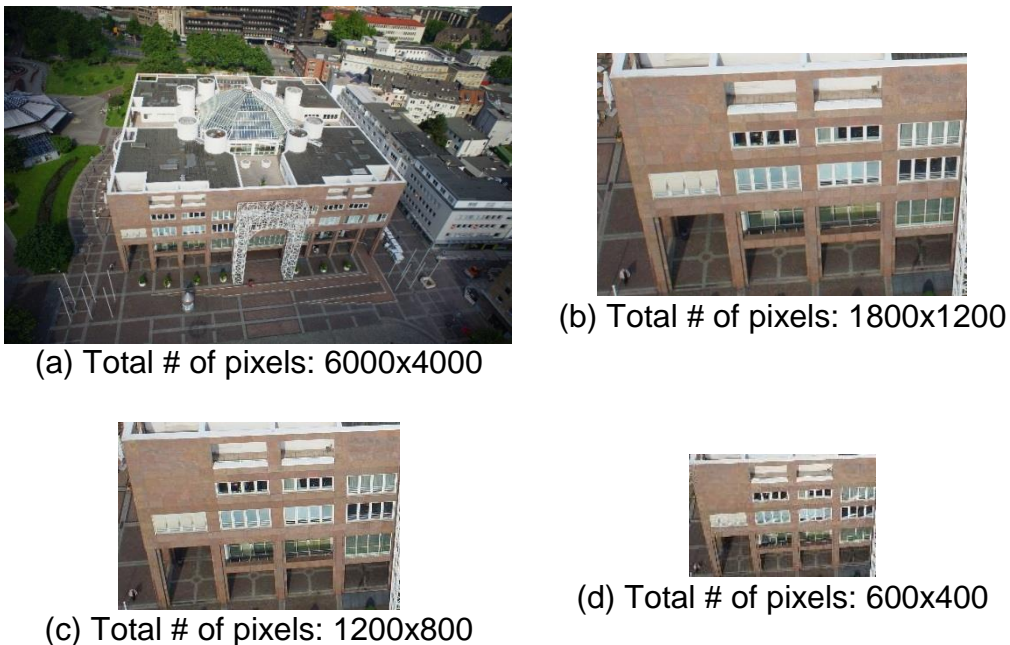
Figure 3.1: Down-sampling process applied to the UAV images. (a) UAV image at original scale, a region in UAV image with (b) down-sampling rate of 30% applied, (c) down-sampling rate of 20% applied, and (d) down-sampling rate of 10% applied.

In the second step, a search window is manually defined to limit the search space in the quite large oblique coverage compared to the relatively small area covered by the UAV images. For that purpose, for each UAV image, the region covered by the UAV images are cropped out from the aerial oblique image by taking into account the common overlap area between the two datasets (Fig. 3.2). Note that, in this thesis, it is assumed that no information is given for the related dataset, and therefore, this search window is manually selected. However, the search window can be approximately set to limit the search space within airborne oblique image with the help of the (initial) orientation information, if exists, for both datasets, as shown in [54].

(a)   (b)
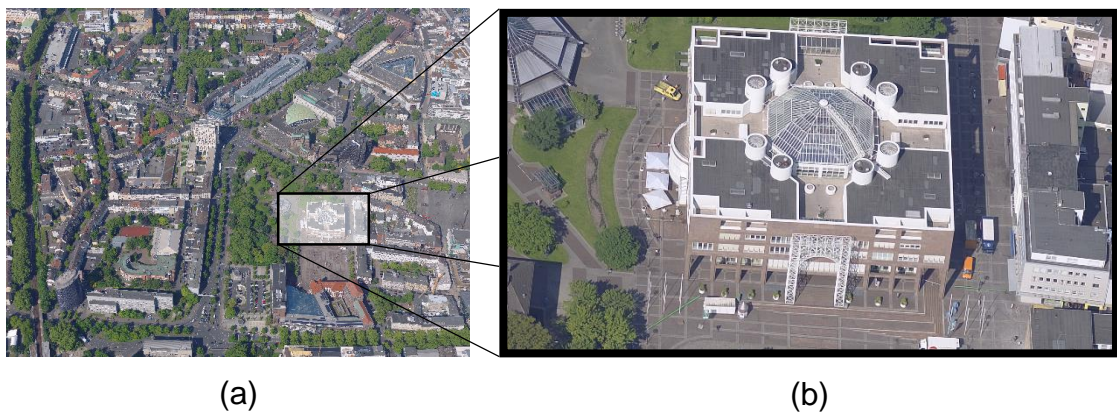
Figure 3.2:   (a) An oblique image and the scene covered by the UAV image in Fig. 3.1a, (b) cropped area from the aerial oblique images.

3.2.   Feature Detection

A detailed comparison of the performance of different point-based feature descriptors and matching of aerial oblique and UAV images were performed in [47]. In that study, the A-KAZE [39] descriptor was tested against SIFT [31], SURF [32], KAZE [38], BRISK [55], BRIEF [34], and provided the best results for all cases. Therefore, based on the results achieved in the previous study in [47], the descriptor A-KAZE is selected as the base feature detection strategy in this thesis.

A-KAZE principally makes use of KAZE features developed in [38]. However, the non-linear diffusion process utilized in KAZE relies on a computationally expensive numerical solution (Additive Operator Splitting [56]) of a large system of linear equations. For that reason, A-KAZE significantly speeds up the process by replacing the numerical solution with a faster numerical solution scheme (Fast Explicit Diffusion [57]). In the A-KAZE framework, this scheme is also implemented in a pyramidal

framework to increase the processing speed of feature detection in nonlinear scale spaces.

The scale spaces of A-KAZE are illustrated in Fig. 3.3. In KAZE, anisotropic diffusion is performed considering the input image size in all diffusion steps, and the smoothing realized after the diffusion process preserves the details, i.e. edges, corners. However, A-KAZE utilizes a pyramidal scale space (Fig. 3.3) when compared to KAZE. The sizes of the images in each pyramid level (a.k.a. octave) reduce by a factor of 2 while moving upper level of the pyramidal framework (Fig. 3.4). Similar to KAZE, the diffusion process in each level of the pyramid preserves the details during anisotropic smoothing. The anisotropic diffusion for an image $I(\mathbf{x})$ and at a time t is defined as

$$\frac{\partial I}{\partial t} = div(c(\mathbf{x}, t). \nabla I) \qquad , \qquad (3.1)$$

where c is the conductivity function, ∇ is the gradient operator, and div denotes the divergence. The function c controls the diffusion process all over the image by applying a function, e.g. as the one proposed in [58]:

$$c(\mathbf{x}, t) = g(|\nabla I_\sigma(\mathbf{x}, t)|) = \frac{1}{1 + \left(\frac{|\nabla I_\sigma|}{k}\right)^2} \qquad , \qquad (3.2)$$

where function ∇I_σ is the gradient of the image I in which Gaussian smoothing is applied, σ and k denote the standard deviation of Gaussian smoothing, and the contrast factor that manages the extent of the diffusion applied, respectively. A-KAZE also includes a histogram thresholding method (70% percentile of the gradient histogram) to obtain automatically the contrast factor k.
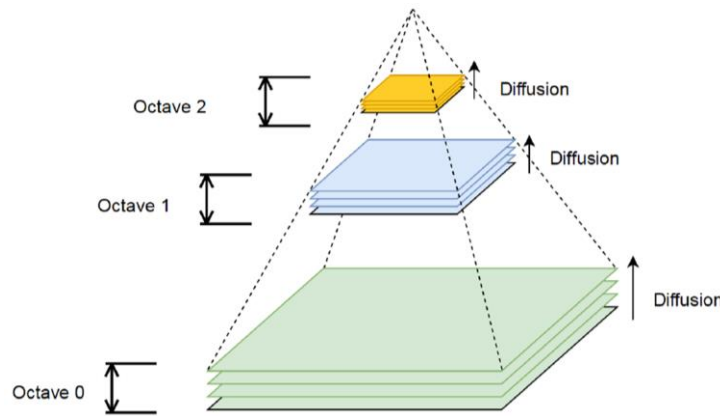
Figure 3.3:   The scale spaces of A-KAZE consisting 3 octaves and 4 sub-levels [59].
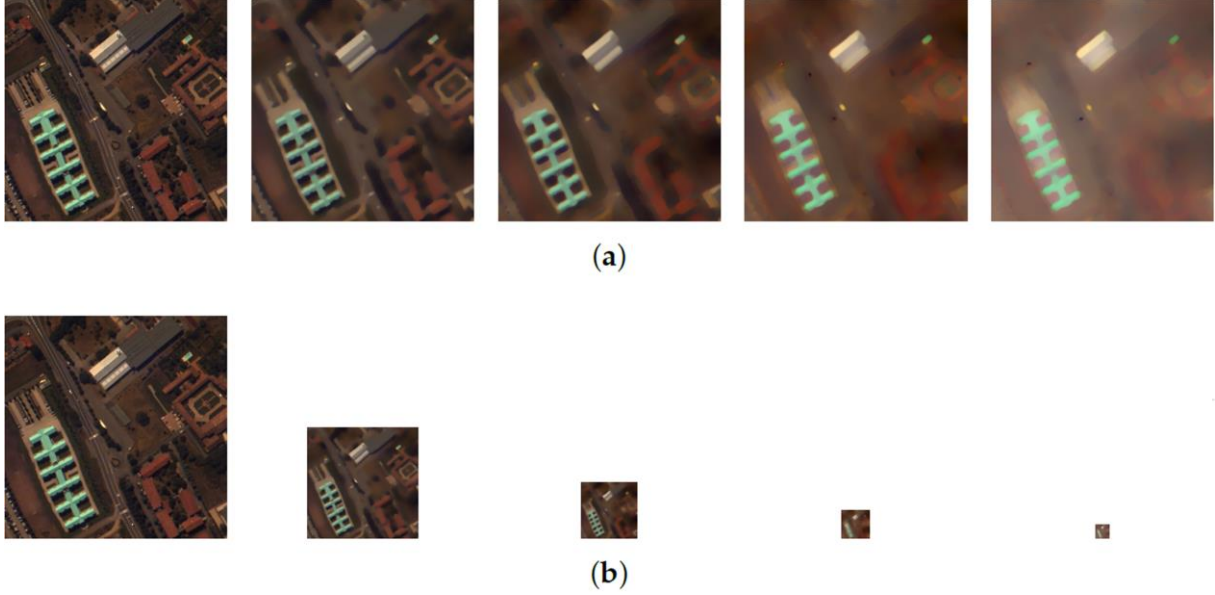
Figure 3.4:   The difference between the scale spaces of (a) KAZE, and (b) A-KAZE. Note that all images represent only the first sub-level of each octave [60].

In the next step A-KAZE detects the keypoints (or interest points), which are points that will be used for the matching step. Keypoints may have certain requirements, e.g. independence of position, robustness against image transformations, and scale independence [60]. A-KAZE extracts keypoints with the help of the determinant of the Hessian matrix (H) at the different scale levels (i) including the second order derivatives (I_{xx}, I_{yy}, I_{xy}) of the input image I:

$$I_H^i = \sigma_{i,norm}^2\left(I_{xx}I_{yy} - I_{xy}^2\right) \qquad , \qquad (3.3)$$

where *norm* denotes the normalized smoothing based on the current state of the sub-level of the octave processed. Thereafter, all keypoints are collected by thresholding the Hessian matrix at each level through a user-defined threshold (a.k.a. detector response threshold, t_{dr}). Finally, A-KAZE utilizes a modified version of Local Difference Binary (LDB) [61] that exploits 3 bits per comparison (two gradient orientations for x and y, and one intensity information) to generate the descriptor. Compared to LDB, modification includes additional sampling information from nonlinear scale space, and also estimating the main orientation rotation and accordingly rotating the grid of LDB (Fig. 3.5).

(a) LDB binary test (b) M-LDB binary test
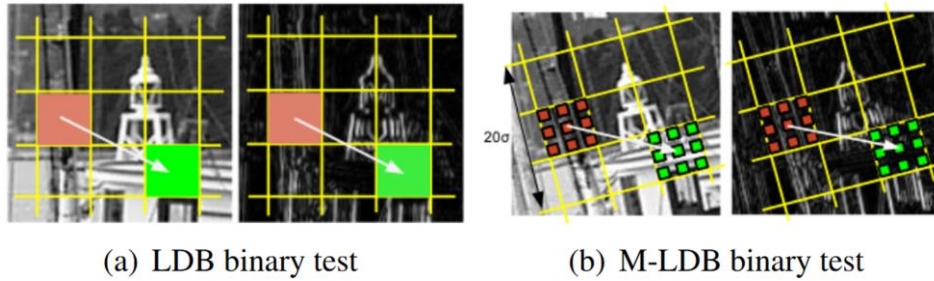
Figure 3.5: (a) LDB, and (b) modified-LDB [39].

3.3. Iterative Approach to Construct Tentative Matches

In this stage, the main tasks are (*i*) to increase the number of unique point-based features and (*ii*) to locate tentative matches as many as possible. For that purpose, two strategies are promoted, and finally, an iterative scheme is proposed.

To accomplish the first task, i.e. increasing the number of unique point-based features, it is realized that more unique keypoints can be detected when different values of the detector response threshold (t_{dr}) parameter of the A-KAZE are defined. Therefore, multiple threshold values are defined (see Chapter 4.4 for the details on this parameter), and applied for the range of the detector response threshold through an iterative processing. The threshold values are increased through an increment parameter (t_d) which is identified after rigorously evaluating a range of different threshold values.

For the second task, i.e. retrieving the tentative matches as many as possible, relaxed Nearest Neighbor Distance Ratio (NNDR) values are tested. The NNDR matching technique also known as Lowe's ratio test, was proposed by D. Lowe [31]. NNDR is relying on the fact that the correct matches need to have the closest matching similarity significantly closer than the closest incorrect match to accomplish reliable matching. For false matches, there will likely be several other false matches within comparable matching similarities [31] (see the details of NNDR computation in Appendix A.1). Therefore, utilizing low NNDR values (e.g. ≤ 0.7) lead to rejecting a number of true matches together with the mismatches; thus, preserving only a small number of matches in the output. In this thesis, different NNDR values are evaluated to maximize the number of initial matches (see Chapter 4.4 for the details on this parameter).

During the construction of tentative matches, for each different detector response threshold, A-KAZE keypoints are detected from the aerial oblique and UAV images, independently. Thereafter, the detected keypoints in both datasets are compared and evaluated using the Hamming distance [62]. Hamming distance efficiently compares every descriptor in the oblique aerial image with the descriptors found from the UAV image using XOR instruction on bit sets, in which the matching is performed.

The left-right consistency check, also known as cross check test, is also applied to filter erroneous matches found. It is based on the mutual matching hypothesis: only the keypoints whose left-to-right (i.e. oblique-to-UAV) and right-to-left (i.e. UAV-to-oblique) matching correspondences stand for the same best matching descriptors in all possible matching candidates can be accepted as valid matching keypoint pairs. Therefore, left-right consistency check removes the inconsistent feature matches amongst the input images (see the details of left-right consistency check computation in Appendix A.2).

Once the validation of feature matches is completed, the matches approved are appended to the initial matching list. After all the required iterations are completed, only the unique matches between the oblique and UAV images are exposed, and finally labeled as tentative matches (Fig. 3.6).

The pseudo-code to construct tentative matches is provided in Algorithm 1.

| (a) | (b) |
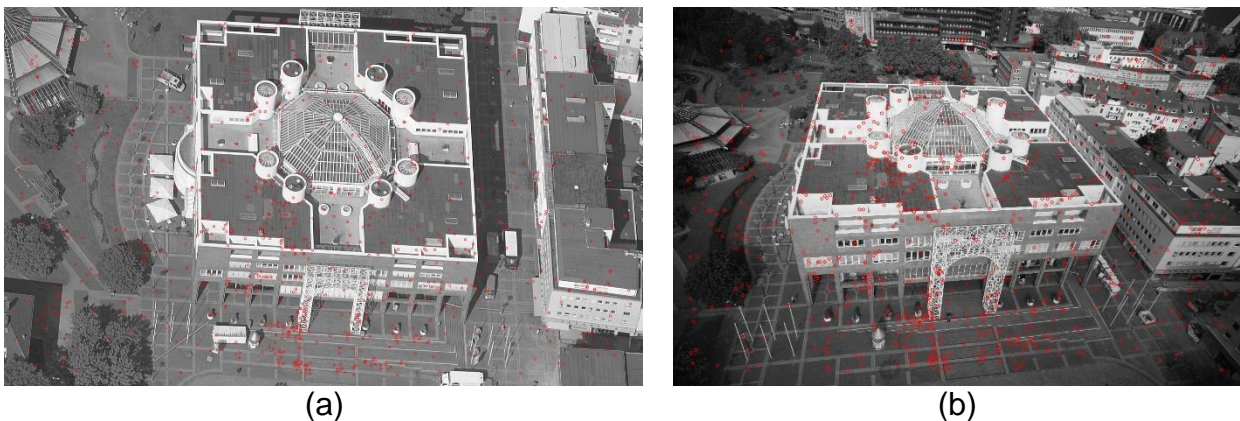
Figure 3.6:   The detected features (in red color). (a) Oblique image #5 and (b) UAV image #1.

Algorithm 1: Iterative Approach to Construct Tentative Matches

---

Input One oblique and one UAV image

Output Initial matching list between the oblique and the UAV images

1. Initialize A-KAZE algorithm parameters and define detector response threshold value (d) and the increment level (t_d)

2. Initialize NNDR value (ratio)

3. Initialize empty matching list

4. **for** *each detector response threshold ($d+t_d$)*

    Detect A-KAZE keypoints for input images using the current response threshold

    Perform feature matching between images using Hamming distance

    Apply left-right consistency check to the matches found

    Validate matches with NNDR threshold

    Append validated matches through updating the matching list

5. Find unique matches in the matching list and label them as tentative matches

---

3.4. Identifying Putative Matches

In this section, three different strategies are implemented to extract putative matches from the tentative matching list, and as a common check for all strategies, epipolar constraint is applied. However, the fundamental matrix (F) (the matrix that encodes the epipolar geometry between any two images of the same scene) relating the corresponding points between the oblique and UAV images is required to enforce the epipolar constraint between two images. If fundamental matrix is known, given a point in oblique image, the epipolar line which the point lies on in the UAV image (or vice versa) can be computed by multiplying the point with the fundamental matrix (see the details of epipolar constraint computation in Appendix A.3). However, the fundamental matrix is commonly not available, and must be calculated from an unreliable correspondence set, i.e. tentative matches, including a large number of false matches. Therefore, robust approaches are necessary to pick out a reliable correspondence set within the tentative matching list.

The main motivation of the strategies implemented is that the fundamental matrix in each strategy is estimated via several robust approaches available in different libraries

of the Python; thus, a comparative performance investigation amongst different approaches becomes expressive. In this thesis, we preferred RANSAC, LMedS & RANSAC, and Graph-Cut RANSAC approaches for a comparison. However, other estimators such as MSAC (M-estimator SAmple and Consensus) [63] and MLESAC (Maximum Likelihood Estimation SAmple and Consensus) [64] are also popular in the literature. Note also that, although MLESAC approach is not exclusively tested in this thesis, the unary term of Graph-Cut RANSAC relies on the MLESAC approach.

In this thesis, all matching points are equally weighted during the estimation of the fundamental matrix. However, the fundamental matrix can also be computed by providing different weights for the matching points (e.g. higher weights can be assigned to the points in the central part of the oblique image than the points closer to the horizon part of the oblique image since they are more reliable). However, such a weight assignment requires the image orientations beforehand, and in this thesis, it is assumed that no information is given for the related datasets.

3.4.1. Strategy #1 (RANSAC)

In this strategy, fundamental matrix is estimated using Random Sampling Concensus (RANSAC) through "findFundamentalMat" function of the OpenCV library (see the details of RANSAC in Appendix A.4).

A critical parameter, maximum distance (d_{max}), of the epipolar constraint controls the inlier/outlier decision of this strategy. This user-based input parameter (see Chapter 4.4 for the details on this parameter) defines the maximum orthogonal distance from a point to an epipolar line in pixels. If the calculated distance is higher than this value for a point, the point is considered as an outlier; thus, it is not used during the computation of the fundamental matrix.

The output is a mask array of N (size of the points from the first image) elements, and in this array, all inliers and outliers points are set to 0 and 1, respectively. Thus, this information can be utilized to select the putative matches amongst the tentative matching list.

Algorithm 2: Strategy #1 (RANSAC) to Identifying Putative Matches

---

Input Tentative Matching List

Output Putative matches between the oblique and the UAV images

1. Estimate F through RANSAC approach
2. Define maximum distance threshold (d_{max}) for epipolar constraint
3. **for** *each matching candidate in input*

    Apply epipolar constraint with the threshold, d_{max}

    Return a mask array
4. Find inliers in the matching list using mask array
5. Label inlier matches as putative matches

3.4.2. Strategy #2 (LMedS + RANSAC)

In this strategy, an initial fundamental matrix is estimated with the Least Median of Squares (LMedS) approach (see the details of LMedS in Appendix A.5), based on the hypothesis that 50% or more points in the input tentative matching list belongs to inliers. According to this initial fundamental matrix, epipolar constraint is enforced, and mask array whose values refer to early separation of inliers/outliers are generated. Next, the tentative matching list is updated based on the matches labelled as inliers, and the fundamental matrix is re-estimated with the RANSAC approach. "findFundamentalMat" function of the OpenCV library is also employed for estimation methods utilized in this strategy.

Algorithm 3: Strategy #2 (LMedS + RANSAC) to Identifying Putative Matches

---

Input Tentative Matching List

Output Putative matches between the oblique and the UAV images

1. Estimate F through LMedS approach
2. Define maximum distance threshold (d_{max}) for epipolar constraint
3. **for** *each matching candidate in input*

    Apply epipolar constraint with the threshold, d_{max}

    Return a mask array

    Find inliers in the matching list using mask array

4. Re-estimate the F using the inliers found after step 3 through RANSAC approach

5. Utilize the same maximum distance threshold (d_{max}) for epipolar constraint

6. **for** *each matching candidate in updated matching list*

      Apply epipolar constraint with the threshold, d_{max}

      Return a mask array

7. Find inliers in the matching list using mask array

8. Label inlier matches as putative matches

3.4.3. Strategy #3 (GC-RANSAC)

In this strategy, fundamental matrix is estimated using Graph-Cut RANSAC (GC-RANSAC) through "findFundamentalMat" function of the OpenCV "pygcransac". Graph-Cut RANSAC method benefits from a maximum likelihood process, i.e. MLESAC, and considers the point-to-point proximity (see the details of GC-RANSAC in Appendix A.6). Similar to the previous strategies, epipolar constraint is forced; a mask array is created, and this information is exploited to select the putative matches amongst the tentative matching list.

Algorithm 4: Strategy #3 (GC-RANSAC) to Identifying Putative Matches

Input Tentative Matching List

Output Putative matches between the oblique and the UAV images

1. Estimate F through GC-RANSAC approach

2. Define maximum distance threshold (d_{max}) for epipolar constraint

3. **for** *each matching candidate in input*

      Apply epipolar constraint with the threshold, d_{max}

      Return a mask array

4. Find inliers in the matching list using mask array

5. Label inlier matches as putative matches

---

4. DATASET AND EXPERIMENTS

This chapter presents the study area, and related dataset, the assessment strategy, and the evaluations related to the experiments conducted. The dataset utilized in this thesis are oblique and UAV images acquired over a part of Germany. The rigorous evaluations of the critical parameters required by the strategies implemented are also presented.

4.1. Study Area

The study area is affiliated with the "Image Orientation" benchmark, provided by the International Society for Photogrammetry and Remote Sensing (ISPRS) Scientific Initiative and EuroSDR[1], and covers a part of Dortmund, Germany. The aim of the benchmark is twofold [65]: (i) to evaluate and test the performance of different approaches developed for the calibration and orientation of images acquired by different platforms, and (ii) to assess the performance of different approaches developed for image matching and dense point cloud generation. In the context of the benchmark, three different test areas made available for researchers, Dortmund-Centre (Germany), Dortmund, Zeche Zollern (Germany), and Zurich (Switzerland). In this thesis, the area covering the Dortmund-Centre is selected to evaluate the strategies implemented. This area is composed of dense urban environment containing tall modern buildings together with the historical constructions such as churches and public buildings (Fig. 4.1). Frame in yellow color in Fig. 4.1 defines the entire area of the Dortmund-Centre benchmark.

4.2. Dataset

The yellow boundary in Fig. 4.1 defines the entire region acquired by the airborne oblique image acquisition system IGI PentaCam, whereas the rectangular region in red frame illustrates the region covered by the UAV flights. The small region labels the

---

[1] EuroSDR is a not-for-profit organization linking National Mapping and Cadastral Agencies with Research Institutes and Universities in Europe for the purpose of applied research in spatial data provision, management and delivery.

Dortmund town hall, the place where the oblique and UAV images in this region are exploited in this thesis:

Figure 4.1:   Dortmund-Centre study area [65]. Rectangles with yellow and red boundaries define the area of the oblique and UAV campaigns.

- **Oblique Images:** A total 1260 images were acquired with Pentacam IGI (flown by AeroWest - Dortmund) on May 19th, 2014. The ground sampling distance is 10 cm in nadir images and it varies from 8 to 12 cm for the oblique views. The overlap for the nadir images is 75%-80% at along/across-track directions while it is 80% for oblique images [65].

In this thesis context, a single side (from the west view) of the Rathaus (town hall) building (also one of the benchmark buildings in the area) is purposefully selected to evaluate the performance of the strategies implemented. The main reason for selecting this building is that it is composed of many different repetitive patterns which turn into a challenging scenario for any kind of feature matching strategy. Both the selected oblique images and the benchmark Rathaus building (within the red colored rectangles) are visible in Fig. 4.2.

- **UAV Images:** A total of 1073 images were acquired with a multi-rotor DIJ S800 on June 8th, 2014. Images were taken with the drone cover the Rathaus (town hall) and Stadthaus buildings together with the obelisk and the surrounding square area. Both nadir and oblique images were collected on the area during 4 different flights and an image GSD varying from 1 to 3 cm was adopted in each flight. A Sony Nex-7 camera was mounted on the drone [65].

Two UAV images utilized in this thesis are shown in Fig. 4.3.

Image ID: '005_018_148000201'

(a) Oblique Image #1
(Image size 1159 x 558 pixels)

Image ID: '005_017_148000202'

(b) Oblique Image #2
(Image size 1267 x 648 pixels)
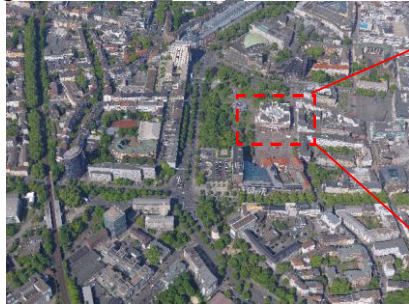
Image ID: '005_016_148000203'

(c) Oblique Image #3
(Image size 1397 x 732 pixels)

Image ID: '005_015_148000204'

(d) Oblique Image #4
(Image size 1553 x 854 pixels)

Image ID: '005_014_148000205'

(e) Oblique Image #5
(Image size 1744 x 1032 pixels)

Figure 4.2:   (a-e) Aerial oblique images (from West view). The red colored regions illustrate the benchmark Rathaus building.

(a) ID: '7111' – UAV Image #1   (b) ID: '7148' – UAV Image #2

(Image size 6000 x 4000 pixels)   (Image size 6000 x 4000 pixels)

Figure 4.3:   Selected UAV images of the Rathaus building.

4.3.   Assessment Strategy

In order to evaluate the accuracies of different matching strategies, the computed putative matches in Section 3 are divided into two categories, True Positive (TP), and False Positive (FP), by checking the automatically matched point features with the aid of manual inspection. If a feature match correctly represents the same point between the oblique and UAV images, it is labeled as a True Positive. Analogously, a False Positive is a feature match that does not indicate any correct correspondence during the visual evaluation. Thus, in this thesis, the well-known Precision (a.k.a. correctness) measure is utilized to interpret the quality of the matching:

$$Precision \ = \ \frac{\|TP\|}{\|TP\| + \|FP\|}$$ (Eq. 4.1)

where $\| . \|$ indicates the number of matches assigned to each class. As well known, another common measure frequently used to evaluate the performance is the number of False Negatives (FNs). This measure provides important information about a feature match that exists in reality but cannot be found by the automated matching strategies. However, due to the iterative approach (e.g. different detector response thresholds $d+t_d$, see Algorithm 1), a large number of unique point-based keypoints are found and computed (see Fig. 3.6), and unfortunately, it become impossible to generate a manually prepared reference matching list that can be used to infer FNs from the results.

4.4. Parameter Evaluation

The parameters necessary to initialize the proposed strategies are presented in Table 4.1. A number of tests on different parameters are performed to select the best parameter configuration, and the effects of each parameter on the matching performances of oblique images (#1 - #5) and UAV image #1 are investigated using the measures defined in (Eq. 4.1). Thereafter, oblique images (#1 - #5) and UAV image #2 are tested with the parameters fixed. In this thesis, only the main parameters are evaluated, whereas certain values that would normally not be changed for different datasets remain fixed, see Table 4.1.

Table 4.1:   Parameter test settings defined for the strategies implemented.

Stage	Parameters	{Tested} / "Selected"
3.1 Initial Processing	Down-sampling ratio (r_{ds}) (%)	{0, 10, "20", 30}
3.2 Feature Detection	Descriptor size	"486" bits
	# of Channels in the descriptor	"3"
	Multiscale detection factor (# of Octaves)	"4" (default)
	Scale levels (# of Octave Layers)	"4" (default)
	Diffusivity type	Perona-Malik [58]
3.3 Construction of Tentative Matches	Increment parameter (t_d)	"5×10^{-4}"
	A-KAZE detector response threshold value (t_{dr})	"1×10^{-7}, 5×10^{-4}, 10×10^{-4}, 15×10^{-4}, 20×10^{-4}"
	Nearest Neighbor Distance Ratio (r_{nnd})	{0.6, 0.65, 0.7, 0.75, 0.8, "0.85"}
3.4 Identifying Putative Matches	Maximum distance to epipolar line in pixels (d_{max})	{1, 2, "3"}
	Level of confidence (probability)	"99%" (default)

The first test investigates the effect of the scale differences between the oblique and UAV images. During the tests, first, oblique and UAV images are matched with their original scales, without applying any down-sampling process (i.e. $r_{ds} = 0$). The results presented in Table 4.2 indicate that the scale differences between the oblique and UAV images during matching are critical; thus, all strategies reported only a small number of feature matches along with a number of FPs. The results achieved with down-

sampling rates of 10% and 20% demonstrate significant increases of TPs (Tables 4.3 and 4.4). However, the results presented in Table 4.5 confirm that the down-sampling rate of 30% noticeably reduce the precision ratios computed. Note also that, the numbers of FPs remain more or less on the same level for all experiments within each strategy (i.e. #1, #2, and #3) implemented. According to these results, the down-sampling rate is fixed to 20% since this ratio provide the highest number of TPs while generating comparable number of FPs.

Table 4.2:   UAV image is not down-sampled (i.e. rate 0%). The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	19	4	82.60	22	4	84.62	85	18	82.52
Pair 2-1	18	4	81.82	23	2	92.00	88	29	75.21
Pair 3-1	28	2	93.33	31	2	93.94	106	28	79.10
Pair 4-1	18	1	94.74	17	3	85.00	94	26	78.33
Pair 5-1	10	4	71.43	5	7	41.67	52	21	71.23

Table 4.3:   UAV image is down-sampled at rate 10%. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	69	2	97.18	67	6	91.78	85	15	85.00
Pair 2-1	53	2	96.36	52	4	92.86	77	12	86.52
Pair 3-1	44	7	86.27	39	4	95.12	65	9	87.84
Pair 4-1	26	5	83.87	28	4	87.50	46	11	80.70
Pair 5-1	11	7	61.11	10	9	52.63	15	13	53.57

Table 4.4:   UAV image is down-sampled at rate 20%. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	209	9	95.87	241	1	99.59	272	12	95.77
Pair 2-1	213	4	98.16	245	7	97.22	276	14	95.17
Pair 3-1	227	3	98.70	282	2	99.30	322	10	96.99
Pair 4-1	280	9	96.89	311	8	97.49	360	12	96.77
Pair 5-1	117	5	95.90	187	2	98.94	226	13	94.56

Table 4.5:   UAV image is down-sampled at rate 30%. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	124	11	91.85	178	5	97.34	254	21	92.36
Pair 2-1	162	4	97.59	204	5	97.61	317	22	93.51
Pair 3-1	158	11	93.49	175	9	95.11	302	23	92.92
Pair 4-1	100	6	94.34	138	5	96.50	334	28	92.27
Pair 5-1	115	10	92.00	127	5	96.21	343	19	94.75

The descriptor size of A-KAZE is set to full size, 486 bits, as this descriptor size slightly provides the best results for matching in overall sense [39]. The number of channels in the descriptor is set to 3 (intensity, x derivative, and y derivative), as this input is shown to significantly improve the matching precision when compared with 1 channel case, i.e. the intensity [39]. For the parameters related to feature detection, we preferred to utilize default values: multiscale detection factor and scale levels are both set to 4, and for the diffusivity parameter, one of the two conductivity functions introduced in [58] is utilized.

During the construction of tentative matches, the aim is to increase the number of unique keypoints with different values of the detector response thresholds (t_{dr}). We performed repeated and varied attempts to select the best values. According to our investigation, we first initialize the t_{dr} parameter with a very small value (i.e. 1×10^{-7}) to focus on weak evidences as descriptor. Next, we increment this threshold using $t_d = 5 \times 10^{-4}$, up to a certain value (i.e. 20×10^{-4}) which the number of unique matches found on images dramatically reduces to a minor level (see Fig. 4.4).
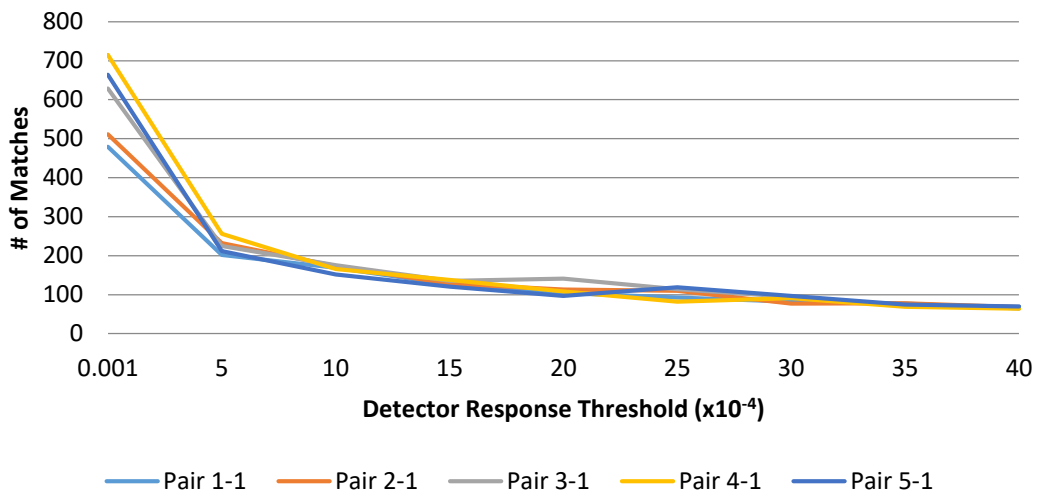
(a)
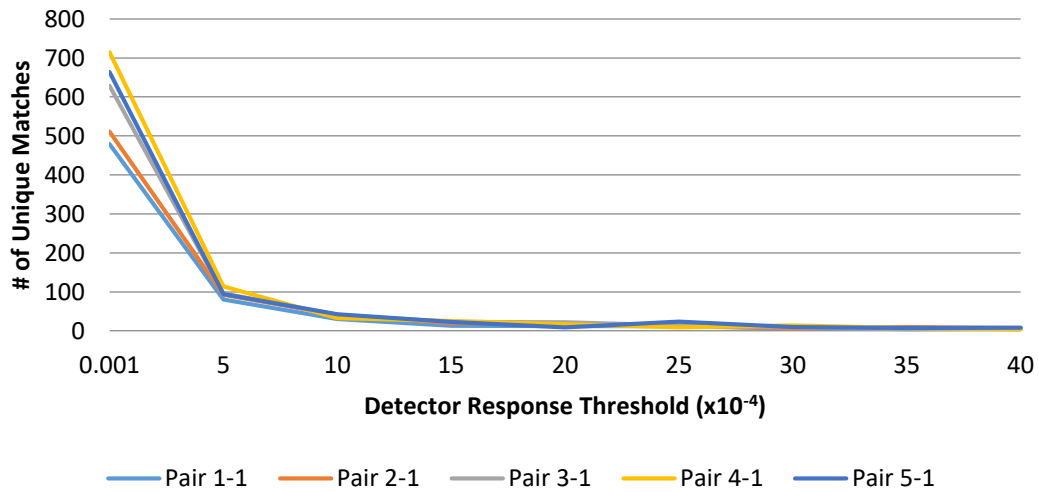
(b)

Figure 4.4: (a) Number of matches, and (b) number of unique point matches found during the construction of tentative matches for different detector response thresholds (t_{dr}).

The numbers of TPs and FPs for different Nearest Neighbor Distance Ratio (r_{nnd}) thresholds during the generation of tentative matches are presented in Tables 4.6 - 4.10. As clearly observable from Table 4.6, a threshold of 0.6 provides perfect scores for matching by completely eliminating all FPs. However, this threshold throws out a large number of TPs as well, e.g. check the number of TPs found with a relaxed threshold in Table 4.10. However, for this scenario, discarding a large number TPs allows computing a fundamental matrix, F, that solely localized to certain part of the image. Thus, this significantly reduces the possibility of computing a reliable fundamental matrix F considering the whole image (see the Section 5 for more information on this issue), whose accurate computation might be very useful for different subsequent tasks to be done after feature matching, e.g. orientation, edge matching, dense image matching etc. During our experiments, for r_{nnd} threshold values larger than 0.85, FPs are found to increase noticeably in the whole image domain. Therefore, in this thesis, r_{nnd} threshold value is fixed to 0.85 for all experiments (Table 4.4).

Table 4.6:   Nearest Neighbor Distance Ratio (r_{nnd}) threshold value set as 0.60 in each method. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	45	0	100	42	0	100	46	0	100
Pair 2-1	40	0	100	34	0	100	43	0	100
Pair 3-1	43	0	100	37	0	100	46	0	100
Pair 4-1	35	0	100	30	0	100	35	0	100
Pair 5-1	23	0	100	23	0	100	28	0	100

Table 4.7: Nearest Neighbor Distance Ratio (r_{nnd}) threshold value set as 0.65 in each method. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	68	0	100	54	1	98.18	72	0	100
Pair 2-1	64	0	100	61	0	100	65	0	100
Pair 3-1	72	0	100	64	0	100	83	0	100
Pair 4-1	62	1	98.41	64	0	100	73	0	100
Pair 5-1	44	1	97.78	36	1	97.30	44	2	95.65

Table 4.8: Nearest Neighbor Distance Ratio (r_{nnd}) threshold value set as 0.70 in each method. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	91	0	100	81	0	100	99	0	100
Pair 2-1	89	0	100	82	0	100	93	1	98.94
Pair 3-1	107	1	99.07	92	0	100	116	2	98.31
Pair 4-1	105	0	100	92	0	100	114	1	99.13
Pair 5-1	56	0	100	55	1	98.21	61	2	96.83

Table 4.9: Nearest Neighbor Distance Ratio (r_{nnd}) threshold value set as 0.75 in each method. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	123	0	100	121	0	100	142	2	98.61
Pair 2-1	120	1	99.17	117	0	100	134	2	98.53
Pair 3-1	162	0	100	165	0	100	177	1	99.44
Pair 4-1	166	0	100	149	0	100	185	2	98.93
Pair 5-1	84	1	98.82	86	0	100	93	3	96.88

Table 4.10: Nearest Neighbor Distance Ratio (r_{nnd}) threshold value set as 0.80 in each method. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	178	0	100	179	0	100	202	2	99.02
Pair 2-1	168	1	99.41	156	1	99.36	187	2	98.94
Pair 3-1	194	2	98.98	202	1	99.51	232	4	98.31
Pair 4-1	223	1	99.55	219	0	100	253	5	98.06
Pair 5-1	129	3	97.73	147	3	98.00	161	8	95.27

During the selection of putative matches, first, maximum distance to epipolar line in pixels (d_{max}) must be provided, and the results related to this experiment are presented in Tables 4.11 - 4.13. As strongly anticipated, increasing the d_{max} threshold also amplifies both the TPs and FPs. Nevertheless, the increase in the number of TPs is significantly higher than the increase in the number of FPs; therefore, this parameter is set to 3 pixels. Second, during RANSAC iterations, at least one outlier free set of matching list must be sampled; therefore, the probability (confidence) is set to the highest value such as 0.99 (the default).

Table 4.11: The maximum distance (d_{max}) from a point to an epipolar line is set as 1 pixel in each method. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	123	1	99.19	145	1	99.32	209	1	99.52
Pair 2-1	106	1	99.07	150	2	98.68	224	3	98.68
Pair 3-1	118	0	100	163	1	99.39	281	5	98.25
Pair 4-1	160	5	96.97	193	1	99.48	307	9	97.15
Pair 5-1	61	4	93.85	118	1	99.16	170	2	98.84

Table 4.12: The maximum distance (d_{max}) from a point to an epipolar line is set as 2 pixels in each method. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	167	5	97.09	209	1	99.52	262	7	97.40
Pair 2-1	168	3	98.25	219	2	99.10	259	6	97.74
Pair 3-1	182	2	98.91	247	1	99.60	314	6	98.13
Pair 4-1	235	7	97.11	288	1	99.65	348	10	97.21
Pair 5-1	94	6	94.00	166	1	99.40	213	13	94.25

Table 4.13: The maximum distance (d_{max}) from a point to an epipolar line is set as 3 pixels in each method. The other parameters required are set based on Table 4.1.

Pairs	Methods								
	Strategy #1			Strategy #2			Strategy #3		
	TP	FP	Precision (%)	TP	FP	Precision (%)	TP	FP	Precision (%)
Pair 1-1	209	9	95.87	241	1	99.59	272	12	95.77
Pair 2-1	213	4	98.16	245	7	97.22	276	14	95.17
Pair 3-1	227	3	98.70	282	2	99.30	322	10	96.99
Pair 4-1	280	9	96.89	311	8	97.49	360	12	96.77
Pair 5-1	117	5	95.90	187	2	98.94	226	13	94.56

The contribution of the proposed iterative strategy during the construction of tentative matches is seen in Table 4.14. As we see from the table, original A-KAZE approach presented in [39] (with NNDR thresholding) results in a satisfactory number of tentative matches ranging from 261 to 471. However, if we apply the left-right consistency check (see the details of left-right consistency check computation in Appendix A.2) to remove erroneous matches amongst these results, this constraint removes a significant number of (incorrect) matches (ranging between 48% - 75%). However, thanks to the Algorithm 1 presented in this thesis, the total number of tentative matches are considerably increased ($\approx$ 2x compared to A-KAZE results in [39]). These results clearly

prove the importance of the proposed iterative strategy performed with different detector response thresholds in Section 3.3.

Table 4.14: The contribution of the proposed iterative strategy during the construction of tentative matches. The parameters required for Algorithm 1 are set based on Table 4.1.

Pairs	Total # of Tentative Matches		
	Results of A-KAZE [39] (with NNDR)	Results of A-KAZE [39] (with NNDR) and Left-Right Consistency Check	Results of the Algorithm 1 (A-KAZE + NNDR + Left-Right Consistency Check)
Pair 1-1	261	136 (eliminated ≈ 48%)	614 (increased ≈ 235%)
Pair 2-1	325	147 (eliminated ≈ 55%)	683 (increased ≈ 210%)
Pair 3-1	348	142 (eliminated ≈ 59%)	811 (increased ≈ 233%)
Pair 4-1	407	125 (eliminated ≈ 69%)	904 (increased ≈ 222%)
Pair 5-1	471	117 (eliminated ≈ 75%)	833 (increased ≈ 177%)

The implementation and processing was performed in Python version 3.7. The experiments were performed on a computer with a quad core Intel Xeon CPU @ 3.00GHz and 32 GB RAM. The number of pixels in each test image is provided in Figures 4.2 and 4.3. The time elapsed for each part of the methodology is presented in Table 4.15.

The total processing time of 5 images took approximately 50 seconds with an average processing time of nearly 10 seconds (Table 4.15). Approximately 80% of the total processing time is spent during the construction of the tentative matches during the iterative approach (cf. Section 3.3). On the other hand, the feature detection step constitutes 20% of the total processing time (cf. Section 3.2). Identification of putative matches (cf. Section 3.4) seems to the fastest step (at most 4% of the total processing time). The results indicate that our approach runs nearly 13-18 times (Table 4.16) slower than the A-KAZE method [39]. This is obviously due to reason that an iterative framework is utilized during the generation of tentative matches which eventually increases the number of candidate correspondences to be matched.

Table 4.15: Elapsed time of each section of the proposed approach (the average processing times of all matching pairs (#1-#5) are calculated).

Process (Section)	Elapsed Time seconds (%)
Feature Detection (Section 3.2)	1.94 (≈19-20%)
Iterative Approach to Construct Tentative Matches (Section 3.3)	7.76 (≈77-79%)
Identifying Putative Matches (Section 3.4)	
• Strategy #1 (RANSAC) • Strategy #2 (LMedS + RANSAC) • Strategy #3 (GC-RANSAC)	0.09 (≈1%) 0.11 (≈1%) 0.36 (≈4%)
Total Processing	
• Section 3.2 + Section 3.3 + Strategy #1 • Section 3.2 + Section 3.3 + Strategy #2 • Section 3.2 + Section 3.3 + Strategy #3	9.79 (%100) 9.81 (%100) 10.06 (%100)

Table 4.16: The processing time required during the construction of tentative matches.

Pairs	Total Processing Time Required (seconds)		
	A-KAZE [39] (with NNDR)	A-KAZE [39] (with NNDR) and Left-Right Consistency Check	Algorithm 1 (A-KAZE + NNDR + Left-Right Consistency Check)
Pair 1-1	0.44	0.46	5.95
Pair 2-1	0.52	0.54	8.32
Pair 3-1	0.58	0.60	9.49
Pair 4-1	0.69	0.72	12.32
Pair 5-1	0.79	0.87	15.73

5. RESULTS & DISCUSSION

This chapter provides all information related to the results of the proposed matching framework. First, the numerical and visual results are presented. Next, the reasons related to the false positive matches are discussed in detail. Finally, the results computed are compared with the matching approaches selected from the literature considering the same test area.

5.1. Results for Oblique Images (#1 - #5) vs. UAV Image #1

The overall results computed for the matching of oblique images (#1 - #5) and UAV image #1 are presented in Table 5.1. According to Table 5.1, the matching results of the novel feature matching framework seem to be robust, and the final results are convincing. According to the numerical results presented, the precision ratios range between 94.56% and 99.59%. The best precision ratios in an overall sense are computed for Strategy #2 in which the Least Median of Squares (LMedS) and RANSAC approaches are successively utilized. In that case, the worst FP is computed for the pair *1-4* with a total of 8 FPs. Additionally, all FPs detected using Strategy #2 are comparatively lower than the number of FPs detected for the other two strategies (expect for the pair *1-2*, Strategy #1).

On one hand, the best results are achieved for the pair *3-1* when the results of all strategies are taken into account. This is an expected result since image patch cropped out from the large oblique image is the closest one to the center area (see Fig. 4.2c); and therefore, this patch is the most similar patch to the UAV image in terms of the viewing direction amongst the available five image patches. On the other hand, worst numerical results are computed for the pairs *1-1* and *5-1*. These results are also due to the viewing direction differences between the oblique and UAV images as these two combinations hold the largest viewing differences. However, even for such view point differences, all computed results provided ≈ 95% precision ratio or higher.

Strategy #3 which utilizes a Graph-Cut RANSAC procedure provides the highest number of TPs for all cases. On average, nearly 40 and 80 TP improvements are observed compared to the results of Strategies #1 and #2, respectively. In this

Table 5.1:   The results of matching between oblique images (#1-#5) and UAV Image #1.

UAV Image #1

UAV Image

Oblique Images

Oblique Image #1

Pair *1-1*	Strategy #1		Strategy #2		Strategy #3	
	TP	FP	TP	FP	TP	FP
Total #	209	9	241	1	272	12
Precision	95.87%		99.59%		95.77%	

Oblique Image #2

Pair *2-1*	Strategy #1		Strategy #2		Strategy #3	
	TP	FP	TP	FP	TP	FP
Total #	213	4	245	7	276	14
Precision	98.16%		97.22%		95.17%	

Oblique Image #3

Pair *3-1*	Strategy #1		Strategy #2		Strategy #3	
	TP	FP	TP	FP	TP	FP
Total #	227	3	282	2	322	10
Precision	98.70%		99.30%		96.99%	

Oblique Image #4

Pair *4-1*	Strategy #1		Strategy #2		Strategy #3	
	TP	FP	TP	FP	TP	FP
Total #	280	9	311	8	360	12
Precision	96.89%		97.49%		96.77%	

Oblique Image #5

| Pair *5-1*	Strategy #1		Strategy #2		Strategy #3	
	TP	FP	TP	FP	TP	FP
Total #	117	5	187	2	226	13
Precision	95.90%		98.94%		94.56%	

manner, we can state that the highest TP performance is strikingly achieved by Strategy #3. For sure, this result is also expected since this strategy benefits from a graph-cut energy minimization framework considering point-to-point proximity to perform the inlier selection of RANSAC procedure. However, as a negative side effect,

this strategy pays the price of the increase of TPs by also increasing the FPs. Nevertheless, it is believed that the numbers of computed FPs are tremendously low for a point matching scenario between an oblique and UAV image.

The numerical results given in Table 5.1 are also investigated visually by marking the matches found in both images for all strategies tested (Fig. 5.1 – 5.3). At this point, we principally analyze and present the point matches in terms of their distribution amongst the input pairs since this information is crucial to computing an accurate Fundamental Matrix (F) which represents the geometric relationship between the image pair. If a good distribution of matching is not presented, then all subsequent processing (e.g. epipolar image generation, image matching, 3D reconstruction etc.) utilizing the F matrix would become unsuccessful. Therefore, evenly distributed matching correspondences must be favored.

As seen from the results of Strategy #1 in Fig. 5.1a and 5.1b, most of the point matches found for pairs *1-1* and *2-1* are located on the rooftop of the Rathaus building. Only a small number matches belong to the other parts of the images: (i) a few of them are on the top-left corner of the oblique image, and (ii) a small number of matches are detected on the other building visible on the right-side of the Rathaus building. The scenario is slightly different for the pair *3-1* (Fig. 5.1c) where a number of correct matches are observed in front of the Rathaus building on the ground level. However, a major change is observed for the pair *4-1* (Fig. 5.1d). In that case, nearly half of the matches are found on the ground level, and a large area of the rooftop of the Rathaus building does not reveal any matches. Finally, for the pair *5-1*, almost all of the matches belong to the ground level, and as a result, the matches are accumulated only a specific part of the input images. Note also that pair *5-1* is the most challenging pair of all image pairs which presents the highest perspective differences between the oblique and UAV images.

The trend observed above is also valid for the outputs of the Strategies #2 and #3 illustrated in Fig. 5.2 and Fig. 5.3, respectively. The main difference observed is that the numbers of matches are increased significantly which eventually supports better matching distributions throughout the images. The differences between the Strategies #2 and #3 are comparable; however, it is believed that the distributions of point correspondences of Strategy #3 are slightly better due to the increase in the numbers of TP matches.
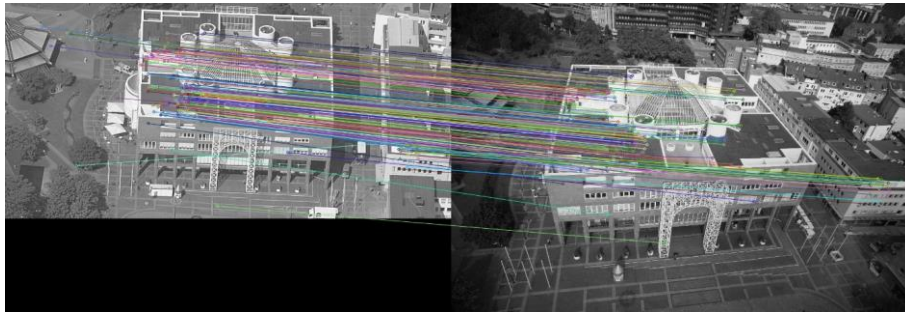
(a) Pair *1-1* – Total matches: 218 (TP: 209 – FP: 9)
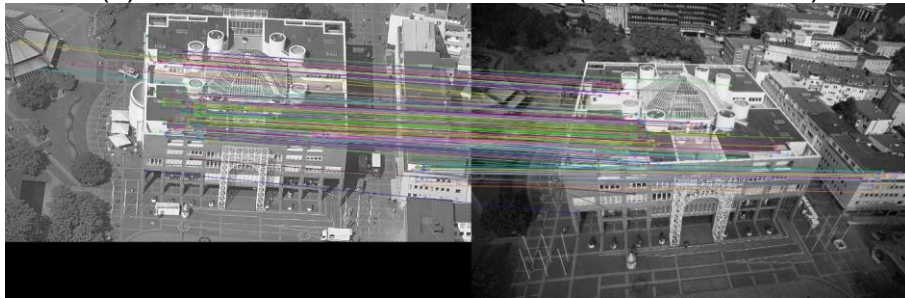
(b) Pair *2-1* – Total matches: 217 (TP: 213 – FP: 4)
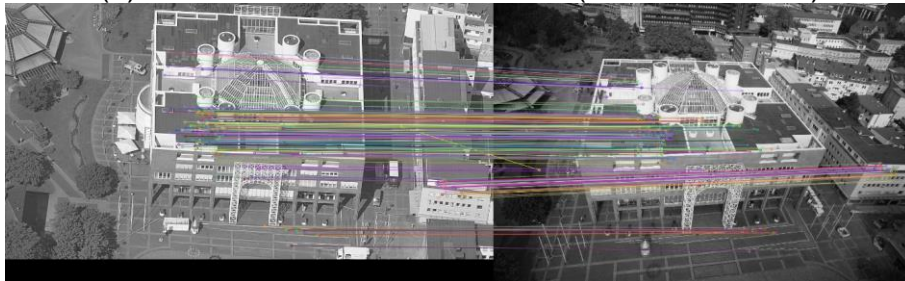
(c) Pair *3-1* – Total matches: 230 (TP: 227 – FP: 3)

(d) Pair *4-1* – Total matches: 289 (TP: 280 – FP: 9)
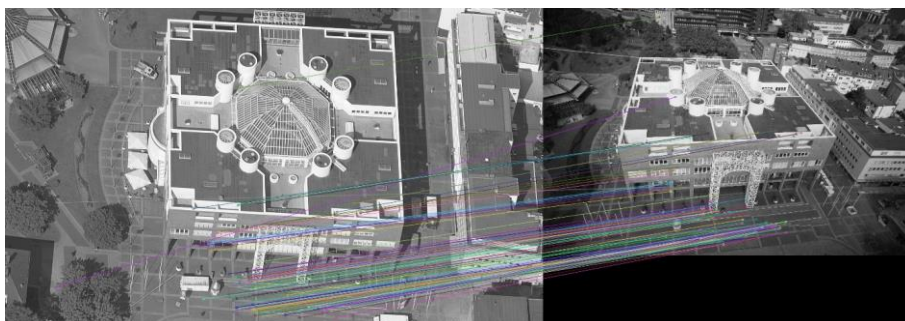
(e) Pair *5-1* – Total matches: 122 (TP: 117 – FP: 5)

Figure 5.1:   Matching results of oblique and UAV image pairs for Strategy #1.
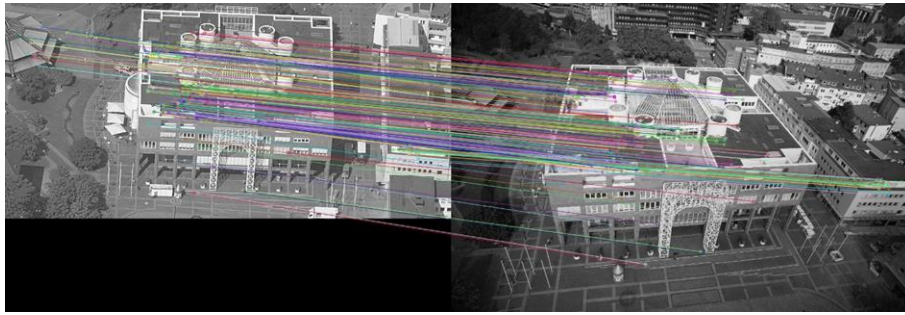
(a) Pair *1-1* – Total matches: 242 (TP: 241 – FP: 1)
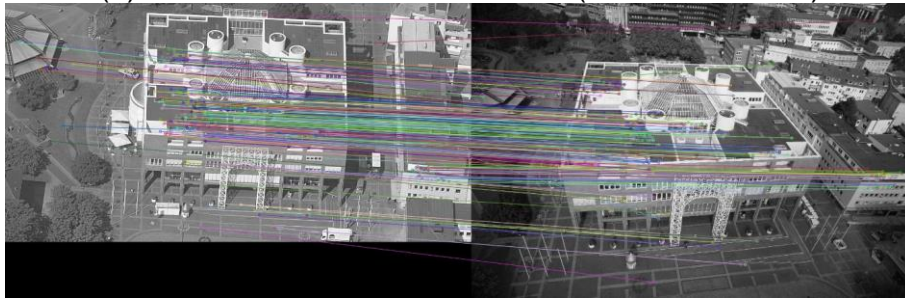
(b) Pair *2-1* – Total matches: 252 (TP: 245 – FP: 7)

(c) Pair *3-1* – Total matches: 284 (TP: 282 – FP: 2)

(d) Pair *4-1* – Total matches: 319 (TP: 311 – FP: 8)

(e) Pair *5-1* – Total matches: 189 (TP: 187 – FP: 2)

Figure 5.2:   Matching results of oblique and UAV image pairs for Strategy #2.
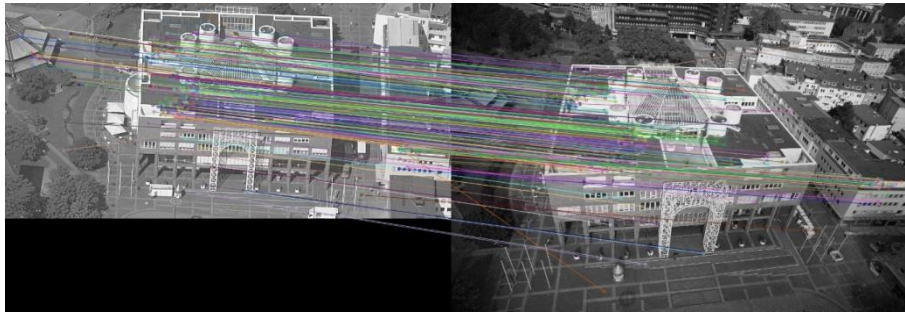
(a) Pair *1-1* – Total matches: 284 (TP: 272 – FP: 12)
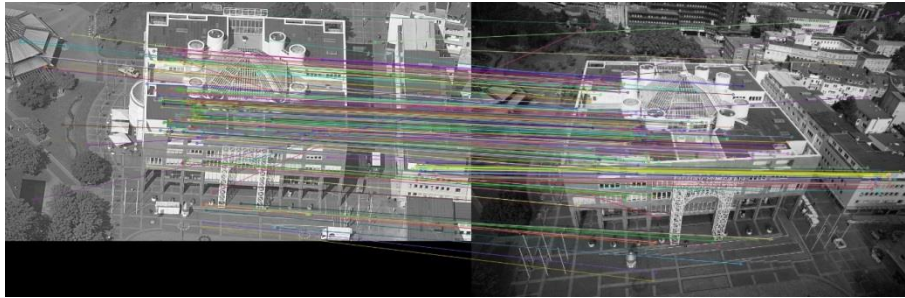
(b) Pair *2-1* – Total matches: 290 (TP: 276 – FP: 14)
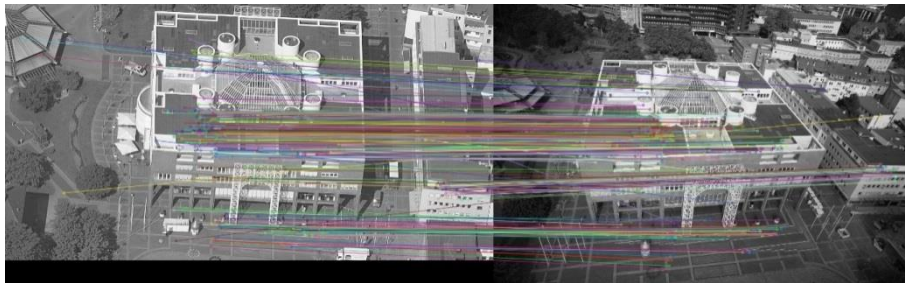
(c) Pair *3-1* – Total matches: 332 (TP: 322 – FP: 10)

(d) Pair *4-1* – Total matches: 372 (TP: 360 – FP: 12)
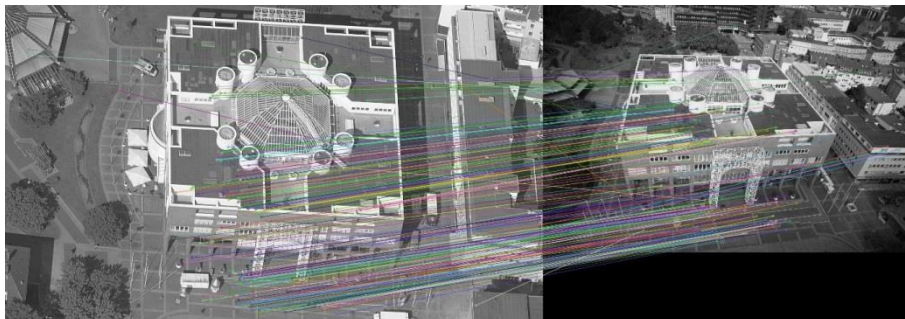
(e) Pair *5-1* – Total matches: 239 (TP: 226 – FP: 13)

Figure 5.3:   Matching results of oblique and UAV image pairs for Strategy #3.

It is also observed that fewer numbers of matches are observed on the facade of the Rathaus building in each case. Although the west-side facade of the Rathaus building has a good texture and nice contrast for point matching, most of the point matches in this context are presumably lost after the NNDR thresholding due to the repetitive design of the facade pattern.

Fig. 5.4 proves the importance of well-distributed correspondences. As seen from Fig. 5.4a, b, although 222 matches (with 117 TPs) are detected, lack of a good distribution throughout the scene caused incorrect computation of the epipolar geometry; thus the epipoles and the epipolar lines are inaccurate. However, better epipolar geometry is estimated using the outputs of the strategies #2 and #3 (see Fig. 5.4c-f).

5.2. Results for Oblique Images (#1 - #5) vs. UAV Image #2

The overall results computed for the matching of Oblique Images (#1 - #5) and UAV Image #2 are presented in Table 5.2. Similar to the previous results presented, the matching results of the novel feature matching framework are robust even for this test. The precision ratios range between 95.10% and 99.67% as given in Table 5.2. In general, once again, the best precision ratios are computed for Strategy #2. In that case, the worst FP is computed for the pair *1-2* with a total of 6 FPs. Besides, all FPs detected using the three strategies are in the same order (except for the pair *5-2*, Strategy #3).

The best results are achieved for the pair *4-1* in which the results of all strategies are computed to be higher than 98%. For that pair, evenly distributed point correspondences all around the rooftop of the Rathaus building are detected for all strategies (see Figs. 5.5 - 5.7). As a general observation, for all strategies, the point correspondences are located mainly over the rooftop; only a marginal number of correspondences are detected in other parts of the image. However, once again, the largest number of point correspondences, and therefore the best matching point distributions all around the images (see Figs. 5.8) are found by the graph-cut energy minimization framework (i.e. the Strategy #3). In spite of this fact, this strategy suffers from the number of FPs detected, e.g. pair *5-2* with 19 FPs (Table 5.2), and thus, an additional post-processing step to further eliminating the erroneous correspondences seems to be essential for this approach.
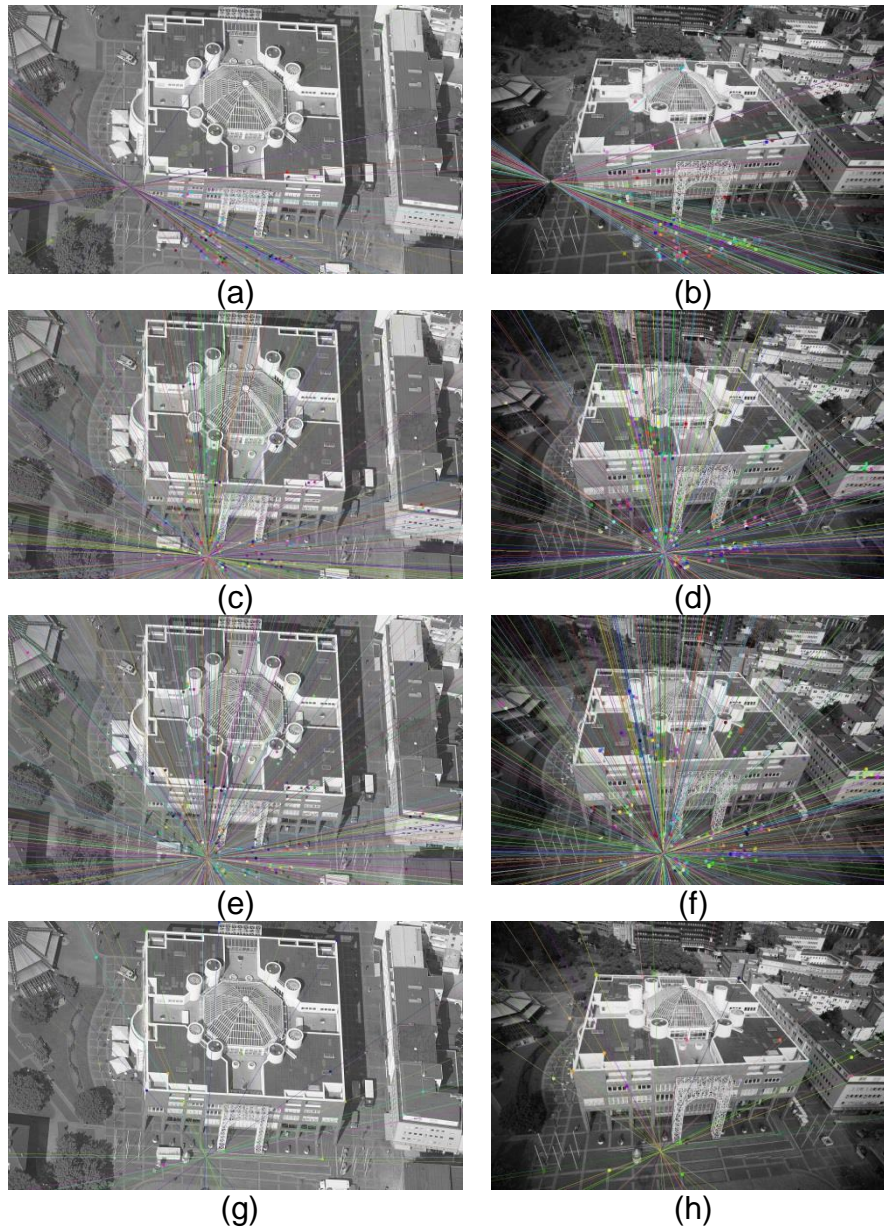
Figure 5.4: The illustration of epipolar geometry computed from (a, b) Strategy #1, (c, d) from Strategy #2, (e, f) from Strategy #3, and (g, h) from manually collected (well-distributed) points. All results belong to Pair *5-1* (first column: oblique image #5, second column: UAV image #1). Each color refers to the epipolar line of corresponding matching points in both images.

Table 5.2:   Matching results of oblique images (#1 - #5) with UAV Image #2 obtained with all strategies.

UAV Image

UAV Image #2

Oblique Images

Oblique Image #1

Pair *1-2*	Strategy #1		Strategy #2		Strategy #3	
	TP	*FP*	*TP*	*FP*	*TP*	*FP*
Total #	143	7	163	3	201	7
Precision	95.33%		96.45%		96.63%	

Oblique Image #2

Pair *2-2*	Strategy #1		Strategy #2		Strategy #3	
	TP	*FP*	*TP*	*FP*	*TP*	*FP*
Total #	242	3	249	6	310	7
Precision	98.78%		97.65%		97.79%	

Oblique Image #3

Pair *3-2*	Strategy #1		Strategy #2		Strategy #3	
	TP	*FP*	*TP*	*FP*	*TP*	*FP*
Total #	251	6	285	4	344	9
Precision	97.66%		98.62%		97.45%	

Oblique Image #4

Pair *4-2*	Strategy #1		Strategy #2		Strategy #3	
	TP	*FP*	*TP*	*FP*	*TP*	*FP*
Total #	353	6	360	2	406	8
Precision	98.33%		99.45%		98.06%	

Oblique Image #5

Pair *5-2*	Strategy #1		Strategy #2		Strategy #3	
	TP	*FP*	*TP*	*FP*	*TP*	*FP*
Total #	267	5	303	1	369	19
Precision	98.16%		99.67%		95.10%	
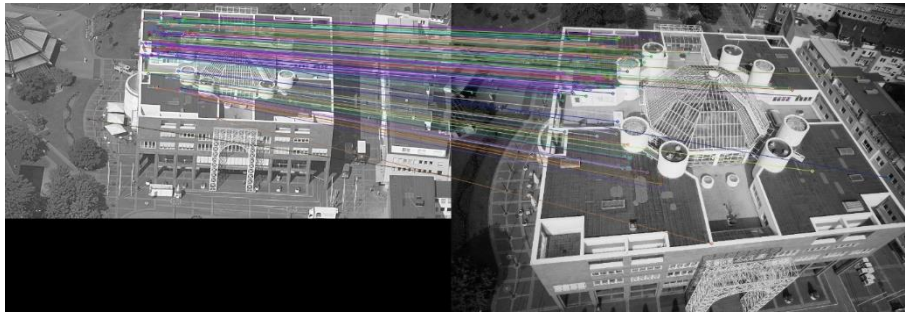
(a) Pair *1-1* – Total matches: 150 (TP: 143 – FP: 7)
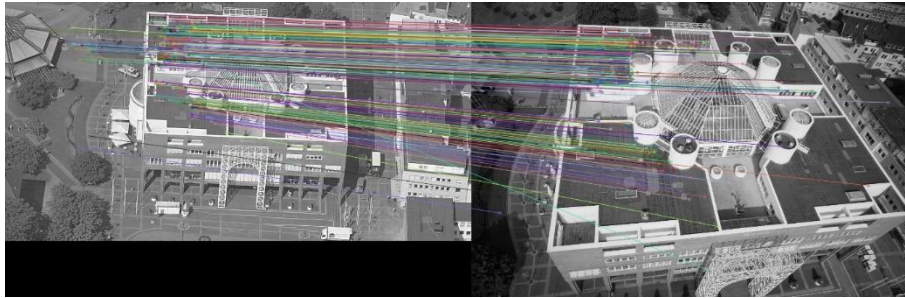
(b) Pair *2-1* – Total matches: 245 (TP: 242 – FP: 3)
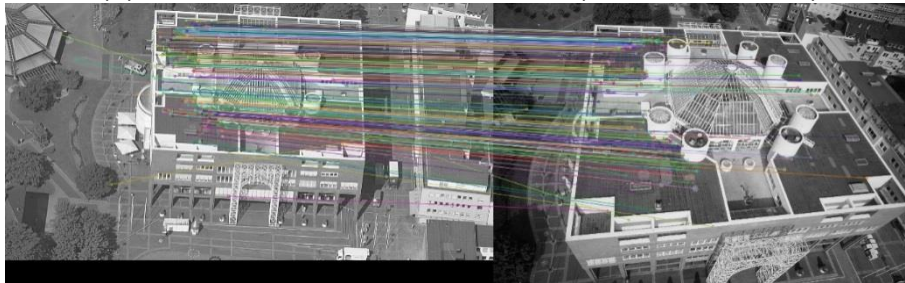
(c) Pair *3-1* – Total matches: 257 (TP: 251 – FP: 6)
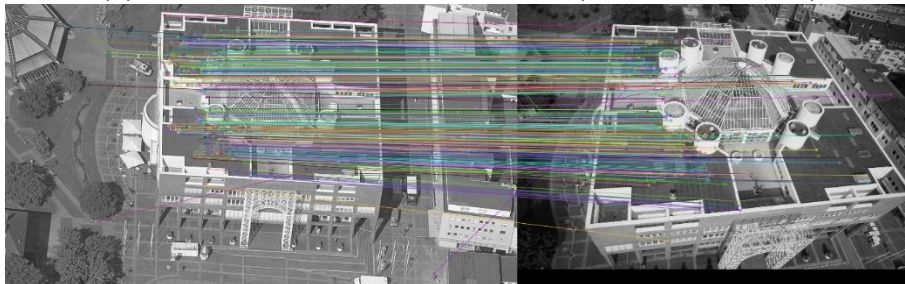
(d) Pair *4-1* – Total matches: 359 (TP: 353 – FP: 6)
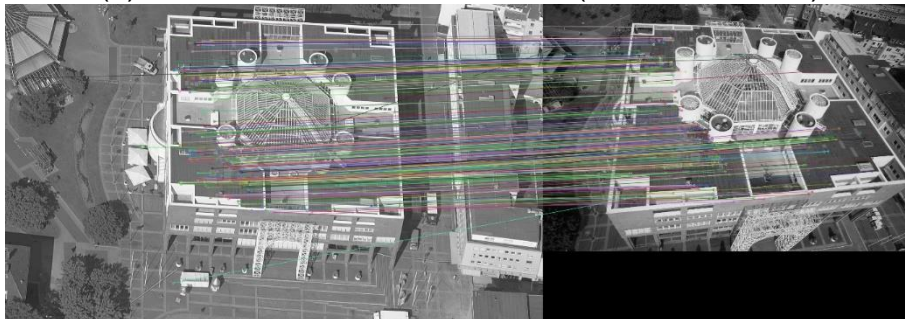
(e) Pair *5-1* – Total matches: 272 (TP: 267 – FP: 5)

Figure 5.5:   Matching results of oblique and UAV image pairs for Strategy #1.
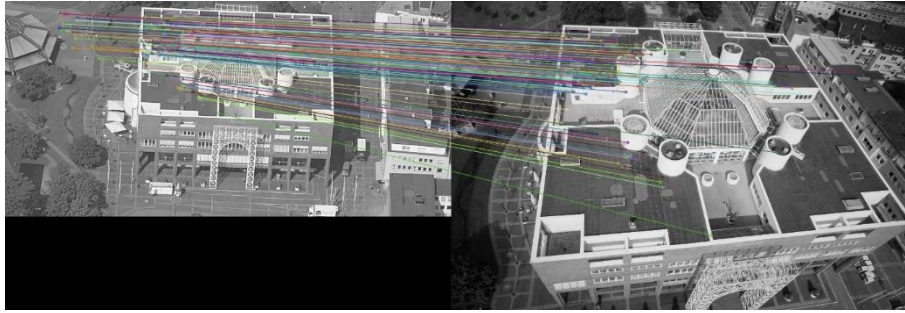
(a) Pair *1-1* – Total matches: 166 (TP: 163 – FP: 3)
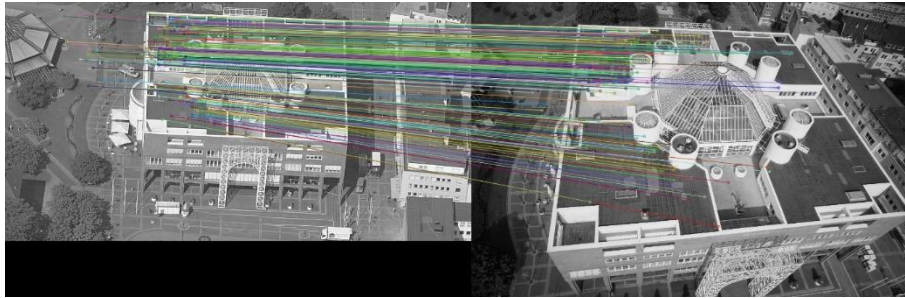
(b) Pair *2-1* – Total matches: 255 (TP: 249 – FP: 6)
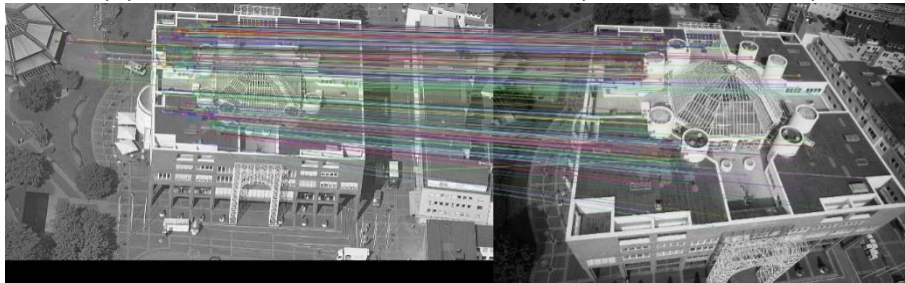
(c) Pair *3-1* – Total matches: 289 (TP: 285 – FP: 4)

(d) Pair *4-1* – Total matches: 362 (TP: 360 – FP: 2)

(e) Pair *5-1* – Total matches: 304 (TP: 303 – FP: 1)

Figure 5.6:   Matching results of oblique and UAV image pairs for Strategy #2.
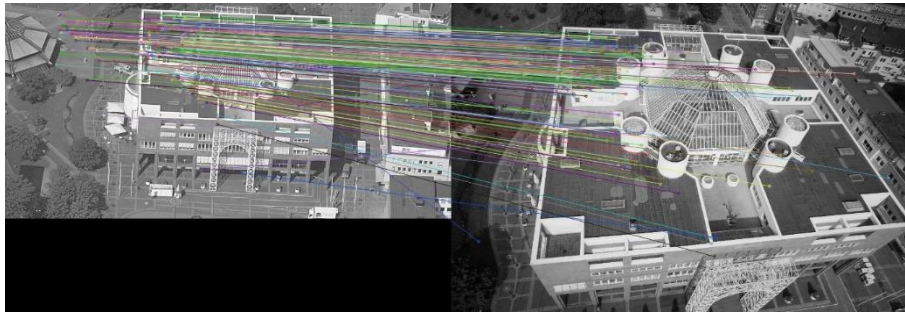
(a) Pair *1-1* – Total matches: 208 (TP: 201 – FP: 7)
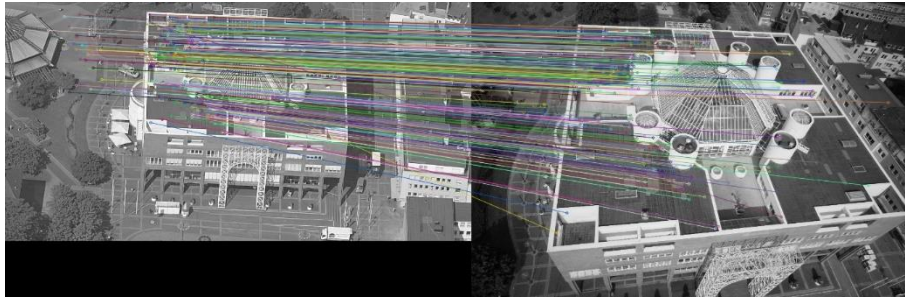
(b) Pair 2-*1* – Total matches: 317 (TP: 310 – FP: 7)
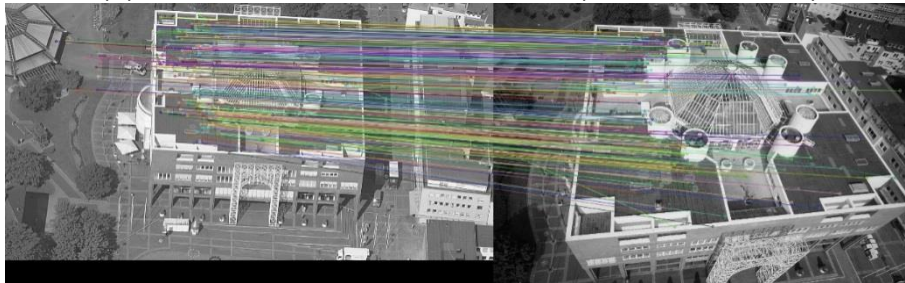
(c) Pair *3-1* – Total matches: 353 (TP: 344 – FP: 9)
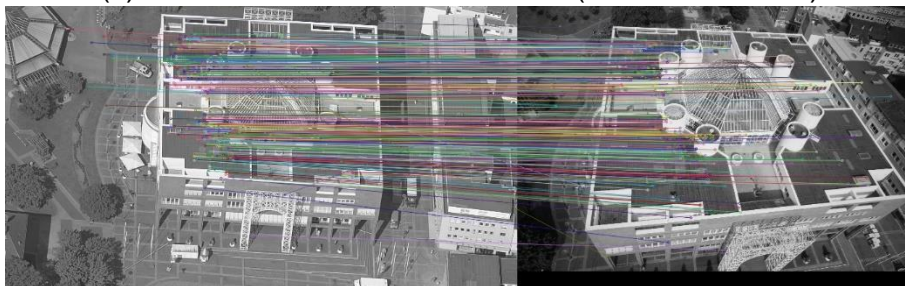
(d) Pair *4-1* – Total matches: 414 (TP: 406 – FP: 8)

(e) Pair *5-1* – Total matches: 388 (TP: 369 – FP: 19)

Figure 5.7:   Matching results of oblique and UAV image pairs for Strategy #3.

5.3.   Discussion of the False Correspondences

Although successful results are achieved for the proposed framework, there are several reasons for the correspondences that are found to be incorrect. In this respect, the related reasons can be classified into two broad cases:

- Repetitive patterns
- Accidental correspondence

The major problem in this context is still the repetitive textures and patterns. During the test performed, the repetitive patterns found to be occurring on building elements like windows on a facade of a building (especially on the Rathaus building), trees, bushes or tiles on the ground level (Fig. 5.8). Although NDDR and left-right consistency check measures are applied to mitigate such cases, apparently there is still room for further research on this topic.

The other case is the errors that occur due to the accidental similarity of (completely) different regions (Fig. 5.9a). Most of these cases are removed through the epipolar constraint; however, several erroneous matches still exist which also satisfy this constraint by chance. One other reason for this type of error is due to the similarity of the point neighborhood in different images caused by mainly perspective viewing (Fig. 5.9b).

5.4.   Comparison with Previous Studies

In this section, the results of the proposed iterative approach are compared with the results of the matching approach presented by Onyango et al. [47]. They proposed a matching method for the UAV and aerial oblique imagery to perform a relative orientation for UAV images via already georeferenced oblique images. In their study, they also evaluated the same benchmark dataset utilized in this thesis, and promoted the usage of multiple homography information. Besides, A-KAZE operator is also implemented for the point-based feature detection. Thus, for the sake of completeness of the comparison conducted, the results of A-KAZE approach
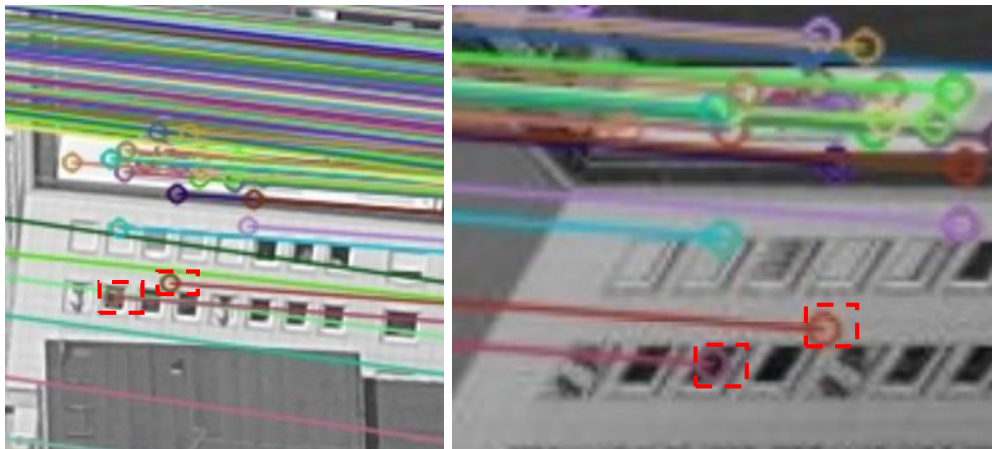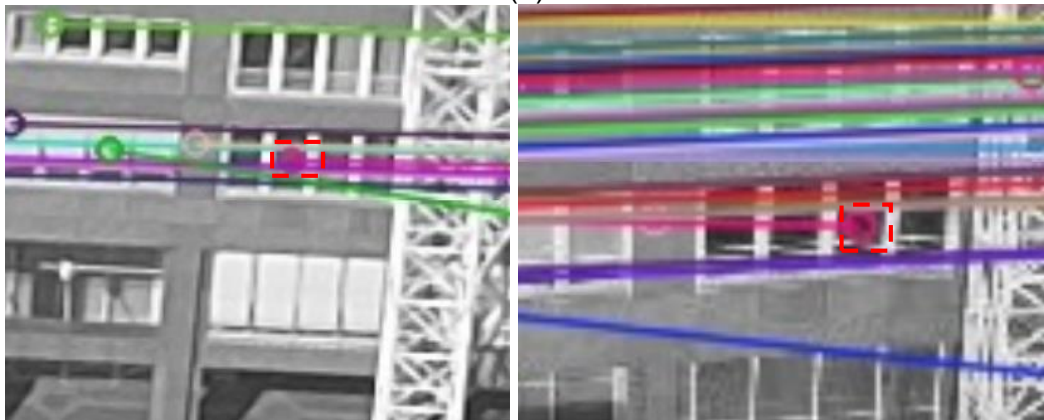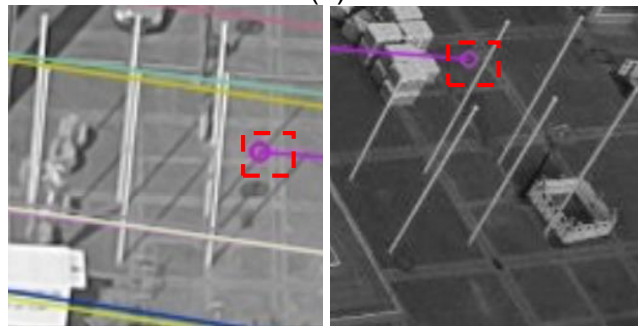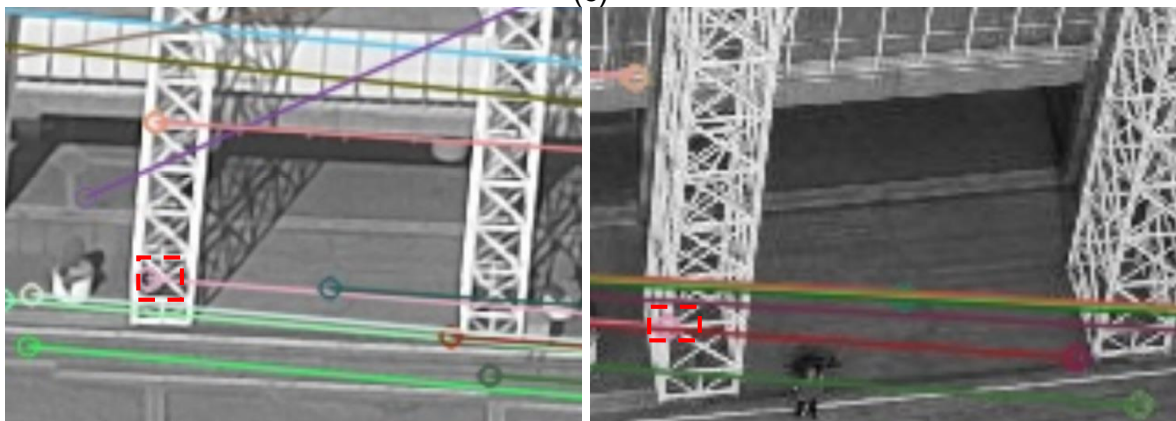
Figure 5.8: Examples of incorrect point correspondences due to repetitive patterns.
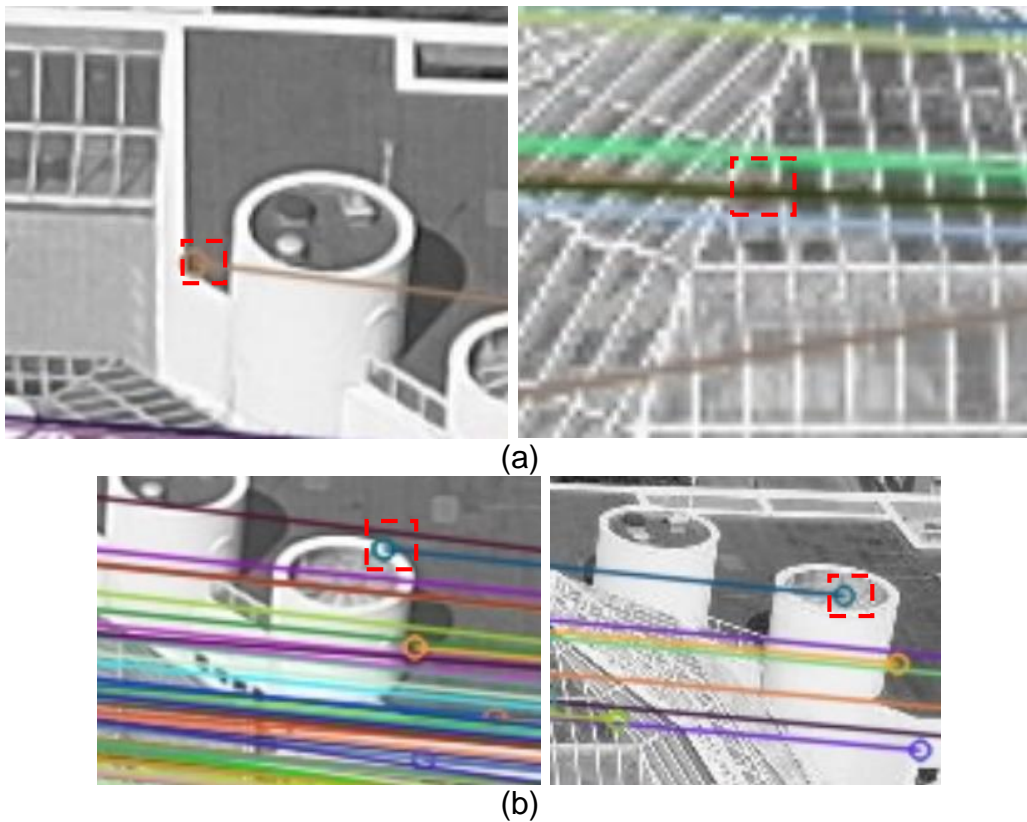
Figure 5.9: Examples of incorrect point correspondences due to accidental correspondence.

presented in [39] (also with left-right consistency check applied) are also provided in this part.

The comparative results with respect to the TPs, FPs and precision ratios of the approaches for the two pairs consisting of Rathaus building are provided in Table 5.3. Based on the results given in Table 5.3, A-KAZE approach in [39] provides too many FPs (i.e. 290). Nonetheless, applying the left-right consistency check significantly reduces the number of FPs; but, also reduces the number of TPs detected. For these two image pairs, the proposed framework in this thesis provided significantly better precision results compared to the results of the approach in [39]. For the approach conducted by Onyango et al. [47], FP values are not provided, and therefore, only a comparison based on TPs are carried out. Considering the results, the proposed framework provided significantly higher TPs, especially for the pair *4-2*. In that case, the proposed framework increased the number of TPs more than 2.6 times than the number of TPs provided by the approach presented in Onyango et al. [47]. Such comparisons reveal that the proposed framework in this thesis provides superior

performance as it has direct ability to increase the number of TPs while suppressing the number of FPs. As a result, a better point matching performance can be reached.

Table 5.3:   Matching results of the developed approach and the other approaches for the two pairs of the Rathaus building scene.

Image Pair	A-KAZE [39] (with NNDR)			A-KAZE [39] (with NNDR) and Consistency Check Applied			Onyango et al. [47]			Proposed Framework (Strategy #3)		
	TP	FP	Pre. (%)	TP	FP	Pre. (%)	TP	FP	Pre. (%)	TP	FP	Pre. (%)
Pair 4-1	117	290	28.8	74	51	59.2	229	NP	-	360	12	96.8
Pair 4-2	163	291	35.9	105	84	55.0	155	NP	-	406	8	98.1

*NP: not provided

6. CONCLUSIONS AND RECOMMENDATIONS

In this chapter, the conclusions derived from the introduced framework are summarized, and the recommendations for future studies are presented.

6.1. Conclusions

In this thesis, a feature matching framework is proposed to guiding point matching between airborne oblique and UAV imagery. The framework presented in this thesis consists of three steps. First, a feature detection step applying A-KAZE descriptor is utilized to collect keypoints. Second, an initial matching step is developed to identify tentative matches between the oblique and UAV images. In the third and final matching step, putative matches are extracted from the previously collected tentative matches. In this thesis, the area covering the Dortmund-Centre from a benchmark dataset is selected to evaluate the results of the proposed framework.

The following conclusions are reached from the results achieved for the proposed point matching framework:

- The results of the proposed point matching framework are found to be robust for the test carried out. All precision ratios are found to be better than 94.5% and therefore, the final results can be expressed as convincing for an output of a point matching task between an airborne oblique and a UAV image.
- A-KAZE descriptor is known to be scale independent [39]. However, rigorous evaluations state that down-sampling operation has an obvious impact on the computed matching results. Therefore, the down-sampling process might still be required before any further operation, and it seems to be critical to achieve large number of consistent correspondences between the airborne oblique and the UAV image.
- It is found that the proposed iterative strategy noticeably increased (≈ 2x compared to A-KAZE results) the number of tentative matches. Therefore, a combination of the outputs of different detector response thresholds seems to be a straightforward way to increase the number of initial matches.
- According to the results, the best precision ratios in an overall sense are computed for Strategy #2 in which the Least Median of Squares (LMedS) and

RANSAC approaches are successively utilized. However, the highest numbers of TPs for all cases are computed for the Strategy #3 utilizing a graph-cut energy minimization framework, and resulting nearly 40 and 80 TP improvements on average compared to the results of Strategies #1 and #2, respectively.

- Considering the distribution of point matches, Strategies #1 in which a classical RANSAC approach is utilized provided the worst results. The distributions of the output correspondences between the Strategies #2 and #3 are found to be comparable; however, it is believed that the outputs of Strategy #3 are slightly better due to the increase in the numbers of TP matches.

- Although successful results are achieved for the proposed framework, repetitive patterns and accidental matching of irrelevant features are still the major concerns. FPs due to repetitive patterns are mostly observed on facades of buildings, trees, bushes or tiles on the ground level. Perspective viewing differences are one of the major reasons of blunders that exist in the output.

- The analyses with oblique images having different perspective views denote that the performance of matching highly depends on viewing differences between the oblique and UAV images. If a satisfactory matching performance is required, the viewing differences between the datasets should not be large.

- The proposed framework runs 13-18x slower than the A-KAZE method. This is obviously due to brute force matching applied and the iterative framework developed during the generation of tentative matches to increase the number of candidate correspondences to be matched.

- For the two image pairs evaluated for comparison, the proposed framework developed in this thesis provided significantly better results compared to the results of the previous studies evaluated. Such comparisons reveal that the proposed framework in this thesis provides superior performance by achieving a nice balance between the TPs and FPs detected.

6.2. Recommendations

The followings are suggested for further studies:

- A final post-processing step constitutes a topic for further development, especially for the outputs of Strategy #3 in which the highest FPs are observed. In this way, the precision ratios computed can be further increased.

- The randomness within the proposed framework can be investigated. All strategies eventually rely on a RANSAC approach that includes an internal random point selection procedure. Although the level of confidence (probability) is set to the highest value of 99% in all test conducted, there may still be an effect of randomness in the outputs, and this effect can be further revealed.

- A-KAZE descriptor is known to be rotation invariant [39]; however, no explicit test on this issue is performed. Therefore, new tests revealing the performance of matching with respect to different rotations of oblique and UAV images can be performed.

- Increasing the number of test cases with different airborne oblique and UAV images should be performed to better understand and further improve the proposed framework and the related parameters. In the end, this will also help to provide more reliable comparative results compared to previous studies in the same context.

- The initial processing stage of the proposed framework can be easily conducted in an automatic manner. First, if the ground sampling distances of both datasets are known a priori, the related down-sampling ratio can be computed and specifically applied. Second, the search window can be approximately set to limit the search space within airborne oblique image with the help of the (initial) orientation information for both datasets, as shown in [54].

- Local Difference Binary (LDB) [61] exploits 3 bits per comparison (two gradient orientations for x and y, and one intensity information) to generate the descriptor. Therefore, the effect of color information on the computed results can be investigated.

- Another interesting test using the proposed framework would be to assessing the performance with the terrestrial images acquired. In that case, the perspective differences would be extremely large, and such evaluations constitute another topic for further development.

60

- One of the drawbacks of the proposed framework is the processing speed. For large databases brute force matching can be replaced with efficient search strategies like Kd-trees. Besides, the iterations can be reformulated and adapted in a parallel processing scheme. In this way, this deficiency can be mitigated.

- The advances in deep learning allow finding good feature representations for feature extraction tasks. In this context, a future work would be to adapting and/or expanding the current framework with deep network architecture.

- The outputs of this framework can be utilized in very different areas and applications, e.g. for improving the dense matching, 3-D reconstruction, surface analysis etc. Besides, the proposed approach might also be tested for tie point generation within the blocks generated from either oblique or UAV images, or both.

REFERENCES

[1]   Zhuo, X., Koch, T., Kurz, F., Fraundorfer, F., Reinartz, P., Automatic UAV Image Geo-Registration by Matching UAV Images to Georeferenced Image Data, Remote Sensing, 9(4): p. 376, **2017**.

[2]   UAV - Airborne Imaging, https://www.sccssurvey.co.uk/surveying-equipment-instruments/uav-solutions-airborne-imaging.html (Access Date: **March 9, 2020**).

[3]   Aasen, H., Honkavaara, E., Lucieer, A., Zarco-Tejada, P. J., Quantitative Remote Sensing at Ultra-High Resolution with UAV Spectroscopy: A Review of Sensor Technology, Measurement Procedures, and Data Correction Workflows, Remote Sensing, 10(7): p. 1091, **2018**.

[4]   Murtiyoso, A., Remondino, F., Rupnik, E., Nex, F., Grussenmeyer, P., Oblique Aerial Photography Tool for Building Inspection and Damage Assessment, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(1): p. 309, **2014**.

[5]   Lemmens, M., A Survey of Features and Systems Digital Oblique Aerial Cameras (1), GIM International-the Worldwide Magazine for Geomatics, 28(4): p. 20-25, **2014**.

[6]   Barsanti, S.G., D. Gherdevich, D. Degrassi, Historic and archaeological itineraries for the discovery of Friuli during the Lombard period,in CAA 2012, **2011**.

[7]   Imam, E., Aerial Photography and Photogrammetary, Remote Sensing and GIS, **2018**.

[8]   Rupnik, E., F. Nex, and F. Remondino, Oblique Multi-Camera Systems-Orientation and Dense Matching Issues, **2014**.

[9]   Remondino, F. and M. Gerke, Oblique Aerial Imagery–A Review, **2015.**

[10]   Colomina, I., Blázquez, M., Molina, P., Parés, M.N, Wis, M., Towards a New Paradigm for High-Resolution Low-Cost Photogrammetry and Remote Sensing, in ISPRS XXI Congress, Beijing, China, **2008**.

[11]   Nex, F. and M. Gerke, ISPRS/EuroSDR Benchmark Multi-Platform Photogrammetry: Preliminary Results, in EuroSDR/ISPRS Workshop on Oblique Cameras and Dense Image Matching, 19-20 October 2015, Southampton, United Kingdom, **2015**.

[12]   Nex, F. and F. Remondino, UAV for 3D Mapping Applications: A Review, Applied Geomatics, 6(1): p. 1-15, **2014**.

[13]   Fixed Wing Versus Rotary Wing for UAV Mapping Applications, https://www.questuav.com/media/case-study/fixed-wing-versus-rotary-wing-for-uav-mapping-applications/, (Access Date: **March 20, 2020**).

[14]   Fonseca, L.M. and B. Manjunath, Registration Techniques for Multisensor Remotely Sensed Imagery, PE & RS- Photogrammetric Engineering & Remote Sensing, 62(9): p. 1049-1056, **1996**.

[15]   Valgaerts, L., Bruhn, A., Mainberger, M., Weickert, J., Dense Versus Sparse Approaches for Estimating the Fundamental Matrix, International Journal of Computer Vision,96(2): p. 212-234, **2012**.

[16]   Scharstein, D. and R. Szeliski, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, International Journal of Computer Vision, 47(1-3): p. 7-42, **2002**.

[17]   Zitova, B. and J. Flusser, Image Registration Methods: A Survey, Image and Vision Computing, 21(11): p. 977-1000, **2003**.

[18]   Hirschmuller, H. and D. Scharstein, Evaluation of Stereo Matching Costs on Images with Radiometric Differences, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(9): p. 1582-1599, **2008**.

[19]   Ansar, A., A. Castano, and L. Matthies, Enhanced Real-Time Stereo Using Bilateral Filtering, in Proceedings, 2nd International Symposium on 3D Data Processing, Visualization and Transmission, IEEE, p. 455-462, **2004**.

[20]   Viola, P. and W.M. Wells III, Alignment by Maximization of Mutual Information, International Journal of Computer Vision, 24(2): p. 137-154, **1997**.

[21]   Hirschmuller, H., Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **2005**.

[22]   Hirschmuller, H., Stereo Processing by Semiglobal Matching and Mutual Information, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2): p. 328-341, **2007**.

[23]   Ummenhofer, B., Zhou, H., Uhrig, J., Mayer, N., Ilg, E., Dosovitskiy, A., Brox, T., Demon: Depth and Motion Network for Learning Monocular Stereo, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, **2017**.

[24]   Kisku, D. R., Tistarelli, M., Sing, J. K., Gupta, P., Face Recognition by Fusion Of Local and Global Matching Scores Using DS Theory: An Evaluation with Uni-Classifier and Multi-Classifier Paradigm, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, **2009**.

[25]   Weingarten, J., G. Gruener, and R. Siegwart, A Fast and Robust 3D Feature Extraction Algorithm for Structured Environment Reconstruction, **2003**.

[26]   Krish, K., Heinrich, S., Snyder, W. E., Cakir, H., Khorram, S., Global Registration of Overlapping Images Using Accumulative Image Features, Pattern Recognition Letters, 31(2): p. 112-118, **2010**.

[27]   Han, B. and L. Davis, Object Tracking by Adaptive fFeature Extraction, in 2004 International Conference on Image Processing (ICIP'04), **2004**.

[28]   Weerakoon, T., K. Ishii, and A.A.F. Nassiraei, Geometric Feature Extraction from 2D Laser Range Data for Mobile Robot Navigation, in 2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS), **2015**.

[29]   Awad, A.I. and M. Hassaballah, Image Feature Detectors and Descriptors, Studies in Computational Intelligence, Springer International Publishing, Cham, **2016**.

[30]   Project 2: Local Feature Matching, https://www.cc.gatech.edu/classes/AY2016/cs4476_fall/results/proj2/html/pbhutada6/index.html, (Access Date: **March 29, 2020**).

[31]   Lowe, D.G., Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 60(2): p. 91-110, **2004**.

[32]   Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., Speeded-up Robust Features (SURF), Computer Vision and Image Understanding, 110(3): p. 346-359, **2008**.

[33]   Morel, J.-M. and G. Yu, ASIFT: A New Framework for Fully Affine Invariant Image Comparison, SIAM Journal on Imaging Sciences, 2(2): p. 438-469, **2009**.

[34]   Calonder, M., Lepetit, V., Strecha, C., & Fua, P., BRIEF: Binary Robust Independent Elementary Features, in European Conference on Computer Vision, Springer, **2010**.

[35]   Madeo, S. and M. Bober, Fast, Compact, and Discriminative: Evaluation of Binary Descriptors for Mobile Applications, IEEE Transactions on Multimedia, 19(2): p. 221-235, **2016**.

[36]   Rublee, E., Rabaud, V., Konolige, K., Bradski, G., ORB: An Efficient Alternative to SIFT or SURF, in 2011 International Conference on Computer Vision, IEEE, **2011**.

[37]   Rosten, E. and T. Drummond, Machine Learning for High-Speed Corner Detection, in European Conference on Computer Vision, Springer, **2006**.

[38]   Alcantarilla, P.F., A. Bartoli, and A.J. Davison, KAZE Features, in European Conference on Computer Vision, Springer, **2012**.

[39]   Alcantarilla, P.F., J. Nuevo, and A. Bartoli, Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, IEEE Trans. Patt. Anal. Mach. Intell., 34(7): p. 1281-1298, **2011**.

[40]   Barath, D. and J. Matas, Graph-Cut RANSAC, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, **2018**.

[41]   Mishkin, D., J. Matas, and M. Perdoch, MODS: Fast and Robust Method for Two-View Matching, Computer Vision and Image Understanding, 141: p. 81-93, **2015**.

[42]   Bansal, M., K. Daniilidis, and H. Sawhney, Ultrawide Baseline Facade Matching for Geo-Localization, in Large-Scale Visual Geo-Localization, Springer, p. 77-98, **2016**.

[43]   Chung, Y.-C., T.X. Han, and Z. He, Building Recognition Using Sketch-Based Representations and Spectral Graph Matching, in 2009 IEEE 12th International Conference on Computer Vision, IEEE, **2009**.

[44]   Matas, J., Chum, O., Urban, M., Pajdla, T., Robust Wide-Baseline Stereo from Maximally Stable Extremal Regions, 22(10): p. 761-767, **2004**.

[45]   Verykokou, S. and C. Ioannidis, A Photogrammetry-Based Structure from Motion Algorithm Using Robust Iterative Bundle Adjustment Techniques, ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4, **2018**.

[46]   Altwaijry, H. and S.J. Belongie, Ultra-wide Baseline Aerial Imagery Matching in Urban Environments, in BMVC, **2013**.

[47]   Onyango, F. A., Nex, F., Peter, M. S., & Jende, P., Accurate Estimation of Orientation Parameters of UAV Images Through Image Registration with Aerial Oblique Imagery, International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 42, **2017**.

[48]   O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., Walsh, J., Deep Learning vs. Traditional Computer Vision, in Science and Information Conference, Springer, **2019**.

[49]   Image Registration: From SIFT to Deep Learning, https://www.sicara.ai/blog/2019-07-16-image-registration-deep-learning, (Access Date: **July 16, 2019**).

[50]   Fischer, P., A. Dosovitskiy, and T. Brox, Descriptor Matching with Convolutional Neural Networks: A Comparison to SIFT, arXiv preprint arXiv:1405.5769, **2014**.

[51]   Yang, Z., T. Dan, and Y.Yang, Multi-temporal Remote Sensing Image Registration Using Deep Convolutional, IEEE Access, 6, p. 38544-38555, **2018**.

[52]   Nassar, A., Amer, K., ElHakim, R., ElHelw, M., A Deep CNN-Based Framework for Enhanced Aerial Imagery Registration with Applications to UAV Geolocalization, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, **2018**.

[53]   Katole, A. L., Yellapragada, K. P., Bedi, A. K., Kalra, S. S., & Chaitanya, M. S., Hierarchical Deep Learning Architecture for 10k Objects Classification, **2015**.

[54]   Onyango, F.A., Multi-Resolution Automated Image Registration, Master of Science Thesis, Faculty of Geo-information Science and Earth Observation, University of Twente, **2017**.

[55]   Leutenegger, S., M. Chli, R.Y. Siegwart, BRISK: Binary Robust Invariant Scalable Keypoints, In 2011 International Conference On Computer Vision, IEEE, **2011**.

[56]   Weickert, J., B.T.H. Romeny, and M.A. Viergever, Efficient and Reliable Schemes for Nonlinear Diffusion Filtering, IEEE Transactions on Image Processing, 7(3): p. 398-410, **1998**.

[57]   Grewenig, S., J. Weickert, and A. Bruhn, From Box Filtering to Fast Explicit Diffusion, in Joint Pattern Recognition Symposium, Springer, **2010**.

[58]   Perona, P. and J. Malik, Scale-space and Edge Detection Using Anisotropic Diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7): p. 629-639, **1990**.

[59]   Kalms, L., K. Mohamed, and D. Göhringer, Accelerated Embedded Akaze Feature Detection Algorithm on FPGA, in Proceedings of the 8th International Symposium on Highly Efficient Accelerators and Reconfigurable Technologies, p. 1-6, **2017**.

[60]   Ordóñez, Á., F. Argüello, and D.B. Heras, Alignment of Hyperspectral Images Using KAZE Features, Remote Sensing, 10(5): p. 756, **2018**.

[61]   Yang, X. and K.-T. Cheng, LDB: An Ultra-Fast Feature for Scalable Augmented Reality on Mobile Devices, in 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE, **2012**.

[62]   Hamming, R.W., Error Detecting and Error Correcting Codes, The Bell System Technical Journal, 29(2): p. 147-160, **1950**.

[63] Hartley, R. and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, **2003**.

[64] Torr, P.H., A. Zisserman, MLESAC: A New Robust Estimator with Application to Estimating Image Geometry, Computer Vision and Image Understanding, 78(1): p. 138-156, **2000**.

[65] Nex, F., Remondino, F., Gerke, M., Przybilla, H. J., Bäumker, M., & Zurhorst, A., ISPRS Benchmark for Multi-Platform Photogrammetry, ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 2, **2015**.

[66] Yan, H., Lv, G., Ren, X., Dong, X., Improved Nearest Neighbor Distance Ratio for Matching Local Image Descriptors, in International CCF Conference on Artificial Intelligence, Springer, **2018**.

[67] Trucco, E. and A. Verri, Introductory Techniques for 3-D Computer Vision, Vol. 201, Prentice Hall Englewood Cliffs, **1998**.

[68] Fischler, M.A. and R.C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Communications of the ACM, 24(6): p. 381-395, **1981**.

[69] Raguram, R., J.-M. Frahm, and M. Pollefeys, A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus, in European Conference on Computer Vision, Springer, **2008**.

[70] Derpanis, K.G., Overview of the RANSAC Algorithm, Image Rochester NY, 4(1): p. 2-3, **2010**.

[71] Rousseeuw, P.J., Least Median of Squares Regression, Journal of the American Statistical Association, 79(388): p. 871-880, **1984**.

[72] Barreto, H., An Introduction to Least Median of Squares, Chapter Contribution to Barreto and Howland, Econometrics via Monte Carlo Simulation, **2001**.

[73] Shertzer, K.W. and M.H. Prager, Least Median Of Squares: A Suitable Objective Function for Stock Assessment Models?, Canadian Journal of Fisheries and Aquatic Sciences, 59(9): p. 1474-1481, **2002**.

APPENDIX

The strategies utilized to constrain the matches and removing the outliers after stereo image matching are described in this part.

1.   Nearest Neighbor Distance Ratio

The NNDR technique also known as Lowe's ratio test, was proposed by D. Lowe [31] and performed to eliminate keypoints in the query image which do not have any correct match in the training database because these keypoints were not detected in the training images or they can be originated from the background clutter.

For a feature descriptor in the reference image, if the distance ratio between the first and second nearest neighbor matches in the target image is below a threshold, DR and D1 are matched [66], i.e.,

$$\frac{||DR - D1||}{||DR - D2||} < t_{ratio}$$

,   (Eq. 1)

where;

i.   DR: a feature descriptor in the reference image;
ii.   D1: the nearest descriptor to DR in the target image;
iii.   D2: the second nearest descriptor in the target image;
iv. $||DR - D1||$: Euclidian distance of DR to D1;
v. $||DR - D2||$: Euclidian distance of DR to D2;
vi. t_{ratio}: threshold value for the ratio computed.

NNDR technique finds only one and best match for a feature in the reference image. A low NNDR value is interpreted as D1 can be a good match while a high value of NNDR demonstrates D1 can be an incorrect or ambiguous match. Thus, the pre-defined NNDR threshold value must provide a good separation between correct and false matches.

2.   Left-Right Cross Check

The cross check technique is a good alternative to the NNDR test mentioned above. Two features in both image feature sets should match each other to be chosen by this technique, and this process provides consistent matching results by returning only those features. The cross check test chooses matches (F_A and F_B) so that;

- F_B is the best match for F_A in I_B,
- and F_A is the best match for F_B in I_A.

where;

      i. F_A: the feature in the image-A,

      ii. F_B: the feature in the image-B,

      iii. I_A: the image-A,

      iv. I_B: the image-B.

3.   Epipolar Constraint

If a homogeneous point in 3D-space X is imaged as x in the first view, and x′ in the second, then the image points satisfy the below relation;

$$x'^T * F * x = 0 \qquad . \qquad (Eq.\ 2)$$

In this equation F is the fundamental matrix and it is the algebraic representation of the epipolar geometry. It is a 3 × 3 homogeneous matrix of rank 2, which encodes the epipolar geometry of two views and defined in terms of pixel coordinates.

Given a point in one view, one can be find out the epipolar line which the point lies on in the second view by multiplying with fundamental matrix.

For any point x in the first image, the corresponding epipolar line is;

$$l' = F * x \qquad . \qquad (Eq.\ 3)$$

Similarly,

$$l = F^T * x' \qquad , \qquad (Eq.\ 4)$$

represents the epipolar line corresponding to x′ in the second image [67].

4.   RANSAC

The RAndom SAmple Consensus (RANSAC) algorithm was introduced by Fischler and Bolles [68] as an iterative parameter estimation approach designed to deal with outlier points in the input dataset. RANSAC estimates a relation that fits the input data while classifying the data into inliers and outliers, hence it can be evaluated as an outlier detection method [69].

The main algorithm steps are summarized as follows [70]:

Algorithm RANSAC

---

1. Given a set of data points, the minimum number of points required to determine a model are selected randomly.
2. An initial model is determined with these randomly selected points.
3. Points from the input dataset which fit the initial model with a pre-defined tolerance *d* are searched and initial point set is enlarged with those points. The enlarged point set is called as a consensus set and all points in this set assumed as inlier points.
4. If the ratio of the number of inlier points to the total number of points within the set exceed a threshold *t*, a new model is estimated by using the consensus set.
5. Otherwise, if the ratio is below the threshold, steps 1 to 4 are repeated for a pre-defined number of iteration value *N*.

RANSAC has been used in numerous robust estimation problems in computer vision, since it has a high capability of tolerating a great number of outliers within the input data [69].

5.   Least Median of Squares (LMedS)

The least median of squares (LMedS) method was proposed by Rousseeuw [71] in 1984 as an alternative estimator to the ordinary least squares (LS) method. The aim of the least square estimation is to find parameters of a given equation which is the best fit for the data points. The main idea of this technique is to minimize squares of the residuals (*r*) of the points from the fitted model;

$$\text{minimize} \sum_{i=1}^{n} r_i^2 \qquad .$$ (Eq. 5)

This technique became a cornerstone in statistics but yet it has been subject to criticism due to its lack of robustness to outliers. To overcome the drawbacks of the LS method many algorithms have been proposed such as least mean squares. In the case of computer vision, the given dataset is generally noisy meaning that there is a large portion of outliers and least mean squares often gives a poor fit. To solve this problem one can be use the LMedS estimation.

In this approach the sum is replaced by the median of the squared residuals for introducing a robust estimator to contamination in the dataset;

$$\text{minimize } med\ r_i^2 \qquad . \qquad \text{(Eq. 6)}$$

LMedS is a more robust fitting technique than least mean squares. It should come as no surprise that an estimator using the median would be less sensitive to extreme values than least mean squares, which is related to the average [72].

LMedS estimator is resisting the effect of nearly 50% of contamination in the data. This is the percentage of outliers that can be tolerated while estimator still returning a good fit [73].

6.   Graph-Cut RANSAC

In [40], a RANSAC approach utilizing the spatial coherence of inliers and outliers in a dataset was proposed. Compared to the classical RANSAC approach, the major improvement proposed was the integration of the neighborhood information, i.e.

$$\text{E}(L) = \text{U}(L) + \lambda\,\text{V}(L) \qquad . \qquad \text{(Eq. 7)}$$

E is the energy, U and V represent the unary and smoothness terms, $L \in \{0,1\}^{|P|}$ is a labeling where $L_p \in L$ is the label of point $p \in P$, |P| denotes the number of points in P, and λ parameter controls the level of smoothing. For binary cases, the globally optimal labeling can be computed by searching the minimum of the energy term E with respect to the labels L using the graph-cut minimization.

The unary term in Eq. 7 is computed as

$$U(L) = \sum_{p \in P} \|L_p\|_K \qquad , \qquad \text{(Eq. 8)}$$

where

$$\|L_p\|_K = \begin{cases} 1 - K(\delta, \epsilon) & if \ L_p = 1 \\ K(\delta, \epsilon) & if \ L_p = 0 \end{cases} \qquad . \qquad \text{(Eq. 9)}$$

In Eqs. 8-9, *K* denotes the Gaussian kernel function, $K(\delta, \epsilon) = e^{-\left(\frac{\delta}{\sqrt{2}\epsilon}\right)^2}$ where δ is the distance measuring the point-to-model assignment cost. If δ=0 for a point, i.e. the point perfectly fits the model, thus, $\|L_p\|_K$ term equals to zero if the point is an inlier. The smoothness term is computed as

$$V(L) = \sum_{p,q \in G} \begin{cases} 1 & if \ L_p \neq L_q \\ \frac{K(\delta_p, \epsilon) + K(\delta_q, \epsilon)}{2} & if \ L_p = L_q = 0 \\ 1 - \frac{K(\delta_p, \epsilon) + K(\delta_q, \epsilon)}{2} & if \ L_p = L_q = 1 \end{cases} \qquad , \qquad \text{(Eq. 10)}$$

where (*p*, *q*) is an edge of neighborhood graph G between points *p* and *q*. The algorithms related to the energy minimization of E(*L*) in Eq. 7 can be found in [40].

Publications

Bas, S., and Ok, A.O., A New Productive Framework for Point-based Matching of Airplane Oblique and UAV-based Images, 2020 (submitted manuscript)